Mathematical Segmentation and Denoising of Images with Applications to MRI and Diffusion Tensor MRI

PhD Thesis

Oddvar Christiansen

Department of Mathematics
University of Bergen

November 2007

Dedicated to my parents who meant so much to me.

Preface

This dissertation is submitted as a partial fulfilment of the requirements for the degree Doctor of Philosophy (PhD) at the University of Bergen. The dissertation consists of two parts. In the first part we give a short introduction, describing the field and the motivation behind the articles. The second part consists of six articles which are the main contribution to this dissertation. Four of the articles are already accepted for publication.

Acknowledgements

I would like to thank my supervisor Prof. Hans-Munthe Kaas. You have guided me through my master and PhD and have always believed in me. You always had time to help and I could not have got a better supervisor.

I would also like to thank my co-supervisor Prof. Xue-Cheng Tai. You introduced me to the field of mathematical imaging and have helped me a lot. Also thanks for inviting me to Singapore for three weeks during your stay there.

Thanks to my girlfriend Ingrid and my sister Hilde-Mari for always being supportive. Also thanks to Ingrid for reading this thesis and correcting errors.

Thanks to Dr. Johan Lie for being a great friend and colleague. We have worked together on four projects and you have learnt me a lot about image processing. I think we make a great team. Also thanks to Ørjan Bergmann who joined us on the SA-DCT project. You are a remarkable programmer.

Thanks to Prof. Tony Chan for hosting me at University of California, Los Angeles for three months during the winter of 2006 and two weeks in November 2006. A special thank to Tin-Man Lee for taking good care of me during my stay in LA.

Thanks to Ass.-Prof. Bernhard Burgeth for hosting me and Johan at Universität des Saarlandes for one week during the spring of 2007.

Thanks to Prof. Arvid Lundervold for being a motivator and an important link between mathematics and medicine. Also thanks for providing some of the MRI data used in this thesis.

Thanks to Assoc.-Prof. Renate Grüner for providing access to the 3T MR scanner at Haukeland University Hospital, Bergen.

Thanks to all the people in the image processing group (BBG) in Bergen.

Bergen, 4. November 2007
Oddvar Christiansen

Contents

Part I

Background

Introduction and Outline

In medicine digital images play a vital role in both research and clinical work. Using advanced techniques like Magnetic Resonance Imaging (MRI), Computertomografi (CT), Positron Emission Tomography (PET) or X-ray, images of the inside of the body can be made in vivo. In order to extract as much information as possible from the data, advanced image processing techniques are used. 3D techniques can be used to make advanced visualization, segmentation can be used to separate the different tissue parts, feature extraction can be used for automatic diagnostics and noise removal methods can be used to improve the image quality.

This is just a few of the possibilities image processing offers. However, many of the techniques are closely related. In order to make a proper visualization, the different body parts need to be segmented and in order to make a good segmentation, noise needs to be removed.

A relatively new MRI modality is Diffusion Tensor Imaging (DTI) [5]. Using this new modality, it is possible to construct a model of the nerve fibers in the human brain. Tracing these fibers the wiring of the brain can be studied. This is however not a trivial task. By construction the DTI images are corrupted by a large amount of noise, making it hard to decide the location of the fibers. Using different noise removal methods it is possible to remove or reduce the noise in the DTI data, and the main part of this thesis is devoted to such methods [7, 32, 33, 54].

In addition to the noise removal algorithms we propose an efficient method for image segmentation [34, 88]. This method is based on intensity based segmentation, meaning that the intensity in the image is used to divide it into different regions. In general these methods are closely related to noise removal algorithms and can be seen as a combined noise removal and projection algorithm.

All except for one of the proposed methods are so-called variational methods. In general the variational methods are derived from a cost functional, which needs to be minimized. A common way to solve the minimization problem is to find the corresponding Euler-Lagrange equation and use the method of steepest descent. This leads to a partial differential equation (PDE), which is solved to steady state.

In addition to the variational methods, we propose a transformed based denoising method [7]. This method uses a shape-adaptive discrete cosine transform (SA-DCT) to transform the image into the frequency domain. In the frequency domain small coefficients are thresholded, which suppress the majority of the noise. The proposed transform based method is compared to the PDE based method.

An image can be viewed as a function sampled on a mesh. Usually this mesh is rectangular (2D) or cubic (3D), and in two dimensions the elements are called

pixels and in three dimensions the elements are called *voxels*. In practice the image domain is discrete. However, when developing mathematical models it is customary to view the domain as continuous. Thus mathemtically we assume that the image is a continous mapping from the image domain Ω to the real numbers $\mathcal{R}$, i.e.

$$f : \Omega \to \mathcal{R}. \tag{1}$$

The function f can be scalar-, vector- or tensor-valued.

We will assume that the image is corrupted by additive noise, i.e.

$$u = f + \sigma. \tag{2}$$

Here f is the true image, σ is the noise level and u is the observed image. The noise is Gaussian with zero mean and variance σ^2.

The introduction is outlined as follows: In Chapter 1 we give an overview of MRI and DTI. In Chapter 2 we give an introduction to PDE based image segmentation. In Chapter 3 we give an introduction to PDE and transformed based denoising methods. In Chapter 4 we give a short summery of the six papers and in the last chapter we give a conclusion of the achieved results.

Chapter 1

MRI and Diffusion Tensor MRI

The main application for many of the methods presented in this thesis is in the field of Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI). In this chapter we give a short overview of MRI and DTI.

1.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging is an imaging technique used primarily in medicine to produce high quality images of the inside of the human body. It is based on the principles of Nuclear Magnetic Resonance (NMR), and was earlier known as Nuclear Magnetic Resonance Imaging (NMRI). However, because of the negative associations with the word nuclear in the late 1970's it was changed to MRI.

The magnetic resonance phenomenon was discovered independently by Felix Bloch [9] and Edward Purcell [80] in 1946, and in 1952 both were rewarded the Nobel Prize in physics for the work. The first magnetic resonance imaging on small test tube samples was performed by Paul Lauterbur [50] in 1973, and in 2003 he was rewarded the Nobel Prize in Medicine for his work. MRI using phase and frequency encoding was proposed by Richard Ernst et al. [48, 49] in 1975, and is the basis of the current MRI techniques. In 1991 Richard Ernst got the Nobel Prize in Chemistry for his work.

The MR scanner is a tube surrounded by a giant circular magnet, see Figure 1.1. The patient is placed on a movable bed, which is inserted into the magnet. The magnet creates a strong homogenic magnetic field that aligns the protons of hydrogen atoms, which are then exposed to a beam of radio waves. This makes the various protons of the body spin around its own axis, and produce a signal which is detected by the receiver of the MR scanner. The receiver information is processed by a computer, and an image is produced. In Figure 1.2 an MR image of the authors brain is shown.

Figure 1.1: The image shows a MAGNETOM Trio, A Tim System 3T from Siemens.

1.2 Diffusion tensor imaging

Diffusion Tensor Imaging was introduced in 1994 by Basser et al. [5], and during the last decade it has been extensively studied [2, 51, 69, 4, 99]. This new modality enables the measurement of the diffusion of water in tissue. The main application is the imaging of the white matter in the brain, where the location and orientation of the nerve fibers can be measured. It has clinical applications in ischemia and is used in research on deceases like multiple sclerosis, dyslexia, schizophrenia and trauma [19, 47, 60, 71]. DTI also has tremendous implications to brain research, as it makes it possible to trace how fibers are connected in the brain, yielding a map of how the brain is wired [3, 70, 87, 100].

Even though diffusion tensor imaging was not proposed before 1994, the measurement of self-diffusivity of water using magnetic resonance was reported as early as 1954 [16]. Stejskal and Tanner [85, 86] made methodical improvements to this diffusion measurement, and together with the development of MRI this led to the diffusion tensor imaging.

The DTI images are matrix valued. From a series of at least 6 direction sensitive MR acquisitions a 3×3 symmetric positive definite matrix D can be calculated for each voxel. This matrix describes the 3-dimensional shape of diffusion. Since the matrix $D \in R^{3\times3}$ is symmetric and positive definite it can be decomposed as

$$D = V\Lambda V^{-1}, \tag{1.1}$$

where V is an orthogonal matrix containing the eigenvectors of D, and Λ is a

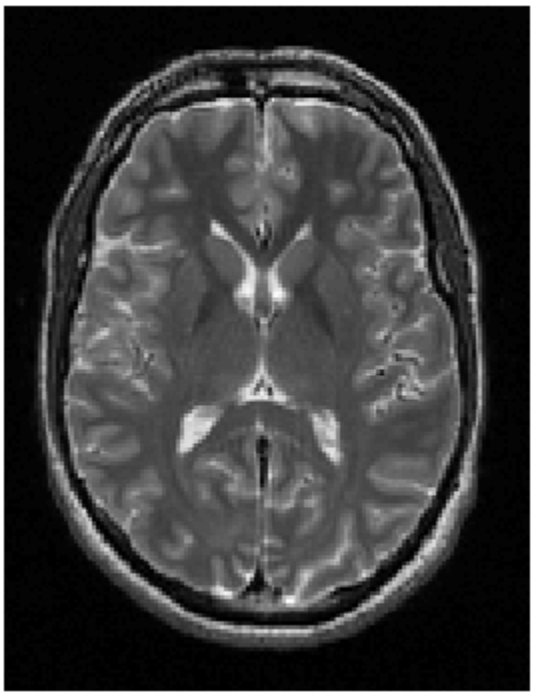
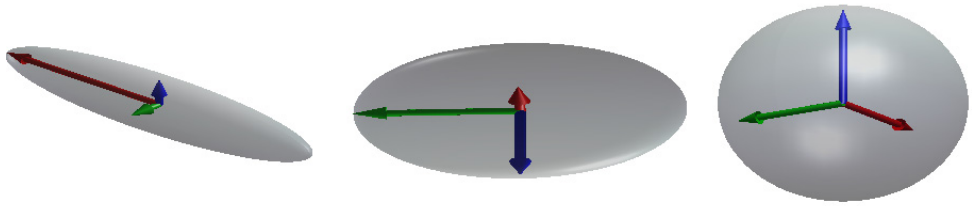


Figure 1.2: The image shows a slice of a T2 weighted MR scan of the authors brain.

diagonal matrix with positive elements containing the eigenvalues of D. We may look at the diffusion matrix as a hyperellipse, where the eigenvectors $\{V_i\}_{i=1}^3$ span the ellipsoid, and the corresponding eigenvalues $\{\lambda_i\}_{i=1}^3$ determine the length of each semi axis. It is customary to arrange the eigenvalues in decreasing order, i.e. $\lambda_1 \geq \lambda_2 \geq \lambda_3$.

The physical interpretation of the diffusion ellipsoid is that the length of each semi axis determines the ability of water molecules to diffuse along the given direction. It is customary to divide the diffusion into three categories. If the diffusion along one axis is much larger than along the other two axes we have a cigar shaped diffusion, see Figure 1.3(a). In this case one eigenvalue is much larger than the other two eigenvalues ($\lambda_1 >> \lambda_2 \approx \lambda_3$), and the main diffusion of water is along V_1 corresponding to λ_1. If the diffusion along one axis is much smaller than along the other two axes ($\lambda_1 \approx \lambda_2 >> \lambda_3$), we have a disc-shaped diffusion, see Figure 1.3(b). In this case the diffusion is restricted to the plane



(a) (b) (c)

Figure 1.3: Three different types of diffusion. (a) Cigar shaped diffusion. (b) Disc shaped diffusion. (c) Spherical diffusion.

spanned by V_1 and V_2. If the diffusion along all axes is almost equal $(\lambda_1 \approx \lambda_2 \approx \lambda_3)$ we have a sphere-shaped diffusion, see Figure 1.3(c). In this case there is no prominent direction of the diffusion. It is customary to refer to the last case as isotropic diffusion, whereas the cigar- and disc-shaped diffusions are anisotropic diffusions.

1.2.1 Tensor estimation

The DTI data received from the MR scanner is K direction-specific measurements $\{S_k\}_{k=1}^K$, where each S_k is a 3D MR image. In addition there is a nonweightened image S_0 called the baseline image. In order to calculate the tensor D we apply the Stejskal-Tanner equation [85, 86],

$$S_k = S_0 e^{-bg^{(k)^T} D g^{(k)}}, \quad k = 1, 2, ..., K. \tag{1.2}$$

Here $g^{(k)} \in R^3$ is the direction associated with S_k, and $b > 0$ is a scalar depending on acquisition time and strength of the magnetic field [91].

Since $D \in R^{3\times 3}$ is symmetric it has 6 degrees of freedom, and at least 6 direction-specific measurements are necessary in addition to the baseline image. However, to increase the image quality it is customary to increase the number of directions (typically 12-50), and solve (1.2) as a minimization problem. This will reduce the noise in the tensor, but the increased number of directions will increase the scanner time for the patient. A different strategy to improve the image quality is to increase the number of images in each direction. Thus rather than increasing the number of directions we instead increase the number of images from each of the 6 necessary directions. For example we can take 20 images from a single direction, and then average these images to produce a better result. Again this will increase the scanner time for the patient. In the papers [7, 32, 33, 54] we propose different mathematical methods for noise removal from DTI images. The idea is

that instead of increasing the scanner time we instead use the low quality images and post-process them to improve the image quality.

One way to solve (1.2) is to use least square minimization [8, 99]. Taking the logarithm on both sides we get the least squares minimization problem

$$\min_D \sum_{k=1}^{K} (\ln(S_k) - \ln(S_0) + b g^{(k)^T} D g^{(k)})^2 = \min_D \sum_{k=1}^{K} (c_k + g^{(k)^T} D g^{(k)})^2, \quad (1.3)$$

where $c_k = \frac{\ln S_k - \ln S_0}{b}$. Defining

$$G^{(k)} = \left[(g_1^{(k)})^2 \;\; 2g_1^{(k)} g_2^{(k)} \;\; 2g_1^{(k)} g_3^{(k)} \;\; (g_2^{(k)})^2 \;\; 2g_2^{(k)} g_3^{(k)} \;\; (g_3^{(k)})^2 \right], \quad (1.4)$$

and letting d be the vector

$$d = \left[D_{11} \;\; D_{12} \;\; D_{13} \;\; D_{22} \;\; D_{23} \;\; D_{33} \right], \quad (1.5)$$

we get the well known form of the linear least square problem

$$\min_{d_k} \sum_{k=1}^{K} (c_k + G^{(k)} d_k)^2 = \min_d ||c + Gd||^2. \quad (1.6)$$

Here G is the $n \times 6$ matrix with row number k equal $G^{(k)}$.

1.2.2 Anisotropic measures

In highly structured tissue like the white matter in the brain, diffusion is highly anisotropic. Thus the anisotropy is an important property in diffusion tensor images. In order to measure the anisotropy, several different measures have been proposed [99]. Among the most popular are the relative anisotropy (RA) and the fractional anisotropy (FA) [6]. Both of these methods are based on the normalized variance of the eigenvalues. An advantage of these measures is that they can be calculated without first calculating the eigenvalues. We have that the relative anisotropy (RA) is given as

$$\begin{aligned} RA &= \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_1 - \lambda_3)^2}}{\lambda_1 + \lambda_2 + \lambda_3} \\ &= \frac{\sqrt{3}}{\sqrt{2}} \frac{|D - \frac{1}{3}\text{trace}(D) I|}{\text{trace}(D)} \end{aligned} \quad (1.7)$$

and the fractional anisotropy as

$$\begin{aligned} FA &= \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_1 - \lambda_3)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \\ &= \frac{\sqrt{3}}{\sqrt{2}} \frac{|D - \frac{1}{3}\text{trace}(D) I|}{|D|}. \end{aligned} \quad (1.8)$$

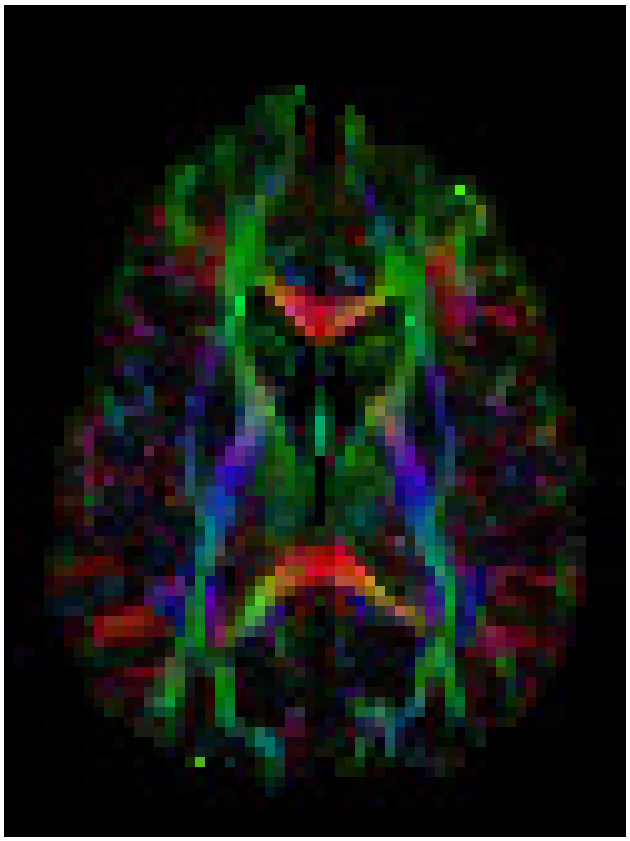


Figure 1.4: The image shows a slice of a color coded FA plot of the authors brain. The visualization is done using the DTI software DTIStudio [68].

The trace is calculated as the sum of the diagonal elements, the norm is calculated as the square root of the sum of the squared elements of the tensor and I is the identity matrix. The scaling ensures that the measures are in the range from 0 to 1.

In Figure 1.4 we show a slice of a color coded FA image of the authors brain. The diffusion directions are color coded as described by Pajevic and Pierpaoli [78].

Chapter 2

Segmentation

In computer vision, segmentation is the process of dividing a digital image into different regions. The purpose of this is to represent the image in a manner which is easier to analyze. For example if a computer is going to read the number plate from an image of a car it needs to separate the plate from the rest of the car.

In general this is done by grouping regions where the pixels have similar characteristics, like color, intensity or texture. There are several ways to perform a segmentation of an image, and among popular ones are methods based on clustering, histograms, region-growing and PDEs [25, 44]. In this thesis we will focus on intensity based PDE methods.

2.1 PDE based image segmentation

In this section we start with an introduction to the level set formulation, which is fundamental for the development of PDE based image segmentation. We then present one of the most well known PDE segmentation algorithms, the CV segmentation, developed by Chan and Vese [28]. Finally we present the Piecewise Constant Level Set Method (PCLSM) developed by Lie et al. in a series of papers [55, 56, 58].

2.1.1 Level Set Formulation

The level set method was developed by Osher and Sethian [77] as a simple and versatile method for computing and analyzing the motion of an interface Γ in two or three dimensions. Since its introduction it has become an important tool in many branches of applied mathematics. For an overview of the field we refer to the book of Osher [76], the book of Sethian [84] and the book of Chan [25].

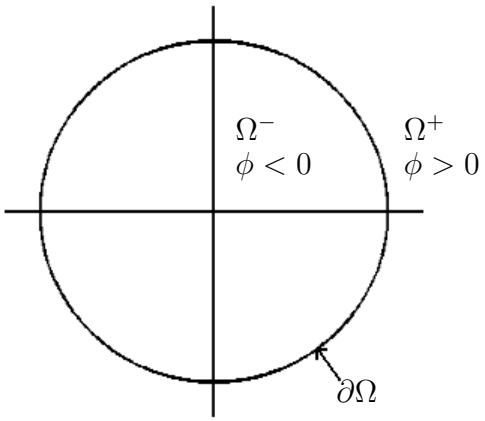


Figure 2.1: Implicit representation of the curve $x^2 + y^2 = 1$. The interface $\partial\Omega$ is implicitly defined as the zero isocontour of the function.

Their idea was to represent the interface by an implicit given function. This is easier than working directly on an explicit definition of the interface. As an example, consider the unit circle displayed in Figure 2.1. Instead of working with an explicit interface definition, which would require all points on the curve to be specified, we instead use the implicit definition $\phi(\vec{x}) = x^2 + y^2 - 1$. The interface $\partial\Omega = \{\vec{x} \mid |\vec{x}| = 1\}$ can now be defined by the $\phi(\vec{x}) = 0$ isocontour. The exterior region $\Omega^+ = \{\vec{x} \mid |\vec{x}| > 1\}$ is defined by $\phi(\vec{x}) > 0$ and the interior region $\Omega^- = \{\vec{x} \mid |\vec{x}| < 1\}$ is defined by $\phi(\vec{x}) < 0$

It should be noted that for the circle it would be easy to give an explicit definition, however for general curves this can be a hard task. In general this would require a parametrization of the curve. Note also that for complicated interfaces no analytical representation is usually known, and in these cases discretization must be used for both the explicit and implicit definitions.

It is also common to require the implicit function $\phi(\vec{x})$ to be a signed distance function. This prevents the function from becoming too steep or flat, and also possess several new properties [76].

Suppose that the velocity of each point on the implicit surface is given as $\vec{V}(\vec{x}) =< u, v, w >$. Given this velocity field we wish to move all the points on the surface $\phi(\vec{x}) = 0$ according to the velocity field. This can be done using the

partial differential equation

$$\frac{\partial \phi}{\partial t} + \vec{V} \cdot \nabla \phi = 0, \tag{2.1}$$

also known as the convection equation. In the context of level set modelling, this equation is also sometimes referred to as the level set equation.

If the velocity field $\vec{V}$ depends directly on the level set function ϕ, it is referred to as a self generated velocity field. A popular self generated velocity field is the motion by mean curvature (MMC). In this case the velocity field is such that the interface moves in the normal direction with a velocity proportional to its curvature, i.e. $\vec{V} = -b\kappa\vec{N}$, where $b > 0$ is a constant, κ is the curvature and $\vec{N}$ is the normal vector. In general the curvature is given as

$$\kappa = \nabla \cdot \Big(\frac{\nabla \phi}{|\nabla \phi|}\Big), \tag{2.2}$$

however in the case where the level set function ϕ is a signed distance function the curvature reduces to $\kappa = \Delta\phi$.

One of the strong sides of the level set representation is that the level curve can merge or break up with no special treatment [76].

2.1.2 Chan-Vese Segmentation

Active contour models are well known in image segmentation. The basic idea in active contour models is to evolve a curve in order to detect an object in an image u_0. Initially we have a curve around the object to be detected, and the curve then moves normal to itself and stops at the boundary of the object. This is done by associating a parametrized curve $\mathcal{C}(I) : [0,1] \rightarrow R^2$ and minimize the energy given by

$$E(\mathcal{C}) = \alpha \int_0^1 |\mathcal{C}'(s)|^2 \, ds + \beta \int_0^1 |\mathcal{C}''(s)|^2 \, ds - \lambda \int_0^1 |\nabla u_0(\mathcal{C}(s))|^2 \, ds, \tag{2.3}$$

where α, β and λ are positive parameters. The first two terms control the smoothness of the contour, while the third term is an edge detector attracting the contour towards the object in the image.

The idea was initially presented by Kass et al. [45] and has become very popular. The model has later been extended to level sets [18, 66].

The active contour models [18, 45, 66] all use an edge detector to stop the evolving curve at the boundary of the object. In 1999 Chan and Vese [28, 27] presented a new active contour model without the edge detector. Instead they used

a stopping criteria based on the Mumford-Shah segmentation technique [74]. The basic Mumford-Shah is to minimize the energy given by

$$E(u,\Gamma) = \int_{\Omega} |u-u_0|^2\,dx + \beta|\Gamma| + \nu \int_{\Omega\setminus\Gamma} |\nabla u|^2\,dx, \tag{2.4}$$

where β and ν are positive parameters. Here Ω is the image region, and Γ is a closed subset of Ω given by the union of a finite number of curves. It represents the edges in the image u_0. The length of Γ is given as $|\Gamma|$. The goal is to find a piecewise smooth approximation u to u_0, with discontinuities only along Γ. In (2.4) the first term is a fidelity term ensuring that the recovered image u is close to the original image u_0. The two last terms are regularizers measuring the curve length and the smoothness of u in $\Omega \setminus \Gamma$.

Based on (2.4), Chan and Vese proposed to minimize the following energy functional for two phase (object and background) segmentation

$$\begin{aligned} E(c_1,c_2,\Gamma) &= \int_{\text{inside }\Gamma} |u_0-c_1|^2\,dx + \int_{\text{outside }\Gamma} |u_0-c_2|^2\,dx \\ &+ \beta|\Gamma| + \nu\cdot\text{Area(inside }\Gamma). \end{aligned} \tag{2.5}$$

Here c_1 and c_2 are constants defined by; c_1 = mean(u_0) inside Γ and c_2 = mean(u_0) outside Γ. The curve Γ is implicitly defined through a level set function satisfying

$$\begin{cases} \phi(x) > 0 & \text{if x is inside } \Gamma, \\ \phi(x) = 0 & \text{if x is at } \Gamma, \\ \phi(x) < 0 & \text{if x is outside } \Gamma. \end{cases}$$

The idea is that inside the object ϕ should be positive, and outside the object ϕ should be negative, thus the zero level set $\phi = 0$ defines the boundary between the object and the background. In (2.5) the two first terms are fidelity terms, while the two last terms are regularizers.

The length of Γ and the area inside Γ can be defined as

$$|\Gamma| = \int_{\Omega} |\nabla H(\phi(x))|\,dx = \int_{\Omega} \delta(\phi(x))|\nabla\phi(x)|\,dx, \tag{2.6}$$

$$\text{Area}\{\phi \geq 0\,\} = \int_{\Omega} H(\phi(x))\,dx. \tag{2.7}$$

Here $H(\phi)$ denotes the Heaviside function

$$H(\phi) = \begin{cases} 1, & \phi > 0, \\ 0, & \phi \leq 0, \end{cases}$$

and δ is the delta function

$$\delta(\phi) = \begin{cases} 1, & \phi = 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Thus (2.5) can be rewritten as

$$\begin{aligned} E(c_1, c_2, \phi) &= \int_\Omega |u_0 - c_1|^2 H(\phi)\, dx + \int_\Omega |u_0 - c_2|^2 (1 - H(\phi))\, dx \\ &+ \beta \int_\Omega \delta(\phi) |\nabla \phi|\, dx + \nu \int_\Omega H(\phi)\, dx. \end{aligned} \tag{2.8}$$

Parametrizing the descent direction by an artificial time $t > 0$, the functional can be minimized using a gradient descent method on ϕ

$$\phi_t = -\frac{\partial E}{\partial t}. \tag{2.9}$$

This gives the following PDE to solve

$$\phi_t = \delta_\epsilon \Big[-(u_0 - c_1)^2 + (u_0 - c_2)^2 - \beta \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} - \nu \Big], \tag{2.10}$$

with proper initial and boundary conditions [28]. Here δ_ϵ denotes a smooth approximation to δ.

One problem with the original Chan-Vese model is that it is not able to distinguish between more than two phases, i.e. object and background. If the image has several objects we need a modified model. In [93] Chan and Vese proposed a multiphase model based on their old model. The idea is very simple. Instead of applying only a single level set they proposed to use several level sets. In fact $\log_2 N$ level sets can represent N phases. The principle is that two level sets can identify four regions by the four possibilities $\phi_i > 0$, $\phi_i < 0$, $i = 1, 2$. More details can be found in [93].

2.1.3 PCLSM Segmentation

In a series of papers Lie et al. propose a different model for the multiphase image segmentation [55, 56, 58]. While the multiphase Chan-Vese model requires $\log_2 N$ level sets to identify N phases, the model proposed by Lie et al. only uses a single level set. The model is called Piecewise Constant Level Set Method (PCLSM) and use different levels in the level set curve to identify different regions.

Assume that we need to identify N regions $\{\Omega_i\}_{i=1}^N$ which form a partition of Ω. In order to find the regions, we want to find a piecewise constant function which takes values

$$\phi = i \text{ in } \Omega_i, \quad i = 1, 2, \ldots, N. \tag{2.11}$$

The discontinuities of ϕ give the curves that separate the regions. Associated with ϕ we define the characteristic functions ψ_i for Ω_i as

$$\psi_i = \frac{1}{\alpha_i} \prod_{\substack{j=1 \\ j \neq i}}^{N} (\phi - j) \text{ with } \alpha_i = \prod_{\substack{k=1 \\ k \neq i}}^{N} (i - k). \tag{2.12}$$

Consequently the characteristic functions ψ_i will have the property

$$\psi_i(\mathrm{x}) = \begin{cases} 1 & \text{if } x \in \Omega_i \\ 0 & \text{elsewhere} \end{cases}, \tag{2.13}$$

as long as (2.11) holds.

The length of the boundary of Ω_i is given by the relation

$$|\partial \Omega_i| = \int_{\Omega} |\nabla \psi_i| dx. \tag{2.14}$$

By linearly combining these characteristic functions we are able to build a cartoon or a piecewise constant image,

$$u = \sum_{i=1}^{n} c_i \psi_i. \tag{2.15}$$

This is a piecewise constant function, and $u = c_i$ in Ω_i if ϕ is as given in (2.11).

In order to guarantee that the level set function ϕ takes the values as in (2.11), the following constraint function is defined

$$K(\phi) = (\phi - 1)(\phi - 2) \cdots (\phi - N) = \prod_{i=1}^{N} (\phi - i). \tag{2.16}$$

Requiring $K(\phi) = 0$ at convergence ensures that ϕ only takes integer values.

Based on this Lie et al. proposed to solve the following functional for the multiphase segmentation of an image u_0:

$$\min_{\substack{\mathbf{c},\, \phi \\ K(\phi)=0}} \Big\{ F(\mathbf{c}, \phi) = \int_{\Omega} |u - u_0|^2 dx + \beta \sum_{i=1}^{n} \int_{\Omega} |\nabla \psi_i| dx \Big\}. \tag{2.17}$$

In the above, β is a nonnegative parameter controlling the regularizing, u is a piecewise constant function depending on ϕ and $\boldsymbol{c}$, as in (2.15). The first term

of (2.17) is a least square functional, measuring how well the piecewise constant image u approximates u_0. The second term is a regularizer measuring the length of the edges in the image u_0.

Other applications of the model can be found in [52, 53] and a binary version of the method is presented in [57].

In [34, 88] we have proposed fast methods for the solution of (2.17). These methods are based on operator splitting [61, 62, 98], and increase the computational efficiency compared to the augmented Lagrangian methods presented in [55, 56, 58]. Other computational efficient methods have also been proposed [89].

Chapter 3

Image Denoising

Denoising is the task of removing unwanted noise from a signal. The presence of noise in images is unavoidable. It may be introduced during formation, recording or transmission. Further processing of the image often requires that the noise must be removed or at least reduced. Even a small amount is harmful when high accuracy is required. Over the last decades, a variety of methods have been proposed. We have filtering methods, transform based methods, variational methods and techniques based on the solution of partial differential equations (PDEs). For an overview we refer to the textbook of Chan [25].

There are several types of image noise. If pixels are lost during transmission we get what is known as salt and pepper noise. The lost pixels are randomly set to black or white pixels and thus the color bears no relation to the color of the surrounding pixels, see Figure 3.1(b). Usually this type of noise will only affect a small number of image pixels. Methods based on median filtering is usually effective on this kind of noise [21].

In Gaussian noise, an amount of noise is added to every part of the picture. Each pixel in the image will be changed from its original value by a small amount, see Figure 3.1(c). The noise has a probability density function as a Gaussian distribution. This is the most common noise and can be produced by the thermal agitation of charged carriers (usually the electrons) inside an electrical conductor.

In this chapter we review several methods for the removal of Gaussian noise from images. In addition we show how several of these methods can be extended and used for denoising of diffusion tensor images [7, 32, 33, 54].

All the visualization and numerical examples in this section is done in Matlab [90].

(a) (b) (c)

Figure 3.1: The figure shows two types of image noise. (a) The noise free image. (b) The image corrupted by salt and pepper noise. (c) The image added Gaussian noise.

3.1 PDE denoising

Partial differential equations are used extensively in image processing and equations similar to the heat equation can be used for noise removal. Traditionally, linear models have been used for PDE denoising. One approach is to use a Gaussian filter, or equivalently solving the heat-equation with the noisy image as input-data, i.e. a linear, 2nd order PDE-model. For some purposes this kind of denoising is adequate. One big advantage of linear noise removal models is the speed. But a disadvantage of the linear models is that they are not able to preserve edges in a good manner, i.e. edges are smeared out. Nonlinear models on the other hand can handle edges in a much better way than linear models can. One popular model for nonlinear image denoising is the Total Variation (TV)-filter, introduced by Rudin, Osher and Fatemi [81]. This filter is good at preserving edges, but smoothly varying regions in the input image are transformed into piecewise constant regions in the output image. Using the TV-filter as a denoiser leads to solving a 2nd order nonlinear PDE.

3.1.1 Laplacian

A simple model for denoising of an image u would be to solve the Laplacian regularization functional with a L_2 fidelity functional. That is we minimize the energy given by

$$E_L(u, \beta) = \int_\Omega |\nabla u|^2 \, dx + \frac{\beta}{2} \int_\Omega (u - f)^2 \, dx, \tag{3.1}$$

where β is a positive parameter controlling the regularization. The first term is a regularization functional measuring the smoothness of the image. The last term is

a least square functional measuring the fitness of the estimated solution compared to the input data.

By differentation we get the Euler-Lagrange equation corresponding to the minimization problem (3.1),

$$\frac{\partial E_L}{\partial t} = -\nabla^2 u + \beta(u - f). \tag{3.2}$$

At a minimum of the functional we have

$$\frac{\partial E_L}{\partial t} = 0. \tag{3.3}$$

There are several ways to solve (3.3). We can for example solve it directly with a fixed-point iteration, however this is usually tedious to carry out numerically. A simpler and common way is to parametrize the descent direction by an artificial time $t > 0$ and use a gradient descent method on u

$$u_t = -\frac{\partial E_L}{\partial t}. \tag{3.4}$$

This gives the following PDE to solve

$$\begin{aligned} u_t &= \nabla^2 u - \beta(u - f) \text{ in } \Omega, \\ \frac{\partial u}{\partial n} &= 0 \text{ on } \partial\Omega, \\ u(x, 0) &= f, \end{aligned} \tag{3.5}$$

which is the heat equation with an additional fidelity term controlling the regularization. The boundary conditions are so-called natural boundary conditions, i.e. they appear naturally in the minimization problem [42].

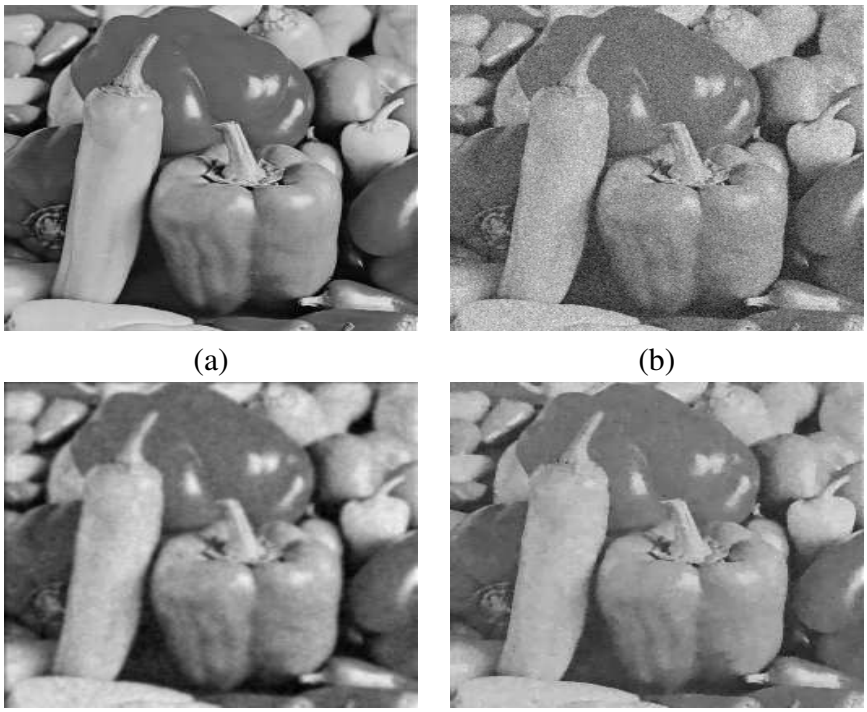
In Figure 3.2(c) we show the performance of the method on a test image. The problem with this model is that the edges are smeared out. This is due to the Laplacian operator ∇^2. We have an isotropic diffusion which smooths equally in all directions, and thus also the edges. Formally the solution space for (3.5) is the Sobolev space $W^{1,2}(\Omega)$ which is defined by

$$W^{1,2}(\Omega) = \{u \ : \ u \in L_2(\Omega), \nabla u \in L_2(\Omega)^d\}, \tag{3.6}$$

where d is the dimension of the domain.

3.1.2 ROF

The problem with the Laplacian model is that it smooths the edges. As can be seen in Figure 3.2(c) we have blurred the image in the regularization process. In



(a) (b)

(c) (d)

Figure 3.2: The figure shows denoising of the "Peppers" image using the Laplacian and the ROF model. (a) The noise free image. (b) The image added Gaussian noise. (c) The denoised image using the Laplacian. (d) The denoised image using the ROF model.

order to fix this problem we need a model which allows for discontinuities in the solution.

In 1992 Rudin, Osher and Fatemi [81] proposed the ROF model based on the total variation (TV)-norm. The TV-norm for scalar valued data is defined as

$$TV[u] = \int_{\Omega} |\nabla u| \, dx \tag{3.7}$$

The ROF model is one of the most famous and powerful variational and PDE based image denoising models. The method removes the noise, while preserving the edges in images. The minimization problem is given as

$$\min_{u} E_{ROF}(u) = \int_{\Omega} |\nabla u| \, dx + \frac{\beta}{2} \int_{\Omega} (u - f)^2 \, dx. \tag{3.8}$$

As in the Laplacian energy, β controls the regularization, the first term is a regularization functional and the last term is a fidelity term. The difference is that

the first term has been replaced with the total variation norm. The advantage of this is that the solution is in the space of functions with bounded variation (BV). This is an ideal choice for images since BV allows sharp discontinuities (edges). Formally the space BV is defined as

$$BV = \{u \ : \ u \in L^1(\Omega) \ , \ \int_\Omega |\nabla u| < \infty)\}. \tag{3.9}$$

Thus, BV functions amount to L^1 functions with bonded TV-norm [23].

By differentation we get the Euler-Lagrange equation

$$\frac{\partial E_{ROF}}{\partial t} = -\nabla \cdot \frac{\nabla u}{|\nabla u|} + \beta(u - f). \tag{3.10}$$

Using steepest descent marching with artificial time t we get the following PDE to solve

$$\begin{aligned} u_t &= \nabla \cdot \frac{\nabla u}{|\nabla u|} - \beta(u - f) \ \text{ in } \Omega, \\ \frac{\partial u}{\partial n} &= 0 \ \text{ on } \partial\Omega, \\ u(x, 0) &= f. \end{aligned} \tag{3.11}$$

Figure 3.2(d) shows the performance of the TV denoising on a test image. Even though the ROF model is superior to the Laplacian, it has a few drawbacks. It has problems with staircasing, i.e. artificial steps or discontinuities are introduced in the denoised image. The problem occurs in the homogeneous regions in the image. To overcome this problem Blomgren et al. [11] proposed to let the exponent in the regularization term depend on the image. They proposed the functional

$$E(u) = \int_\Omega |\nabla u|^{p(|\nabla u|)} \, dx + \frac{\beta}{2} \int_\Omega (u - f)^2 \, dx, \tag{3.12}$$

where p is a decreasing function such that $p(0) = 2$ and $p(x) = 1$ for all x greater than some M. This adapts the smoothing to the gradient in the image. The idea is to use isotropic diffusion in homogeneous regions and TV diffusion near edges. Similar approaches can be found in [30, 83]. Higher order methods to reduce staircasing can be found in Lysaker et al. [63, 64].

In the ROF model (3.8) L_2 norm is used in the fidelity term. Usually this is the best choice, however in some situation L_1 norm can produce better results [22].

3.1.3 Color TV

The ROF model is only suited for denoising of scalar valued (grey scaled) images. Since color images consist of three different color channels (red, green and blue)

we need a model which can handle this. One choice is to denoise each color channel separately using the ROF model, however this has shown to produce poor results due to the lack of coupling between the color channels [10].

In 1996 Blomgren and Chan [10] proposed a total variation norm for vector valued functions. This made it possible to apply total variation regularization on color and other vector valued images. The vector valued TV-norm is defined as

$$TV[\mathbf{u}] = \sqrt{\sum_{i=1}^{m} \big(TV[u_i]\big)^2}, \tag{3.13}$$

where u_i is vector element i. For color images the elements are the color channels. Based on the vector valued TV-norm Blomgren and Chan proposed the following regularization functional for vector images

$$E(\mathbf{u}) = TV[\mathbf{u}] + \frac{\beta}{2} \sum_{i=1}^{m} \int_{\Omega} (u_i - f_i)^2 \, dx \tag{3.14}$$

As with the standard ROF model the minimization problem can be solved using steepest descent marching with artificial time t:

$$\frac{\partial u_i}{\partial t} = \alpha_i \nabla \cdot \frac{\nabla u_i}{|\nabla u_i|} - \beta(u_i - f_i) \text{ in } \Omega, \tag{3.15}$$

$$\frac{\partial u_i}{\partial n} = 0 \text{ on } \partial\Omega, \tag{3.16}$$

$$u_i(x, 0) = f_i. \tag{3.17}$$

with

$$\alpha_i = \frac{TV[u_i]}{TV[\mathbf{u}]}. \tag{3.18}$$

The weight α_i in (3.15) acts as a coupling between the channels. The coupling has the implication that a channel with large TV will be smoothed more than a channel with a small TV. This ensures that the model does not wipe out the weaker channels. Thus the coupling balances how much each channel is smoothed.

It should be noted that for color images better results can be achieved by applying the ROF model on the CB and HSV color models [26].

3.1.4 Matrix TV

In diffusion tensor imaging noise is a common problem. The MRI data from the scanner contains measurement noise which degrades the quality of the images. In order to use the ROF model on DTI data we need to extend it to matrix data. As

explained in chapter 1, DTI data contains a 3×3 symmetric and positive definite matrix D in every voxel. In [32] we have proposed an extension of the color TV model for denoising of matrix fields. The matrix valued TV-norm of a matrix D is defined as

$$TV[D] = \sqrt{\sum_{ij} \left(TV[d_{ij}]\right)^2}, \tag{3.19}$$

where d_{ij} denotes the elements of the matrix. Since the diffusion tensor matrix D must be symmetric and positive definite we need to incorporate this into the model. This can easily be done if we write D as the product

$$D = LL^T, \tag{3.20}$$

where L is a lower triangular matrix. This is called the Cholesky factorization and has the implications that D will be symmetric, positive definite and have orthogonal eigenvectors. These properties are exactly what we need.

Based on this we propose to solve the following minimization problem for regularization of matrix valued DTI data:

$$\min_{\ell_{kl}} \left\{ \sqrt{\sum_{ij} \left(\text{TV}[d_{ij}(\ell_{kl})]\right)^2} + \frac{\beta}{2} \sum_{ij} \int_\Omega \left[d_{ij}(\ell_{kl}) - \hat{d}_{ij}\right]^2 \right\}, \tag{3.21}$$

where $\{kl\} \in \{11, 21, 31, 22, 23, 33\}$. Here $d_{ij}(\ell l_{kl})$ are the matrix elements of D as a function of its Cholesky factors L, $\hat{d}_{ij}$ denotes the elements of the tensor estimated from the noisy data and β is a positive parameter controlling the regularization. Using steepest descent marching we get the following PDEs to solve

$$\frac{\partial \ell_{kl}}{\partial t} = \sum_{ij} \alpha_{ij} \nabla \cdot \left(\frac{\nabla d_{ij}}{|\nabla d_{ij}|} \right) \frac{\partial d_{ij}}{\partial \ell_{kl}} - \beta \sum_{ij} \left(d_{ij} - \hat{d}_{ij} \right) \frac{\partial d_{ij}}{\partial \ell_{ij}}, \tag{3.22}$$

with

$$\alpha_{ij} = \frac{\text{TV}[d_{ij}]}{\text{TV}[D]}. \tag{3.23}$$

As in the vector case the weight term α_{ij} acts as a coupling between the elements in the matrix. This balances the smoothing and ensures that the model does not wipe out the weaker elements.

This work is related to the functional proposed by Wang et al. [94]. However, while Wang et al. regularize the elements of the lower triangular matrix L we still regularize the elements of the full diffusion tensor D. We claim that regularizing the elements of D is more direct than regularizing the elements of L. In addition there is no coupling between the matrix elements in Wang's method. More details and numerical results can be found in [32].

It should be noted that the noise in DTI data is not Gaussian, but Rician [65]. The reason for this is that the DTI data is recorded in the frequency domain and results in complex data. If the real and imaginary components of the signal are assumed to have Gaussian noise, the resulting magnitude image will have Rician distributed noise. However, it is still customary to evaluate the models using Gaussian noise. To confirm that the models also handle Rician noise we always perform tests on real data.

3.1.5 Scale Space formulation

The PDE denoising methods presented so far have been regularization functionals on the form:

$$E(u) = R(u) + \frac{\beta}{2}F(u, f), \tag{3.24}$$

where $R(u)$ is a geometric regularization functional measuring the smoothness of the estimated solution and $F(u, f)$ is a fidelity functional measuring the fitness of the estimated solution . The positive scalar β balances between the regularization term and the fidelity term and hence controls the smoothness of the solution. Solving the minimization problem using steepest decent marching results in a PDE which is solved to steady state.

A different strategy is to remove the fidelity term and instead solve the PDE a given number of time steps. Thus instead of controlling the smoothness of the solution with the scaling of the fidelity term, the number of time steps controls the smoothness. This is usually referred to as scale space formulation, since every time step correspond to a given scale of the image. Initially we have the finest scale with noise and fine details present in the image. As time increases we get a coarser scale where the noise and fine details are erased.

A famous scale space model is the Perona-Malik (PM) [79], which is given as

$$u_t = \nabla \cdot \big(g(|\nabla u|)\nabla u\big), \tag{3.25}$$

where $g(\cdot)$ has to be a non-negative monotonically decreasing function with $g(0) = 1$. Notice that for $g(\cdot) = 1$ equation (3.25) reduces to the heat equation, $u_t = \nabla^2 u$, and thus will be highly diffusive. The expression inside the function, $|\nabla u|$, acts as an edge estimator. Since $g(\cdot)$ is monotonically decreasing this will ensure that the diffusion will mainly take place in the interior regions and will not affect the region boundaries since the magnitude of $|\nabla u|$ is large here.

There are many choices of $g(\cdot)$. If we set

$$g(|\nabla u|) = \frac{1}{|\nabla u|}. \tag{3.26}$$

we get a scale space version of the ROF model, i.e. without the fidelity term. However, in general the scale space PDEs do not correspond to a given variational problem.

Originally Perona and Malik applied two different functionals. The first,

$$g(|\nabla u|) = e^{-(|\nabla u|/K^2)}, \tag{3.27}$$

privileges high-contrast edges over low contrast ones. The second,

$$g(|\nabla u|) = \frac{1}{1 + \left(\frac{|\nabla u|}{K}\right)^2}, \tag{3.28}$$

privileges wide regions over smaller ones. K is a constant depending on the noise level. In addition to denoising, these models are also used for edge detection and multiscale representations of images.

Weickert and his coworkers have also contributed significantly to the field of scale space imaging [95, 96, 97]. They have studied isotropic and anisotropic diffusion and introduced the framework of structured tensors for regularization of matrix valued data.

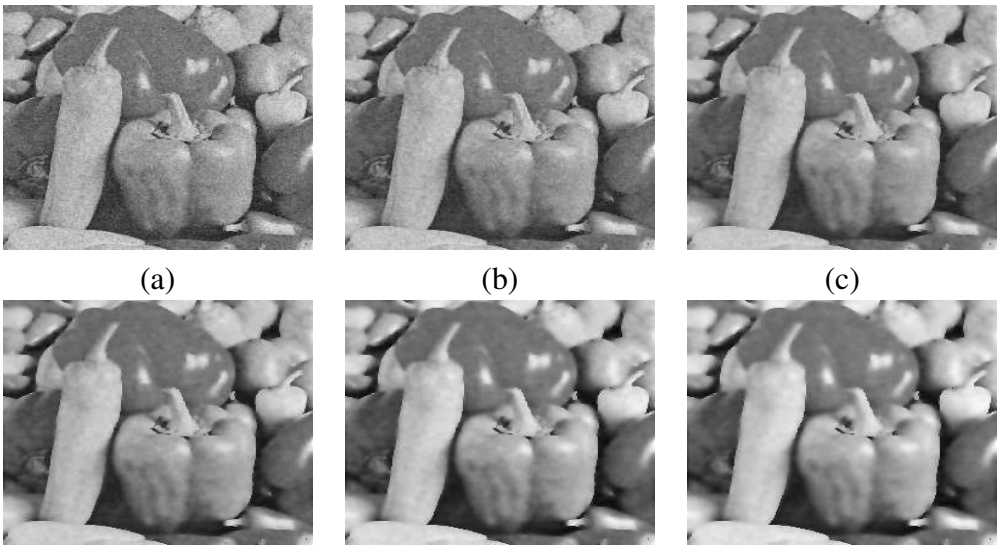
In Figure 3.3 we show a multiscale representation of a noisy test image. Notice that as time evolves the noise and fine details are erased. Thus if the goal is to remove the noise, the challenge is to find a suitable stopping time. The choice of stopping time is crucial. If we stop too early the image is still noisy, on the other if we stop too late the image will be oversmoothed. The stopping time is related to the noise level in the image, and models for finding good stopping criteria can be found in [72, 73].

3.1.6 Algebraic Method

In 2007 Burgeth et al. [14, 15] proposed an operator-algebraic approach for filtering of matrix fields. They developed a generic framework to find matrix valued counterparts of the scalar valued PDEs. The framework was successfully applied to several PDEs.

The framework exploits operator-algebraic properties of (symmetric) matrices to establish truly matrix-valued PDEs. The underlying idea is that to a certain extent symmetric matrices can be regarded as a generalization of real numbers. This is used to generalize notations like functions of matrices, derivatives and gradients to a matrix-valued setting.

Define $\mathrm{Sym}_n(\mathbb{R})$ as the subset of symmetric matrices and $\mathrm{Sym}_n^+(\mathbb{R})$ as the set of positive definite matrices.



(a) (b) (c)

(d) (e) (f)

Figure 3.3: The figure shows the scale space using the Perona-Malik model (3.27). The time is increasing from (a) to (f). Notice that fine details are disappearing during the flow.

The definition of a function of a symmetric matrix U is defined as

$$g(U) = V^T \text{diag}(g(\lambda_1), ..., g(\lambda_n))V, \tag{3.29}$$

where $U = V^T \text{diag}(g(\lambda_1), ..., g(\lambda_n))V$ is the spectral decomposition of U and g is a real function.

The product of two symmetric matrices is in general not symmetric. Several symmetric products are available, one is the Jordan product

$$A \bullet B = \frac{1}{2}(AB + BA). \tag{3.30}$$

The matrix-valued version of a difference quotient leads to a component-wise definition of spatial or temporal partial derivatives [14, 15]

$$\overline{\partial}_\omega U(\omega_0) = \lim_{h \to 0} \frac{U(\omega_0 + h) - U(\omega_0)}{h} = \left(\partial_\omega u_{ij}(\omega_0)\right)_{ij}. \tag{3.31}$$

Other linear operators like the arithmetic mean and convolution with a kernel are also interpreted component-wise in the matrix setting.

The set of spatial partial derivatives forms the spatial **gradient** for matrix fields $\overline{\nabla} U \in (\text{Sym}_n(\mathbb{R}))^d$

$$\overline{\nabla} U(x) = \left[\overline{\partial}_{x_1} U(x), \overline{\partial}_{x_2} U(x), \cdots, \overline{\partial}_{x_d} U(x)\right]^T. \tag{3.32}$$

The generalized divergence is defined similarly by

$$\overline{\text{div}}\ (A(x))^\top := \sum_{i=1}^{d} \overline{\partial}_{x_i} A_i(x)$$

for $A(x) := \big(A_1(x), \ldots, A_d(x)\big)$. The definition of a **Laplacian** for a matrix field U is straight forward:

$$\overline{\Delta} U := \sum_{i=1}^{d} \overline{\partial}_{x_i}^2 U$$

Finally, there is a notion of p-length in the module $(\text{Sym}_n(\mathbb{R}))^d$ realized for a vector $W = (W_1, \ldots, W_d) \in (\text{Sym}_n(\mathbb{R}))^d$ by

$$|W|_p := \sqrt[p]{|W_1|^p + \cdots + |W_d|^p} \in \text{Sym}_n^+(\mathbb{R}).$$

Using the framework the scale space version of the Laplacian

$$u_t = -\Delta u \tag{3.33}$$

can easily be generalized to the matrix PDE

$$\overline{\partial}_t U = -\overline{\Delta} U. \tag{3.34}$$

The scale space version of the ROF model

$$u_t = -\text{div}\left(\frac{\nabla u}{|\nabla u|}\right) \tag{3.35}$$

can similarly be generalized to a matrix valued setting by

$$\overline{\partial}_t U = -\overline{\text{div}}\left(\frac{1}{|\overline{\nabla} U|} \bullet \overline{\nabla} U\right), \tag{3.36}$$

where $|\overline{\nabla} U|$ stands for the 2-length of $\overline{\nabla} U$, that is, $|\overline{\nabla} U| := |\overline{\nabla} U|_2 = \sqrt{\sum_{i=1}^{d} |U_i|^2}$. The term $\frac{1}{|\overline{\nabla} U|}$ can be interpreted as the inverse of $|\overline{\nabla} U|$.

3.1.7 Primal-Dual Method

So far we have not addressed the most challenging task in image denoising; the computational speed. Usually the data processed is huge, especially when dealing with 3D DTI data. Thus it is important to design fast algorithms, or else the computation can take days.

A major problem with the ROF model is the highly nonlinear term, which causes convergence difficulties. As presented in section 3.1.2 the scalar ROF model can be solved through the PDE

$$u_t = \nabla \cdot \frac{\nabla u}{|\nabla u|} - \beta(u - f) = 0, \text{ in } \Omega. \tag{3.37}$$

A problem occurs at points where the term $|\nabla u|$ vanishes, since (3.37) degenerates at such points. To overcome this problem it is common practice to regularize the problem and instead solve the slightly modified problem

$$u_t = \nabla \cdot \frac{\nabla u}{\sqrt{|\nabla u|^2 + \epsilon}} - \beta(u - f) = 0, \text{ in } \Omega. \tag{3.38}$$

This equation is solved to steady state by explicit integration in time. However, severe restrictions on the size of the time-step, in addition to first order convergence in time, make this straight forward approach impractical for large-scale problems. The main problem is the highly nonlinear term $\frac{\nabla u}{\sqrt{|\nabla u|^2+\epsilon}}$. More details on the convergence behaviour can be found in the article by Chan et al. [24]. Here they also propose a primal dual method to overcome this problem. The dual variable

$$w = \frac{\nabla u}{|\nabla u|} \tag{3.39}$$

is introduced and the equivalent system of nonlinear partial differential equations

$$\begin{aligned} |\nabla u| w - \nabla u &= 0, \\ -\nabla \cdot w + \beta(u - f) &= 0, \end{aligned} \tag{3.40}$$

is solved using Newton's method. The advantage of this method is that no division by zero is explicitly performed in the formulation. This makes the problem better behaved and larger time steps can be used. This reduces the computational time.

3.1.8 Dual Method

It is possible to further extend the primal-dual method. The problem (3.40) depends on both the primal variable u and the dual variable w, hence the name primal-dual method. As described by Carter [17] it is possible to extend this problem further to a full dual problem. The total variation of u can be written in the following equivalent way

$$\text{TV}(u) = \int_\Omega |\nabla u| dx = \max_{|w| \leq 1} \int_\Omega u \left(\nabla \cdot w\right) dx. \tag{3.41}$$

Using this we can transform the ROF model (3.8) into the equivalent min/max problem

$$\min_u \max_{|w|\leq 1} \int_\Omega u(\nabla \cdot w)dx + \frac{\beta}{2}\int_\Omega (u-f)^2 dx. \tag{3.42}$$

The functional in (3.42) is convex in u and concave in w, and the region $\{w : |w| \leq 1\}$ is bounded [17]. Therefore we can interchange the order of the min and max operators to obtain

$$\max_{|w|\leq 1} \min_u \int_\Omega u(\nabla \cdot w)dx + \frac{\beta}{2}\int_\Omega (u-f)^2 dx. \tag{3.43}$$

Differentiating this expression with respect to u, and setting it to zero, we solve the innermost minimization problem, and write u as a function of w

$$\nabla \cdot w + \beta(u-f) = 0 \Rightarrow u = f - \frac{1}{\beta}\nabla \cdot w. \tag{3.44}$$

By substitution of (3.44) into (3.43) we get

$$\max_{|w|\leq 1} \int_\Omega (f - \frac{1}{\beta}\nabla \cdot w)\nabla \cdot w dx + \frac{\beta}{2}\int_\Omega (f - \frac{1}{\beta}\nabla \cdot w - f)^2 dx. \tag{3.45}$$

Rearranging terms and setting $\alpha = \frac{1}{2\beta}$ we get

$$\max_{|w|\leq 1} \int_\Omega \big(-\alpha(\nabla \cdot w)^2 + f\nabla \cdot w\big)dx. \tag{3.46}$$

This problem is known as the dual problem corresponding to the primal problem (3.8). In [17] the problem was solved directly using an interior-point primal-dual method with different relaxation methods.

3.1.9 Chambolle

A more efficient way to solve the dual problem was proposed by Chambolle [20]. Notice that we can write (3.46) in the following equivalent way

$$\min_{|w|\leq 1} \int_\Omega \big(\alpha(\nabla \cdot w)^2 - f\nabla \cdot w + f^2\big)dx = \min_{|w|\leq 1} ||\gamma\nabla \cdot w - f||, \tag{3.47}$$

where $\gamma = 2\alpha = 1/\beta$. Here we have added the constant function f and changed the sign into a minimization problem; none of these operations alter the solution. This is exactly the same problem as the one presented in the work of Chambolle, however he used the framework of convex optimization to deduce the minimization problem. Thus the dual problem of Carter is the same problem as the one

solved by Chambolle, however Chambolle found a very clever way to solve the problem. The corresponding Lagrangian system for the minimization problem (3.47) is

$$-\left(\nabla(\gamma\nabla\cdot w - f)\right)_{i,j} + \lambda_{i,j} w_{i,j} = 0, \tag{3.48}$$

where $\lambda_{i,j}$ are Lagrange multipliers associated to the constraint in (3.47) [75]. Chambolle now made the observation that we only have two situations; either $\lambda_{i,j} > 0$ and $|w_{i,j}| = 1$, or $|w_{i,j}| < 1$ and $\lambda_{i,j} = 0$. But in any case

$$\lambda_{i,j} = |(\nabla(\gamma\nabla\cdot w - f))_{i,j}|. \tag{3.49}$$

Thus we get the following gradient descent algorithm

$$w_{i,j}^{n+1} = w_{i,j}^{n} + \tau\big((\nabla(\nabla\cdot w^n - f/\gamma))_{i,j} - |(\nabla(\nabla\cdot w^n - f/\gamma))_{i,j}| w_{i,j}^{n+1}\big), \tag{3.50}$$

so that

$$w_{i,j}^{n+1} = \frac{w_{i,j}^{n} + \tau(\nabla(\nabla\cdot w^n - f/\gamma))_{i,j}}{1 + |\tau(\nabla(\nabla\cdot w^n - f/\gamma))_{i,j}|}. \tag{3.51}$$

Chambolle also proved that the algorithm always will converge as long as $\tau \leq 1/8$. Once the dual variable w^* is found from the steady state of (3.51) we easily get the primal variable u from the relation

$$u = f - \gamma\nabla w^* \tag{3.52}$$

The dual method has shown to increase the convergence speed considerably compared to the primal method [20].

3.1.10 Dual Matrix

As already mentioned computational efficiency is crucial when dealing with DTI or large scale matrix valued images. Since the dual formulation has shown to increase the convergence speed of the scalar ROF model significantly it would be beneficial to develop a dual method for regularization of matrix valued data. In [33] we propose to combine the operator algebraic framework of Burgeth et al. and the dual formulation to get a dual operator algebraic method for regularization of matrix valued data.

As explained in section 3.1.6 the framework of Burgeth et al. can be used to find matrix valued counterparts of the scalar valued PDEs. Using this framework we propose the following matrix valued projection scheme

$$\begin{aligned} W_{i,j,k}^{n+1} &= \left(W_{i,j,k}^{n} + \tau(\nabla(\overline{\text{div}}\left(W_{i,j,k}^{n}\right) - F/\gamma))_{i,j,k}\right) \\ &\bullet \left(I + |\tau(\nabla(\overline{\text{div}}\left(W_{i,j,k}^{n}\right) - F/\gamma))_{i,j,k}|\right)^{-1} \end{aligned} \tag{3.53}$$

for the regularization of matrix valued data. This projection scheme can be used as an algorithm by simply specifying the initial condition W^0 and a suitable stopping criteria. Thus the algorithm can be stated as

A Matrix Projection Algorithm

1. Set $W^0 = 0$, $\tau = 1/8$ and $n = 0$.
2. Find W^{n+1} from equation (3.53).
3. Check for convergence (e.g. $\max |W^{n+1} - W^n| < 1/100$).
   - *Either* set $n \leftarrow n + 1$ and repeat step (2) and (3),
   - *or* set $D = F - \gamma \overline{\text{div}} \left(W^{n+1} \right)$ and *stop*.

In [33] we show that proposed dual method increase the convergence speed substantially. Approximately $1/10$ of the number of iterations are needed for the dual formulation compared to the primal formulation.

3.1.11 Inverse Scale Space

Until now we have discussed two different types of PDEs for image denoising, the regularization methods and the scale space methods. In between these two approaches we have the inverse scale space formulation.

As explained in section 3.1.5 the traditional scale space formulation starts with a fine scale image and evolves this into coarser scales using a PDE. In the inverse scale space formulation we reverse this and start with the coarse scale image and evolve this into finer scales, see Figure 3.4. Even though this perhaps is more of theoretical than practical interest the field has got a lot of attention [13, 43, 59, 82].

For scalar images the standard variational flow is given as

$$u_t = R(u) - \beta(u - f) \tag{3.54}$$
$$u(x, 0) = f, \tag{3.55}$$

with appropriate boundary conditions. Here $R(u)$ is a geometric regularization functional and f is the initial data. In the ROF case the regularization functional will be

$$R(u) = \nabla \cdot \frac{\nabla u}{|\nabla u|}. \tag{3.56}$$

In order to reverse the flow Burger et al. [13] proposed the relaxed inverse

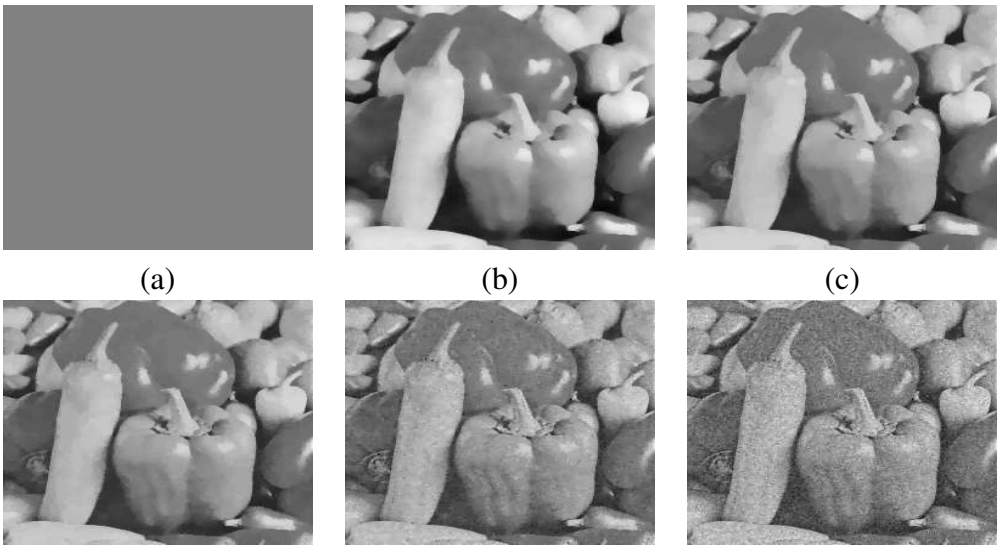


Figure 3.4: The figure shows a relaxed inverse scale space flow. The time is increasing from (a) to (f). Notice that fine details are added during the flow and eventually the noise is added too.

scale space flow in the following way

$$\begin{aligned} u_t &= R(u) - \beta(u - f - v), && (3.57)\\ v_t &= -\alpha(u - f), && (3.58)\\ u(x,0) &= mean(f), && (3.59)\\ v(x,0) &= 0, && (3.60) \end{aligned}$$

where $\alpha \leq \frac{\beta}{5}$ is a relaxation parameter. This flow will evolve the solution from a smooth initial condition $mean(f)$ towards the data f. Theoretical proves for the convergence of the relaxed flow can be found in [59].

3.1.12 Matrix inverse scale space

No known attempts have been made to develop inverse scale spaces for matrix valued images. During the *visualization and processing of tensor fields* workshop in Dagstuhl, Johan Lie and Bernhard Burgeth initiated a collaboration, and together we have developed a novel method for matrix inverse scale spaces [54]. As in the dual matrix case we apply the framework of Burgeth [14, 15] to make a matrix version of the scalar PDEs.

Here we will only focus on the ROF case, however the framework can be applied to different regularizers $R(u)$. Applying the framework of Burgeth we get the matrix version of the regularizer (3.56)

$$\overline{R}(u) = -\overline{\text{div}}\left(\frac{1}{|\overline{\nabla}U|} \bullet \overline{\nabla}U\right). \tag{3.61}$$

The relaxed inverse flow can now be written for matrix valued images in the following way:

$$\begin{aligned} \overline{\partial}_t U &= \overline{R}(U) - \beta(U - F - V), \\ \overline{\partial}_t V &= -\alpha(U - F), \\ U(x,0) &= \text{simp}(F) \\ V(x,0) &= 0, \end{aligned} \tag{3.62}$$

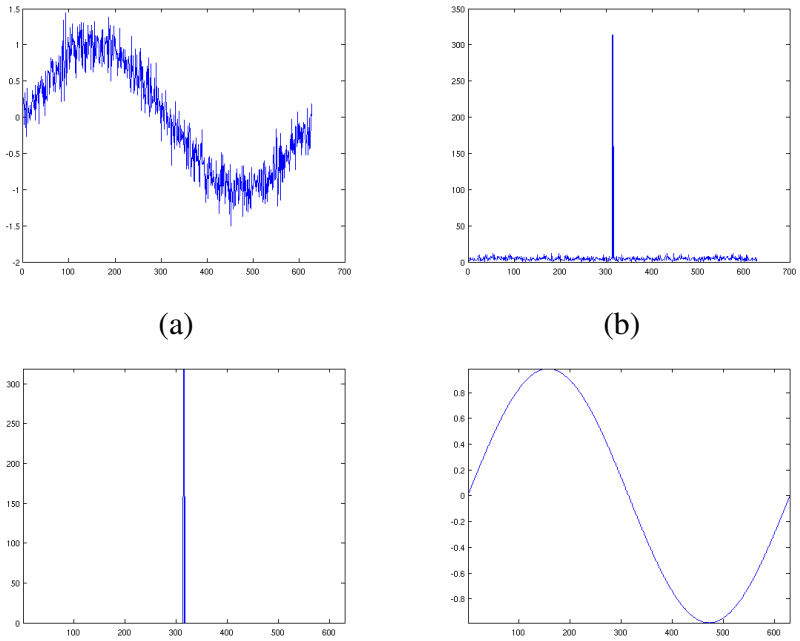
with $\alpha \leq \beta/5$ and U, F and V being matrix fields. The expression simp(F) stands for a simplified version of F. This might be the arithmetic mean of the matrix field F or the solution of another evolution process such as linear, Perona-Malik- or TV- diffusion.

More details can be found in [54]. Here we also show numerical examples which confirm that the inverse scale space concept can be transferred to matrix fields via the operator algebraic framework.

3.2 Transform based denoising

Until now we have only considered variational and PDE based denoising. A completely different strategy is transform based denoising. Transformed based methods do not try to regularize the image directly in the spatial domain, but transforms the image into a different domain before the denoising is done. Usually a frequency transform like wavelet, Fourier or Gabor transform is used [29, 31, 35, 36, 37, 38].

The idea is quite simple. Frequency transforms have the advantage that a smooth signal can be represented by a limited number of frequency coefficients. However, Gaussian or white noise will distribute uniformly in the frequency domain, but with a much lower energy than the signal components. This is reasonable since from the definition white noise is uniformly distributed in the frequency domain. In Figure 3.5(b) we show the Fourier transform of a noisy sine curve. Notice the spike which represents the sine curve, and the small uniformly distributed coefficients which represents the noise. This noise can easily be removed by thresholding all the small coefficients. Usually this is done by hard thresholding,



(a) (b)

(c) (d)

Figure 3.5: The figure illustrates denoising of a sine curve using hard thresholding in the Fourier domain. (a) A sine curve added uniformly distributed noise. (b) The frequency spectrum of the noisy sine curve (zero-frequency component is shifted to center of spectrum). (c) The frequency spectrum after hard thresholding with threshold level $\delta = 20$. (d) The recovered sine curve

i.e.

$$y_{hard} = \begin{cases} x(t) & |x(t)| > \delta \\ 0 & |x(t)| \leq \delta \end{cases},$$

where $x(t)$ is the function being thresholded and δ is the threshold level. This means that all function values with an absolute value below or equal to the threshold level are replaced by 0, see Figure 3.6.

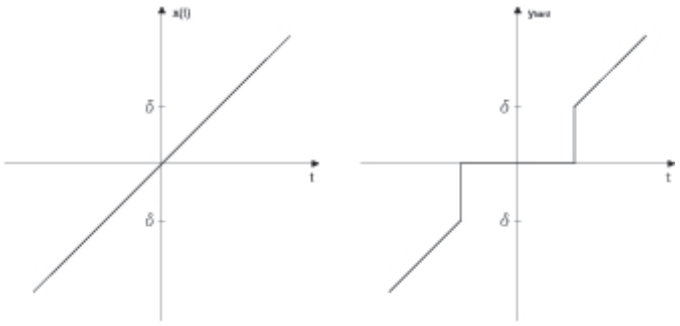


Figure 3.6: Hard thresholding the function $x(t)$.

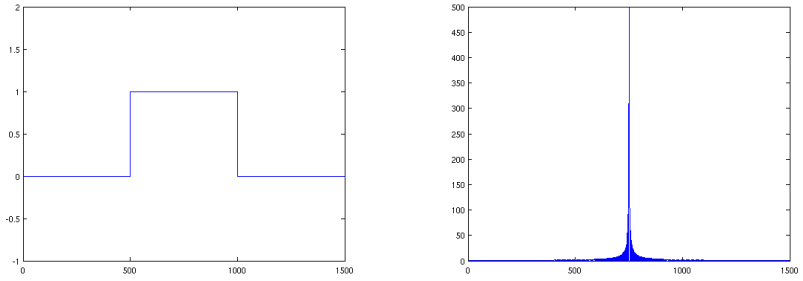
In Figure 3.5(c) we show the result of hard thresholding the noisy frequency spectrum. After the thresholding is done, the signal is transformed back to the spatial domain, see Figure 3.5(d).

Transformed based denoising can be summarized in the following three step algorithm:

- Transform the noisy data into the frequency domain.
- Threshold the resulting coefficients, thereby suppressing those coefficients containing noise.
- Transform back into the original domain.

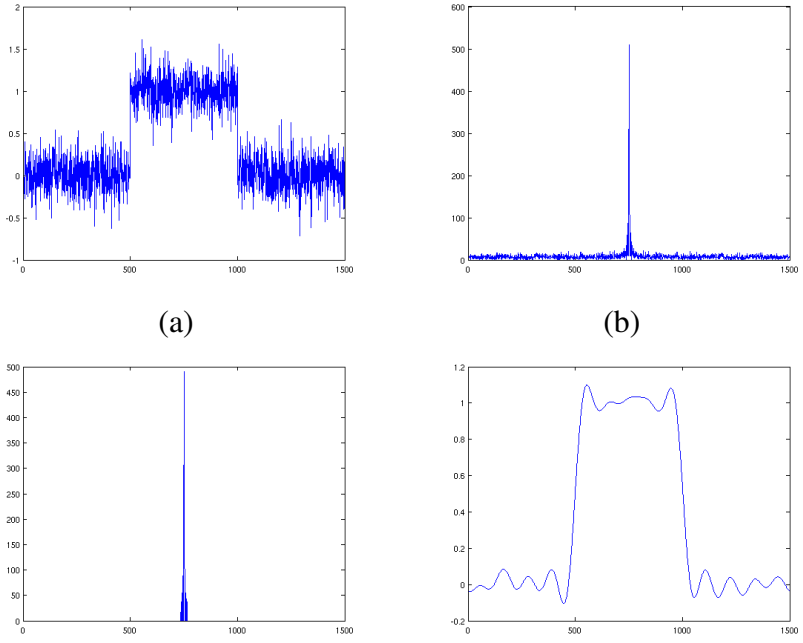
Transform based methods are quite popular and can be used for denoising of a wide range of signals, like images, audio signals, geophysical data and infrared spectra [1, 29, 31, 92]. Among popular methods are the VisuShrink [36] and the SureShrink [37]. However, VisuShrink is known to produce overly smoothed images.

The problem with images is that they in general are not smooth. Images contain edges, which as in the PDE case cause problems. The reason for this is that the wavelet or Fourier transform of an edge contains high frequency components with low amplitude, see Figure 3.7. If a crude thresholding is applied, these coefficients will be erased and result in ringing artifacts, see Figure 3.8. Thus finding a good threshold is a crucial task, however this is not easy. VisuShrink uses a global threshold, i.e. the threshold is constant for the entire image. This will produce overly smoothed images since edge coefficients are lost. In order to denoise images the threshold has to be adaptive. In regions containing edges we need a small threshold, whereas in regions containing no edges a larger threshold can be applied. The SureShrink estimates the threshold from the energy in each subband



(a) (b)

Figure 3.7: (a) A pulse and (b) the corresponding frequency spectrum (zero-frequency component is shifted to center of spectrum).



(a) (b)

(c) (d)

Figure 3.8: The figure illustrates denoising of a pulse using hard thresholding in the Fourier domain. (a) A pulse added uniformly distributed noise. (b) The frequency spectrum of the noisy pulse (zero-frequency component is shifted to center of spectrum). (c) The frequency spectrum after hard thresholding with threshold level $\delta = 30$. (d) The recovered pulse

and works better with images. Chang et al. have proposed a method based on Bayesian estimation [29].

Usually wavelet based transforms decompose the image into several subbands, ranging from coarse to detailed scale. Thresholding is then performed in each subband separately. In Fourier based methods a 2D window of fixed size is slided over the image and each region is denoised separately. This crude segmentation results in poor denoising compared to the wavelet based methods, and for a long time wavelet based transforms have by far been the most superior ones. However, recently a new method based on adaptive windows has been proposed [39, 40, 41, 46].

3.2.1 SA-DCT

The problem with the fixed window size is that the image is partitioned into regions without taking into consideration the shape of the image. Many of the regions will thus contain edges which will be destroyed in a thresholding procedure. To fix this problem we need to pick regions which are as homogeneous as possible. In this way the regions do not contain edges, and we can apply thresholding without destroying the edges in the image.

The proposed method relies on local polynomial approximation (LPA) and intersection of confidence intervals (ICI) to span adaptive regions around every pixel in the image. The idea is that the region spanned around a pixel should contain pixels that have similar intensities. This results in homogeneous regions without edges. These regions are then transformed into the frequency domain using a 2D shape-adaptive discrete cosine transform (SA-DCT). The noise can now be removed by thresholding each region. DCT is used since this is a simplified version of the Fourier transform, i.e. we have no phase information.

The region around a pixel is spanned from the 8 lines indicated in Figure 3.9. The length of each line is chosen such that all the pixels along the line have similar intensities. The method of intersection of confidence intervals is applied to detect if there is a significant change of the intensity along a line. When the length of every line is found the region is closed by joining neighbouring endpoints of the vertices by line segments. The resulting region is then transformed to the frequency domain and thresholded.

Since we span a region around every pixel we have an extensive region-overlap, i.e. we have an overcomplete basis. To reconstruct the image we weight every region together. The weights depend on the size and mean variance of the region.

The method is shown to produce very good results. Another advantage is that it is local in nature, i.e. every region can be processed separately. This makes it very easy to parallelize.

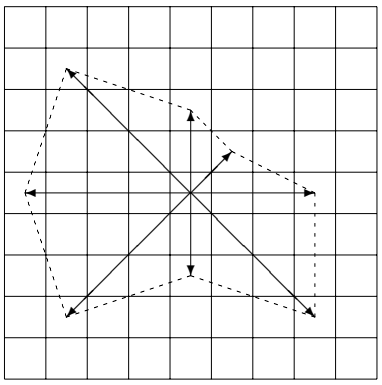


Figure 3.9: Illustration of the eight directions used to span a region around a pixel.

The state-of-the-are results makes the method attractive for denoising of 3D scalar valued images as well as 3D matrix valued images. As a comparison to the PDE methods we have extended the framework for denoising of both 3D scalar valued and 3D matrix valued data [7].

The regions are now 3D and every region is spanned from 26 lines. The length of every line is decided from the ICI rule and the region is closed by joining neighbouring endpoints of the vertices by triangles. Every region is then transformed with a 3D shape-adaptive DCT and thresholded. This method can be applied directly to 3D data like MRI. In order to denoise DTI we apply the Cholesky factorization $D = LL^T$ and apply the 3D algorithm to each element L.

More details can be found in [7]. Here we also show numerical examples on both synthetical and real MRI and DTI data. The experiments indicate that the performance of the method is similar to the proposed PDE methods.

Chapter 4

Summary of papers

In this chapter we will give a short summary of the six papers included in this thesis. The papers are arranged chronologically, based on when the work was performed. The majority of the work have been outlined in the introduction, however for completeness we will summarize the results.

Summary of Paper A

Image Segmentation Using Some Piecewise Constant Level Set Methods with MBO Type of Projection

Xue-Cheng Tai, Oddvar Christiansen, Ping Lin and Inge Skjælaaen

In a series of papers Lie et al. [55, 56, 57, 58] have proposed a new method for image segmentation. The method is called Piecewise Constant Level Set Method (PCLSM) and is based on the Mumford-Shah image segmentation. Two different variants of the method is proposed. A binary version which uses $\log_2 N$ level set functions to identify N regions and a version which uses a single level set function to identify arbitrary number of regions.

In order to solve the PCLSM a variational problem needs to be minimized. This is done by solving the corresponding Euler-Lagrange equation using a steepest decent method. However, this is computational heavy. In order to speed up the computation we in this paper propose to solve the problem using operator splitting [61, 62, 98].

In the paper we show how sequential and parallel splitting can be used to split the Euler-Lagrange equation in two parts. The first of these problems can be solved efficiently using a semi-implicit Picard iteration and exact solvers for tri-diagonal matrices. The second problem can be solved as a projection similar to the MBO scheme [67].

We show numerical results on both synthetical and real images and we compare the results from the sequential and parallel splitting. With this new method we have reduced the computational time considerably compared to applying a steepest decent method on the Euler-Lagrange equations. However, the scheme is reported to be sensitive to the time step, which makes the task of finding optimal regularization parameters a harder task.

Summary of Paper B

Fast implementation of piecewise constant level set methods

Oddvar Christiansen and Xue-Cheng Tai

This is an extension of Paper A. In the previous paper we proposed to use sequential and parallel splitting to split the Euler-Lagrange equations in two parts. The first part was then solved using semi-implicit Picard iteration and exact solvers for tri-diagonal matrices. The second part was solved using a projection similar to the MBO scheme [67].

In this paper we propose to solve the second part in a different way. The second part is essentially a constraint evolving the level set function towards integer values. The effect of the constraint is controlled through a penalization parameter. Instead of using the MBO scheme as in Paper A we apply Newtons method. We show how we can ensure that Newton converges to a unique solution by choosing the time step and the penalization parameter properly. In order to ensure a proper evolution of the constraint, the proposed algorithm modifies the constraint parameter during the iterations. Initially the constraint has little impact, and during the iterations the effect of the constraint is gradually increased.

We show numerical results on both synthetical and real images. The numerical results indicate that the proposed method is not sensitive to the time step. This makes it easier to find good regularization parameters.

Summary of Paper C

Total Variation Regularization of Matrix Valued Images

Oddvar Christiansen, Tin-Man Lee, Johan Lie, Usha Sinha and Tony F. Chan

In this paper we generalize the total variation (TV) restoration model of Rudin, Osher and Fatemi (ROF) [81] into a matrix valued method for denoising of DTI data. The proposed method is an extension of the color TV method proposed by Blomgren and Chan [10].

DTI data are matrix valued and every voxel contains a 3×3 diffusion matrix D. This matrix is symmetric and positive definite. The eigenvalues and eigenvectors span a hyperelipse which models the diffusion in the given voxel. We propose to represent the matrix D by the Cholesky factorization

$$D = LL^T. \tag{4.1}$$

This ensures that the matrix is symmetric and positive definite.

In order to regularize the matrix valued data we propose to solve the following minimization problem

$$\min_{\ell_{kl}}\left\{\sqrt{\sum_{ij}\left(\mathrm{TV}[d_{ij}(\ell_{kl})]\right)^2}+\frac{\beta}{2}\sum_{ij}\int_\Omega\left[d_{ij}(\ell_{kl})-\hat{d}_{ij}\right]^2\right\}, \tag{4.2}$$

where $\{kl\} \in \{11, 21, 31, 22, 23, 33\}$. Here $d_{ij}(\ell l_{kl})$ are the matrix elements of D as a function of its Cholesky factors L, $\hat{d}_{ij}$ denotes the elements of the tensor estimated from the noisy data and β is a positive parameter controlling the regularization. In the model the first term is regularization functional, while the second term is a fidelity term.

In the paper we derive the Euler-Lagrange equations for the minimization problem (4.2). We show numerical examples demonstrating denoising of both synthetical and real DTI data. In addition we present an evaluation of the denoised tensor field.

Summary of Paper D

Shape-Adaptive DCT for Denoising of 3D Scalar and Tensor Valued Images

Ørjan Bergman, Oddvar Christiansen, Johan Lie and Arvid Lundervold

In this paper we propose a transformed based method for denoising of 3D scalar and matrix valued data. The idea behind transformed based methods is to transform the data into a more suitable domain and do the denoising there. Usually wavelet or Fourier based transforms are used. The advantage of these frequency transforms is that Gaussian noise will distribute uniformly with a low amplitude in the frequency domain. Thus the noise can be removed by thresholding small coefficients.

Traditionally, wavelet based methods have been the superior ones for denoising of images. However, recently Foi et al. [39, 40, 41, 46] have proposed a new method based on shape-adaptive DCT. This method has shown to produce very good results and in this paper we extend it to 3D scalar and tensor valued images.

A problem when denoising images is that the edges result in small coefficients in the frequency domain. If these coefficients are thresholded we will get ringing artifacts. In order to avoid this problem Foi et al. propose to use adaptive windows. The idea is to choose regions which are as homogeneous as possible. Around every pixel a statistical method called intersection of confidence intervals (ICI) is used to span a region containing similar pixels. This region is then transformed using shape-adaptive DCT and thresholded.

In the paper we show how these adaptive regions can be extended to 3D and denoised using a 3D shape-adaptive DCT. We show numerical examples on both scalar and matrix valued data and compare the results to PDE denoising. The numerical examples indicate that the proposed method achieves results comparable to the PDE methods. An advantage with the proposed method is that it is local in nature, i.e. every region can be processed independently. This makes the method very suitable for parallel processing.

Summary of Paper E

An Operator Algebraic Inverse Scale Space Method for Matrix Images

Johan Lie, Bernhard Burgeth and Oddvar Christiansen

During the last few years considerable research have been done to study multi-scale properties of images using partial differential equations. In traditional scale space methods we start with a fine scale image and evolve this into coarser scales using a PDE. In the inverse scale space formulation this is reversed, and we start with the coarse scale and then evolve the image towards the finer scales. This is perhaps more of theoretical than practical interest, but the field has got a lot of attention [13, 43, 59, 82].

To our knowledge no known attempts have been done to extend the framework of inverse scale space to matrix valued images. In this paper we present a method to extend the relaxed inverse scale space flow to matrix valued images. The work is a result of a collaboration initiated during the *visualization and processing of tensor fields* workshop in Dagstuhl.

Previously Burgeth et al. [14, 15] have presented a novel framework to find matrix valued counterparts of scalar valued PDEs. In this paper we show how this framework can be used to generalize the inverse scale space flow to matrix valued images. We perform numerical experiments on synthetical and real diffusion fields, which confirm that the inverse scale space concept can be transferred to matrix valued data.

Summary of Paper F

A Dual Operator Algebraic Method for Regularization of Matrix Valued Images

Oddvar Christiansen, Johan Lie and Bernhard Burgeth

A crucial part when dealing with image processing is the computational speed. This is especially true when dealing with 3D matrix valued DTI data. These data sets are huge, and if the algorithm is not computational efficient it can take several hours or even days to complete the computation. If the method should be used clinical it is obvious that the computational time must be kept at a minimum.

In this paper we propose a method which is very computational efficient. The method is based on the dual method presented by Carter [17] and the framework of Chambolle [20].

Following [17] the ROF model can be transformed into a dual problem which is better behaved than the original problem, i.e. no division by zero is explicitly performed in the Euler-Lagrange equations. Applying the method of Chambolle [20] the dual problem can be solved in a very efficient way.

The theory is developed in the scalar case and the resulting problem is a PDE. This PDE is transformed into its matrix valued counterpart using the framework of Burgeth et al. [14, 15]. In the paper we show numerical examples which confirm that the proposed method increases the convergence rate by a factor of roughly 10. This makes the proposed method feasible for clinical use.

Chapter 5

Conclusion

In this thesis we have presented new image processing techniques for segmentation and denoising of scalar and matrix valued data. The majority of the methods are variational methods leading to partial differential equations. However, a new transform based method for denoising is also presented. The main application for the methods are in the field of Diffusion Tensor Imaging and Magnetic Resonance Imaging, but other applications can also be found.

A segmentation method which improves the computational speed of the methods of Lie et al. [55, 56, 58] is proposed. When working with huge data sets as images it is important to develop methods which are computational efficient. In the original work of Lie et al. the variational problem was solved using steepest descent on the corresponding Euler-Lagrange equations. The steepest descent algorithm is not computationally efficient and we propose an operator splitting method for a more efficient solution of the problem.

We also propose a new method for denoising of matrix valued data. This method can be used for denoising of diffusion tensor images. It is a well known fact that these images contain a large amount of noise. A traditional way to increase the quality of the data is to increase the number of directions used in the recording, or to take several images from each direction and average these. However, both these methods increase the scanner time for the patient. We instead propose to use low quality images and improve the quality of these by image processing techniques. We propose a method which is a natural extension of the color TV method of Blomgren and Chan [10]. The resulting method is a matrix valued version of the ROF model of Rudin, Osher and Fatemi [81]. This method is solved using steepest descent on the corresponding Euler-Lagrange equations. In the numerical examples, processed low quality DTI images are compared to high quality DTI images. The tests show that we are able to improve the quality of the DTI images considerably.

Even though the proposed matrix valued regularization method produces ex-

ceptionally good results, it is not yet suitable for clinical use. The 3D DTI data are huge and the computational time is several hours. In order to make the method useful for clinical research, the method of steepest descent must be replaced with faster methods.

To improve the computational time even further we proposed a dual method for regularization of matrix valued data. The method applies the recently proposed framework of Burgeth et al. [14, 15] to make a matrix valued counterpart of the dual ROF model. Numerical results show that the proposed method reduces the number of iterations by a factor of 10 compared to standard gradient methods. This makes the proposed method useful for clinical research.

A transform based method for matrix valued images is also proposed. Spanning adaptive regions around every voxel we are able to apply frequency denoising without destroying the edge information. The proposed method is local, i.e. every region can be denoised separately. This makes the method very suitable for parallel processing. Numerical experiments indicate that the efficiency of the method is comparable to the PDE methods.

The framework of Burgeth et al. is also used to make an inverse scale space method for matrix valued images. With this new method we are able to flow from a simplified matrix field towards a more detailed matrix field.

The dual framework has shown very promising results and is a very interesting topic for further research. Just recently Chan et al. [12] have proposed a dual extension of the color TV model. It would be interesting to extend our matrix model and compare the efficiency against the operator algebraic dual method.

Acronyms

BV Bounded Variation

CT Computertomografi

CV Chan-Vese Model for image denoising

DCT Discrete Cosine Transform

FA Fractional Anisotropy

ICI Intersection of Confidence Intervals

LPA Local Polynomial Approximation

MBO Merriman, Bence and Osher Scheme

MMC Motion by Mean Curvature

MR Magnetic Resonance

MRI Magnetic Resonance Imaging

NMR Nuclear Magnetic Resonance

NMRI Nuclear Magnetic Resonance Imaging

PDE Partial Differential Equation

PET Positron Emission Tomography

PCLSM Piecewise Constant Level Set Method

PM Perona-Malik

RA Relative anisotropy

ROF Rudin, Osher and Fatemi Model for image denoising

SA-DCT Shape Adaptive Discrete Cosine Transform

TV Total Variation

Bibliography

[1] B.K. Alsberg, A.M. Woodward, M.K. Winson, J. Rowland, and D.B. Kell. Wavelet denoising of infrared spectra. *Analyst*, 122(7):645–652, 1997.

[2] R. Bammer. Basic principles of diffusion-weighted imaging. *European Journal of Radiology*, 45:169–184, 2003.

[3] R. Bammer, B. Acar, and M.E. Moseley. In vivo mr tractography using diffusion imaging. *European Journal of Radiology*, 45:223–234, 2003.

[4] P.J. Basser. New Histological and Physiological Stains Derived from Diffusion-Tensor MR Images. *Annals of the New York Academy of Sciences*, 820(1):123–138, 1997.

[5] P.J. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, 1994.

[6] P.J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B*, 111(3):209–219, 1996.

[7] Ø. Bergmann, O. Christiansen, J. Lie, and A. Lundervold. Shape adaptive DCT for tensor valued images. *To appear in Journal of Digital Imaging*, 2006.

[8] Ø. Bergmann, A. Lundervold, and T. Steihaug. Generating a synthetic diffusion tensor dataset. *Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on*, pages 277–281, 2005.

[9] F. Block. Nuclear induction. *Physical Review*, 7 and 8:460–474, 1946.

[10] P. Blomgren and T.F. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.

[11] P. Blomgren, T.F. Chan, P. Mulet, and C.K. Wong. Total variation image restoration: numerical methods and extensions. *Image Processing, 1997. Proceedings., International Conference on*, 3, 1997.

[12] X. Bresson and T.F. Chan. Piecewise constant level set methods for multiphase motion. CAM-UCLA 07-25, 2007.

[13] M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Communications in Mathematical Sciences*, 4(1):179–212, 2006.

[14] B. Burgeth, S. Didas, L. Florack, and J. Weickert. A Generic Approach to Diffusion Filtering of Matrix-Fields. *Universität des Saarlandes, Fachrichtiung 6.1 Mathematik, Preprint*, 191, 2007.

[15] B. Burgeth, S. Didas, L. Florack, and J. Weickert. A Generic Approach to the Filtering of Matrix Fields with Singular PDEs. *Preprint*, 2007.

[16] H.Y. Carr and E.M. Purcell. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review*, 94(3):630–638, 1954.

[17] J.L. Carter. *Dual Methods for Total Variation-Based Image Restoration*. PhD thesis, University of California, Los Angeles, 2002.

[18] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours. *Image Processing, Numerische Mathematik*, 66:1–31, 1993.

[19] M. Cercignani, M. Inglese, E.G. Pagani, M. Comi, and M. Filippi. Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis. *AJNR Am J Neuroradiol.*, 22(5):952–958, 2001.

[20] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[21] R.H. Chan, C.W. Ho, and M. Nikolova. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *Image Processing, IEEE Transactions on*, 14(10):1479–1485, 2005.

[22] T.F. Chan and S. Esedoglu. Aspects of total variation regularized L_1 function approximation. *SIAM Journal of Applied Mathematics*, 65(5):1817–1837, 2005.

[23] T.F. Chan, S. Esedoglu, F. Park, and A. Yip. *Handbook of Mathematical Models in Computer Vision*, chapter Recent Developments in Total Variation Image Restoration. Springer, 2006.

[24] T.F. Chan, G.H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6), 1999.

[25] T.F. Chan and Shen J. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Siam, 2005.

[26] T.F. Chan, S.H. Kang, and J. Shen. Total Variation Denoising and Enhancement of Color Images Based on the CB and HSV Color Models. *Journal of Visual Communication and Image Representation*, 12(4):422–435, 2001.

[27] T.F. Chan and L.A. Vese. An active contour model without edges. In *Scale-Space Theories in Computer Vision, Lect. Notes in Comput. Sci.*, volume 1682, pages 141–151, 1999.

[28] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.

[29] S.G. Chang, B. Yu, and M. Vetterli. Adaptive wavelet thresholding for image denoising and compression. *Image Processing, IEEE Transactions on*, 9(9):1532–1546, 2000.

[30] Y. Chen, S. Levine, and M. Rao. Variable exponent, linear growth functionals in image restoration. *SIAM J. Appl. Math. To appear*, 2005.

[31] O. Christiansen. Time-frequency analysis and its applications in denoising. *Master thesis*, 2002.

[32] O. Christiansen, T.M. Lee, J. Lie, U. Sinha, and T.F. Chan. Total variation regularization of matrix valued images. *International Journal of Biomedical Imaging, special issue on Mathematics in Biomedical Imaging*, (June 2007).

[33] O. Christiansen, J. Lie, and B. Burgeth. A dual operator algebraic method for regularization of matrix valued images. *To be submitted to Inverse Problems and Imaging*, 2007.

[34] O. Christiansen and X.C. Tai. Fast implementation of piecewise constant level set methods. In *Proceedings of the International Conference on PDE-Based Image Processing and Related Inverse Problems, CMA, Oslo, August 8-12, 2005*, Mathematics and Visualization. Springer, 2006.

[35] D.L. Donoho. De-noising by soft-thresholding. *Information Theory, IEEE Transactions on*, 41(3):613–627, 1995.

[36] D.L. Donoho and I.M. Johnstone. Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika. Also Tech. Report, Department of Statistics, Stanford University*, 81(s 425):425–455, 1992.

[37] D.L. Donoho and I.M. Johnstone. Adapting to Unknown Smoothness Via Wavelet Shrinkage. *Journal of the American Statistical Association*, 90(432), 1995.

[38] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet Shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(2):301–369, 1995.

[39] A. Foi. *Anisotropic nonparametric image processing: theory, algorithms and applications*. PhD thesis, Dipartimento di Matematica, Politecnico di Milano, April 2005.

[40] A. Foi, K. Dabov, V. Katkovnik, and K. Egiazarian. Shape-adaptive DCT for denoising and image reconstruction. *Proceedings of SPIE*, 6064:203–214, 2006.

[41] A. Foi, V. Katkovnik, and K. Egiazarian. Pointwise shape-adaptive DCT as an overcomplete denoising tool . *Proc. 2005 Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2005*, pages 164–170, 2005.

[42] I.M. Gelfand and S.V. Fomin. *Calculus of variations*. Revised English edition translated and edited by Richard A. Silverman. Prentice-Hall Inc., Englewood Cliffs, N.J., 1963.

[43] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE transactions on image processing*, 11(7), 2002.

[44] R.C. Gonzalez and R.E. Woods. *Digital image processing*. Addison-Wesley Reading, Mass, 1987.

[45] M. Kass, A. Witkin, and D. Terzopulos. Snakes: Active contour models. *Int. J. Comput. Vision*, 1:321–331, 1988.

[46] V. Katkovnik, K. Egiazarian, and J. Astola. Adaptive Window Size Image De-noising Based on Intersection of Confidence Intervals (ICI) Rule. *Journal of Mathematical Imaging and Vision*, 16(3):223–235, 2002.

[47] T. Klingberg, M. Hedehus, E. Temple, T. Salz, J.D.E. Gabrieli, M.E. Moseley, and R.A. Poldrack. Microstructure of Temporo-Parietal White Matter

as a Basis for Reading Ability Evidence from Diffusion Tensor Magnetic Resonance Imaging. *Neuron*, 25(2):493–500, 2000.

[48] A. Kumar, D. Welti, and R.R. Ernst. Imaging of macroscopic objects by nmr-fourier-zeugmatography. *Naturwissenhaft*, 62(34), 1975.

[49] A. Kumar, D. Welti, and R.R. Ernst. Nmr-fourier-zeugmatography. *J. Magn. Reson.*, 18(69), 1975.

[50] P.C. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.

[51] D. Le Bihan, J.F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534 – 546, 2001.

[52] H. Li and X.C. Tai. Piecewise constant level set method for interface problems. CAM-UCLA 06-05, 2006.

[53] H. Li and X.C. Tai. Piecewise constant level set methods for multiphase motion. CAM-UCLA 06-17, 2006.

[54] J. Lie, B. Burgeth, and O. Christiansen. An operator algebraic inverse scale space method for matrix images. *Submitted to proceedings of the 2nd Dagstuhl workshop on Visualization and Processing of Tensor Fields.*, 2007.

[55] J. Lie, M. Lysaker, and X.C. Tai. A piecewise constant level set framework. In P. Neittaanmaki, T. Rossi, S. Korotov, E. Onate, , and D. Knorzer, editors, *European Congress on Computational Methods in Applied Sciences and Engineering ECCOMAS 2004*, 2004.

[56] J. Lie, M. Lysaker, and X.C. Tai. Piecewise constant level set methods and image segmentation. In Ron Kimmel, Nir Sochen, and Joachim Weickert, editors, *Scale Space and PDE Methods in Computer Vision: 5th International Conference, Scale-Space 2005*, volume 3459 of *Lecture Notes in Computer Science*. Springer-Verlag GmbH, 2005.

[57] J. Lie, M Lysaker, and X.C. Tai. A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE Transactions on Image Processing*, 15(5):1171–1181, 2006.

[58] J. Lie, M. Lysaker, and X.C. Tai. A variant of the level set method and applications to image segmentation. *AMS Mathematics of Computation*, 75:1155–1174, 2006.

[59] J. Lie and J.M. Nordbotten. Inverse scale spaces for nonlinear regularization. *Journal of Mathematical Imaging and Vision*, 27(1):41–50, 2007.

[60] K.O. Lim, M. Hedehus, M. Moseley, A. De Crespigny, E.V. Sullivan, and A. Pfefferbaum. Compromised white matter tract integrity in schizophrenia inferred from diffusion tensor imaging. *ARCHIVES OF GENERAL PSYCHIATRY*, 56(4):367–374, 1999.

[61] T. Lu, P. Neittaanmaki, and X.C. Tai. A parallel splitting up method and its application to navier-stoke equations. *Applied Mathematics Letters*, 4:25–29, 1991.

[62] T. Lu, P. Neittaanmaki, and X.C. Tai. A parallel splitting up method for partial differential equations and its application to navier-stokes equations. *RAIRO Math. Model. and Numer. Anal.*, 26:673–708, 1992.

[63] M. Lysaker, A. Lundervold, and X.C. Tai. Noise removal using a fourth-order pde with application to mri in space and time. *IEEE Transactions on image processing*, 12(12):1579–1590, 2003.

[64] M. Lysaker, A. Lundervold, and X.C. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, 2003.

[65] A. Macovski. Noise in MRI. *Magn Reson Med*, 36(3):494–7, 1996.

[66] R. Malladi, J.A. Sethian, and B. Vemuri. A topology independent shape modelling scheme. *Proc. SPIE Conf. Geom. Methods. Vision II*, 2031:246–258, 1993.

[67] B. Merriman, J.K. Bence, and S. Osher. Motion of multiple junctions: a level set approach. *J. Comput. Phys.*, 112(2):334–363, 1994.

[68] Mori and coworkers. Dti-studio.

[69] S. Mori and P.B. Barker. Diffusion magnetic resonance imaging: Its principle and applications. *The Anatomical Record (NEW ANAT.)*, 257:102–109, 1999. Feature Article.

[70] S. Mori and P.C.M. van Zijl. Fiber tracking: principles and strategies - a techical review. *NMR Biomed*, 15:468–480, 2002.

[71] M.E. Moseley, J. Kucharczyk, J. Mintorovitch, Y. Cohen, J. Kurhanewicz, N. Derugin, H. Asgari, and D. Norman. Diffusion-weighted mr imaging of acute stroke: correlation with t2- weighted and magnetic susceptibility-enhanced mr imaging in cats. *American Journal of Neuroradiology*, 11:423–429, 1990.

[72] P. Mrázek. *Nonlinear Diffusion for Image Filtering and Monotonicity Enhancement*. PhD thesis, Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University, 2001.

[73] P. Mrázek. Monotonicity enhancing nonlinear diffusion. *Journal of Visual Communication and Image Representation*, 13(1):313 – 323, 2002.

[74] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and assosiated variational problems. *Comm. Pure Appl. Math*, 42, 1989.

[75] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, New York, 1999.

[76] S. Osher and R. Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2003.

[77] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.

[78] S. Pajevic and C. Pierpaoli. Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: Application to white matter fiber tract mapping in the human brain. *Magnetic Resonance in Medicine*, 42:526–540, 1999.

[79] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.

[80] E.M. Purcell, H.C. Torrey, and R.V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69:37–38, 1946.

[81] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithm. *Physica D.*, 60:259–268, 1992.

[82] O. Scherzer and C. Groetsch. Inverse scale space theory for inverse problems. In M. Kerckhove, editor, *Scale-Space and Morphology in Computer Vision : Third International Conference, Scale-Space 2001*, volume 2106 of *Lecture Notes in Computer Science*. Springer-Verlag GmbH, 2001.

[83] P. Schultz, E.M. Bollt, R. Chartrand, S. Esedoglu, and K.R. Vixie. Graduated adaptive image denoising: local compromise between total variation and isotropic diffusion. *submitted to the Journal of Advances in Computational Mathematics*, 2007.

[84] J.A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 1999.

[85] E.O. Stejskal. Use of spin echoes in a pulsed magnetic-field gradient to study anisotropic, restricted diffusion and flow. *The Journal of Chemical Physics*, 43(10):3597–3603, 1965.

[86] E.O. Stejskal and J.E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics*, 42(1):288–292, 1965.

[87] B. Stieltjes, W.E. Kaufmann, P.C.M. van Zijl, K. Fredericksen, G.D. Pearlson, M. Solaiyappan, and S. Mori. Diffusion tensor imaging and axonal tracking in the human brainstem. *NeuroImage*, 14:723–735, 2001.

[88] X.C. Tai, O. Christiansen, P. Lin, and I. Skjælaaen. Image Segmentation Using Some Piecewise Constant Level Set Methods with MBO Type of Projection. *International Journal of Computer Vision*, 73(1):61–76, 2007.

[89] X.C. Tai and C.H. Yao. Image segmentation by piecewise constant mumford-shah model without estimating the constants. CAM-UCLA 06-18, 2006.

[90] The Mathworks. MatLab, The Language of Technical Computing.

[91] P. Tofts. *Quantitative MRI of the Brain: Measuring Changes Caused by Disease*. John Wiley and Sons, 2003.

[92] D.O. Trad and J.M. Travassos. Wavelet filtering of magnetotelluric data. *Geophysics*, 65:482, 2000.

[93] L.A. Vese and T.F. Chan. A multiphase level set framevork for image segmentation using the mumford and shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[94] Z. Wang, B.C. Vemuri, and T. Chen, Yand Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex dwi. *IEEE Transactions on Medical Imaging*, 23(8):930 – 939, 2004.

[95] J. Weickert. A review of nonlinear diffusion filtering. *Lecure Notes in Computer Science*, 1252:3 – 28, 1997. Proceedings of the First International Conference on Scale-Space Theory in Computer Vision.

[96] J. Weickert. Applications of nonlinear diffusion in image processing and computer vision. *Acta Math. Univ. Comenian. (N.S.)*, 70(1):33–50, 2000.

[97] J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. Universitat des Saarlandes, Fachrichtung 6.1 Mathematik, Preprint No. 58, 2002.

[98] J. Weickert, B. H. Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE TRans. Image Process.*, 7:398–409, 1998.

[99] C.F. Westin, Maier S.E., Mamata. H., A. Nabavi, F.A. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6:93–108, 2002.

[100] L. Zhukov and A.H. Barr. Oriented tensor reconstruction: Tracing neural pathways from diffusion tensor mri. IEEE Visualization 2002, 2002.