

Medical Image Processing: Mathematical Modelling and Numerical Resolution

Emanuele Schiavi, Juan Francisco Garamendi, and Adrián Martín

Abstract Medical image processing is an interdisciplinary research field attracting expertise from applied mathematics, computer sciences, engineering, statistics, physics, biology and medicine. In this context we shall present an introduction to basic techniques and concepts as well as more advanced methods to promote interests for further study and research in the field.

1 Introduction

Medical image processing is a growing field in medicine and mathematics which aims to improve the diagnostic power of some acquisition data modalities such as MRI, fMRI, PET, MEG, CT, etc. This leads to improved treatment control and therapies. In this work we shall consider some digital image processing related mathematical problems such as filtering, denoising and segmentation of digital images with a particular view to medical image processing and restoration. Our research is based at Fundación CIEN-Fundación Reina Sofía, Madrid, Spain

<http://www.fundacionreinasofia.es/ES/Paginas/home.aspx>

E. Schiavi (✉)

Departamento de Matemática Aplicada, Ciencia e Ingeniería de los Materiales y Tecnología Electrónica, Universidad Rey Juan Carlos, Madrid, Spain
e-mail: emanuele.schiavi@urjc.es

J.F. Garamendi

Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Barcelona, Spain

A. Martín

Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

where a interdisciplinary group of scientists coming from different areas and institutions is working on biomarkers for neurological diseases such as Alzheimer and Parkinson.

These lecture notes cover some basic aspects of the mathematical modelling but they also aim to introduce the reader to the most recent techniques and numerical algorithms. A very straightforward and applied introduction to the field can be found in [11] where basic, routinary algorithms are implemented. A more advanced introduction to the theoretical material we shall consider in this work is described in the book by Chan and Shen [5] where the mathematical foundations of modern image processing and low-level computer vision are presented, bridging contemporary mathematics with state-of-the-art methodologies in modern image processing. An interesting medical images processing overview can be found in <http://www.math.wisc.edu/~angenent/preprints/medicalBAMS.pdf> and a more general, geometric approach to PDE image processing is in the book by Osher [15].

2 Digital Image Processing

Digital image processing is a recent and challenging branch of applied mathematics which develops models and numerical algorithms for Filtering, Denoising, Deblurring, Edge-enhancing, Segmentation, Registration, Tracking, Impainting, Smoothing, Compression, Features Extraction and Pattern Recognition. The great improvement in computational power as well as the design of specific patient tailored acquisition modalities which took place in the last decade have motivated the implementation of advanced mathematical theories to the pre-processing analysis and the statistical post-processing interpretation of huge amounts of possibly multimodal patient data. In short, fast and accurate mathematical analysis are both possible and necessary paradigms, contrary to the past view where fast but very approximated results were looked for. In this section we shall briefly introduce the reader to the key steps of the mathematical analysis focusing on the models and results which made possible the evolution and implementation of numerical algorithms for the resolution of the PDE that appear in image processing and enhancement. We shall consider two basic approaches. In the first one we shall see how it is possible to filter an image using directly an evolution diffusion equation. Then we shall move to the variational approach in order to solve the energy minimization problems which arise when we try to solve the associated inverse problems and their Tikhonov regularization. This introduces the need for nonlinear operators which include the very famous Total Variation Model by Rudin, Osher and Fatemi (1987), see [17].

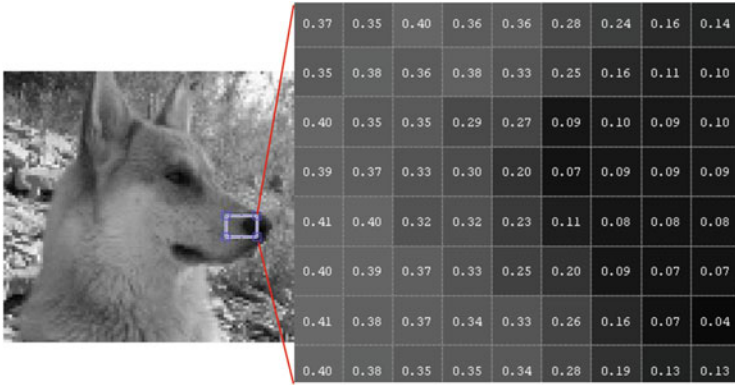


Fig. 1 A detail of the image showing the matricial coding

2.1 Linear Filtering and Convolution

The basic material of this section covers linear diffusion filtering and its relations with gaussian smoothing. This introduces the use of partial differential equations into image processing. More advanced properties of this linear approach like scale-space properties and its applications, generalizations and limitations can be found in http://www.lpi.tel.uva.es/muitic/pim/docus/anisotropic_diffusion.pdf. Here we briefly introduce some concepts. Digital images are commonly defined as matrices of scalars for grayscale images or as vectors for multimodality and/or multichannel images as well as simple multichannel colour RGB images. In a discrete setting images are then $u = (u_{i,j}), 1 \leq i, j \leq N, u_{i,j} \in [0, 1]$ or $u_{i,j} \in [0, 255]$ 2D discrete, bounded signals (Fig. 1). In the variational framework we shall adopt a continuous world view so that a grayscale image is a real valued function $u : \Omega \rightarrow \mathbb{R}$ on an open set $\Omega \subset \mathbb{R}^2$. A color image is a vector-valued function $u : \Omega \rightarrow \mathbb{R}^3$ on an open set $\Omega \subset \mathbb{R}^2$ which maps into RGB color space.

In fact the digital images can also be organized into functional and algebraic structures such as multichannel images where different data acquisition modalities can be grouped to form a unique vectorial description of the image.

These matrices can be seen as the values of a distribution (generalized function) $u_0(x)$ defined on an open and bounded 2D or 3D domain Ω x being a pixel (2D) or a voxel (3D). This allows a functional analytic setting for image-processing problems and in particular, for the design of digital processing algorithms through partial differential equations (PDEs) models. More recent and advanced acquisition techniques in Magnetic Resonance, such as scalar Diffusion Weighted Images (DWI) or Diffusion Tensor Images (DTI) provide 3D volumes of tensorial data, a sort of matrices of matrices which inform about the (anisotropic) movement of water molecules through the fibers of the white matter of the brain (Figs. 2 and 3).

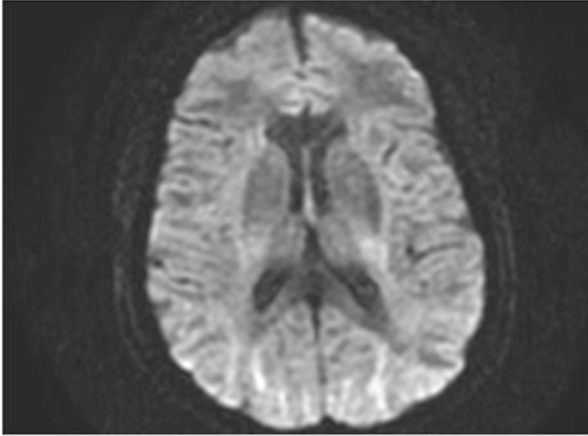


Fig. 2 A DW-MR image courtesy of Fundación Reina Sofia, Centro de Alzheimer, Madrid

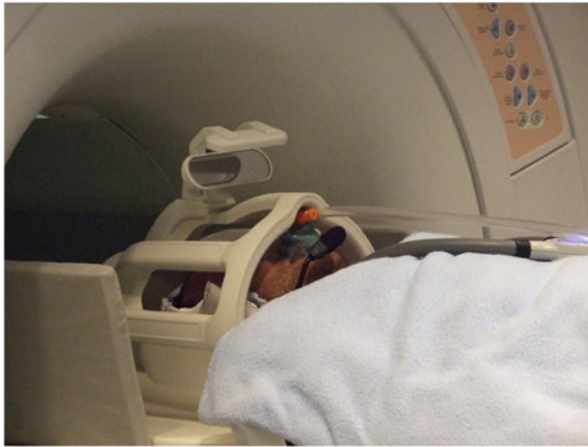


Fig. 3 The MR scanner of General Electric 3T Signa at Fundación Reina Sofia, Centro de Alzheimer, Madrid (Research agreement with General Electric). Image courtesy of CIEN Foundation

Filtering is a technique for modifying or enhancing an image. For example, you can filter an image to emphasize certain features or remove other features. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement. In the continuous case we can understand the analogy between filtering and convolution by means of the heat equation which is a linear diffusion equation (Figs. 4 and 5).

Let \mathbf{J} be a flux of any scalar magnitude such as intensity of the signal, temperature or concentration of a chemical substance. The flux is generated by local differences in the intensity and we have $\mathbf{J} = -D\nabla u$ where D is a tensor characterizing the

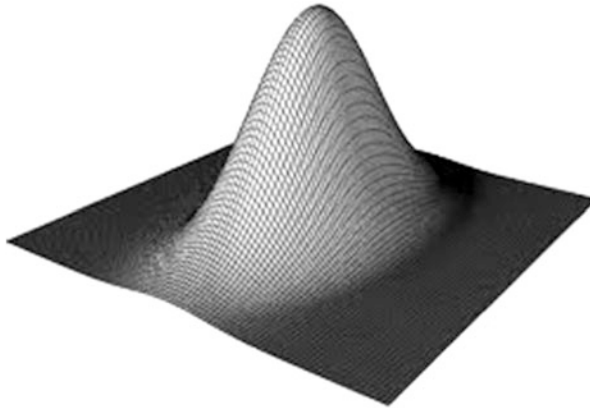


Fig. 4 A plot of a 2D gaussian function

possible anisotropy of the diffusion. In the isotropic case $D = I_d$, the identity matrix. The mass conservation equation states that, without sources or sinks, the local variation of the magnitude of u is caused by the divergence of the flux, $\partial_t u = -\text{div}\mathbf{J}$ which is

$$\partial_t u = \text{div}(\nabla u) = \Delta u.$$

Let n denote the spatial dimension and consider the Cauchy problem

$$\begin{cases} \partial_t u = \Delta u, & \mathbb{R}^n \times (0, +\infty) > 0, \\ u(x, 0) = u_0(x), & \mathbb{R}^n \end{cases}$$

associated to the initial data $u_0(x)$.

If we assume that $u_0(x) = \delta_0$, the Delta function located at $x = 0$ the explicit solution (or Gauss kernel) of the Cauchy problem is:

$$G(x, t) = \frac{e^{-|x|^2/4t}}{(4\pi t)^{n/2}}$$

where the gaussian is represented in Fig. 4.

The solution of the original problem can be expressed in terms of the convolution:

$$u = G * u_0 = \int_{\mathbb{R}^n} G(x - y)u_0(y)dy.$$

Defining $\sigma = \sqrt{2t}$ we see that the solution of our problem is given by the convolution of the initial data with a gaussian function with *standard deviation* σ

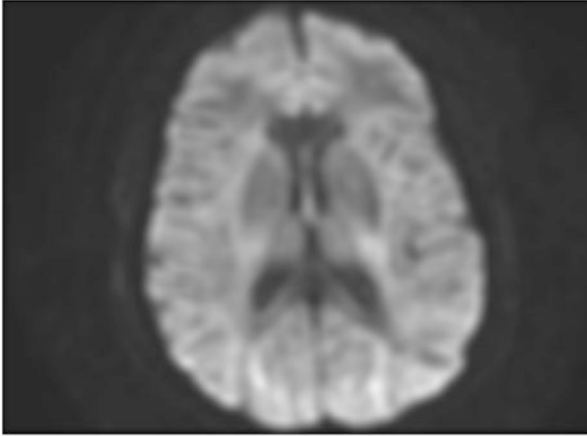


Fig. 5 The oversmoothed, blurred image obtained by convolution

(the width of the gaussian kernel) which corresponds to a linear diffusion process during exactly $T = \sigma^2/2$ where σ^2 is the estimated *variance* of the noise affecting the data. In the discrete case filtering is a neighborhood operation, in which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. Linear filtering of an image is accomplished through an operation called convolution. Convolution is a neighborhood operation in which each output pixel is the weighted sum of neighboring input pixels (Fig. 5).

A fundamental property of the convolution operation is that it regularizes the data and, even with $u_0 \in L^1(\mathbb{R})$ we have $G * u_0 \in C^\infty(\mathbb{R})$ for any $t > 0$. This clearly is a poor result in image processing because this low pass filter smooths out all the high frequencies of the image, where noise and details are involved. The need for nonlinear filtering became readily evident.

2.2 Nonlinear Filtering

It has been introduced into the digital imaging community through the *intriguing* model proposed by Perona and Malik, in [16]. Details about the theoretical difficulties associated to this *forward-backward* nonlinear diffusion model can be found in http://www.lpi.tel.uva.es/muitic/pim/docus/anisotropic_diffusion.pdf. The associated PDE is

$$\partial_t u = \operatorname{div}(g(\nabla u)\nabla u)$$

with $u(x, 0) = u_0(x)$ and

$$g(s^2) = \frac{1}{1 + s^2/\lambda^2}, \quad \lambda > 0.$$

The consideration of the 1D case

$$\partial_t u = \partial_x (g(u_x)u_x)$$

with flux function

$$\Phi(s) = sg(s^2) = \frac{s}{1 + s^2/\lambda^2}$$

reveals that

$$\Phi'(s) \geq 0 \quad |s| \leq \lambda, \quad \Phi'(s) < 0 \quad |s| > \lambda$$

and the equation

$$\partial_t u = \partial_x (\Phi(u_x)) = \Phi'(u_x)u_{xx}$$

has *negative* diffusion when the gradient is big, e.g. near the edges of the image. Despite of this, the numerical resolution of this equation introduces *numerical* diffusion which stabilizes the solution and the model provide quite good results. A simple and straightforward introduction to nonlinear diffusion and related algorithms in MATLAB can be found in <http://staff.science.uva.nl/~rein/nldiffusionweb/nldiffusioncode.pdf>. The original and detailed analysis of nonlinear diffusion and anisotropy is in the excellent book by Weickert http://www.lpi.tel.uva.es/muitic/pim/docus/anisotropic_diffusion.pdf.

2.3 Modelling Medical Images Processing and Restoration

Digital image denoising and segmentation are basic problems in image processing and computer vision which can be dealt with in the variational framework. Roughly speaking this amounts to the minimization of an energy functional defined in a suitable functional space. The minima of the functional can be characterized as the weak solutions of the associated Euler–Lagrange equations which are, typically, nonlinear second order elliptic partial differential equations. These nonlinearities are necessary in order to avoid oversmoothing as predicted by the general linear elliptic regularity theory. This introduces both, mathematical and numerical difficulties in the analysis of such models and makes the implementation of efficient numerical methods challenging.

We shall review some aspects of what is called the Tikhonov Regularization for ill-posed inverse problems. This introduces a General Regularization Model which can be justified by means of a Bayesian formulation. In fact many of the tasks encountered in image processing can be considered as problems in statistical inference. In particular, they fit naturally into a Bayesian framework:

$$\log p(u|f) \propto \log p(f|u) + \log p(u)$$

and a MAP (Maximum A Posteriori) estimation of u is:

$$\max_u \{\log p(f|u) + \log p(u)\}$$

where $p(f|u) = \exp(-H(u, f))$ is the likelihood term and $p(u) = (1/\lambda) \exp(-J(u))$ is the *prior*. Following this Bayesian modelling approach we consider the minimization problem

$$\min_{u \in BV(\Omega)} J(u) + \lambda H(u, f) \quad (1)$$

where $J(u)$ is the convex nonnegative Total Variation regularization functional

$$J(u) = |u|_{BV} = \int_{\Omega} |Du| \quad (2)$$

and the data fidelity term (modelling gaussian noise) is

$$H(u, f) = \int_{\Omega} |f - u|^2 dx.$$

The term $\int_{\Omega} |Du|$ denotes the Total Variation of u with Du being its generalized gradient (a vector bounded Radon measure). When $u \in W^{1,1}(\Omega)$ we have $\int_{\Omega} |Du| = \int_{\Omega} |\nabla u| dx$. The λ parameter in (1) is a scale parameter tuning the model. In this (weak) setting it is a very common and useful approach to describe images as distributions.

One popular model for image denoising is the Rudin, Osher and Fatemi's (ROF) model, where we seek for a distribution u in the space of the Bounded Variation ($BV(\Omega)$) distributions, which is the solution to the following nonlinear minimization problem.

Given $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ which represents the data, minimize the (strictly convex) energy

$$E(u) = \int_{\Omega} |Du| + \frac{1}{2\lambda} \int_{\Omega} |f - u|^2 dx \quad (3)$$

where Ω is a Lipschitz domain (the unit square or a cube for the sake of simplicity) and $f \in L^{\infty}(\Omega)$ is the image affected by Gaussian white noise. Due to the fact that

the functional in (2) is not differentiable at the origin we introduce the notion of the subdifferential of $J(u)$ at a point u by

$$\partial J(u) = \{p \in BV(\Omega)^* \mid J(v) \geq J(u) + \langle p, v - u \rangle\}$$

for all $v \in BV(\Omega)$, to give a (weak and multivalued) meaning to the Euler–Lagrange equation associated to the minimization problem. Using variational calculus and convex analysis the associated Euler–Lagrange Equation is then

$$\lambda \partial J(u) + (u - f) \ni 0$$

which is a multivalued equation which reflects the non differentiability of the TV operator. The proper setting for such multivalued equations is in terms of variational inequalities which can be deduced from the so called *Complementary Formulation*. Typically this difficulty is avoided using the approximating minimization problems

$$J(u_\epsilon) = \int_{\Omega} \sqrt{|\nabla u_\epsilon|^2 + \epsilon} dx + \frac{1}{2\lambda} \int_{\Omega} |f - u_\epsilon|^2 dx \tag{4}$$

with Euler–Lagrange Equation

$$-\lambda \operatorname{div} \left(\frac{\nabla u_\epsilon}{\sqrt{|\nabla u_\epsilon|^2 + \epsilon}} \right) + (u_\epsilon - f) = 0.$$

It is standard to look for a solution to (3) [and (4)] by solving a related nonlinear parabolic equation using a pseudo-time-stepping algorithm in order to approximate the steady-state configuration $u(x)$. This approach, known as (primal) gradient descent, has two serious drawbacks: the approximating problems have *continuous* solutions u_ϵ which are unfeasible in medical imaging because different organs and subcortical structures are characterized by discontinuities; moreover, the numerical method is slowly convergent. An elegant and brilliant solution to these problems can be found in Chambolle [3]. A deep theoretical study of this kind of *linear* energy functionals is considered in [1].

In what follows we shall describe some advanced models that our group has proposed and applied in the last years.

3 Advanced Models

In this section we shall present some advanced models for image segmentation and denoising. Notice that image denoising can be considered as a pre-processing step previous to the segmentation task. A PDE approach to image segmentation is based on the celebrated Mumford and Shah model, [14]. When piecewise constant

solutions of the Mumford-Shah model are considered we have a *minimal partition problem* and a huge literature is concerned with the analysis of such a model [6]. Here we shall consider an anisotropic version of the Mumford and Shah functional which has been proposed in [8,9] for multichannel and multiphase image segmentation.

Let \bar{f} be a vector valued function such as $\bar{f} \in L^\infty(\Omega; \mathbb{R}^M)$ defined on a bounded open domain $\Omega \subset \mathbb{R}^D$, where each scalar component $f_i(x) : \Omega \rightarrow \mathbb{R}$ is a channel. Let \bar{u} be a vector valued piecewise constant function such as $\bar{u} = \sum_1^N \bar{c}_i \chi_i$ with $\bar{c}_i \in \mathbb{R}^M$ and χ_i the characteristic functions of the domain partition. Then we can perform multiclass (N classes) and multichannel (M channels) image segmentation minimizing

$$J(\mathbf{C}, \Gamma) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \|\bar{c}_j - \bar{c}_i\|_{L^p(\Omega; \mathbb{R}^M)} \int_{\Gamma_{ij}} d\mathcal{H}^{D-1} + \frac{1}{2\lambda} \sum_{i=1}^N \int_{\Omega_i} |\bar{c}_i - \bar{f}|^2 dx \tag{5}$$

where

$$\|\bar{c}_j - \bar{c}_i\|_{L^p(\Omega; \mathbb{R}^M)} = \left[\sum_{m=1}^M |c_{j,m} - c_{i,m}|^p \right]^{1/p}.$$

In our current numerical implementation we choose $p = 2$. Notice that the functional is expressed in terms of a matrix \mathbf{C} with components $c_{i,j}$ which reflect the different values of the piecewise solution as well as in terms of a curve Γ along which the solution is discontinuous. The key idea of our method relies on the strong analogy between this anisotropic Mumford and Shah functional (AMS) and the ROF model we introduced before. In fact, in the class of piecewise constant functions both energies coincides. This suggests that the minima of the AMS model can be obtained thresholding the ROF minima. To show an application of these ideas we consider a four classes segmentation problem, as in MRI brain segmentation where white and gray matter, together with liquid and background are the relevant classes. Let $u_{rof} \in BV(\Omega) \cap [0, 1]$ be the minimum of the ROF functional (3). If we threshold this solution by mean of a vector $\bar{t} \in \mathbb{R}^3$ we generate a piecewise constant approximation of u_{rof} for every \bar{t} in form $\bar{c}(\bar{t}) \cdot \bar{\chi}(\bar{t})$. The problem is then to minimize the Anisotropic Mumford Shah energy (5) finding the best threshold $\bar{t} \in \mathbb{R}^3$ for the solution of the ROF model (3). This can be accomplished by using a genetic algorithm (notice that the problem is not convex) where the search is restricted to simple functions $u_{\bar{t}}(x) \in SBV(\Omega)$ taking, for a.e. $x \in \Omega$, the (possibly re-ordered) values $\bar{c} = (c_1, c_2, c_3, c_4)$ as defined by formula (6) which we shall deduce below. Let $\bar{\chi} = \{\chi_i\}_{i=1}^4$ be a given partition. Then

$$\frac{\partial J}{\partial c_j} = \left(\sum_{i=1, i \neq j}^4 \frac{(j-i)}{|j-i|} |\Gamma_{i,j}| \right) - \frac{1}{\lambda} \int_{\Omega_j} f dx + \frac{|\Omega_j|}{\lambda} c_j, \quad j = 1, \dots, 4$$

and the functional $J(\bar{c}, \bar{\chi})$ is optimized by the choice

$$c_j = \bar{f}^j + \frac{\lambda}{|\Omega_j|} \left(\sum_{i=1, i \neq j}^4 \frac{(i-j)}{|i-j|} |\Gamma_{ij}| \right), \quad j = 1, \dots, 4 \tag{6}$$

where \bar{f}^j are the local averaged data values as predicted by the partition:

$$\bar{f}^j = \int_{\Omega_j} f dx / |\Omega_j|, \quad j = 1, \dots, 4$$

Moreover we have:

$$\sum_{j=1}^4 c_j |\Omega_j| = \sum_{j=1}^4 \int_{\Omega_j} f dx + \lambda \sum_{j=1}^4 \left(\sum_{i=1, i \neq j}^4 \frac{(i-j)}{|i-j|} |\Gamma_{ij}| \right) = \int_{\Omega} f dx. \tag{7}$$

This implies that, if we calculate u as the (unique) minimum of $J(u)$ in (3) and we threshold u by means of a threshold vector $\bar{t} = (t_1, t_2, t_3) \in \mathbb{R}^3$, then we generate a partition $\bar{\chi} = (\chi_1, \chi_2, \chi_3, \chi_4)$ (defining $\Omega_i = \{x \in \Omega / t_{i-1} \leq u(x) < t_i\}$) and, using formula (6) for the best constants, an optimal representation of u for the given partition in form $u = \bar{c} \cdot \bar{\chi}$. Notice that a relabeling is performed to ensure the ordering of the optimal constants once the threshold \bar{t} is applied. More details in this procedure can be found in [8].

The numerics are performed using the dual formulation of the problem. This provides a convenient framework to solve the multiphase systems associated with the minimization of the AMS functional. More advanced staggered schemes are proposed in [10]. Segmentation results with different values of the λ parameter are presented below. Figure 6 shows the segmentation of a brain phantom slice with three levels of added noise with different values of the parameter.

Finally, we segmented real MRI images acquired at Fundación Reina Sofía in Madrid. Figure 7 shows the result of the automatic segmentation with two different values of the parameter. Both results are visually correct, while the λ parameter allows to obtain segmentations at different scales of detail.

More sophisticated results can be obtained when segmenting FA color code DT-MR images as we show in figures below. A brief introduction to this kind of scalar MR images which are obtained from tensorial data is presented in the next section (Figs. 8 and 9).

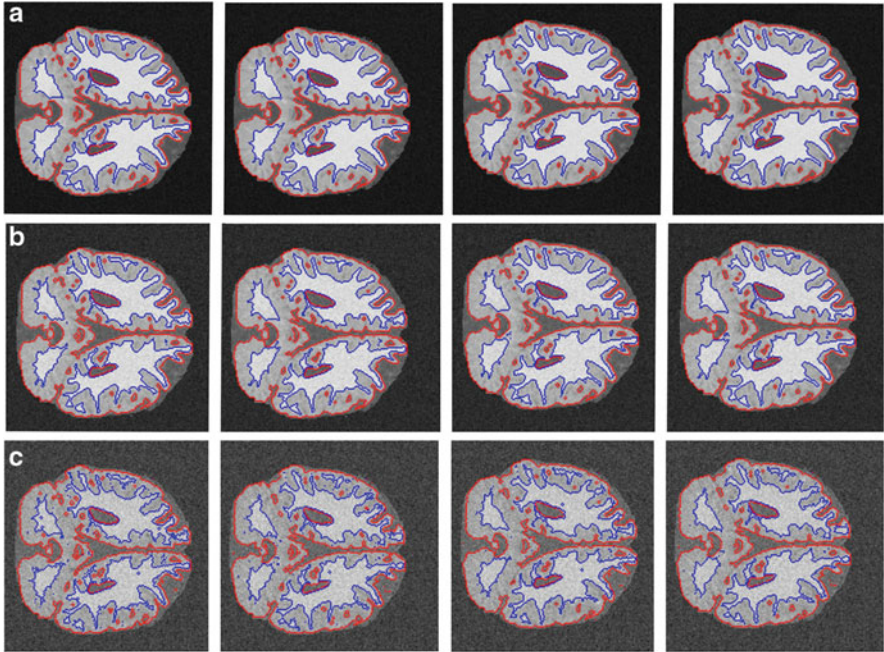


Fig. 6 Segmentation of the same slice of a phantom with different noise levels and different values of λ . From left to right, results with λ values 0.08, 0.09, 0.1 and 0.11. (a) Phantom with 5 % noise; (b) phantom with 10 % noise; (c) phantom with 20 % noise

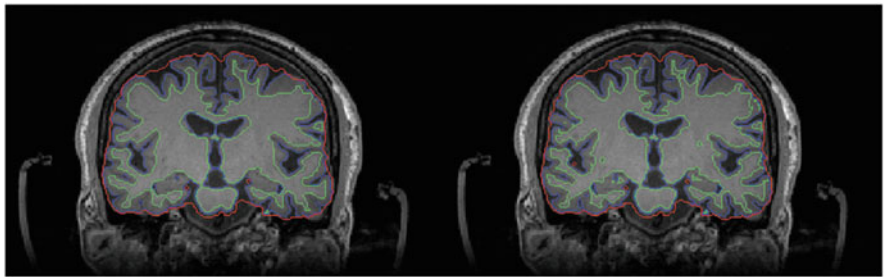


Fig. 7 Segmentation of real MRI brain data with two values of the λ parameter. Left: $\lambda = 0.12$; Right: $\lambda = 0.08$

3.1 MRI Denoising

We now step forward in the modelling exercise. In fact, accurate MRI noise modelling is a fundamental issue in medical image processing which leads naturally to the assumption that MR magnitude images are corrupted by Rician noise which is a signal dependent noise. Indeed this noise is originated in the computation of the

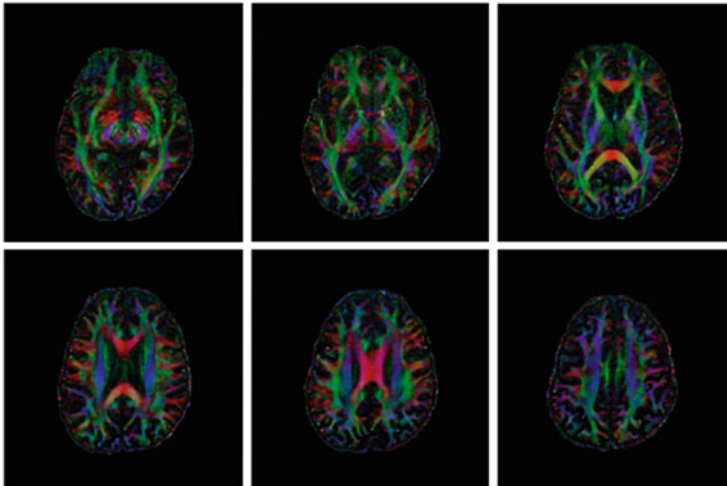


Fig. 8 A color code Fractional Anisotropy (FA) image which is obtained computing the eigenvalues of the Diffusion Tensor Image (DTI) reconstructed from the Diffusion Weighted Images (DWI) acquired at the Hospital Reina Sofia

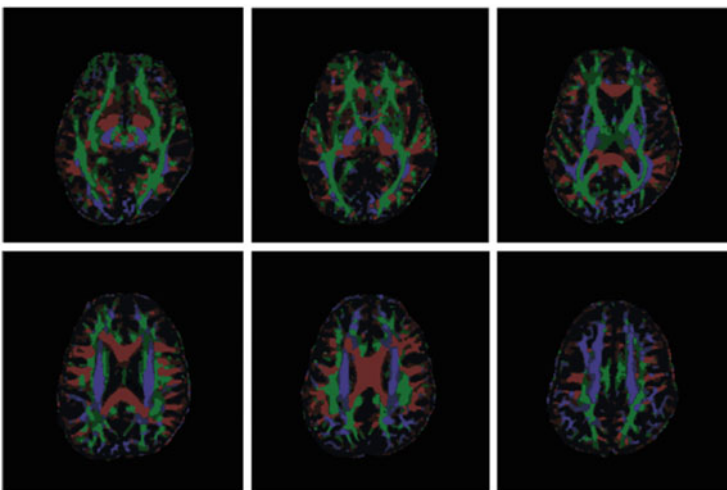


Fig. 9 The obtained segmentation. Notice that we segment the directions along which the fibers propagate in the brain

magnitude image from the real and imaginary images, that are obtained from the inverse Fourier Transform applied to the original raw data. This process involves a non-linear operation which maps the original Gaussian distribution of the noise to a Rician distribution. Nevertheless it is usually argued that this bias does not affect seriously the processing and subsequent analysis of MR images so that a

(identically distributed and signal independent) Gaussian noise is modelled. This assumptions fails when low signal-to-noise ratio are considered. With this purpose we consider, in a variational framework, a denoising model for MR Rician noise contaminated images proposed in [12] which combines the Total Variation semi-norm with a Rician data fitting term.

The data term $H(u, f)$ is a fitting functional which is nonnegative with respect to u for fixed f . To model Rician noise $H(u, f)$ has been deduced previously in [2] in the context of diffusion tensor MR images. The Rician likelihood term is of the form:

$$H(u, f) = \frac{1}{2\sigma^2} \int_{\Omega} u^2 dx - \int_{\Omega} \log I_0 \left(\frac{uf}{\sigma^2} \right) dx \tag{8}$$

where σ is the standard deviation of the Rician noise of the data and I_0 is the modified zeroth-order Bessel function of the first kind. It can be shown that functional (8) is possibly non-convex depending on the data f , λ and σ . Using (1), (2) and (8) the minimization problem is formulated as: Fixed λ and σ and given a noisy image $f \in L^\infty(\Omega)$ recover $u \in BV(\Omega) \cap L^\infty(\Omega)$ minimizing the energy:

$$J(u) + \lambda H(u, f) = \int_{\Omega} |Du| + \frac{\lambda}{2\sigma^2} \int_{\Omega} u^2 dx - \lambda \int_{\Omega} \log I_0 \left(\frac{uf}{\sigma^2} \right) dx. \tag{9}$$

When the functional in (9) is considered for minimization, the variational approach leads to the resolution of a nonlinear multivalued PDE elliptic equation which is the Euler Lagrange equation for optimization. In fact the first order optimality condition reads

$$\partial J(u) + \lambda \partial_u H(u, f) \ni 0 \tag{10}$$

with (Gâteaux) differential

$$\partial_u H(u, f) = \frac{u}{\sigma^2} - \frac{I_1(uf/\sigma^2)}{I_0(uf/\sigma^2)} \frac{f}{\sigma^2} \tag{11}$$

where I_1 is the modified first-order Bessel function of the first kind and verifies $0 \leq I_1(\xi)/I_0(\xi) < 1, \forall \xi > 0$. As we introduced before, this gives rise to a number of interesting theoretical problems when the Total Variation operator is considered as *a priori*, because the energy functional is not differentiable at the origin (i.e. $\nabla u = \bar{0}$) and regular approximated problems must be solved. A number of mathematical difficulties is associated with the multivalued formulation (10) and a regularization of the diffusion term $\text{div}(\nabla u/|\nabla u|)$ in form $\text{div}(\nabla u/|\nabla u|_\epsilon)$, with $|\nabla u|_\epsilon = \sqrt{|\nabla u|^2 + \epsilon^2}$ and $0 < \epsilon \ll 1$ is implemented to avoid degeneration of the equation where $\nabla u = \bar{0}$. Using this approximation it is possible to give a (weak) meaning to the following formulation: Fixed λ, σ and (small) ϵ and given

$f \in L^\infty(\Omega) \cap [0, 1]$ find $u_\epsilon \in W^{1,1}(\Omega) \cap [0, 1]$ solving

$$-\operatorname{div} \left(\frac{\nabla u}{|\nabla u|_\epsilon} \right) + \frac{\lambda}{\sigma^2} \left(u - \left[I_1 \left(\frac{uf}{\sigma^2} \right) / I_0 \left(\frac{uf}{\sigma^2} \right) \right] f \right) = 0$$

which we write in form

$$-\operatorname{div} \left(\frac{\nabla u_\epsilon}{|\nabla u_\epsilon|_\epsilon} \right) + \frac{\lambda}{\sigma^2} [u_\epsilon - r_\epsilon(u_\epsilon, f)] f = 0 \tag{12}$$

complemented with Neumann homogeneous boundary conditions $\partial u_\epsilon / \partial n = 0$ and where, for notational simplicity, we introduced the nonlinear function

$$r_\epsilon(u_\epsilon, f) = I_1(u_\epsilon f / \sigma^2) / I_0(u_\epsilon f / \sigma^2).$$

This is a nonlinear (in fact quasilinear) elliptic problem that we solve with a gradient descent scheme until stabilization (when $t \rightarrow +\infty$) of the evolutionary solution to steady state, i.e. a solution of the elliptic problem (12) which is a minimum of the approximating energy functionals

$$\begin{aligned} E_\epsilon(u_\epsilon) &= J_\epsilon(u_\epsilon) + \lambda H(u_\epsilon, f) = \\ &= \int_\Omega j_\epsilon(u_\epsilon) dx + \lambda \int_\Omega h(u_\epsilon) dx = \\ &= \int_\Omega \sqrt{|\nabla u_\epsilon|^2 + \epsilon^2} dx + \frac{\lambda}{2\sigma^2} \int_\Omega u_\epsilon^2 dx - \lambda \int_\Omega \log I_0 \left(\frac{u_\epsilon f}{\sigma^2} \right) dx. \end{aligned} \tag{13}$$

When $\epsilon \rightarrow 0$ we have $u_\epsilon \rightarrow u$, $J_\epsilon(u_\epsilon) \rightarrow J(u)$ and the energies in (9) and (13) coincide.

The gradient descent approach amounts to solve the associated nonlinear parabolic problem:

$$\frac{\partial u_\epsilon}{\partial t} = \operatorname{div} \left(\frac{\nabla u_\epsilon}{|\nabla u_\epsilon|_\epsilon} \right) - \frac{\lambda}{\sigma^2} [u_\epsilon - r_\epsilon(u_\epsilon, f)] f \tag{14}$$

complemented with Neumann homogeneous boundary conditions $\partial u_\epsilon / \partial n = 0$ and initial condition $u_\epsilon(0, x) = u_0^\epsilon(x)$ whose (weak) solution stabilizes (when $t \rightarrow +\infty$) to the steady state of (12), i.e. a minimum of (13) which approximates, for ϵ sufficiently small, a minimum of the energy functional (9). A direct gradient descent method has been used in [12] in order to validate the model assumption of Rician noise. This approach is found to be inherently slow because a stabilization at the steady state is needed. Also, that scheme is finally explicit and very small time steps have to be used to avoid numerical oscillations. Here we present a framework to solve numerically and efficiently the gradient descent scheme (gradient flow)

associated to the Rician energy minimization problem introducing a semi-implicit formulation. Details can be found in [13].

Using a simple Euler discretization of the time derivative, stationary problems of the ROF type [17] are deduced. This allows to use the well known dual formulation of the ROF model proposed in [3] to speed up the computations. As a by-product of this approach the exact Total Variation operator can be computed and this provides accuracy of the solution in so far truly (discontinuous) bounded variation solutions are numerically approximated. In fact we considered the approximated Euler–Lagrange equation (12) associated to the minimization of the energy (9). This is a modelling approximation and we can get rid of it. We argue as follows. Considering the original Euler–Lagrange equation associated to the energy (9) we have (with abuse of notation for the diffusive term)

$$-\operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) + \frac{\lambda}{\sigma^2} [u - r(u, f)f] = 0 \quad (15)$$

with $r(u, f) = I_1(uf/\sigma^2)/I_0(uf/\sigma^2)$. A rigorous treatment of Eq. (15) should follow the multivalued formalism of (10).

Using again a gradient descent scheme we have to solve the parabolic problem:

$$\frac{\partial u}{\partial t} = \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) - \frac{\lambda}{\sigma^2} [u - r(u, f)f] \quad (16)$$

together with Neumann homogeneous boundary conditions $\partial u/\partial n = 0$ and initial condition $u(0, x) = u_0(x)$. For comparison purposes we used $u_0(x) = u_0^\epsilon(x)$ in all numerical tests.

Using forward finite differences for the temporal derivative in (16) and a semi-implicit scheme where only the term depending on the ratio of the Bessel's functions is delayed, results in the numerical scheme:

$$\left(1 + \tau \frac{\lambda}{\sigma^2} \right) u^{n+1} = u^n + \tau \left(\operatorname{div} \left(\frac{\nabla u^{n+1}}{|\nabla u^{n+1}|} \right) + \frac{\lambda}{\sigma^2} r(u^n, f)f \right) \quad (17)$$

where the diffusive term is (formally) exact and implicitly considered. Defining $\beta = (\tau\lambda)/\sigma^2$, $\gamma = (1 + \beta)/\tau$ and

$$g^n = \left(\frac{1}{1 + \beta} \right) u^n + \left(\frac{\beta}{1 + \beta} \right) r(u^n, f)f \quad (18)$$

we can write:

$$-\operatorname{div} \left(\frac{\nabla u^{n+1}}{|\nabla u^{n+1}|} \right) + \left(\frac{1}{\gamma} \right) (u^{n+1} - g^n) = 0 \quad (19)$$

which is the Euler–Lagrange equation of a ROF energy functional.

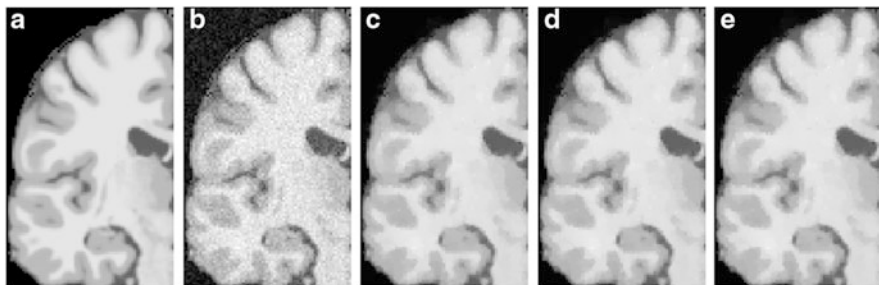


Fig. 10 The original image and the contaminated phantom are shown in (a) original phantom and (b) noisy for $\sigma = 0.05$. The denoised images obtained with the R-ROF-Dual, R-ROF-Primal-Dual, R-Primal-Dual algorithms and for the parametric values $\sigma = 0.05, \lambda = 0.075$ are presented in the sub-plots (c) R-ROF-D denoised, (d) R-ROF-PD denoised and (e) R-PD denoised, respectively

$$E_n(u) = \int_{\Omega} |Du| + \left(\frac{1}{2\gamma}\right) \int_{\Omega} (u - g^n)^2 dx \tag{20}$$

for any positive integer $n > 0$, with (artificial) time $t_n = n\tau$. Hence, at each gradient descent step τ , we can solve a ROF problem associated to the minimization of the energy (20) in the space $BV(\Omega) \cap [0, 1]$. This problem is mathematically well-posed and it can be numerically solved by very efficient methods, when formulated using well known duality arguments in [3] or primal-dual algorithms in [4, 18].

In our study we first compared different algorithms using synthetic brain images from the BrainWeb Simulated Brain Database¹ at the Montreal Neurological Institute. The original phantoms were artificially contaminated with Rician noise considering the data as a complex image with zero imaginary part and adding random Gaussian perturbations to both the real and imaginary part, before computing the magnitude image (Fig. 10).

Apart from the modelling exercise and the implementation details of the algorithms presented above, our main interest relies in the application to real brain images. In the following we present some preliminary results we obtained in [13] for Diffusion Weighted Magnetic Resonance Images (DW-MRI) denoising. The DW-MRI are images acquired in order to obtain a Diffusion Tensor Image (DTI). Accurate denoising of the DW-MRI is crucial for a good DTI reconstruction because of their characteristic very low SNR, [2].

Diffusion Tensor Imaging is becoming one of the most popular methods for the analysis of the white matter (WM) structure of the brain, where some alterations can be found from early stages in some degenerative diseases. This technique measures Brownian motion (random motion) of the water molecules in the brain, which is assumed to be isotropic when it is not restricted by the surrounding structure. In the WM regions, which contain densely packed fibre bundles, they cause an anisotropic diffusion of water molecules along the perpendicular directions to them. At each

¹Available at <http://www.bic.mni.mcgill.ca/brainweb>.

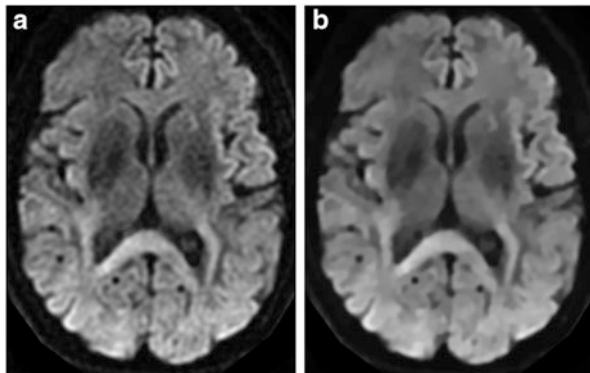


Fig. 11 A slice of the original Diffusion Weighted Image corresponding to the $(1, 0, 0)$ gradient direction and the corresponding denoised image. (a) Original; (b) denoised with $\lambda = \sigma/2$

voxel of a DTI the water diffusion is represented by a symmetric 3×3 tensor, where the information of the preferred directions of the motion and the relevance of these directions is found in the eigenvectors and the eigenvalues of the tensor. These tensorial data can be represented as different scalar measurements, one of them is the Fractional Anisotropy (FA) of the tissue, which is defined as

$$FA = \sqrt{\frac{3 \left((\hat{\lambda} - \lambda_1)^2 + (\hat{\lambda} - \lambda_2)^2 + (\hat{\lambda} - \lambda_3)^2 \right)}{2 (\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$$

where the λ_i are the eigenvalues of the tensor and $\hat{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$. The FA values vary from 0, (when the motion in the voxel is completely isotropic) to 1 (totally anisotropic). For the reconstruction of the DTI a set of DWI has to be acquired, scanning the tissue in different directions of the space. At least six DWI volumes are needed in order to be able to calculate the DTI, which is a positive defined matrix. The noise present into the DWI scalar images can generate small, negative eigenvalues. Increasing the number of directions along which the brain is scanned improves the image quality but at the expenses of a longer acquisition time. The importance of pre-processing the DW Images previously to the DTI reconstruction is then two-fold: to improve the DT image quality through accurate Rician denoising so allowing shorter scanning time.

The data we used consist of a DW-MR brain volume provided by Fundación CIEN-Fundación Reina Sofía which was acquired with a 3T General Electric scanner equipped with an 8-channel coil. The DW images have been obtained with a single-shot spin-eco EPI sequence (FOV = 24 cm, TR = 9,100, TE = 88.9, slice thickness = 3 mm, spacing = 0.3, matrix size = 128×128 , NEX = 2). The DW-MRI data consists on a volume obtained with $b = 0/\text{mm}^2$ and 15 volumes with $b = 1,000 \text{ s}/\text{mm}^2$.

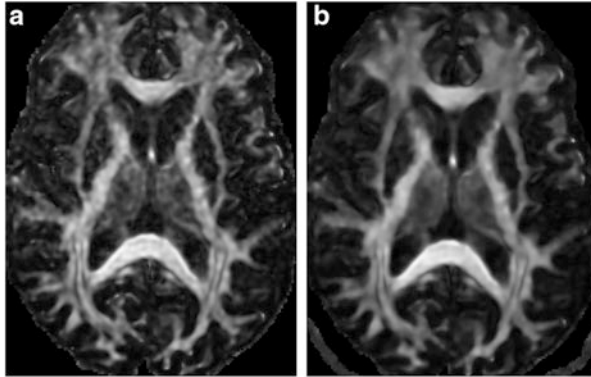


Fig. 12 A slice of the Fractional Anisotropy estimated from the Tensor Image. *Dark colour* corresponds to values near zero (isotropic regions) and *bright color* corresponds to values near one (anisotropic regions). (a) From original DWI data; (b) from denoised DWI data with $\lambda = \sigma/2$

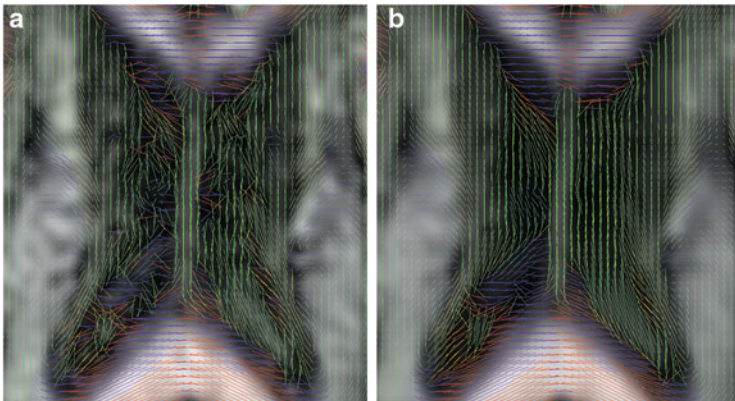


Fig. 13 A detail of the first eigenvectors of the DTI over the FA image. The color is based on the main orientation of the tensorial data. *Red* means right-left direction, *green* anterior-posterior and *blue* inferior-superior. Fibres with an oblique angle have a color that is a mixture of the principal colors and dark color is used for the isotropic regions. (a) From original DWI data; (b) from denoised DWI data with $\lambda = \sigma/2$

These DW-MR images, which represent diffusion measurements along multiples directions, are denoised with the proposed method previously to the Diffusion Tensorial Image reconstruction, which was done with the 3d Slicer tools.² In Fig. 11a we show a slice of the original DWI data corresponding to the (1, 0, 0) gradient direction where the affecting noise is clearly visible. The complete DW-MRI data volume is denoised using the proposed method. The Rician noise

²Free available in <http://www.slicer.org/>.

standard deviation (σ) has been estimated for each slice of each gradient direction while we used a value of $\lambda = \sigma/2$ for the denoising. The slice resulting from the denoising process is shown in Fig. 11b. It can be observed how noise has been removed in the denoised images but the details and the edges have been fully preserved, as we should expect when the exact TV model is solved. The effect of this denoising process over the reconstructed tensor and their derived scalar measurements (obtained with the 3d Slicer tools) is presented in Figs. 12 and 13. Figure 12 shows a Fractional Anisotropy image where the structures and details are clearly enhanced if the DW-MRI volume is denoised previously. When finer details are considered the denoising step is yet more crucial. For instance in Fig. 13 the main eigenvector of the tensor is represented, where the noise on the original DWI data cause inhomogeneities (see Fig. 13a) in the eigenvectors field which are product of the noise (Fig. 13b).

Acknowledgements The authors wish to thank to all the Research Institutions involved into the Alzheimer Project (Fundación CIEN-Fundación Reina Sofía, Lab. de Neuroimagen. Centro de Tecnología Biomédica (UPM-URJC) Universidad Rey Juan Carlos) and very specially to the MICINN Spanish Minister for Science and Innovation for supporting Project TEC2012-39095-C03-02. Finally the first author wish to thank to the organizers of the XV Escuela Hispano-Francesa Jacques-Louis Lions for inviting him as a lecturer.

References

1. Andreu-Vailló, F., Caselles, V., Mazón, J.M.: Parabolic quasilinear equations minimizing linear growth functionals. In: Progress in Mathematics. Birkhauser, Basel (2004). ISBN 3-7643-6619-2
2. Basu, S., Fletcher, T., Whitaker, R.: Rician noise removal in diffusion tensor MRI. In: MICCAI 2006. Lecture Notes in Computer Science, vol. 4190, pp. 117–125. Springer, Heidelberg (2006)
3. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**(1–2), 89–97 (2004)
4. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
5. Chan, T., Shen, J.: Image processing and analysis: variational, PDE, wavelet, and stochastic methods. In: Society for Industrial and Applied Mathematics (2005). ISBN: 089871589X
6. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Process.* **10**(2), 266–277 (2001)
7. Garamendi, J.F., Malpica, N., Martel, J., Schiavi, E.: Automatic segmentation of the liver in CT using level sets without edges. In: Proceedings of Iberian Conference on Pattern Recognition and Image Analysis, Lecture Notes in Computer Science, vol. 4477, pp. 161–168. Springer, Berlin, Heidelberg, New York (2007). ISBN-10 3-540-72846-5
8. Garamendi, J.F., Malpica, N., Schiavi, E.: A fast anisotropic Mumford-Shah functional based segmentation. In: Lecture Notes in Computer Science, vol. 5524, pp. 322–329. Springer, Berlin, Heidelberg (2009). <http://www.springerlink.com/content/v2u2nq61502hw50x/>
9. Garamendi, J.F., Malpica, N., Schiavi, E. (2009). Multiphase systems for medical image region classification. In: Mathematical Models in Engineering, Biology and Medicine. Conference on Boundary Value Problems, Santiago de Compostela, Spain, 16–19 September 2008, p. 104
10. Garamendi, J.F., Gaspar, F.J., Malpica, N., Schiavi, E.: Box relaxation schemes in staggered discretizations for the dual formulation of total variation minimization. *IEEE Trans. Image Process.* **22**, 2030–2043 (2013)

11. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing using MATLAB, 2nd edn. Gatesmark Publishing (2009). ISBN: 978-0-9820854-0-0
12. Martín, A., Garamendi, J.F., Schiavi, E.: Iterated rician denoising. In: IPCV'11, pp. 959–963. CSREA Press (2011)
13. Martin, A., Garamendi, J.-F., Schiavi, E.: MRI TV-rician denoising. *Commun. Comput. Inf. Sci.* **357**, 255–268 (2013)
14. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* **42**(5), 577–685 (1989)
15. Osher, S., Paragios, N.: *Geometric Level Set Methods in Imaging, Vision, and Graphics*. Springer, New York (2003). ISBN: 0387954880
16. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990)
17. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenomena* **60**(1), 259–268 (1992)
18. Zhu, M., Chan, T.: An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration. UCLA CAM Report 08–34 (2008)