Fractional Anisotropic Diffusion for Noise Reduction in Magnetic Resonance Images

Jonathan M Blackledge and Matthew D Blackledge

Abstract—We extend the method of anisotropic diffusion for noise reduction in digital images to the case when the diffusion processes are non-Gaussian and Lévy distributed. This yields a fractional diffusion equation characterised by the Lévy index. A solution to this equation is considered and a numerical algorithm developed. The algorithm is then applied as a case study to the problem of reducing noise in magnetic resonance imaging. The focus of this study is on diffusion weighted images which have low signal-to-noise ratios.

Index Terms—Anisotropic Diffusion, Lévy Processes, Fractional Diffusion, Noise Reduction, Magnetic Resonance Imaging

I. INTRODUCTION

NOISE refers to a variety of unwanted disturbances due to measuring and recording errors of all types and interference from external sources. All signals and images have some degree of noise present in them. The amplitude of the noise may vary considerably. Also, depending on the type of imaging system used, noise may be confined to a range of frequencies or exist over the entire spectrum of the image. In the latter case, the noise is referred to as white noise in analogy with white light which is composed of a range of different frequencies (in the visible part of the electromagnetic spectrum). Noise which is confined to a band of frequencies is sometimes referred to as coloured noise. The contamination of signals and images by noise has important consequences for all types of processing.

The aim of a noise reduction algorithm is primarily to enhance the visual quality of an image by eliminating features which are random and uncorrelated. In general, noise tends to corrupt the higher frequency content of most images where the energy of the data spectrum is usually low. Thus, one way of reducing noise is by attenuating the high frequency components of the data over a range of frequencies which can be selected and adjusted by the user to provide optimum results. This can be achieved by applying a lowpass filter. A well known and commonly used lowpass filter is a Gaussian function which is related to classical diffusion processes. This is because the Point Spread Function of an image generated by the diffusion of light is a Gaussian function, a result that can be derived by considering the propagation and interaction of light to be based on a random walk process. The underlying equation for the intensity of light is then given by the diffusion equation.

The analysis of noise in imaging is associated with the use of statistical imaging methods derived from stochastic field theory. Statistical modelling techniques are used in an attempt to classify an image or regions of an image that are of statistical significance. It is an approach that is broadly based on an analysis of the Probability Density Function (PDF) of grey and/or colour levels just as Fourier based image processing is based on an analysis of the spectrum of an image. This includes approaches to deriving the PDF of an image from stochastic field models - statistical optics.

The statistics of an image can be classified into two main types, those associated with a coherent and those of an incoherent image. The statistics of an incoherent image are variable, i.e. the PDF of an incoherent image varies considerably from one image to another. However, the statistics of a coherent image have a common form that is characterized by a PDF of a negative exponential type. Thus, the PDF of a coherent image has a characteristic ‘shape’ whereas the shape of the PDF of an incoherent image is arbitrary. In both cases, statistical methods can be used to process and/or analyse images by computing different statistical parameters for the image as a whole or by applying a moving window to segment the image statistically. In the latter case, information on the variations of a statistic across an image can provide a means for its classification.

Statistical analysis ideally requires a model for the physical behaviour of the random variable(s) that is derived from basic principles. In the case of statistical signal and image analysis, this typically involves modelling the scattered field in terms of its interaction with an ensemble of ‘scattering sites’ based on an assumed stochastic process. If the density of these scattering sites is low enough so that multiple scattering is minimal, then we can apply weak scattering approximation methods to develop a model for the intensity of a wavefield interacting with a random but weak scatterer. However, strong scattering processes are considerable more difficult to model.

In this paper, we review the relationship between diffusion processes in general and the diffusion of light by multiple scattering. After considering forward an inverse solutions for diffusion in a homogeneous medium, it is shown how the inclusion of an inhomogeneous Diffusivity function can be used to derive a noise reduction algorithm - anisotropic diffusion. This approach is then extended to the case of fractional anisotropic diffusion which is the central theme of this paper representing an original contribution to the field.

A case study is given which compares the use of applying...
anisotropic with fractional anisotropic diffusion algorithms for the problem of reducing noise in a Magnetic Resonance Imaging (MRI). Noise reduction in MRI is of particular importance with regard to diffusion MRI when the signal-to-noise ratio is relatively low. This is due to the relatively short duration that is required to obtain Diffusion Weighted Images (DWI) through the application of diffusion sensitizing gradients which induces signal attenuation. Multiple image averaging is often used to counteract this effect but only at the expense of longer acquisition times. It is shown that a model based on fractional anisotropic diffusion can provide a faster solution to noise reduction, given the iterative nature of the algorithms used.

II. THE DIFFUSION EQUATION

Let \( p(x) \) denote the Probability Density Function (PDF) associated with the position in a one-dimensional space \( x \) where a particle can exist as a result of a ‘random walk’ generated by a sequence of ‘elastic scattering’ processes (with other like particles). Also, assume that the random walk takes place over a time scale where the ‘environment’ does not change. Suppose we consider an infinite concentration of such particles at a time \( t = 0 \) to be located at the origin \( x = 0 \) and described by a perfect spatial impulse, i.e., a delta function \( \delta(x) \). Then the characteristic Impulse Response Function of the ‘random walk system’ at \( t = \tau \) is given by

\[
u(x, \tau) = \delta(x) \odot_x p(x) = p(x)
\]

where \( \odot_x \) denotes the correlation integral over \( x \). Thus, if \( u(x, t) \) denotes a macroscopic field at a time \( t \) such as the concentration of a canonical assemble of particles all undergoing the same random walk process, then the field at \( t + \tau \) will be given by

\[
u(x, t + \tau) = u(x, t) \odot_x p(x)
\]

Here, the field \( u \) represents the concentration of particles per unit length but in a three-dimensional space, the field \( u(r, t + \tau) \) (the concentration of particles per unit volume) will be given by

\[
u(r, t + \tau) = u(r, t) \odot_r p(r)
\]

where \( r \) denotes the three dimensional spatial vector.

Equation (1) represents the ‘master equation’ for the classical diffusion problem. The derivation of the classical diffusion equation

\[
\frac{\partial}{\partial t} u(r, t) = D \nabla^2 u(r, t)
\]

where \( \nabla^2 \) is the Laplacian operator given by

\[
\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}
\]

and \( D \) is the diffusivity is given in Appendix I and relies on the application of a Taylor series to approximate the field \( u(r, t + \tau) \). This is the basis for the original derivation of the diffusion equation by Albert Einstein [1], a derivation that is independent of the functional form of the PDF \( p(r) \) relying only upon the normalisation condition

\[
\int_{-\infty}^{\infty} p(r) dr = 1
\]

However, there is another way of deriving equation (2) from equation (1) that is informative and relies on an application of the convolution theorem. If \( p(r) \) is symmetric, then the correlation and convolution integrals become equivalent and we can write equation (1) as

\[
u(r, t + \tau) = u(r, t) \odot_r p(r), \quad p(r) = p(-r)
\]

where \( \odot_r \) denotes the convolution integral over \( r \). Thus, using the convolution theorem, we have

\[
U(k, t + \tau) = U(k, t) P(k)
\]

where \( U \) and \( P \) are the Fourier transforms of \( u \) and \( p \) given by

\[
U(k, t + \tau) = \int_{-\infty}^{\infty} u(r, t + \tau) \exp(-ik \cdot r) dr
\]

and

\[
P(k) = \int_{-\infty}^{\infty} p(r) \exp(-ik \cdot r) dr
\]

respectively, and where \( k \) is the three-dimensional spatial frequency vector. Suppose we consider a PDF \( p(r) \) that is normally or Gaussian distributed, i.e.

\[
p(r) = \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{3}{2}} \exp[-(r^2/2\sigma^2)], \quad \int_{-\infty}^{\infty} p(r) dr = 1
\]

where \( \sigma \equiv |r| \) and \( \sigma \) is the Standard Deviation. Then the Characteristic Function \( P(k) \) is also a Gaussian function given by

\[
P(k) = \exp(-\sigma^2 k^2/2)
\]

where \( k = |k| \) and since

\[
P(k) \sim 1 - \frac{\sigma^2 k^2}{2}, \quad \sigma \to 0
\]

we can then write

\[
\frac{U(k, t + \tau) - U(k, t)}{\tau} = -\frac{\sigma^2}{2\tau} k^2 U(k, t)
\]

But

\[
\int_{-\infty}^{\infty} \nabla^2 u(r, t) \exp(-ik \cdot r) dr = -k^2 U(k, t)
\]

so that as \( \tau \to 0 \) we obtain the equation

\[
\frac{\partial}{\partial t} u(r, t) = D \nabla^2 u(r, t)
\]

where \( D = \sigma^2/2\tau \). This approach to deriving the diffusion equation relies on specifying the characteristic function \( P(k) \) and upon the conditions that both \( \sigma \) and \( \tau \) approach zero, thereby allowing \( D \) to be of arbitrary value. This is the basis for the approach considered in Section VII with regard to a derivation of the anomalous or fractional diffusion equation.
III. OPTICAL DIFFUSION

Diffusion processes are not usually considered to be compatible with optics which, at a fundamental level, considers the propagation and scattering of waves with dielectric materials. In the scalar wave theory of light, the underlying equation is the wave equation

\[ \nabla^2 u(r, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} u(r, t) = 0 \]  

(3)

where \( u \) is the scalar wavefield with measurable optical properties given by the intensity \( | u |^2 \) and \( c \) is the speed of light which may vary according to the dielectric properties of a material. However, the scalar wave theory of optics considers single interactions of light waves with dielectrics under the Born or Kirchhoff approximations depending on whether volume or surface scattering effects are considered, respectively. In either case, multiple scattering effects are ignored under the ‘weak’ scattering approximation. However, if light is taken to undergo multiple interactions with a complex of scatterers, then the effect of strong multiple scattering becomes a principal issue. Moreover, if we think in terms of a single ray of light scattering from one scatterer to the next with a random direction and path length, then an analogy can be made with the concept of a particle undertaking a random walk. In this context, we can consider the intensity of a light wavefield \( I \), taken to be a complex of light rays, to undergo the same process of diffusion as an ensemble of particles with a concentration \( u \). This supposes that, under strong scattering condition, the wave equation (3) is replaced with the diffusion equation (1). How is this possible? In other words, how is it possible to transform the wave equation to the diffusion equation in a systematic way under certain self-consistent conditions?

With regard to equation (3), let

\[ u(x, y, z, t) = \phi(x, y, z, t) \exp(i\omega t) \]

where it is assumed that field \( \phi \) varies significantly slowly in time compared with \( \exp(i\omega t) \) and note that

\[ u^*(x, y, z, t) = \phi^*(x, y, z, t) \exp(-i\omega t) \]

is also a solution to the wave equation. Differentiating

\[ \nabla^2 u = \exp(i\omega t) \nabla^2 \phi, \]

and

\[ \frac{\partial^2}{\partial t^2} u = \exp(i\omega t) \left( \frac{\partial^2}{\partial t^2} \phi + 2i\omega \frac{\partial \phi}{\partial t} - \omega^2 \phi \right) \]

\[ \simeq \exp(i\omega t) \left( 2i\omega \frac{\partial \phi}{\partial t} - \omega^2 \phi \right) \]

when

\[ \left| \frac{\partial^2 \phi}{\partial t^2} \right| < < 2\omega \left| \frac{\partial \phi}{\partial t} \right| \]

Under this condition, the wave equation reduces to

\[ (\nabla^2 + k^2)\phi = \frac{2ik \frac{\partial \phi}{\partial t}}{c} \]

where \( k = \omega/c \). However, since \( u^* \) is also a solution,

\[ (\nabla^2 + k^2)\phi^* = -\frac{2ik \frac{\partial \phi^*}{\partial t}}{c} \]

and thus,

\[ \phi^* \nabla^2 \phi - \phi \nabla^2 \phi^* = \frac{2ik}{c} \left( \phi^* \frac{\partial \phi}{\partial t} + \phi \frac{\partial \phi^*}{\partial t} \right) \]

which can be written in the form

\[ \nabla^2 I - 2\nabla \cdot (\phi \nabla \phi^*) = \frac{2ik}{c} \frac{\partial I}{\partial t} \]

where \( I = \phi \phi^* \). Let \( \phi \) be given by

\[ \phi(r, t) = A(r, t) \exp(ik \hat{n} \cdot \mathbf{r}) \]

where \( \hat{n} \) is a unit vector and \( A \) is the amplitude function. Differentiating, and noting that \( I = A^2 \), we obtain

\[ \hat{n} \cdot \nabla A = \frac{2 \frac{\partial A}{\partial t}}{c} \]

or

\[ \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) A(x, y, z, t) = \frac{2}{c} \frac{\partial}{\partial t} A(x, y, z, t) \]

which is the unconditional continuity equation for the amplitude \( A \) of a wavefield

\[ u(r, t) = A(r, t) \exp[i(k \hat{n} \cdot \mathbf{r} + \omega t)] \]

where \( A \) varies slowly with time. The equation

\[ \nabla^2 I - 2\nabla \cdot (\phi \nabla \phi^*) = \frac{2ik}{c} \frac{\partial I}{\partial t} \]

is valid for \( k = k_0 - ik \) (i.e. \( \omega = \omega_0 - ikc \)) and so, by equating the real and imaginary parts, we have

\[ D\nabla^2 I + 2\text{Re}[\nabla \cdot (\phi \nabla \phi^*)] = \frac{\partial I}{\partial t} \]

and

\[ \text{Im}[\nabla \cdot (\phi \nabla \phi^*)] = -\frac{k_0}{c} \frac{\partial I}{\partial t} \]

respectively where \( D = c/2\kappa \), so that under the condition

\[ \text{Re}[\nabla \cdot (\phi \nabla \phi^*)] = 0 \]

we obtain

\[ D\nabla^2 I = \frac{\partial I}{\partial t} \]

This is the diffusion equation for the intensity of light \( I \). The condition required to obtain this result can be justified by applying a boundary condition on the surface \( S \) of a volume \( V \) over which the equation is taken to conform. Using the divergence theorem

\[ \text{Re} \int_V \nabla \cdot (\phi \nabla \phi^*) d^3 r = \text{Re} \int_S \phi \nabla \phi^* \cdot \hat{n} d^2 r \]

\[ = \int_S (\phi_r \nabla \phi_r + \phi_i \nabla \phi_i) \cdot \hat{n} d^2 r \]

Now, if

\[ \phi_r(r, t) \nabla \phi_r(r, t) = -\phi_i(r, t) \nabla \phi_i(r, t), \quad r \in S \]

then the surface integral is zero and

\[ D\nabla^2 I(r, t) = \frac{\partial}{\partial t} I(r, t), \quad r \in V \]
This boundary condition can be written as

$$\frac{\nabla \phi}{\nabla \phi_0} = -\tan \theta$$

where $\theta$ is the phase of the field $\phi$ which implies that the amplitude $A$ of $\phi$ is constant on the boundary (i.e. $A(r,t) = A_0$, $r \in S, \forall t$), since

$$\frac{\nabla A_0 \cos \theta(r,t)}{\nabla A_0 \sin \theta(r,t)} = -\frac{A_0 \sin \theta(r,t) \nabla \theta(r,t)}{A_0 \cos \theta(r,t) \nabla \theta(r,t)}$$

$$= -\tan \theta(r,t), \quad r \in S$$

IV. DIFFUSION BASED OPTICAL IMAGING

Suppose we record the intensity $I$ of a light field in the $xy$-plane for a fixed value of $z$. Then for $z = z_0$ say,

$$I(x,y,t) \equiv I(x,y,z_0,t)$$

so that

$$\frac{\partial}{\partial t} I(x,y,t) = D \nabla^2 I(x,y,t)$$

(4)

Let this two-dimensional diffusion equation be subject to the initial condition

$$I(x,y,0) = I_0(x,y)$$

Then, at any time $t > 0$, it can be assumed that light diffusion is responsible for blurring the image $I_0$ and that as time increases, the image becomes progressively more blurred. For the infinite domain, the Green’s function solution to equation (4) is [2]

$$I(x,y,t) = \frac{1}{4\pi Dt} \exp \left[ -\frac{(x^2 + y^2)^2}{4Dt} \right] \otimes_{x,y} I_0(x,y)$$

(5)

where $\otimes_{x,y}$ denotes the convolution integral over $(x,y)$, respectively. This result can, for example, be used to model the diffusion of light through an optical diffuser. An example of such an effect is given in Figure 1 which shows a light source (the ceiling light of a steam room) imaged through air and then through steam and a simulation of the effect based on equation (5) using a Gaussian blurring filter with a 34 pixel radius, i.e. a Gaussian lowpass filter with a variance of 34 pixels or $Dt = 17$ pixels.

Steam effects light by scattering it a large number of times through the complex of small water droplets from which (low temperature) steam is composed. The high degree of multiple scattering that takes place allows us to model the transmission of light through steam in terms of a ‘diffusive’ rather than a ‘propagative’ process. The initial condition $I_0$ denotes the initial image which is, in effect, and with regard to Figure 1, the image of the light source obtained in air. As observed in Figure 1, the details associated with the light source are blurred through the convolution of the object function $I_0$ with the Gaussian Point Spread Function, a function that is characteristic of optical diffusion in general.

A. Inverse Solution

Developing inverse solutions in imaging are fundamental to a wide range of image based technologies [3]. Given the diffusion based model for an image compounded in equation (4), the associated inverse problem is to find $I_0$ from $I$ at some time $t > 0$. Consider the case in which we record the diffused image $I$ at a time $t = T$. Taylor expanding $I$ at $t = 0$ we can write

$$I(x,y,0) = I(x,y,T) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} T^n \left[ \frac{\partial^n}{\partial t^n} I(x,y,t) \right]_{t=T}$$

We note that, from equation (4), that

$$\frac{\partial^2 I}{\partial t^2} = D \nabla^2 \frac{\partial I}{\partial t} = D^2 \nabla^4 I$$

$$\frac{\partial^3 I}{\partial t^3} = D \nabla^2 \frac{\partial^2 I}{\partial t^2} = D^3 \nabla^6 I$$

etc. so that, in general, we can write

$$\left[ \frac{\partial^n}{\partial t^n} I(x,y,t) \right]_{t=T} = D^n \nabla^{2n} I(x,y,T)$$

Substituting this result into the series for $I(x,y,0) = I_0(x,y)$ given above, we get

$$I_0(x,y) = I(x,y,T) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} (DT)^n \nabla^{2n} I(x,y,T)$$

so that, for $DT = 1$,

$$I_0(x,y) = I(x,y) - \nabla^2 I(x,y) + \frac{1}{2!} \nabla^4 I(x,y) - \frac{1}{3!} \nabla^6 I(x,y) + ...$$

and, for $DT \ll 1$,

$$I_0(x,y) = I(x,y) - \frac{\sigma^2}{2} \nabla^2 I(x,y)$$

(6)

where $\sigma = \sqrt{2DT}$ is the Standard Deviation.

B. Digital Filtering

Consider equation (6) in terms of a discrete system where

$$I(x,y) \rightarrow I(i,j) \equiv I_{ij}$$

Applying a centre differencing scheme,

$$\nabla^2 I(x,y) \rightarrow I(i+1,j) + I(i-1,j) + I(i,j+1) + I(i,j-1) - 4I_{ij}$$

(7)

and hence, in discrete form, equation (6) becomes

$$I^0_{ij} = I_{ij} - \frac{\sigma^2}{2} \nabla^2 I_{ij}$$
The application of the filter given by equation (6) for the Gaussian blurred image (above) and a de-blurred image (below) after applying the filter given in Figure 2. An example of the application of this filter is given in Figure 2.

Given the simplicity of the process (i.e. application of a linear operation. Applying this operation to a digital image $I_{ij}$ is the same convolving the image with the two-dimensional array (the Finite Impulse Response of FIR filter)

$$
\begin{pmatrix}
0 & 1 & 0 \\
1 & -4 & 1 \\
0 & 1 & 0
\end{pmatrix}
$$

Hence, computing $I^0_{ij}$ is equivalent to convolving $I_{ij}$ with the FIR filter

$$\frac{\sigma^2}{2} \begin{pmatrix}
0 & -1 & 0 \\
-1 & 4 + 2\sigma^{-2} & -1 \\
0 & -1 & 0
\end{pmatrix}$$

An example of the application of this filter is given in Figure 2. Given the simplicity of the process (i.e. application of a $3 \times 3$ FIR filter), the method provides an effective image de-blurring technique if and only if the degradation of the image conforms to a light diffusion (strong scattering) model and that the diffusion is relatively weak, i.e. $DT << 1$.

V. ANISOTROPIC DIFFUSION

Equation (4) and the general Green's function solution given by equation (5) assumes that the Diffusivity $D$ is isotropic over the (infinite) spatial domain. As shown in Appendix I, this requires that the diffusion tensor $D$ is diagonally symmetric and that the Diffusivity is homogeneous. With regard to using the diffusion equation (4) for image analysis, the isotropic and homogeneity conditions limit the applications to the stationary case. Anisotropic diffusion is concerned when the Diffusivity is non-isotropic and inhomogeneous. In the latter case, we are interested in an application of the model

$$\frac{\partial}{\partial t} I(x, y, t) = \nabla \cdot [D(x, y)\nabla I(x, y, t)]$$

or, after expanding the right hand side,

$$\frac{\partial}{\partial t} I(x, y, t) = \nabla D(x, y) \cdot \nabla I(x, y, t) + D(x, y)\nabla^2 I(x, y, t)$$

If, in addition, it can be assumed that the gradient $\nabla D$ is 'weak', then, under the condition

$$\|\nabla D(x, y) \cdot \nabla I(x, y, t)\| << \|D(x, y)\nabla^2 I(x, y, t)\|$$

we can consider the inhomogeneous diffusion equation

$$\frac{\partial}{\partial t} I(x, y, t) = D(x, y)\nabla^2 I(x, y, t)$$

(8)

While it is possible to consider a Green's function solution to equation (8), in the context of developing a numerical algorithm for processing a digital image, we can adopt a simple iterative solution by forward differencing in time and centre differencing in space. This approach yields an iterative digital filter of the form (for $k = 0, 1, 2, ..., N$)

$$I^{k+1}_{ij} = I^k_{ij} + \Delta D_{ij} \begin{pmatrix}
0 & 1 & 0 \\
1 & -4 & 1 \\
0 & 1 & 0
\end{pmatrix} \otimes_{i,j} I^k_{ij}$$

(9)

where $\otimes_{i,j}$ denotes the two-dimensional convolution sum over elements $(i, j)$, $\Delta$ is the time step and $I^0_{ij}$ is the initial condition.

In the context of processing a digital image $I^0_{ij}$, application of equation (9) depends critically upon the generation the array $D_{ij}$ from the image $I^0_{ij}$. With regard to using equation (9) for noise reduction, the value of $D_{ij}$ at any pixel location $(i, j)$ can be used to control the degree of diffusion, i.e. the degree of blurring in the locality of $(i, j)$. For correlated features in an image that are edge dominant and are taken to be relatively noise free, in terms of their high pixel value contrast with regard to low pixel value background noise, we require the degree of diffusion to decrease. This is accomplished by reducing the value of $D_{ij}$ in regions of an image that are edge dominant and increasing the value of $D_{ij}$ in those regions that are background noise dominant. In the absence of any a priori information on the Diffusivity $D_{ij}$, a method of achieving this is to apply an edge detector to the image $I^0_{ij}$ to obtain an output $E_{ij}$, say, and compute

$$D_{ij} = 1 - E_{ij}$$

where it is noted that $D_{ij} \geq 0\forall(i, j)$ and $E_{ij} \geq 0\forall(i, j)$. There are a range of edge detection filters that can be applied in this case.
VI. EDGE DETECTION

Edge detection is a method of segmenting an image into regions of discontinuity. In other words, it allows the user to observe those features of an image where there is a more or less abrupt change in grey level or texture - indicating the end of one region in the image and the beginning of another. Like other methods of image analysis, edge detection is sensitive to noise. For this reason, detected edges can occur in places where the transition between regions is not abrupt enough or else edges can be detected in regions of an image where the texture is uniform.

Edge detection makes use of differential operators to detect changes in the gradients of the grey levels. It is divided into two main categories: (i) first order edge detection; (ii) second order edge detection. First order edge detection is based on the use of a first order derivative whereas second order edge detection is based on the use of a second order derivative, in particular, the Laplacian $\nabla^2$. In this section we only consider an overview of first order edge detection as second order edge detection is not compatible with the application of generating a Diffusivity array $D_{ij}$.

To detect edges in an image we aim to highlight or emphasise changes in the value of the pixels. Derivative operations are ideally suited for this purpose. The first derivative, $\partial/\partial x$, shows extremes at an edge and the second derivative, $\partial^2/\partial x^2$, crosses the zero axis where the edge has its steepest gradient.

A. First Order Edge Detection

First order edge detection is based on computing the gradient of an image in $x$ and $y$ and observing the locations in the image where it changes abruptly. If the image is denoted by the function $f(x,y)$ then the basic idea is to compute

$$\nabla f(x,y) = \mathbf{x} \frac{\partial}{\partial x} f(x,y) + \mathbf{y} \frac{\partial}{\partial y} f(x,y)$$

and then display the gradient magnitude defined by

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

A threshold can be applied so that only a fraction of the gradients present in the image are retained. A binary image composed of 1’s and 0’s is then obtained which is a display of all the points in the image where its gradient is larger than the value of the threshold that is applied.

B. Digital Gradients

In many ways the different gradient methods used for edge detection result from attempts to find digital approximations to $\nabla f$. The approximations available are compounded in a class of operators known as digital gradients. A continuous function $f(x)$ can be expanded about a point $x = \Delta x$, say, as a Taylor series

$$f(x + \Delta x) = f(x) + \Delta x \frac{df}{dx}(x) + \frac{(\Delta x)^2}{2!} \frac{d^2 f}{dx^2}(x) + ...$$

If we neglect all the terms which occur after the second term, then we obtain an approximation for the derivative of $f$ at $\Delta x$, given by the difference equation

$$\frac{d}{dx} f(x) = \frac{f(x + \Delta x) - f(\Delta x)}{\Delta x}$$

With partial derivatives we have

$$\frac{\partial}{\partial x} f(x,y) = \frac{f(x + \Delta x,y) - f(x,y)}{\Delta x}$$

and

$$\frac{\partial}{\partial y} f(x,y) = \frac{f(x,y + \Delta y) - f(x,y)}{\Delta y}$$

Thus, with a digital image $f_{ij}$, we can replace the partial derivatives with the differences

$$D_x f_{ij} = f_{i(j+1)} - f_{ij}$$

and

$$D_y f_{ij} = f_{(i+1)j} - f_{ij}$$

These operations are equivalent to convolving $f_{ij}$ with the kernel $(-1,1)$ in the $x$-direction to give $D_x f_{ij}$ and convolving $f_{ij}$ with $(-1,1)$ in the $y$-direction to give $D_y f_{ij}$. The kernels or FIR filters which are used to convolve a digital image in this way are called masks. They are shift invariant operators which allow us to write the former difference equations in the form

$$D_x f_{ij} = D_x \otimes f_{ij},$$

$$D_y f_{ij} = D_y \otimes f_{ij}$$

where

$$D_x = \begin{pmatrix} -1 & 1 \end{pmatrix}$$

and

$$D_y = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

The equivalence of digital gradient operations and discrete convolutions (i.e. FIR filters) with certain types of masks is an important result in the theory and practice of edge detection.

The differencing scheme derived above generates output values for the digital gradients centered at $(i + 1/2, j)$ and $(i, j + 1/2)$. This differencing is actually only one of a number of different arrangements that can be used. For example, to obtain digital gradients centered at $(i, j)$, we use the differencing scheme

$$D_x f_{ij} = \frac{1}{2} [f_{(i+1)j} - f_{(i-1)j}],$$

$$D_y f_{ij} = \frac{1}{2} [f_{ij} + 1] - f_{ij}$$
The angle of gradient $\theta$ digital gradients that operate on a larger pixel array.

Another approach is to consider operator usually require pre-processing to reduce the level of noise inherent in an image. Another approach is to consider digital gradients such as the Roberts gradient is based on approximating first order gradients using cross-differences, where the gradient magnitude is given by

$$G_{ij} = \sqrt{[f_{ij} - f_{(i+1)(j+1)}]^2 + [f_{(i+1)(j)} - f_{i(j+1)}]^2}$$

which is based on application of the masks

$$D_x = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and

$$D_y = \frac{1}{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

However, these masks operate on a relatively small array of pixels and are consequently relatively sensitive to noise. Practical applications of digital gradients such as the Roberts operator usually require pre-processing to reduce the level of noise inherent in an image. Another approach is to consider digital gradients that operate on a larger pixel array.

\[ \begin{align*}
\frac{\partial}{\partial x} f(x, y) &\approx \frac{f(x, y) - f(x + \Delta x, y)}{\Delta x} \\
\frac{\partial}{\partial x} f(x, y) &\approx \frac{f(x - \Delta x, y) - f(x, y)}{\Delta x}
\end{align*} \]

and, hence,

\[ \frac{\partial}{\partial x} f(x, y) \approx \frac{f(x - \Delta x, y) - f(x + \Delta x, y)}{2\Delta x} \]

In this case, the masks are given by

\[ D_x = \frac{1}{2} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \]

and

\[ D_y = \frac{1}{2} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \]

To detect edges independent of orientation, a gradient magnitude operator (a vector operator) is defined as

\[ G = \sqrt{D_x^T D_x + D_y^2 D_y} \]

To simplify the computation (of the square root) an alternative is sometimes used, namely

\[ G = |D_x| + |D_y| \]

The angle of gradient $\theta(G)$ is also a useful quantity and is given by

\[ \theta(G) = \tan^{-1} \left( \frac{D_y}{D_x} \right) \]

There are a number of variants on this theme, that give different emphasis to detecting different types of edges which are discussed below.

\[ \begin{align*}
D_x &= \frac{1}{8} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \\
D_y &= \frac{1}{8} \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}
\]

An example of the output produced by this detector is given in Figure 3.

\[ \begin{align*}
D_x &= \frac{1}{6} \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \\
D_y &= \frac{1}{6} \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}
\]

\[ \begin{align*}
0^\circ : \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \\
45^\circ : \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}
\]

D. The Sobel Edge Detector

The Sobel edge detector is an extension which includes a degree of smoothing to reduce automatically certain artifacts caused by noise. The larger the filter array the more noise reduction will occur with fewer edges being detected, but as the filter becomes too large useful edges may not be detected. The Sobel filter is based on the following digital gradients:

\[ \begin{align*}
G_x &= \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{pmatrix}, \\
G_y &= \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}
\]

E. The Prewitt Edge Detector

The Prewitt edge detector is based on the following digital gradients:

\[ \begin{align*}
G_x &= \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{pmatrix}, \\
G_y &= \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}
\]

F. The Compass Edge Detector

This detector is designed to avoid the sensitivity that a filter has for specific orientation of an edge. A set of different filters is employed detecting specific angles, for the gradient $G_i$. The resulting gradient is then computed as $G = \max\{G_i : i = 1 \text{ to } n\}$. Various kernels can be used. As an example, the first two compass Sobel filters are:

\[ \begin{align*}
0^\circ : \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \\
45^\circ : \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}
\]
Any other filter can be used to create a set of compass filters. A few common ones include the Prewitt filter,

\[
0^\circ : \begin{pmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{pmatrix}, \quad 45^\circ : \begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{pmatrix},
\]

the Kirsch filter

\[
0^\circ : \begin{pmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{pmatrix}, \quad 45^\circ : \begin{pmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{pmatrix},
\]

and the Robinson filter

\[
0^\circ : \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \quad 45^\circ : \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}
\]

G. Nine Dimensional Operators

Given a set of 3 \( \times \) 3 filters, a complete basis is formed with nine orthogonal filters. There is obviously an infinite number of sets to choose from. For completeness, it is worth mentioning at this point one example, namely, a set containing four filters for edge detection \((w_1-w_4)\), four filters for line detection \((w_5-w_8)\) and the last filter, \(w_9\), which is a simple averaging filter as defined below.

\[
\begin{pmatrix}
1 & \sqrt{2} & 1 \\
0 & 0 & 0 \\
-1 & -\sqrt{2} & -1
\end{pmatrix}

\begin{pmatrix}
w_1 \\
w_2 \\
w_3
\end{pmatrix}

\begin{pmatrix}
0 & -1 & -\sqrt{2} \\
1 & 0 & -1 \\
\sqrt{2} & 1 & 0
\end{pmatrix}

\begin{pmatrix}
w_4 \\
w_5 \\
w_6
\end{pmatrix}

\begin{pmatrix}
0 & 1 & 0 \\
-1 & 0 & -1 \\
0 & 1 & 0
\end{pmatrix}

\begin{pmatrix}
w_7 \\
w_8 \\
w_9
\end{pmatrix}

H. The Canny Edge Detector

In developing the filters discussed above, two problems are encountered. As the data elements are discrete, it is not always obvious how to calculate a gradient value and, in the presence of noise, many spurious edges can become apparent. Increasing the size of the filter can alleviate these anomalies. The Canny edge detector is designed to combat some of these problems and consists of three main stages:

1) Gaussian blur the image to reduce the amount of noise and remove speckles within the image. It is important to remove the very high frequency components that exceed those associated with the gradient filter used, otherwise, these can cause false edges to be detected.

2) Gradient detect using one of the above filters, creating two images, one containing the gradient magnitudes \(G\), and another containing the orientation \(\theta(G)\). The most common implementations use a simple symmetric discrete first order derivative.

3) Threshold the gradient magnitudes above a certain minimum threshold value so that only major edges are detected. As well as this minimum low threshold value, a high threshold value is also specified. On any connected line, at least one of the edge points has to exceed this high value. This removes small or insignificant line segments.

By controlling the standard deviation of the Gaussian blurring operation, and the high and low threshold values, most general edges can be detected. If it is known \textit{a priori} what kind of edge is to be detected and the kind of noise that is present in the image, then an alternative filter can be applied instead of the Gaussian filter.

VII. The Fractional Diffusion Equation for Lévy Distributed Processes

In Section II, the classical diffusion equation was derived from equation (1) for a Gaussian PDF \(p(r)\). This is consistent with any system that exhibits random processes that are Gaussian distributed. We now consider the derivation of the fractional diffusion equation for a Lévy PDF. Lévy processes are random walks whose distribution has infinite moments. The statistics of (conventional) physical systems are usually concerned with stochastic fields that have PDFs where (at least) the first two moments (the mean and variance) are well defined and finite. Lévy statistics is concerned with statistical systems where all the moments (starting with the mean) are infinite. Many distributions exist where the mean and variance are finite but are not representative of the process, e.g. the tail of the distribution is significant, where rare but extreme events occur. These distributions include Lévy distributions \([4],[5]\). Lévy’s original approach to deriving such distributions is based on the following question: Under what circumstances does the distribution associated with a random walk of a few steps look the same as the distribution after many steps (except for scaling)? This question is effectively the same as asking under what circumstances do we obtain a random walk that is statistically self-affine. The characteristic function \(P(k)\) of such a distribution \(p(r)\) was first shown by Lévy to be given by (for symmetric distributions only)

\[
P(k) = \exp(-a | k |^\gamma), \quad 0 < \gamma \leq 2 \quad (10)
\]

where \(a\) is a constant and \(\gamma\) is the Lévy index. For \(\gamma \geq 2\), the second moment of the Lévy distribution exists and the sums of
large numbers of independent trials are Gaussian distributed. For example, if the result were a random walk with a step length distribution governed by $p(r)$, $\gamma \geq 2$, then the result would be normal (Gaussian) diffusion, i.e. a Brownian random walk process. For $\gamma < 2$ the second moment of this PDF (the mean square), diverges and the characteristic scale of the walk is lost. For values of $\gamma$ between 0 and 2, Lévy’s characteristic function corresponds to a PDF of the form [6]

$$p(r) \sim \frac{1}{|r|^{1+\gamma}}, \quad x \to \infty$$

**A. Derivation of the Fractional Diffusion Equation**

Lévy processes are consistent with a fractional diffusion equation as defined in [7] and we now show [7]. Consider the evolution equation for a random walk process to be given by equation (1) which, in Fourier space, is

$$U(k, t + \tau) = U(k, t) P(k)$$

for symmetric $p(r)$. From equation (10),

$$P(k) = 1 - a |k|^\gamma, \quad a \to 0$$

so that we can write

$$\frac{U(k, t + \tau) - U(k, t)}{\tau} \sim -\frac{a}{\tau} |k|^\gamma U(k, t)$$

which for $\tau \to 0$ gives the fractional diffusion equation

$$\frac{\partial}{\partial t} u(r, t) = D \nabla^\gamma u(r, t), \quad \gamma \in (0, 2) \quad (11)$$

where $D = a/\tau$ and we have used the Reisz definition of a fractional Laplacian, i.e.

$$\nabla^\gamma u(r, t) = -\frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} |k|^\gamma U(k, t) \exp(i k \cdot r) \, d^3 k \quad (12)$$

This derivation of the fractional diffusion equation reveals its physical origin in terms of Lévy processes as defined in terms of the characteristics function given by equation (10) for $a \to 0$. The approach can be used to derive a fractional partial differential equation for arbitrary PDFs. Applying the correlation theorem to equation (1) we note that

$$U(k, t + \tau) = U(k, t) P^*(k)$$

where the PDF $p(r)$ and the Characteristic Function $P(k)$ may be asymmetric. Then

$$\frac{U(k, t + \tau) - U(k, t)}{\tau} = \frac{1}{\tau} U(k, t) [P^*(k) - 1]$$

so that as $\tau \to 0$ we obtained a generalised anomalous diffusion equation given by

$$\frac{\partial}{\partial t} u(r, t) = \frac{1}{\tau} [u(r, t) \otimes \tau p(r) - u(r, t)]$$

**B. Fractional Anisotropic Diffusion of an Image**

Given equation (11), we consider an equivalent equation to equation (8) for the fractional diffusion of an image $I$, i.e.

$$\frac{\partial}{\partial t} I(x, y, t) = D(x, y) \nabla^\gamma I(x, y, t) \quad (13)$$

However, application of a digital Laplacian as used in equation (9) has no direct equivalent with regard to developing a numerical solution to equation (13). This is because the definition of a fractional Laplacian is based on equation (12). Instead, we consider a modification of equation (13) that includes the Laplacian $\nabla^2$, i.e.

$$I(x, y, t) = D(x, y) \nabla^2 [\nabla^{-2} I(x, y, t)]$$

where

$$\nabla^{-2} I(x, y, t) = \frac{1}{(2\pi)^2}$$

$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \tilde{I}(k_x, k_y, t) \exp(ik_x x) \exp(ik_y y) dk_x dk_y \right)$$

and

$$\tilde{I}(k_x, k_y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y, t) \exp(-ik_x x) \exp(-ik_y y) dx dy$$

In practice, $\nabla^{-2} I$ is computed using a Discrete Fourier Transform to output the digital equivalent operation which we denote as $\nabla^{-2} I_{ij}$. Thus, given equation (9), we consider the following iterative filter: For $k = 0, 1, 2, \ldots, N$

$$I_{ij}^{k+1} = I_{ij}^k + \Delta D_{ij} \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \otimes_{ij} \nabla^{-2} I_{ij}^k \quad (14)$$

Equation (14) is the focus with regard to the principal contribution of this paper and example M-code for the algorithm is given in Appendix III. In the following section, a case study is given on the application this algorithm to the reduction of noise in Magnetic Resonance Imaging.

**VIII. Case Study: Noise Reduction in Magnetic Resonance Imaging**

Magnetic Resonance Imaging (MRI) is a relatively new, non-invasive imaging technique that is rapidly gaining popularity in the medical community. It measures the signal produced by the protons contained in water molecules throughout the body when they are subjected to a range of magnetic and electro-magnetic fields. Protons exhibit the quantum-mechanical property of ‘spin’, which, in a classical sense, may be envisaged as a rotation of the proton around its axis. This rotation provides microscopic nuclear magnetization for each proton whose direction, under normal conditions, is randomly orientated due to thermal fluctuations. However, under the influence of a large external magnetic field, a majority of the proton spin population is aligned parallel to the direction of the applied field, reaching an equilibrium state. Through the application of radio frequency electromagnetic waves at a given energy, it is possible to tilt the spins away from...
the equilibrium position where after they begin to precess in unison (resonance) about the axis of the external field. The oscillating magnetic field produced by the precessing protons is detectable using sensitive coils that are placed close to the surface of the patient. If magnetic field gradients are also used, it is possible to encode the spatial position in the received signal so that images can be constructed by post-processing the data.

One of the advantages of MRI is its ability to provide numerous forms of soft-tissue contrast imaging such as weighted [8], [9], dynamic contrast enhanced [10], perfusion and diffusion weighted imaging [10]. Furthermore, these images can be acquired in any plane using a range of spatial resolutions making the method invaluable for imaging the body, especially in the brain and in oncological applications.

Due to improvements in scanner hardware and imaging sequences [12], Diffusion Weighted MRI (DWI) is now starting to be used routinely in the clinic, as it provides excellent morphological contrast and also quantitative information in the form of an Apparent Diffusion Coefficient (ADC). DWI encodes the rate of diffusion of water molecules in the received image by applying large magnetic field gradients in the imaging sequence. These gradients cause signal loss that depends on both the strength and timings of gradients and also on the diffusion coefficient of the imaged tissue. The equation that governs the amount of signal loss, known as the Stejskal-Tanner relation, is given by [13]

\[
S(r, b) = S(r, 0) \exp[-bD(r)]
\]

where \(D(r)\) is the position dependent diffusion coefficient of the imaged tissue, \(b\) is a constant encapsulating the various parameters of the diffusion sensitizing gradients and \(S(r, b)\) is image signal at position \(r\) at a given \(b\)-value. It is clear from this model that by acquiring 2 or more images at different \(b\)-values, it is possible to estimate the values for \(S(r, 0)\) and \(D(r)\) at each pixel location, making DWI a quantitative technique. A major assumption behind this relationship, however, is that there is a single diffusion coefficient for the imaged material. As most tissues exhibit large degrees of heterogeneity, the term ADC [14] is often used to describe the calculated estimations of \(D\).

Another important application of DWI is providing high contrast images for detection of abnormalities. Tumours, for example, have been shown to have characteristically lower measured diffusion coefficients than background tissues' meaning that a signal at optimised acquisition \(b\)-values [15] is retained compared to healthy background [16]. This form of contrast, along with ADC maps, provides the clinician with an excellent tool for diagnosis and localization of disease [17].

One of the major drawbacks of DWI is that the application of diffusion sensitizing gradients always induces signal attenuation (although the degree of attenuation depends on the diffusion coefficient of the tissue) leading to a loss in the Signal-to-Noise Ratio (SNR) of images. This is often overcome by multiple image averaging at the same \(b\)-value. However, this is at the expense of longer acquisition times, e.g. \(\sim 30 – 40\) minutes for a whole body scan. Furthermore, this reduces the number of possible \(b\)-values that can be acquired leading to reduction in the accuracy of ADC measurements and inhibiting the exploration of newer imaging models such as intra-voxel incoherent motion imaging [18] or stretched exponential fitting [19]. Thus, if new methods can be developed that improve the SNR of MR images, then DWI can be implemented more effectively and over shorter acquisition times. In the following sections we study the effect of applying the anisotropic and then the fractional anisotropic diffusion algorithms for reducing noise in MR images.

A. Noise Reduction of MR Images using Anisotropic Diffusion

Figure 4 show a typical noisy MR image of the Brain. As with all noise reduction algorithms, the principal aim is to reduce the noise while preserving the resolution and fidelity of the image. In this respect, Figure 4 also shows the effect of applying a Gaussian lowpass filter (Gaussian blur) using a 2 pixel radius. This is typical of the effect of applying a lowpass filter where the noise is reduced at the expense of resolution.

![Fig. 4. MR image of the Brain before (above [20]) and after (below) application of a Gaussian lowpass filter with a 2 pixel radius.](image-url)
Figure 5 shows the effect of applying the anisotropic diffusion algorithm given in Appendix II for a Prewitt filter after 5, 10 and 20 iterations using a time step $\Delta = 0.1$.

![Fig. 5. MR image of the Brain before (top-left) and the results of applying the anisotropic diffusion algorithm given in Appendix II for a Prewitt edge detector after 5 (top-right), 10 (bottom-left) and 20 (bottom-right) iterations.](image)

B. Noise Reduction of MR Images using Fractional Anisotropic Diffusion

To illustrate the effect of using the fractional anisotropic diffusion method to reduce noise, we consider the same MR image given in Figure 4. Figure 5 show the outputs of the algorithm provided in Appendix III after 5, 10 and 15 iterations using a time step $\Delta = 0.05$ and a Lévy index $\gamma = 1.98$.

![Fig. 6. MR image of the Brain before (top-left) and the results of applying the fractional anisotropic diffusion algorithm given in Appendix III for a Prewitt edge detector after 5 (top-right), 10 (bottom-left) and 15 (bottom-right) iterations with $\Delta = 0.05$ and $\gamma = 1.98$.](image)

The value of $\gamma$ that is used in this case is critical, and must, in general, be close to 2 as given in the example shown in Figure 6. If $\gamma$ moves too far below 2, the lowpass filter $| k |^{-2(\gamma-1)/2}$ attenuates the high frequency components in the image too severely at each iteration. On the other hand, for values of $\gamma$ close to 2, the number of iterations required to de-noise the image is significantly less than in the application of the non-fractional anisotropic diffusion algorithm, noise reduction being optimal after only 5 iteration as shown in Figure 6.

IX. Conclusions

This paper has been concerned with extending the method of anisotropic diffusion for noise reduction in digital images to the fractional anisotropic diffusion case. The method developed is compounded in equation (14) which reduces to equation (9) as $\gamma \to 2$. This result models noise that is assumed to have been generated through non-Gaussian processes compounded in the symmetric Characteristic Function given by equation (10) for $\alpha \to 0$. In this sense, fractional diffusion is a generalisation of classical diffusion which includes long tail distributions associated with Lévy processes.

With regard to equation (14), the output image depends on the number of iterations $N$ used, the values of $\Delta$ and $\gamma$ and the edge detection filter used to compute $D_{ij}$. It is noted that only values of $\gamma$ close to 2 are suitable for application of equation (14), otherwise, the lowpass filter used to compute $\nabla_s^{\gamma/2}$ attenuates the high frequency components to such an extent that resolution in the output is lost.

Although the approach considered has applicability to digital imaging in general, the case study provided in Section VIII has focused on the application to diffusion weighted magnetic resonance imaging. This is an area in which the signal-to-noise ratio of the image is reduced due to signal attenuation leading to noisy images unless multiple image averaging is used at the expense of longer acquisition times. In turn, this reduces the ability to use diffusion weighted images for three-dimensional motion imaging. The noise reduction algorithm compounded in equation (14) represents a fast and effective way of overcoming this problem with regard to the production of images such as that given in Figure 7.

APPENDIX I

DERIVATION OF THE CLASSICAL DIFFUSION EQUATION

Let $\tau$ be a small interval of time in which a particle moves between $\lambda$ and $\lambda + d\lambda$ with probability $p(\lambda)$ where $\lambda = \sqrt{\lambda_x^2 + \lambda_y^2 + \lambda_z^2}$ and $\tau$ is small enough to assume that the movements of the particle are $\tau$-independent. If $u(\mathbf{r}, t)$ is the concentration (i.e. the number of particles per unit volume) then the concentration at time $t + \tau$ is given by

$$u(\mathbf{r}, t + \tau) = \int_{-\infty}^{\infty} u(\mathbf{r} + \lambda, t)p(\lambda)d\lambda$$

where $d^3\lambda \equiv d\lambda_x d\lambda_y d\lambda_z$. Since $\tau << 1$, we may approximate $u(\mathbf{r}, t + \tau)$ as

$$u(\mathbf{r}, t + \tau) = u(\mathbf{r}, t) + \tau \frac{\partial}{\partial t}u(\mathbf{r}, t)$$
and write $u(r + \lambda, t)$ in terms of the Taylor series

$$u(r + \lambda, t) = u + \lambda_x \frac{\partial u}{\partial x} + \lambda_y \frac{\partial u}{\partial y} + \lambda_z \frac{\partial u}{\partial z}$$

$$+ \frac{\lambda_x^2}{2!} \frac{\partial^2 u}{\partial x^2} + \frac{\lambda_y^2}{2!} \frac{\partial^2 u}{\partial y^2} + \frac{\lambda_z^2}{2!} \frac{\partial^2 u}{\partial z^2}$$

$$+ \lambda_x \lambda_y \frac{\partial^2 u}{\partial x \partial y} + \lambda_x \lambda_z \frac{\partial^2 u}{\partial x \partial z} + \lambda_y \lambda_z \frac{\partial^2 u}{\partial y \partial z} + \ldots$$

However, higher order terms can be neglected since, if $\tau << 1$, then the distance travelled, $\lambda$, must also be small. Equation (I.1) may then be written as

$$u + \frac{\tau}{\partial t} u = \int_{-\infty}^{\infty} u p(\lambda) d^3 \lambda$$

$$+ \int_{-\infty}^{\infty} \left( \lambda_x \frac{\partial u}{\partial x} + \lambda_y \frac{\partial u}{\partial y} + \lambda_z \frac{\partial u}{\partial z} \right) p(\lambda) d^3 \lambda$$

$$+ \int_{-\infty}^{\infty} \left( \frac{\lambda_x^2}{2!} \frac{\partial^2 u}{\partial x^2} + \frac{\lambda_y^2}{2!} \frac{\partial^2 u}{\partial y^2} + \frac{\lambda_z^2}{2!} \frac{\partial^2 u}{\partial z^2} \right) p(\lambda) d^3 \lambda$$

$$+ \frac{\lambda_x \lambda_y}{2!} \frac{\partial^2 u}{\partial x \partial y} + \lambda_x \lambda_z \frac{\partial^2 u}{\partial x \partial z} + \lambda_y \lambda_z \frac{\partial^2 u}{\partial y \partial z} \ldots$$

Assuming that $p(\lambda)$ is normalized we have

$$\int_{-\infty}^{\infty} p(\lambda) d^3 \lambda = 1$$

so that

$$\frac{\tau}{\partial t} u = \int_{-\infty}^{\infty} \frac{\lambda_x^2}{2} \frac{\partial^2 u}{\partial x^2} p(\lambda) d^3 \lambda + \int_{-\infty}^{\infty} \frac{\lambda_y^2}{2} \frac{\partial^2 u}{\partial y^2} p(\lambda) d^3 \lambda$$

$$+ \int_{-\infty}^{\infty} \frac{\lambda_z^2}{2} \frac{\partial^2 u}{\partial z^2} p(\lambda) d^3 \lambda + \int_{-\infty}^{\infty} \lambda_x \lambda_y \frac{\partial^2 u}{\partial x \partial y} p(\lambda) d^3 \lambda$$

$$+ \int_{-\infty}^{\infty} \lambda_x \lambda_z \frac{\partial^2 u}{\partial x \partial z} p(\lambda) d^3 \lambda + \int_{-\infty}^{\infty} \lambda_y \lambda_z \frac{\partial^2 u}{\partial y \partial z} p(\lambda) d^3 \lambda$$

which may be written as a matrix equation of the following form

$$\frac{\partial}{\partial t} u(r, t) = \nabla \cdot D \nabla u(r, t) + V \cdot \nabla u(r, t)$$

where $D$ is the diffusion tensor given by

$$D = \begin{pmatrix}
D_{xx} & D_{xy} & D_{xz} \\
D_{yx} & D_{yy} & D_{yz} \\
D_{zx} & D_{zy} & D_{zz}
\end{pmatrix}$$

where

$$D_{ij} = \int_{-\infty}^{\infty} \frac{\lambda_i \lambda_j}{2\tau} p(\lambda) d^3 \lambda = \frac{1}{2\tau} \langle \lambda_i \lambda_j \rangle$$

and $V$ is a flow vector which describes any drift velocity that the particle ensemble may have and is given by

$$V = \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix}$$

$$V_i = \int_{-\infty}^{\infty} \frac{\lambda_i}{\tau} p(\lambda) d^3 \lambda = \frac{1}{\tau} \langle \lambda_i \rangle$$

Note that as $\lambda_i \lambda_j = \lambda_j \lambda_i$, the diffusion tensor is diagonally symmetric (i.e. $D_{ij} = D_{ji}$). For isotropic diffusion where $\langle \lambda_i \lambda_j \rangle = 0$ for $i \neq j$ and $\langle \lambda_i \lambda_j \rangle = \langle \lambda^2 \rangle$ for $i = j$ and with no drift velocity so that $V = 0$, then

$$\frac{\partial}{\partial t} u(r, t) = \nabla \cdot \begin{pmatrix} 0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix} \nabla u(r, t)$$

$$= D \nabla^2 u(r, t)$$

where

$$D = \int_{-\infty}^{\infty} \frac{\lambda^2}{2\tau} p(\lambda) d^3 \lambda$$
**APPENDIX II**

**M-CODE FUNCTION FOR THE ANISOTROPIC DIFFUSION ALGORITHM**

```matlab
function AD(delta,nit)
%Function: Anisotropic Diffusion Algorithm for noise reduction
%Inputs:
%delta - Time step
%nit - Number of iterations
%Read image into array I
I=imread('name');
I=double(I); %Float image array
I=I./max(max(I)); %Normalise
%Compute image size
[n,m]=size(I);
%Compute mid-points of image for frequency space filtering
nn=1+n/2;
mm=1+m/2;
%Compute edge image
D=Prewitt(I);
%Compute Diffusivity D
D=1-D;
%Compute matrix for Laplacian operator
a(1,1)=0; a(1,2)=1; a(1,3)=0;
a(2,1)=1; a(2,2)=-4; a(2,3)=1;
a(3,1)=0; a(3,2)=1; a(3,3)=0;
%Start iterations
for k=1:nit
  %Transform to frequency space
  I=fftshift(fft2(I));
  for i=1:n
    for j=1:m
      %Compute filter
      Gx(1,3)=1; Gx(2,3)=1; Gx(3,3)=1;
      Gx=Gx/6;
      %Compute y-gradient matrix
      Gy(1,1)=-1;Gy(1,2)=-1;Gy(1,3)=-1;
      Gy(2,1)=0;Gy(2,2)=0;Gy(2,3)=0;
      Gy(3,1)=1;Gy(3,2)=1;Gy(3,3)=1;
      Gy=Gy/6;
      %Convolve input image with Gx
      Bx=conv2(A,Gx,'same');
      %Convolve input image with Gy
      By=conv2(A,Gy,'same');
      %Combine absolute value images
      B=abs(Bx)+abs(By);
      %Normalise
      B=B./max(max(B));
    end
  end
  imshow(I);%Show image
  pause;
end
%Output image
imwrite(I,'name','format');
end
```

**APPENDIX III**

**M-CODE FUNCTION FOR THE FRACTIONAL ANISOTROPIC DIFFUSION ALGORITHM**

```matlab
function FAD(gamma,delta,nit)
%Function: Fractional Anisotropic Diffusion Algorithm for Noise Reduction in images
%Inputs:
%gamma - Levy index < 2
%delta - Time step
%nit - Number of iterations
%Read image into array I
I=imread('name');
I=double(I); %Float image array
I=I./max(max(I)); %Normalise
%Compute image size
[n,m]=size(I);
%Compute mid-points of image for frequency space filtering
nn=1+n/2;
mm=1+m/2;
%Compute edge image
D=Prewitt(I);
%Compute Diffusivity D
D=1-D;
%Compute matrix for Laplacian operator
a(1,1)=0; a(1,2)=1; a(1,3)=0;
a(2,1)=1; a(2,2)=-4; a(2,3)=1;
a(3,1)=0; a(3,2)=1; a(3,3)=0;
%Start iterations
for k=1:nit
  %Transform to frequency space
  I=fftshift(fft2(I));
  for i=1:n
    for j=1:m
      %Compute filter
      Gx(1,3)=1; Gx(2,3)=1; Gx(3,3)=1;
      Gx=Gx/6;
      %Compute y-gradient matrix
      Gy(1,1)=-1;Gy(1,2)=-1;Gy(1,3)=-1;
      Gy(2,1)=0;Gy(2,2)=0;Gy(2,3)=0;
      Gy(3,1)=1;Gy(3,2)=1;Gy(3,3)=1;
      Gy=Gy/6;
      %Convolve input image with Gx
      Bx=conv2(A,Gx,'same');
      %Convolve input image with Gy
      By=conv2(A,Gy,'same');
      %Combine absolute value images
      B=abs(Bx)+abs(By);
      %Normalise
      B=B./max(max(B));
    end
  end
  imshow(I);%Show image
  pause;
end
%Output image
imwrite(I,'name','format');
end
```
\[ x=(i-\text{nn})^2; \quad y=(j-\text{mm})^2; \]
\[ \text{filter}=(x+y)^{((2-\text{gamma})/2)}; \]
\[ \text{if } \text{filter}>0 \]
\[ \text{I}(i,j)=\text{I}(i,j)/\text{filter}; \]
\[ \text{end} \]
\[ \text{end} \]
\[ \text{I}=\text{abs}(\text{ifft2}(\text{I})); \]
\[ \% \text{Convolve with Laplacian and} \]
\[ \% \text{compute the next image} \]
\[ \text{I}=\text{I}+\text{delta}*(\text{D}.*\text{conv2}(\text{I},\text{a},'\text{same'})); \]
\[ \text{I}=\text{I}/\text{max}([\text{max}(\text{I})]); \%	ext{Normalise} \]
\[ \text{imshow}(\text{I}); \%	ext{Show image} \]
\[ \text{end} \]
\[ \% \text{Output image} \]
\[ \text{imwrite}(\text{I},'\text{name}','\text{format'}); \]

\textbf{ACKNOWLEDGMENT}

The work reported in this paper is funded by the Science Foundation Ireland and Cancer Research UK. The authors are grateful for the support of Dublin Institute of Technology (School of Electrical Engineering Systems), the CRUK and EPSRC Cancer Imaging Centre in association with the MRC and Department of Health (England) grant C1060/A10334, and also NHS funding to the NIHR Biomedical Research Centre.

\textbf{REFERENCES}


\textbf{Jonathan Blackledge} graduated in physics from Imperial College in 1980. He gained a PhD in theoretical physics from London University in 1984 and was then appointed a Research Fellow of Physics at Kings College, London, from 1984 to 1988, specializing in inverse problems in electromagnetism and acoustics. During this period, he worked on a number of industrial research contracts undertaking theoretical and computational research into the applications of inverse scattering theory for the analysis of signals and images. In 1988, he joined the Applied Mathematics and Computing Group at Cranfield University as Lecturer and later, as Senior Lecturer and Head of Group where he promoted postgraduate teaching and research in applied and engineering mathematics in areas which included computer aided engineering, digital signal processing and computer graphics. In 1994, Jonathan Blackledge was appointed Professor of Applied Mathematics and Head of the Department of Mathematical Sciences at De Montfort University where he expanded the post-graduate and research portfolio of the Department and established the Institute of Simulation Sciences. From 2002-2008 he was appointed Visiting Professor of Information and Communications Technology in the Advanced Signal Processing Research Group, Department of Electronics and Electrical Engineering at Loughborough University, England (a group which he co-founded in 2003 as part of his appointment). In 2004 he was appointed Professor Extraordinaire of Computer Science in the Department of Computer Science at the University of the Western Cape, South Africa. He currently holds the prestigious Stokes Professorship in ICT under the Science Foundation Ireland programme based in the School of Electrical Engineering Systems, Dublin Institute of Technology and is Distinguished Professor at the Centre for Advanced Studies, Warsaw University of Technology, Poland.

\textbf{Matthew Blackledge} is a PhD research student at the Institute of Cancer Research in the United Kingdom. Having graduated with a degree in Physics from Imperial college, London in 2006, he moved on to pursue a career in medical imaging and completed a Masters degree in Medical physics with distinction from the university of Surrey in 2007. His main research areas include the acquisition and post-processing of diffusion weighted magnetic resonance images focusing on their application for the detection of brain tumours.