Partial Differential Equations in Image Processing and Computer Vision

Joachim Weickert

A selection of publications as partial fulfillment of the requirements for the habilitation degree in computer science

January 2001

Department of Mathematics and Computer Science
University of Mannheim
Germany

Acknowledgements

This collection of publications describes work that has been carried out between 1996 and 2000. It has benefited from fruitful collaborations and inspiring discussions with many colleagues. It is a pleasure to take this opportunity to express my gratitude to everybody who contributed to this work.

Between February 1996 and March 1997 I worked as a postdoctoral researcher at the Image Sciences Institute at Utrecht University (The Netherlands). In this young and dynamic group I had the possibility to learn a lot about medical image analysis, and to experience Bart ter Haar Romeny's enthusiasm for scale-space. I would like to thank him as well as Wiro Niessen, Karel Zuiderveld, Koen Vincken, Antonio López, Michael Abràmoff, Erik Meijering and Max Viergever for fruitful scientific collaborations during my stay in Utrecht and afterwards. Thanks also to Robert Maas, Willem Jacob van Enk, Alfons Salden, Maurits Konings, Luc Florack, Estia van Raaij, Evert-Jan Vonken, and Bart Muyzer for their friendship.

In March 1996 I met Atsushi Imiya (Chiba University, Japan) at a workshop in Dagstuhl (Germany). He introduced me to the fascinating world of early Japanese scale-space research conducted by Taizo Iijima decades before scale-space became popular in America and Europe. From him as well as from Seiji Ishikawa (Kyushu Institute of Technology, Japan) I learned much about this work and Japan in general. This has influenced my way of thinking about life.

From April 1997 to December 1998 I was a member of the Image Group at Copenhagen University (Denmark), where I became interested in information theory as well as in computer vision. My stay has been made possible by grants from Peter Johansen and Jens Arnspang who showed me how pleasant it can be to work in an unhierarchical group in an unhierarchical country. In that time I collaborated with Jon Sporring, Ole Fogh Olsen, and Mads Nielsen, and I enjoyed interesting scientific and nonscientific discussions with Niels Holm Olsen, Ioana Mogensen, Morten Hanehøj, Knud Henriksen, Kristoffer Jensen, and Peter Riber. I would also like to thank Birthe Olsen and Anders Holm for their hospitality in difficult times.

In May 1998 Otmar Scherzer invited me for a week to Linz (Austria). This was the starting point of a fruitful and still ongoing phase of joint work on regularization methods, scale-spaces, and image sequence analysis. Esther Radmoser (University

of Linz) joined us in these activities.

In March 1999, Luis Alvarez, Javier Sánchez (University of Las Palmas, Spain) and I started working on optic flow computations for large displacements, and on 3D reconstructions from stereo image pairs. I look forward to doing further research with them in the future.

As of January 1999 I am with the Computer Vision, Graphics, and Pattern Recognition Group of the University of Mannheim (Germany). Together with Christoph Schnörr, Rita Schieker, Christian Schellewald, Daniel Cremers and Jens Keuchel I have been involved in the exciting phase of starting up this new group. Although it meant very much work for everybody, I am sure that nobody of us would like to miss this time. In the hours where we had time for research, Christoph and I were merging our ideas on optic flow computations, and we were also contributing to Daniel's research on diffusion snakes. I would also like to thank Gabriele Steidl for her encouragement and her interest in my work, and I hope that we have the opportunity for interesting joint projects in the future. This also applies to Ulrich Brüning and Thomas Sonar (Technical University of Braunschweig, Germany) with whom I have joint grant applications. During my time at Mannheim I also collaborated with Hanno Scharr (University of Heidelberg, Germany).

The financial basis for my research was provided by grants from the *Real World Computing Partnership*, the *Danish Research Council*, the *EU–TMR Research Network VIRGO*, and the state of *Baden–Württemberg*. This is gratefully acknowledged.

I would like to thank again all these people and institutions for their contributions to my scientific life. Last but not least, I owe a lot of thanks to my parents Gerda and Norbert for giving me all kinds of help throughout my whole life.

Mannheim, January 2001 Joachim Weickert

Preface

This work is a selection of twelve publications which serve as partial fulfillment of the requirements for the habilitation degree in computer science at the University of Mannheim, Germany. It contains contributions to the fields of digital image processing and computer vision. The methods used are based on partial differential equations.

The included material is a subset of the author's publications. A complete listing can be found on page 351. One selection criterion was to include work that was done after the author's dissertation in 1996. The article in Chapter 6 contains results from the doctoral thesis as well as more recent research. The author's dissertation has been published as a book (*Anisotropic Diffusion in Image Processing*, Teubner Verlag, Stuttgart, 1998), and is not part of this collection. In all included work the author acted either as first author or provided significant contributions as coauthor. Eleven of the included papers have been published or accepted for publication in journals, and one is currently under review.

Each chapter represents one article. A summary of the main contributions and their relations can be found in Chapter 1. The chapters in this collection are – apart from reformattings and a few minor corrections – identical with the versions that have been accepted for publication. As a consequence, it may happen that similar introductory explanations can be found in related chapters. Since each article used the nomenclature that appeared most convenient to its authors, slight deviations between the nomenclature in different chapters may occur.

Contents

Chapter 1

Introduction and Overview

The goal of this chapter is to give an overview of the author's recent scientific contributions to the field of image processing and computer vision methods using partial differential equations (PDEs). We start by justifying the use of this relatively new class of image analysis techniques. Then a short sketch of the state-of-the-art in PDE-based methods in image processing and computer vision is given. Afterwards it is described in more detail in which sense the author's contributions in Chapters 2–13 have created novel scientific knowledge in scale-space analysis, nonlinear diffusion filtering, variational image restoration, and optic flow computation. The chapter is concluded by sketching some related work in which the author was involved and which is not included in this collection.

1.1 Why Partial Differential Equations?

Sometimes the state of image processing and computer vision has been compared with the state of chemistry shortly before the discovery of the periodic system of elements. Indeed, progress in the last decade creates the impression that image processing and computer vision are currently in a transition phase from an "alchemistic" trial-and-error state to a science which is built on a solid mathematical foundation. Sophisticated techniques such as wavelets in signal and image processing [99, 240], unifying geometric concepts for computer vision [120, 123, 162, 236], and geometric algebra for solving problems in machine vision and robotics [223, 364, 365, 366, 367] are important examples that demonstrate this general tendency.

One of the innovative areas in this transition phase consists of methods that are based on *partial differential equations* [87, 116]. Such equations describe relations between an unknown function in several variables and its partial derivatives. They play a fundamental role in physics and engineering, but also in chemistry, biology or even economics. In the context of image processing and computer vision, PDE-

based techniques offer several advantages:

- Many mathematical results with respect to well-posedness are available, and well-understood numerical frameworks exist that describe how stable algorithms can be designed. As a result, PDE-based methods are one of the mathematically best-founded techniques in image processing and computer vision.

- Some PDE-based techniques make use of the Euler-Lagrange equations arising from variational problems where a suitable energy functional is minimized. Such optimization problems offer the advantage of using a conceptually clear formalism without any hidden assumptions: all model assumptions enter the energy functional, and the resulting solution is optimal with respect to the desired criteria.

- The ultimate goal in machine vision is a transition from a signal-based image representation (pixels with grey values) to a symbol-based high-level description of the depicted real-world objects in their 3D configuration. This bottom-up approach requires mechanisms that introduce a hierarchy into the image structure. One such concept is a so-called *scale-space* representation. It embeds the original image into a continuous family of increasingly simplified representations. Work by Alvarez *et al.* [12] has shown that PDEs are the natural language in which scale-space concepts should be formulated.

- Several classical image processing methods have been reinterpreted under such a unifying PDE-based framework. This includes many well-known techniques such as Gaussian convolution, median filtering, dilation or erosion.

- PDE formulations have also led to the discovery of entirely new methods. They can offer more invariances than classical image processing techniques, or describe novel ways of adaptive filtering, shape simplification, image matching, and interactive segmentation. PDEs are also involved in new computer vision methods for image sequence analysis, stereo reconstruction, or shape-from-shading.

- PDE formulations are genuinely continuous. This allows rotationally invariant formulations. Using adequate numerical techniques ensures that the resulting algorithms approximate this invariance very well. This distinguishes them from many purely discrete concepts.

After this general discussion, let us now survey some important PDE techniques for image processing and computer vision.

1.2 Relation to Previous Work

In this section we shall briefly review the current PDE methods in imaging, and we shall sketch how the work in Chapters 2–13 relates to the state-of-the-art in this field. This section is not intended to be complete, but it will give some pointers to seminal papers. For more detailed explanations, further material and a more extensive list of references, the reader is referred to the collections and monographs on PDEs in image processing and computer vision [38, 274, 330, 382, 381, 412] as well as to a survey paper by Deriche and Faugeras [104] and special issues of *IEEE Transactions on Image Processing* (Vol. 7, March 1998) and *Journal of Visual Communication and Image Representation* (Vol. 11, June 2000).

Linear PDE-based image processing techniques became popular in the eighties by the scale-space work of Witkin [436] and Koenderink [213]. They proposed to embed an image into a family of Gaussian-smoothed versions and to investigate their deep structure, i.e. the behaviour of important features (such as edges) over scale. It is well-known that this family can be created by evolving the initial image under a linear diffusion process. Interestingly, linear scale-space had already been axiomatically introduced to imaging by Iijima in 1959 [174]. His numerous contributions to linear scale-space theory remained unknown to the western world until 1997 [420]. They are surveyed in Chapter 2. A detailed treatment of the various aspects of linear scale-space theory can be found in [127, 230, 369, 383] and the references therein.

Nonlinear PDEs for image enhancement started to become an active research area after Perona and Malik's 1987 paper on nonlinear diffusion filtering [301]. They proposed to smooth an image under a space-variant nonlinear diffusion equation that reduces smoothing at edges of the evolving image. An early predecessor of this work was a publication by the Nobel laureate Dennis Gabor who proposed in 1965 a nonlinear deblurring method using forward diffusion along level lines with backward diffusion across them [138, 233]. Essentially the same behaviour can be observed for the Perona–Malik filter. Since it has been realized that the Perona–Malik process in its original formulation is ill-posed to a certain degree [74, 280, 204, 202], Catté *et al.* [74] studied a regularization for which they established existence and uniqueness results. In his monograph [412], the author extended their continuous results and established related theorems in the semidiscrete and fully discrete setting. By means of Lyapunov functionals he showed that these nonlinear methods, which may locally act edge enhancing, can be regarded as smoothing scale-space transformations. Chapter 5 presents an application of these ideas in an information theoretic context, where the scale-space behaviour of Rényi's generalized entropies is studied. Other research has been devoted to the design of more flexible anisotropic diffusion models with a diffusion tensor [91, 404]. One representative of this class is coherence-enhancing anisotropic diffusion. This

method, which allows to smooth along lines and flow-like structures, is discussed in Chapters 6 and 7 for the scalar-valued and vector-valued case, respectively. Recent years have also witnessed efforts to design appropriate efficient schemes for nonlinear diffusion filtering [33, 305, 410]. Chapters 8, 9 and 10 describe the author's contributions where additive operator splittings on sequential and parallel architectures are developed and their embedding into nested iteration schemes is investigated.

Nordström [281] was the first to introduce a bias term in the Perona–Malik process. This enabled him to establish relations to variational image restoration methods (also called regularization methods). Well-posed variants based on convex nonquadratic optimization formulations have been proposed simultaneously by Charbonnier *et al.* [80], Schnörr [345] and Stevenson *et al.* [376]. Under specific parameters settings, these methods approximate total variation regularization, a powerful denoising technique introduced by Rudin *et al.* in 1992 [326]. Some ideas therein are based on Osher and Rudin's earlier work on shock filtering [290] that can be traced back to Rudin's Ph.D. thesis in 1987 [325]. In Chapters 3 and 4, it is shown that there is a very close connection between regularization methods and diffusion scale-spaces. By regarding regularization methods as time discrete diffusion filters, the scale-space theory for nonlinear diffusion filtering is transferred to regularization methods. Thus, regularization methods create scale-spaces where the regularization parameter serves as scale.

Another influential image processing method that has been formulated in terms of energy functionals is the segmentation technique of Mumford and Shah [257]. Their functional, which may be considered as a continuous formulation of the Markov random field model of Geman and Geman [143] and the weak membrane model of Blake and Zisserman [48], has triggered a lot of mathematical research [253]. Recent results by Chambolle indicate even close relationship to the Perona–Malik filter. The Mumford–Shah process has been approximated in the sense of Γ-convergence leading to a functional whose steepest descent equations are given by a set of coupled diffusion–reaction equations [19, 318]. Many related processes have been studied for a number of computer vision problems by Van Gool's group in Leuven [308]. Interestingly, it was also a group from the same department who pioneered nonlinear diffusion–reaction systems for image enhancement already in 1989 [307]. Their methods were based on Turing's pattern formation models.

Although mathematical morphology is an image analysis technique based on algebraic set theory, PDEs for continuous-scale dilations and erosions with convex structuring elements have been found around 1992 by Brockett and Maragos [56], van den Boomgaard [390], Arehart *et al.* [23] and Alvarez *et al.* [12]. The resulting nonlinear PDEs are hyperbolic and can be treated with specific upwind schemes that have been proposed in the context of level set methods [291, 353].

Since morphological filters act only on the level sets of an image, it is not sur-

prising that there is a very fruitful connection between mathematical morphology and level set methods. In the last decade image processing researchers also became interested in morphological processes that move level sets with curvature dependent speed. The use of mean curvature motion in imaging goes back to a 1990 paper by Kimia *et al.* [207] followed by work of Alvarez *et al.* [13]. Later on it became clear that this process can be thought of as the limiting case of iterated median filtering [158]. A closely related process called affine morphological scale-space has been discovered independently and simultaneously by Alvarez *et al.* [12] and Sapiro and Tannenbaum [333]. It can be axiomatically derived as the unique equation with invariance under the special affine group [12]. This has also triggered research on how one can construct flows with other invariances. In 1993, Faugeras was the first to study projective invariant flows [119], followed by work of Olver *et al.* [286], Bruckstein and Shaked [59], and Dibos [108]. It has turned out that Cartan's moving frames are an elegant tool for describing the differential geometry of the different curve evolutions [121]. Another useful framework for unifying several flows for scalar and vector-valued images has been studied by Kimmel, Sochen, and Malladi [209, 362]. They regard these flows as steepest descent methods of energy functionals that have been proposed by Polyakov in the context of string theory.

An interesting application field of flows of mean-curvature type consists of implicit active contour models [67, 238]. An important representative of this class of interactive segmentation tools are geodesic active contour models which have been introduced simultaneously and even at the same conference by Caselles *et al.* [69] and Kichenassamy *et al.* [205]. Recently Leventon *et al.* have shown how one can incorporate statistical shape information into geodesic active contours [224].

Optic flow estimation belongs to the classic fields in computer vision. Already in 1981, Horn and Schunck proposed a variational method for this purpose. The resulting flow field, however, did not respect discontinuities, since a quadratic regularizer was used. In 1983, Nagel proposed an oriented smoothness constraint that prevents blurring across image edges [260]. Through the years much progress has been made regarding the theoretical analysis of this method [262, 263, 342, 361]. Recently it has been discovered that one may regard Nagel's method even as an early method that performs linear anisotropic diffusion filtering [17]. This relation and some modifications that improve Nagel's method significantly in the case of large displacement fields are discussed in Chapter 11. Nonquadratic regularization approaches for optic flow problems have been considered by I. Cohen in 1993 and later on by Schnörr [344], Deriche *et al.* [105] and Kumar *et al.* [221]. They lead to systems of diffusion–reaction equations for the two unknown flow components. Methods of this type typically use spatial regularizers. In Chapter 12 it is demonstrated how one can improve their performance by means of spatio-temporal regularization. Chapter 13 presents a unifying taxonomy for convex regularizers

which includes many earlier models as well as novel ones that are anisotropic and nonlinear.

Besides these regularization ideas, image sequence analysis has also benefited from other PDE-based image processing concepts. Nesi [270], for instance, introduced an optic flow functional that is motivated from the segmentation functional of Mumford and Shah. Another example is object tracking in image sequences. Here one can combine ideas from optic flow estimation with level sets and active contour concepts. Methods of this type have been investigated by Caselles and Coll [68] and Paragios and Deriche [294].

Techniques similar to optic flow estimation may also be used for stereo reconstruction, since they lead to essentially the same correspondence problem. Variational methods along these lines have been considered by Robert *et al.* [322, 321] and Alvarez *et al.* [10]. A level set approach for stereo reconstruction that arises also from a variational principle has been investigated by Faugeras and Keriven [122].

Shape-from-shading is another computer vision problem that has been addressed with PDE methods. Horn and Brooks [172] were among the first to cast this reconstruction problem into a variational framework with an additional smoothness constraint, while Kimmel and Bruckstein [208] considered a level set formulation for a shape-from-shading model leading to an Eikonal equation. Rouy and Tourin [323] obtained uniqueness results for the shape-from-shading problem using the theory of viscosity solutions.

These discussions show that nonlinear PDE methods play an important role in solving key problems in image processing and computer vision. It is thus not surprising that they have also been applied to other problems in this area including for instance histogram equalization [331], blind image restoration [78, 441], corner detection [15], and interpolation [71]. From an application viewpoint, PDE methods have been used e.g. for enhancement of electron microscopy images [134], restoration of old copper plates [112], classification of chrysanthemum leaves [1], computer aided quality control of fabrics [404], and several computer graphics problems [106, 243, 306]. Their main application area, however, continues to be medical imaging; see e.g. [30, 144, 276, 334]. In quite a number of these fields, state-of-the-art results have been obtained which confirm the general usefulness of mathematically sound techniques.

1.3 Overview of the Main Results

The previous section indicated that the author contributed to PDE-based imaging in the fields of scales-space representations, models and algorithms for diffusion filters, and variational methods for image restoration and optic flow computation.

In the sequel we shall discuss these contributions in more detail by sketching the main results of Chapters 2–13.

1.3.1 Foundations and Applications of Scale-Spaces

Early Japanese Scale-Space Results

Let us consider a two-dimensional (scalar-valued) image which is given by a continuous bounded mapping $f : \mathbb{R}^2 \to \mathbb{R}$. One of the most widely used methods for smoothing f is to regard it as the initial state of a homogeneous linear diffusion process:

$$\partial_t u = \partial_{x_1 x_1} u + \partial_{x_2 x_2} u =: \Delta u, \tag{1.1}$$

$$u(x, 0) = f(x) \tag{1.2}$$

with $x := (x_1, x_2)^\top$. Its solution is given by the convolution integral

$$u(x, t) = \begin{cases} f(x) & (t = 0) \\ (K_{\sqrt{2t}} * f)(x) & (t > 0) \end{cases} \tag{1.3}$$

where K_σ denotes a Gaussian with standard deviation σ:

$$K_\sigma(x) := \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \tag{1.4}$$

Linear diffusion filtering is the oldest and best-studied representative of a *scale-space*. In scale-space theory one embeds an image f into a continuous family $\{T_t f \mid t \geq 0\}$ of gradually smoother versions of it. The original image corresponds to the scale $t = 0$ and increasing the scale should simplify the image without creating spurious structures. Since a scale-space introduces a hierarchy of the image features, it constitutes an important step from a pixel-related image representation to a semantical image description. Numerous publications and two successful conferences [381, 274] demonstrate that scale-space theory has evolved into a mature field of image processing and computer vision.

Usually a 1983 paper by Witkin [436] is regarded as the first reference to the scale-space idea. In Chapter 2 it is shown that this concept is at least 24 years older: An axiomatic derivation of linear scale-space has already been presented by Taizo Iijima in a Japanese conference paper from 1959 [174]. Iijima derived linear scale-space in 1D under five axioms: linearity, translation invariance, scale invariance, semigroup property, and preservation of positivity. This has been the starting point of an entire world of linear scale-space research in Japan, which remained unknown in the western world. Possible reasons for this could be the fact that many papers were written in Japanese. However, there are also English papers available in which Iijima describes his scale-space ideas. A more likely reason is

that his results came too early to be appreciated. Thus, the goal of Chapter 2 is to survey important milestones of this unknown early Japanese scale-space research. It has been written in close collaboration with two Japanese image processing researchers. This allowed to analyse a large amount of original literature.

The main focus in this chapter is on scale-space axiomatics. Under a suitable set of axioms, the scale-space evolution (1.1) can be singled out as the unique possibility to create a *linear* scale-space. Chapter 2 presents four Japanese axiomatics for linear scale-space that have been proposed between 1959 and 1981. By comparing them to ten western axiomatics, a survey on the state-of-the-art in this field is given. It turns out that the Japanese axiomatics require only a relatively small set of axioms, and that some of them are much more systematic than most of the western ones, while others are sufficiently simple to be taught in undergraduate courses.

Apart from these pioneering results regarding the axiomatic foundations of linear scale-space, it is shown that Japanese scale-space theory is well embedded in a general framework for pattern recognition and object classification. Also in this context, many Japanese results had been obtained earlier than in the western world. The main application field was optical character recognition (OCR). Already in 1971, Iijima's scale-space based recognition theory has been implemented in hardware in the optical character reader ASPET/71. It was capable of reading 2000 alphanumeric characters per second, and the underlying scale-space framework has been regarded as the reason for its high reliability and robustness. Later on these concepts have also been used as the main algorithm of Toshiba's OCR systems. Successful applications of linear scale-space theory were thus commercially available long before the concept itself was rediscovered in America and Europe.

Another aspect that characterizes Japanese scale-space research was deep structure analysis. By defining so-called stable viewpoints, a topological scale-space tree has been derived in 1985 that resulted in a focus-of-attention method. Later on, some of these ideas have been continued by Makoto Sato, a former Ph.D. student of Iijima.

Iijima has written many papers and several textbooks on his pattern recognition theory that is based on linear scale-space. Chapter 2 provides a list of key references, Japanese ones as well as English ones. Iijima held professorships at the Tokyo Institute of Technology, Tokyo Engineering University, and the Advanced Institute of Science and Technology. In spring 1997, he retired at the age of 72. Since Japanese and western linear scale-space theory evolved with no interaction, it is both interesting and inspiring to compare the results that have been obtained in both cases. Chapter 2 may be regarded as a first step in this direction.

A Scale-Space Framework for Regularization Methods

Another contribution to the foundations of scale-spaces is presented in Chapters 3 and 4, where regularization methods are regarded as time-discrete diffusion scale-spaces.

Partial differential equations can be classified into three main types [87, 116]: parabolic equations behaving in a diffusion-like manner, hyperbolic processes with wave-like character, and elliptic PDEs that can be related to variational problems.

The pioneering work of Alvarez *et al.* [12] has shown that imposing a reasonable set of architectural, invariance and simplification properties automatically leads to scale-spaces that can be described in terms of partial differential equations. It should be noted that, in contrast to our previous discussion on Japanes scale-space research, we do also consider *nonlinear* scale-spaces here.

Examples of PDE-based scale-spaces include parabolic PDEs such as linear and nonlinear diffusion scale-spaces [175, 303, 412], but also curvature scale-spaces like mean-curvature motion [13, 206] and affine morphological scale space [12, 333]. Hyperbolic PDEs with scale-space properties are given by the dilation and erosion equations from continuous-scale morphology [12, 23, 57, 195, 390]. Thus, the question appears whether it makes sense to study elliptic scale-spaces arising from variational principles.

To this end, so-called *regularization methods* are considered. Their basic structure is as follows. Let us consider some rectangular image domain $\Omega \subset \mathbb{R}^2$. In order to obtain a restoration u of some noisy image $f : \Omega \to \mathbb{R}$, an energy functional of type

$$E(u) := \int_\Omega \left((f-u)^2 + \alpha \, \Psi(|\nabla u|^2) \right) \, dx \tag{1.5}$$

is minimized. The first summand encourages similarity between the restored image and the original one, while the second summand rewards smoothness. In the simplest case, the so-called *regularizer* $\Psi(s^2)$ is supposed to be increasing and differentiable in s^2 and convex in s. Moreover, we assume that there exists some constant $c > 0$ such that $\Psi(s^2) \geq cs^2$. The smoothness weight $\alpha > 0$ is called *regularization parameter*. From variational calculus it follows that the minimizer of $E(u)$ satisfies the Euler–Lagrange equation

$$\frac{u-f}{\alpha} \;=\; \operatorname{div} \left(\Psi'(|\nabla u|^2) \, \nabla u \right). \tag{1.6}$$

Following [253, 339, 378] we may regard this elliptic PDE as an approximation to the nonlinear diffusion process

$$\partial_t u \;=\; \operatorname{div} \left(\Psi'(|\nabla u|^2) \, \nabla u \right) \tag{1.7}$$

with initial image f and stopping time α. This shows that regularization methods

approximate diffusion filters. If we choose e.g. the modified total variation regularizer [268]

$$\Psi(|\nabla u|^2) \;:=\; \varepsilon|\nabla u|^2 + \sqrt{\beta^2 + |\nabla u|^2} \tag{1.8}$$

with some small positive numbers ε and β, we end up with the nonlinear diffusivity

$$\Psi'(|\nabla u|^2) \;=\; \varepsilon + \frac{1}{\sqrt{\beta^2 + |\nabla u|^2}}. \tag{1.9}$$

Since this diffusivity becomes small at edges where $|\nabla u|$ is large, we have a diffusion filter that is edge preserving. If we use the Tikhonov regularizer [387]

$$\Psi(|\nabla u|^2) \;:=\; |\nabla u|^2, \tag{1.10}$$

we obtain $\Psi' = 1$. Hence, we approximate linear diffusion filtering.

The author has established a well-posedness and scale-space theory for nonlinear diffusion filters in his monograph [412]. One part of this theory analyses in which sense the filtered image can be regarded as a smoother, simplified version of the original one. Smoothing qualities play a central role in every scale-space evolution. Establishing smoothing properties for nonlinear diffusion filters is neither trivial nor obvious, since some of them may even act edge-enhancing on a local scale.

One simplification property that is applicable here consists of a maximum–minimum principle. Hummel [173] has shown that this is equivalent to the fact that level lines can be traced back in scale. Such a *causality* property [213] is fundamental for scale-spaces, since it allows to connect structures at different scales (e.g. in order to improve their localization).

Another useful concept for expressing the smoothing behaviour of nonlinear diffusion filters arises from the study of their Lyapunov functionals [412]. As a result of these considerations, it has been shown that during the temporal evolution, the L^2 norm of the solution decreases. By Parseval's equality, this proves that the integral of the squared Fourier or wavelet coefficients decreases as well. Lyapunov functionals also imply statistical smoothness results: the image variance and all higher even central moments are decreased by diffusion filters. As an information theoretic consequence of the Lyapunov functionals, the entropy, a measure of uncertainty and missing information, increases during diffusion filtering. Moreover, when the scale parameter t tends to infinity, the filtered image approaches the average grey value of the initial image. All these results do not only hold in the continuous setting, they have also been proved in the spatially discrete and the fully discrete setting.

What has this to do with regularization methods? If one could establish a similar scale-space theory for regularization methods, one would simultaneously have a theory for time-discrete diffusion filters. Such a result would not only be

the first proof that a class of parameter-dependent elliptic processes can create scale-spaces, it would also complete the theory of diffusion filtering.

This has been done in Chapter 3. It is shown that regularization methods are well-posed in the Sobolev space $H^1(\Omega)$ and that they satisfy a causality property in terms of a maximum–minimum principle. A class of Lyapunov functionals is established which implies that the regularized image has a smaller L^2 norm than the original one, that all even central moments become smaller, and that the entropy grows. Moreover, u tends to the average grey value of f for $\alpha \to \infty$. These results hold for a large class of regularizers including Tikhonov regularization [387], the modified total variation regularizations of Ito and Kunisch [193], Nashed and Scherzer [268], Geman and Yang [142] and Chambolle and Lions [76], as well as for Schnörr's nonquadratic regularization [345].

Since regularization is an approximation to diffusion filtering with a single time step, one may ask what happens if the regularization parameter α is replaced by a sequence of regularization parameters $h_1,...,h_k$ with $\sum_{i=1}^{k} h_i = \alpha$. The idea is to start with the original image f, regularize it with parameter h_1, use the result as input for the next regularization with parameter h_2, and so forth until k regularizations have been carried out. This it called *iterated regularization*. If the regularization parameter remains constant, one speaks of *stationary* regularization, while for differing parameters one gets *nonstationary* regularization. The latter case is treated in detail in Chapter 4.

It is clear that iterated regularization gives a better approximation to diffusion filtering than noniterated. Since diffusion filters have to be discretized in time direction anyway, one may even identify diffusion filtering and iterated regularization. Now the question arises whether iterated or noniterated regularization gives better results. In the linear case with Tikhonov iteration, it is proved in Chapter 3 that iterated regularization is more efficient in removing high-frequency noise than noniterated. This was also confirmed in experiments, where three real-world images were degraded by Gaussian noise, and the signal-to-noise ratio of the best restorations was computed. In the nonlinear case with modified total variation regularization, theoretical results are not available so far and experiments remained inconclusive. The signal-to-noise ratio, however, was in all cases better than the one obtained for Tikhonov regularization.

As already mentioned, Chapter 4 is devoted to the case of nonstationary iterative regularization. This is of special interest when one uses adaptive time steps. Typically, one uses small parameters in the beginning and increases them when the process is slowing down. The proofs in Chapter 4 show that all scale-space results that have been established for noniterated regularization do also carry over to iterated regularization methods, even in the nonstationary case. In addition to that, it is shown that the filtered image depends continuously on the sequence of regularization parameters. This constitutes an important stability result.

Another extension in Chapter 4 is concerned with a more general class of diffusion filters and their approximation in terms of iterated regularization methods. The previous convex regularization methods resulted from fully implicit discretizations of diffusion filters. Implicit means that the divergence term is evaluated at the unknown new time level. Discretizing *edge-enhancing* diffusion filters, such as the Perona–Malik filter [301, 303] or its well-posed variants in [74, 412], in an implicit manner would lead to *nonconvex* problems where no satisfactory theory is available. As a remedy one may apply semi-implicit discretizations which treat the nonlinear diffusivity (or the diffusion tensor) in an explicit way using the known results from the old time level, while the remainder is discretized implicitly. This linearization strategy leads to quadratic iterative regularization methods, for which all theoretical results from Chapters 3 and 4 apply. Such a convexification by freezing the nonlinear part also relates this method to the adaptive linearization technique of Geman and Reynolds ([141]; see also [81, 215]) and to the Kačanov method from elasticity theory [137, 165, 419].

Information Measures in Scale-Spaces

Let us now turn our attention to information theoretic aspects of linear and nonlinear diffusion scale-spaces. It has already been mentioned that the entropy is a specific Lyapunov functional for diffusion filters. The goal of Chapter 5 is to analyse the scale-space behaviour of an entire family of entropies, namely Rényi's generalized entropies [314, 316].

To this end we identify an image with its 2D brightness distribution on some rectangular image domain. In a space-discrete setting we may represent such an image with N pixels by a vector $u = (u_1, ..., u_N)^\top$. Rényi's generalized entropies for this distribution are defined by

$$S_\alpha(u) \;:=\; \frac{1}{1-\alpha} \log \sum_{i=1}^{N} u_i^\alpha \tag{1.11}$$

for $\alpha \neq 1$. The limit $\alpha \to 1$ gives the Shannon–Wiener entropy

$$S_1(u) \;:=\; -\sum_{i=1}^{N} u_i \log u_i, \tag{1.12}$$

which can also be considered as part of the continuum. The parameter α is called *information order*.

The generalized entropies have been successfully applied to image processing problems such as local thresholding [54, 55, 327]. Moreover, it is not difficult to see that the generalized entropies, the multifractal spectrum, the grey-value moments, and the grey-value histogram are equivalent representations.

Since scale-spaces are intended to simplify images, it is natural to investigate their simplification properties in terms of information measures. Vice versa, a scale-space extension complements the entropies with spatial information. Two diffusion scale-spaces are studied in Chapter 5, but all results hold for the entire class of diffusion scale-spaces considered in [412]. In the continuous setting, the two scale-spaces under consideration are given by the linear diffusion process

$$\partial_t u \;=\; \Delta u \tag{1.13}$$

and a nonlinear diffusion scale-space of type

$$\partial_t u \;=\; \nabla \cdot (g(|\nabla u|^2)\, \nabla u) \tag{1.14}$$

where the diffusivity decreases at edges [80]:

$$g(|\nabla u|^2) := \frac{1}{\sqrt{1 + |\nabla u|^2/\lambda^2}} \tag{1.15}$$

The following monotony results have been obtained: For fixed t, the generalized entropies $S_\alpha(u(t))$ are decreasing in α. For fixed α, it is proved that they are increasing in t if $\alpha > 0$, constant if $\alpha = 0$, and decreasing if $\alpha < 0$. For $t \to \infty$, all entropies converge to S_0. Regarding smoothness it is shown that they are at least one time continuously differentiable in t as well as in α.

In an experimental section of Chapter 5, possible image processing applications of the scale-space behaviour of Rényi's entropies are studied. By considering characteristic times where the change of entropies by logarithmic scale becomes maximal, global scale analysis is performed. Experiments with linear scale-space indicate that this point of maximal entropy change is proportional to the dominating structure size in the image. This does not only hold for the Shannon–Wiener entropy, but also for the entire family of generalized entropies. A specific feature of the generalized entropies is that they allow to separate the dominating size of foreground and background: Large positive information orders focus on high grey values, while for large negative information orders low grey values are analysed.

In another experiment not only selected information orders are considered, but the entire continuum. By calculating the locations of extremal entropy change in time, and tracking the behaviour of these locations over a continuum of information orders, characteristic curves are created. In analogy with edge detection in linear scale-space they are called *fingerprints*. It is conjectured that these fingerprints are a compact representation of significant image structure. Preliminary experiments suggest that they might be of use for texture analysis, but more detailed investigations are needed here. The localization of the fingerprint lines is more stable for the nonlinear diffusion scale-space than for the linear one.

The theoretical and experimental results in Chapter 5 show that investigating relations between entropy and scale-space is a promising area where much more

research is desirable. Recent follow-up work by Tanaka *et al.* [380] and results by Ferraro *et al.* [124] confirm this impression by giving further evidence of the usefulness of scale-space ideas in conjunction with entropy concepts.

1.3.2 Coherence-Enhancing Anisotropic Diffusion

We have already seen that many scale-spaces can be expressed as diffusion equations. While there is basically only one *linear* scale-space, an entire world of opportunities appears when *nonlinear* equations are admitted as well. In this context we shall study a specific nonlinear diffusion model which has been designed for enhancing flow-like patterns and closing interrupted lines. This so-called coherence-enhancing anisotropic diffusion filter is investigated in an m-dimensional setting in Chapter 6, and its extension to colour images is analysed in Chapter 7.

The work in these chapters has been quite influential. The underlying idea to steer diffusion filters by means of a structure tensor has been used by other researchers for restoring images [209], enhancing ridges [363] and corners [319], visualizing flow fields in computer graphics [306], designing novel shock-capturing algorithms for hyperbolic conservation laws [152], restoring old copper plates [112], processing flame images obtained from laser-induced fluorescence spectroscopy [337], enhancing cell images in 3D electron microscopy [134], and for some medical applications [30, 244]. Shell is using coherence-enhancing diffusion for processing 3D seismic data for oil exploration. Successful applications in this context may lead to significant commercial savings.

Coherence-Enhancing Diffusion of Scalar-Valued Images

Early nonlinear diffusion filters such as the Perona–Malik filter [301, 303] use an equation of type

$$\partial_t u \;=\; \mathrm{div}\,\big(g(|\nabla u|^2)\,\nabla u\big) \tag{1.16}$$

where the scalar-valued diffusivity g is a decreasing function in $|\nabla u|^2$. This ensures that the diffusion process is slowed down at edges where the gradient is large. We will call such models, in which the diffusive flux $j := -g\,\nabla u$ is parallel to ∇u, *isotropic*.[1]

In some applications, it would be desirable to have more flexible diffusion filters. At noisy edges, for instance, one wants to inhibit smoothing *across* the edge, while still permitting smoothing *along* the edge in order to reduce noise. These demands cannot be satisfied with a scalar-valued diffusivity anymore. A diffusion tensor D is needed instead that does not only take into account the magnitude of the gradient (or a regularized version of it), but also its direction. Such a situation,

[1] This terminology is not used in a consistent manner in the image processing literature. Sometimes diffusion filters with *space-variant* diffusivities are already called anisotropic.

where the flux $j = -D\,\nabla u$ is in general not parallel to ∇u, is called *anisotropic* in our terminology. Early anisotropic diffusion models have been proposed by Cottet and Germain [89, 91] and the author [404, 405].

Another motivation for introducing anisotropy into diffusion filters arises from the wish to process 1D features. One might e.g. be interested in closing interrupted lines in a fingerprint image. This requires more sophisticated structure descriptors than the gradient. A good descriptor of local structure is the *structure tensor* [132]

$$J_\rho(\nabla u_\sigma) := K_\rho * (\nabla u_\sigma \nabla u_\sigma^\top) \tag{1.17}$$

where K_ρ denotes a Gaussian with standard deviation ρ, and $u_\sigma := K_\sigma * u$. The eigenvalues $\mu_1 \geq \mu_2$ of J_ρ measure the local contrast along the eigendirections v_1, v_2. The latter ones describe characteristic local structure directions. If one wants to enhance coherent structures, one should smooth mainly along the lowest contrast direction v_2 with a diffusivity λ_2 that increases with respect to the coherence $(\mu_1 - \mu_2)^2$. This can be accomplished by designing D such that it possesses the same eigenvectors v_1, v_2 as J_ρ, and choosing its corresponding eigenvalues as

$$\lambda_1 := \alpha, \tag{1.18}$$

$$\lambda_2 := \begin{cases} \alpha & \text{if } \mu_1 = \mu_2, \\ \alpha + (1-\alpha) \exp\left(\frac{-C}{(\mu_1-\mu_2)^2}\right) & \text{else,} \end{cases} \tag{1.19}$$

where $C > 0$ is a contrast parameter, and the small constant $\alpha > 0$ is needed for theoretical reasons in order to prove well-posedness. This filter has been introduced by the author in [406]. It performs *coherence-enhancing anisotropic diffusion*.

Chapter 6 gives a detailed theoretical analysis of this filter and its generalization to the m-dimensional case. It is shown that it leads to a mathematically well-posed evolution where local extrema are not enhanced and where causality in terms of a maximum–minimum principle holds. A family of Lyapunov functionals is derived, comprising decreasing L^p norms, decreasing even central moments and increasing Shannon–Wiener entropy. For $t \to \infty$, the solution tends to the average grey value of the original image. This shows that coherence-enhancing diffusion filtering creates a nonlinear scale-space.

In order to evaluate its image enhancement qualities, several experiments are presented. They show that, unlike previous filters such as the Cottet-Germain model [91] or mean-curvature motion [13, 207], coherence-enhancing diffusion is capable of closing interrupted lines in fingerprints. This is caused by its use of semilocal information obtained by averaging over some integration scale ρ in the structure tensor. Furthermore, it is shown that coherence-enhancing diffusion gives useful results for some other applications including multiscale visualization of the stripes in nonwoven fabrics and trabecular structures in 2D and 3D CT images of human bones.

Coherence-Enhancing Diffusion of Vector-Valued Images

In Chapter 7, coherence-enhancing diffusion filtering is extended to colour images, and its robustness under degradations by Gaussian noise is demonstrated. Moreover, a detailed discussion of all model parameters is presented, and selection criteria for these parameters are suggested.

The naive way to extend a nonlinear diffusion filter to vector-valued images would be to apply it in each channel separately. Since nonlinear diffusion filters may enhance features, however, this would lead to the problem that these features may be formed at different locations in each level. It is therefore recommendable to introduce some sort of synchronization between the evolutions in the different channels [144]. A natural way to achieve this is to use the same space-variant diffusion tensor for all channels. It should depend on a structure descriptor that simultaneously takes into account the local image structure in all channels. We thus need a generalization of the structure tensor to vector-valued images. Such a generalization has been investigated by Di Zenzo [107]. Its first use for steering a diffusion filter for colour images has been proposed by the author in [405].

Di Zenzo's idea comes down to summing up the structure tensors in each channel. This creates a structure tensor of a vector-valued image. Its orthogonal system of eigenvectors gives the preferred local directions, while its eigenvalues measure the average local contrast along these directions. Refinements of these ideas would consider a weighted mean of the structure tensors, where the weight is adapted to the reliability and the noise level in each channel [60]. In the absence of any *a priori* knowledge, one usually chooses equal weights.

The resulting structure tensor for vector-valued images is then used to steer the joint diffusion tensor in the same way as was described above. Experiments show that this process gives potentially useful results for simplifying wood surfaces, analysing sonograms and microscopic images. Interesting effects can also be obtained by processing paintings by van Gogh. It appears that van Gogh's painting style uses similar effects as coherence-enhancing diffusion filtering: both are emphasizing flow-like structures. It is thus not surprising that coherence-enhancing diffusion has also been used in computer graphics for visualizing flow fields [306]. Here it may be regarded as an alternative to the classical *line integration convolution* technique by Cabral and Leedom [63].

Since coherence-enhancing diffusion filtering requires several parameters, it is necessary to clarify their meaning and their scaling behaviour, if the image size is altered. Such questions are also addressed in Chapter 7, and empirical guidelines for the selection of these parameters are provided. In particular, it is discussed, how one can use a specific Lyapunov functional, namely a convex combination of the variances in each channel, in order to impose an *a priori* stopping criterion for the evolution. If the signal-to-noise ratio of the input image is known, this information

can be used in the stopping criterion. Experiments demonstrate that the suggested criterion yields realistic estimates of the optimal stopping time. Moreover, it is shown that coherence-enhancing diffusion is very robust under additive Gaussian noise. Even if the noise variance in each channel amounts to four times the signal variance, such that the human eye can hardly find oriented structures anymore, coherence-enhancing diffusion continues to smooth along the "correct" direction. This experimentally observed high robustness is caused by the structure tensor: Coherence-enhancing diffusion uses the eigenvectors of the structure tensor and the difference between pairs of its eigenvalues in order to steer the process. Recently it has been shown that these expressions are invariant under isotropic Gaussian noise of zero mean [198]. Another theoretical foundation for coherence-enhancing diffusion of vector-valued images is due to Kimmel *et al.* [209]. They show that it can also be derived as steepest descent method in a Beltrami framework inspired from high-energy physics [209].

1.3.3 Efficient Algorithms

Often the quality of PDE-based methods leaves no reason to complain, but it is argued that they are too slow for some applications. Here one should keep in mind that many image processing practitioners have no specific background in numerical analysis. Hence the simplest – and often slowest – numerical algorithms are applied to tackle a problem. The design of adequate and efficient PDE-based algorithms for image processing and computer vision problems is therefore a very important topic. Such questions are addressed in Chapters 8, 9, and 10, where fast and absolutely stable algorithms for nonlinear diffusion filtering and variational image restoration are developed. In order to keep things as simple as possible, isotropic processes are considered. The basic ideas can also be useful for anisotropic filters, as is sketched in Chapter 6 and in [412].

It should be noted that there is an important difference between image processing and other fields of scientific computing. In most other fields, a diffusion equation is motivated from some underlying physical problem. Hence, a good numerical method aims at approximating it as closely as possible. This may result e.g. in high-order methods and sophisticated error estimators. In image processing, there is no physical problem behind the model, and one is interested in having methods that inherit all *qualitative* properties of a continuous model rather than highly precise, but possibly oscillating schemes. Such qualitative properties may include maximum–minimum principles, smoothing properties in terms of Lyapunov functionals, invariances such as the preservation of the average grey level, and convergence as the diffusion time tends to infinity.

Because of these specific requirements, it may not be a good idea to apply scientific computing programme packages that have been developed and optimized

for other purposes. The need for a specific class of simple and efficient numerical methods was the motivation for the work in Chapters 8–10.

Additive Operator Splitting Schemes

In Chapter 8 a novel class of numerical algorithms for diffusion filtering is introduced. The basic principle is explained for m-dimensional isotropic nonlinear diffusion filters of regularized Perona–Malik type [74]:

$$\partial_t u \;=\; \operatorname{div}\left(g(|\nabla u|^2)\,\nabla u\right) \;=\; \sum_{l=1}^{m} \partial_{x_l}\left(g(|\nabla u_\sigma|^2)\,\partial_{x_l} u\right) \tag{1.20}$$

where g is a smoothly decreasing positive function, and $u_\sigma := K_\sigma * u$ denotes a Gaussian-smoothed version of u. This modification of the Perona–Malik model is more robust under noise, and it allows to prove well-posedness [74, 412].

In matrix–vector notation, a classical explicit finite difference scheme for (1.20) is given by

$$\frac{u^{k+1} - u^k}{\tau} \;=\; \sum_{l=1}^{m} A_l(u^k)\,u^k, \tag{1.21}$$

where $u^k = (u_1^k, ..., u_N^k)^\top$ denotes the grey values of all N pixels at time level k, and τ is the time step size. The matrix–vector product $A_l(u^k)u^k$ approximates $\partial_{x_l}\left(g(|\nabla u_\sigma|^2)\,\partial_{x_l} u\right)$. This scheme can be explicitly solved for the unknown u^{k+1}:

$$u^{k+1} \;=\; \Big(I + \tau \sum_{l=1}^{m} A_l(u^k)\Big)\,u^k, \tag{1.22}$$

where I is the unit matrix of size $N \times N$. Unfortunately, such an explicit scheme is only stable if the time step size is very small. Hence, a large number of small steps are required to reach some "interesting" diffusion time. The explicit scheme is thus simple, but inefficient in many applications.

In order to obtain an algorithm with better stability properties, a so-called *additive operator splitting (AOS)* scheme can be used instead. The basic idea behind AOS schemes is to split up an m-dimensional diffusion problem with time step size τ into m semi-implicit diffusion problems with time step size $m\tau$. These 1D problems are very easy to solve, and the desired m-dimensional solution is the average over all m one-dimensional solutions.

An AOS scheme for (1.20) is given by

$$u^{k+1} \;=\; \frac{1}{m} \sum_{l=1}^{m} \Big(I - m\tau A_l(u^k)\Big)^{-1} u^k. \tag{1.23}$$

While (1.22) and (1.23) are not identical, one can show that they have the same first-order Taylor expansion in τ, and they have the same approximation order

with respect to the continuous equation (1.20). In Chapter 8 it is proved that (1.23) is absolutely stable, i.e. in theory, arbitrary large time time steps may be taken. In practise, one should not overdo it, since the accuracy suffers when τ is chosen too large.

It has already been mentioned that the author has developed a theoretical framework for diffusion scale-spaces not only for the continuous, but also for the algorithmically important discrete setting [412]. In Chapter 8 it is shown that the AOS scheme was designed in such a way that it satisfies all requirements for this discrete scale-space framework. As a consequence, it has the same qualitative properties as the continuous process: causality in terms of a maximum–minimum principle, a large family of image simplifying Lyapunov sequences, convergence to a constant image as $k \to \infty$, and conservation of the average grey value (under suitable boundary conditions).

What is the price one has to pay for these nice properties and for gaining absolute stability?
Linear systems of equations need to be solved. In each AOS step one has to solve m tridiagonal systems

$$\left(I - m\tau A_l(u^k)\right) w_l^{k+1} = u^k \qquad (l = 1, ..., m) \tag{1.24}$$

in order to obtain the desired solution $u^{k+1} = \frac{1}{m} \sum_{l=1}^{m} w_l^{k+1}$. Fortunately, these linear systems are solvable in linear complexity by a simple Gauss algorithm, such that the overall price is rather low: One AOS step takes about twice as long as one explicit step. Under realistic accuracy requirements, however, one may use 20 times larger time step sizes, such that the total efficiency gain is a factor 10. Often only a few AOS iterations are needed for simple denoising tasks. On a current 700 MHz PC, five AOS iterations with a 256×256 image take about 0.3 seconds. Further speed-up is possible by implementing AOS schemes on parallel architectures. This is described in Chapter 9.

Splitting schemes are well-known in the numerical literature; see e.g. [241]. Classical schemes such as ADI and LOD, however, use *multiplicative* splittings. They lead to sequential methods where the final result may depend on the order in which the 1D diffusions are calculated. This means that rotating an image by 90 degrees leads to different results, which is unacceptable for many imaging applications. Such a drawback cannot appear for AOS schemes, since the sum of the 1D results does not depend on the order.

The efficiency and simplicity of AOS schemes has triggered other researchers to use them for medical imaging problems [317], for speeding up algorithms for geodesic active contours [148], and even for numerical simulations of sandpile growth [168].

Parallel Algorithms for Regularization Methods and Diffusion Filters

In Chapter 9, connections between numerical techniques for variational image restoration methods and nonlinear diffusion filters are discussed, and performance results are reported for parallel implementations of the algorithms. The theoretical basis for such a comparison has already been presented in Chapters 3 and 4: Variational image restoration (regularization) may be regarded as fully implicit time discretization of a diffusion filter. Hence, efficient numerical methods for one of these frameworks may also be applied for the other.

For the class of convex regularization methods, Schnörr's nonquadratic regularizer [345] was used. The variational approach was linearized using the Kačanov method [137, 201], which is closely related to the linearization strategy of Geman and Reynolds [141]: an auxiliary function is introduced into the functional which is then minimized in a two-step procedure. It comes down to freezing the nonlinear part in each iteration step and solving a quadratic optimization problem. Discretizing this problem with finite elements leads to a positive definite linear system in each iteration step. It is solved on an SGI Power Challenge using a parallel conjugate gradient algorithm with Block-Jacobi preconditioning. This algorithm has been taken from the PETSc library [31, 32] which is based on the message passing standard MPI [247].

Experiments have shown that the scaling behaviour of the algorithm is very satisfactory. This indicates that the PETSc code is well parallelized. It appears, however, that there is some communication overhead between the finite element programme and the PETSc code which might slow down the total method more than necessary. In future work it is thus intended to refrain from using general packages for solving linear systems. An interesting observation showed that one can speed up the total algorithm significantly by solving the linear system in the inner loop only in an inexact way before switching back to the outer nonlinear loop.

For the class of diffusion filters, equation (1.20) has been used in connection with a diffusivity that allows edge enhancement. Finite differences have been applied in such a way that one obtains an AOS scheme which requires to solve tridiagonal systems of equations.
AOS schemes possess two granularities of parallelism:

- Coarse grain parallelism: Diffusion in different directions can be performed simultaneously on different processors.
- Mid grain parallelism: 1D diffusions along the same direction decouple as well. Performing e.g. diffusion in x-direction on an image of size 400×300 creates 300 independent tridiagonal systems of size 400. They can be distributed to different processors.

This mid grain parallelism has been exploited for implementing a three-dimensional AOS scheme on an SGI Power Challenge. The scaling behaviour was slightly worse than for the variational approach. The total method, however, was quite efficient. The experiments indicate than one can expect a speed-up of one order of magnitude by parallelizing AOS schemes. Motivated from these encouraging results on a shared memory machine, it is planned to study AOS schemes on architectures with distributed memory, such as system area networks with low latency communication.

At first glance it appears as if the two numerical strategies do not have too much in common. In Chapter 9, however, it is shown that they are quite related. AOS schemes are an efficient realization of a semi-implicit discretization, where the nonlinearity is discretized at the old time level and the linear part is approximated in an implicit manner in order to gain absolute stability. Such a linearization strategy for parabolic PDEs finds its analogy in the Kačanov method for elliptic equations. It is also shown that the underlying discretization techniques (finite elements versus finite differences) lead to similar problems: A finite difference Kačanov method creates a linear system of equations that resembles the one arising from a finite element Kačanov scheme. These structural analogies can be used to create a common numerical platform for variational image restoration methods and diffusion filters.

Efficient Image Segmentation Combining Fast PDE Methods with Watersheds

We have seen that AOS schemes allow efficient diffusion filtering of images. The goal of Chapter 10 is to investigate their use in conjunction with a fast morphological segmentation algorithm.

As prototype for the segmentation method, a watershed algorithm is used that is based on a waterfall implementation [118]. A well-known disadvantage of watershed methods is their tendency to create oversegmentations. Much efforts have been spent in the past to minimize this problem by suitable pre- and postprocessing. In Chapter 10, two AOS algorithms are studied as preprocessing tools: the AOS scheme for isotropic nonlinear diffusion from Chapter 8, and a novel AOS algorithm for convex variational image restoration.

The basic idea behind the novel AOS method is as follows. While it is possible to approximate variational image restoration methods by diffusion filters, this approximation becomes worse for larger regularization parameters. An alternative in this case would be to minimize the energy functional

$$E(u) := \int_{\Omega} \left((f-u)^2 + \alpha \, \Psi(|\nabla u|^2) \right) \, dx \tag{1.25}$$

by evolving its steepest descent equation

$$\partial_t u \;=\; \operatorname{div}\left(\Psi'(|\nabla u|^2)\nabla u\right) + \tfrac{1}{\alpha}\,(f-u) \tag{1.26}$$

for $t \to \infty$. A modified AOS scheme for such a diffusion–reaction equation has been derived. It has the structure

$$u^{k+1} \;=\; \frac{1}{m}\sum_{l=1}^{m}\left(I - \frac{m\alpha\tau}{\alpha+\tau}\,A_l(u^k)\right)^{-1}\frac{\alpha u^k + \tau f}{\alpha+\tau}. \tag{1.27}$$

In order to speed up convergence to the steady state, this AOS scheme is embedded into a Gaussian pyramid framework: The image is sampled down to a coarse level, a fixed number of AOS iterations is applied, the result is interpolated and used as input for the next higher level. The AOS iterations with subsequent interpolation are repeated until the finest level is reached.

Although this nested iteration is relatively simple, it proved to be rather successful: Usually 5 iterations per pyramid level are sufficient to give good results. Typical CPU times are around 0.3 seconds for regularizing a 256×256 image on a 700 MHz PC. In the experiments, AOS-based preprocessing is used in combination with a watershed algorithm and postprocessing by region merging. The total CPU time for segmenting a 256×256 image is in the order of 0.5 seconds.

The results show that PDE-based methods are capable of improving the segmentation quality of the watershed algorithm, and that AOS algorithms are sufficiently fast for this task. Preprocessing with an isotropic edge-enhancing diffusion filter gives better results than the convex nonquadratic variational approach if the image is not very much degraded by noise. For noisy images, the convex variational approach is more robust.

1.3.4 Variational Optic Flow Computation

Estimating displacement fields in image sequences is a fundamental problem in computer vision. The resulting so-called optic flow field can be used in a number of applications ranging from robot navigation to second generation video coding.

Variational methods for optic flow computation offer the advantage of creating dense flow fields. In Chapters 11–13, three approaches for casting optic flow problems into a variational framework are described.

Optic Flow Computation for Large Displacement Fields

Since temporal undersampling is a common problem in image sequence analysis, there is a need for optic flow methods that work well also in cases where the displacements between two frames exceed the order of one pixel. Such a method is investigated in Chapter 11.

Let us consider two frames $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$ of an image sequence. In image sequence analysis, it is frequently assumed that structures keep their grey value over time:

$$f_1(x_1, x_2) \;=\; f_2(x_1+u_1,\, x_2+u_2), \tag{1.28}$$

where $u = (u_1, u_2)^\top$ is the space-variant optic flow field. This equation is called *optic flow constraint (OFC)*. Here it is centred in f_1, but sometimes the OFC is also centred in f_2:

$$f_1(x_1-u_1,\, x_2-u_2) \;=\; f_2(x_1, x_2). \tag{1.29}$$

Since it is not possible to determine the two unknowns u_1 and u_2 from a single optic flow constraint, additional assumptions are needed to obtain a unique solution. Often it is assumed that the flow field should be smooth [171]. In 1986, Nagel and Enkelmann [265] proposed to require piecewise smoothness. They suggested to compute the optic flow as minimizer of the functional

$$\begin{aligned} E_{NE}(u) \;=\; & \int_\Omega \Big(f_1(x_1-u_1,\, x_2-u_2) - f_2(x_1, x_2)\Big)^2 \, dx \\ & + \alpha \int_\Omega \left(\nabla u_1^\top D(\nabla f_1) \nabla u_1 + \nabla u_2^\top D(\nabla f_1) \nabla u_2\right) dx, \end{aligned}$$

where $\alpha > 0$ serves as regularization parameter, and $D(\nabla f_1)$ is basically a projection matrix perpendicular to ∇f_1. The idea is thus to supplement the OFC with an oriented smoothness constraint that avoids smoothing across image edges.

This technique forms the basis for the method in Chapter 11. However, it has been modified in several ways:

- The Nagel–Enkelmann method centres the OFC in f_2 and the smoothness constraint in f_1. This inconsistency may create errors for large displacements. The model in Chapter 11 uses thus a formulation where both expressions are centred in f_1.

- The optic flow constraint causes a nonconvex energy functional that may have many local minima. In order to encourage convergence to the global minimum, the method is embedded in a linear scale-space that allows to focus down from coarse to fine scales in small steps.

- Minimizers of the energy functional (1.30) are not invariant under linear brightness changes of the images f_1 and f_2. This can lead to problems in the focusing strategy where the dynamic range of the input images strongly depends on the smoothing. As a remedy, a parameter adaptation strategy has been introduced leading to a functional which is invariant under linear brightness changes.

By considering the steepest descent equations of the energy functional, it is shown that the oriented smoothness constraint creates an anisotropic diffusion process with a diffusion tensor that resembles the one used for edge-enhancing anisotropic diffusion [408]. The Nagel–Enkelmann method can thus be regarded as an early predecessor of anisotropic scale-space ideas.

Chapter 11 also gives a detailed analysis of the diffusion–reaction system resulting from the modified energy functional. In particular, existence and uniqueness results are proved. The equations are discretized with a semi-implicit finite difference scheme, and the resulting linear system of equations is solved using symmetric Gauß–Seidel iterations. The role of all model parameters is explained in detail and it is shown that the results are quite robust under parameter variations. As a consequence, default parameter settings may be used in a number of applications.

Experiments have been carried out with both synthetic and real-world image sequences. It turns out that the consistent centring of OFC and smoothness constraint is just as important as the scale-space focusing strategy in order to obtain good results in case of large displacements. Comparisons with the performance evaluation paper by Barron *et al.* [35] show that the suggested method outperforms all classical optic flow methods with 100 % density that are analysed in this paper. Moreover, displacements of more that 10 pixels can be handled with good accuracy.

Spatio-Temporal Optic Flow Regularization

Let us now turn our attention to the case in which displacements in the optic flow field are small. In this situation it makes sense to express the brightness constancy assumption by a differential version of the optic flow constraint. For an image sequence $f(x_1, x_2, \theta)$, where $(x_1, x_2) \in \Omega$ denotes the location and $\theta \in [0, T]$ is the time, the differential OFC can be written as

$$f_{x_1} u_1 + f_{x_2} u_2 + f_\theta \;=\; 0. \tag{1.30}$$

This differential OFC can be embedded into a variational framework that incorporates an additional smoothness assumption. One can for instance consider the following convex functional [344, 414]:

$$E(u, v) \;:=\; \int_\Omega \left((f_{x_1} u_1 + f_{x_2} u_2 + f_\theta)^2 \;+\; \alpha\, \Psi\left(|\nabla u_1|^2 + |\nabla u_2|^2 \right) \right) dx, \tag{1.31}$$

where Ψ is a convex nonquadratic regularizer that reduces smoothing at discontinuities of the flow field. Since only *spatial* derivatives enter the regularizer, we have an approach that rewards spatial smoothness. This is characteristic for most optic flow regularizers. For such methods, two frames are sufficient to compute

the displacement field. One may minimize this functional by considering its steepest descent equations. They are given by a coupled system of two-dimensional diffusion–reaction equations.

On the other hand, it is evident that image sequences do not only possess spatial but also *temporal* coherence. It would thus be more appropriate to replace spatial regularizers by spatio-temporal ones. This is investigated in Chapter 12. The resulting energy functional can be written as

$$E(u,v) := \int_{\Omega\times[0,T]} \left((f_{x_1}u_1+f_{x_2}u_2+f_\theta)^2 + \alpha\,\Psi\left(|\nabla_3 u_1|^2+|\nabla_3 u_2|^2\right)\right) dx\,d\theta \qquad (1.32)$$

where $\nabla_3 := (\partial_{x_1}, \partial_{x_2}, \partial_\theta)^\top$ denotes the spatio-temporal nabla operator. Minimizing this functional by applying steepest descent, one ends up with a three-dimensional diffusion–reaction system. Two numerical schemes are presented in order to solve this system: a stabilized variant of a simple explicit scheme, and an AOS algorithm that resembles (1.27).

Interestingly, only a few spatio-temporal regularizers have been proposed in the literature so far [44, 259, 264]. It is likely that this has historical reasons: Spatio-temporal variational methods require more memory, since the entire sequence is processed simultaneously. For the typical test sequences in computer vision, this does no longer cause problems when modern PCs or workstations are used. In the spatio-temporal case, the CPU time to calculate the optic flow for the entire sequence is only 50 % higher than in the spatial case.

The experiments in Chapter 12 demonstrate that spatio-temporal regularization clearly outperforms spatial regularization. Outliers are less frequent, the flow field is more homogeneous, and the motion boundaries are more accurate. An evaluation with a synthetic test sequence also shows that the spatio-temporal regularizer has competitive performance: it has the lowest angular error in a long list of classic algorithms that are evaluated by Galvin *et al.* [139].

It should be noted that the spatio-temporal extension that is studied here is not limited to the specific structure of the data term and the smoothness term. Any spatial regularization method for optic flow computation can be transformed into a spatio-temporal variational method: All that one has to do is to extend the integration domain by the time domain, and to replace the spatial nabla operator in the regularizer by a spatio-temporal one. Also in other functionals than the one discussed here, similar improvements have been observed. Spatio-temporal regularization is therefore a general strategy for exploiting the entire image sequence data in order to optimize the local flow estimation.

A Taxonomy for Optic Flow Regularizers

We have already seen that there are fruitful connections between regularization methods and diffusion filters. In Chapter 13, such connections are investigated for the case of optic flow regularizers and diffusion methods for vector-valued images. This leads us to a taxonomy for variational optic flow methods that includes existing methods as well as novel ones.

We restrict ourselves to rotationally invariant convex regularizers, and we use the differential optic flow constraint (1.30) as data term in the energy functional. Convexity enables us to establish well-posedness results and to implement globally convergent numerical methods.

A general variational approach with spatial regularization has the structure

$$E(u) := \int_\Omega \left(\tfrac{1}{2} \left(f_{x_1} u_1 + f_{x_2} u_2 + f_\theta \right)^2 + \alpha \, V(\nabla f, \nabla u) \right) dx. \tag{1.33}$$

where V is a rotationally invariant convex regularizer. It may be chosen in such a way that the smoothing process respects image or flow discontinuities. As is described in the previous section, the spatial functional (1.33) may also be extended to a spatio-temporal energy functional.

The steepest descent equations for (1.33) are given by

$$\partial_t u_1 = \partial_{x_1} V_{u_{1x_1}} + \partial_{x_2} V_{u_{1x_2}} - \tfrac{1}{\alpha} f_{x_1} \left(f_{x_1} u_1 + f_{x_2} u_2 + f_\theta \right), \tag{1.34}$$

$$\partial_t u_2 = \partial_{x_1} V_{u_{2x_1}} + \partial_{x_2} V_{u_{2x_2}} - \tfrac{1}{\alpha} f_{x_2} \left(f_{x_1} u_1 + f_{x_2} u_2 + f_\theta \right). \tag{1.35}$$

This is a diffusion–reaction system in which the underlying diffusion process has the structure

$$\partial_t u_i = \partial_{x_1} V_{u_{ix_1}} + \partial_{x_2} V_{u_{ix_2}} \qquad (i = 1, 2). \tag{1.36}$$

Such equations also appear for diffusion filtering of vector-valued images. We may use this diffusion structure to classify optic flow regularizers. If the diffusion process uses a scalar-valued diffusivity, we call the regularizer *isotropic*, and for a diffusion tensor it is named *anisotropic*. Depending on whether the diffusivity / diffusion tensor is a function of ∇f or ∇u, it is called *image-driven* or *flow-driven*. This shows that image-driven approaches are linear in the diffusion part, while flow-driven ones are nonlinear.

Table 1.1 gives an overview of the different types of optic flow regularizers and vector-valued diffusion filters that can be found in the literature. Interestingly, there is a gap in the table: No energy functional is described that uses anisotropic flow-driven regularizers. Such regularizers would be very useful, though: they could be designed in such a way that smoothing is permitted along flow discontinuities, while it is inhibited across them.

One of the main results in Chapter 13 is the construction of anisotropic flow-driven regularizers. Matrix-valued functions, tensor products and trace operators

Table 1.1: Survey of current optic flow regularizers and the corresponding diffusion processes for vector-valued images. In the linear diffusion cases, formulations for vector-valued images are unusual, and the citations refer to the scalar-valued processes. In the optic flow interpretation, f denotes the image sequence, and u is the optic flow. In the diffusion framework, f may be regarded as the initial state of the evolving vector-valued image u.

regularizer $V(\nabla f, \nabla u)$	diffusion process $\partial_t u_i = \partial_{x_1} V_{u_{ix_1}} + \partial_{x_2} V_{u_{ix_2}}$
homogeneous $\sum_{i=1}^{2} \lvert\nabla u_i\rvert^2$ (Horn/Schunck 1981 [171])	homogeneous $\partial_t u_i = \Delta u_i$ (Iijima 1959 [174])
image-driven, isotropic $g(\lvert\nabla f\rvert^2) \sum_{i=1}^{2} \lvert\nabla u_i\rvert^2$ (Alvarez *et al.* 1999 [11])	linear isotropic $\partial_t u_i = \operatorname{div}\,(g(\lvert\nabla f\rvert^2)\,\nabla u_i)$ (Fritsch 1992 [135])
image-driven, anisotropic $\sum_{i=1}^{2} \nabla u_i^\top D(\nabla f)\,\nabla u_i$ (Nagel 1983 [260])	linear anisotropic $\partial_t u_i = \operatorname{div}\,(D(\nabla f)\,\nabla u_i)$ (Iijima 1962 [176])
flow-driven, isotropic $\Psi\left(\sum_{i=1}^{2} \lvert\nabla u_i\rvert^2\right)$ (Schnörr 1994 [344])	nonlinear isotropic $\partial_t u_i = \operatorname{div}\left(\Psi'(\sum_j \lvert\nabla u_j\rvert^2)\,\nabla u_i\right)$ (Gerig *et al.* 1992 [144])
flow-driven, anisotropic ? ?	nonlinear anisotropic $\partial_t u_i = \operatorname{div}\,(D(\nabla u)\,\nabla u_i)$ (Weickert 1994 [405])

are required in order to design such models. The final result, however, looks both simple and elegant: For every isotropic flow-driven regularizer

$$\Psi\Big(\sum_{i=1}^{2}|\nabla u_i|^2\Big) \;=\; \Psi\Big(\mathrm{tr}\sum_{i=1}^{2}\nabla u_i \nabla u_i^\top\Big), \tag{1.37}$$

an anisotropic counterpart can be constructed by exchanging the role of Ψ and the trace operator:

$$\mathrm{tr}\Psi\Big(\sum_{i=1}^{2}\nabla u_i \nabla u_i^\top\Big). \tag{1.38}$$

In the latter case, Ψ is regarded as a matrix-valued function. More details can be found in Chapter 13. There one can also find an experiment which shows that anisotropic flow-driven regularization performs favourably in comparison to the previous methods in this taxonomy. Moreover, a design principle is introduced that explains how to construct other anisotropic regularizers that are rotationally invariant. Interestingly, in turns out that a variational model that has been proposed by Schnörr in 1994 [344] can be interpreted in this framework. It leads to two strongly coupled anisotropic diffusion–reaction methods that are useful for analysing meteorological images.

After closing the gap in the taxonomy, a unifying variational description is developed in Chapter 13. In comprises all processes in this taxonomy. Well-posedness is proved for this model, both for the spatial and the spatio-temporal case. In this way a general theoretical foundation of convex regularization methods for optic flow computation is established. It is planned to use this taxonomy as a platform for detailed performance evaluations and for future algorithmic research.

1.4 Extensions

This habilitation thesis is a selection of papers that describe the main road of the author's research. This restriction is necessary in order to limit the size of the thesis and to avoid distracting the reader too much from its main aspects. Nevertheless, some related side roads should be mentioned as well. In the last five years, the author has also contributed to the following topics that involve PDE methods in imaging:

- The deep structure of corner detectors in linear diffusion scale-space has been analysed showing that corners may be created in scale-space [370, 371].
- Novel numerical schemes for coherence-enhancing anisotropic diffusion have been introduced. Since they use derivative filters that have been optimized with respect to rotation invariance, angular errors resulting in dissipative effects have been reduced significantly [338, 422].

- Nonlinear diffusion filtering has been applied to a number of medical imaging problems:
  - Segmenting brain images into grey matter, white matter and cerebrospinal fluid was done using a scale-space based segmentation algorithm named hyperstack. Experiments have been performed in the 2D setting [277] as well as in 3D [278, 276].
  - MR sequences depicting human eye movement have been postprocessed with a parallel AOS scheme for nonlinear diffusion filtering [379].
  - Coherence-enhancing anisotropic diffusion turned out to be well-suited for orientation detection in trabecular bones [384, 427].
  - 3D ultrasound data sets have been processed in order to remove speckle noise [385].

  Some of these activities are documented in a recent survey paper on applications of PDEs in medical imaging [424].

- A novel AOS scheme with harmonic averaging has been developed for geodesic active contours [417]. In contrast to a previous AOS scheme for this task [148], no time-consuming reinitializations of the distance function are required.

- Moving objects in a video sequence have been tracked in [219] by combining a structure tensor based optic flow method [43, 197] with geodesic active contours [69, 205].

- The Mumford–Shah functional has been supplemented with a term that models statistical knowledge about learned shapes. The resulting active contour models, called *diffusion snakes*, include *a priori* knowledge about the expected objects [93, 94].

- The modified Nagel–Enkelmann optic flow method described in Chapter 11 has also been used in a variational approach for stereo reconstruction [9, 10].

These contributions show that the usefulness of the developed methods is not limited to the contents presented in Chapters 2–13.

Part I

Foundations and Applications of Scale-Spaces

Chapter 2

Linear Scale-Space Has First Been Proposed in Japan

Joachim Weickert, Seiji Ishikawa, and Atsushi Imiya.
Journal of Mathematical Imaging and Vision,
Vol. 10, No. 3, 237–252, May 1999.

Abstract

Linear scale-space is considered to be a modern bottom-up tool in computer vision. The American and European vision community, however, is unaware of the fact that it has already been axiomatically derived in 1959 in a Japanese paper by Taizo Iijima. This result formed the starting point of vast linear scale-space research in Japan ranging from various axiomatic derivations over deep structure analysis to applications to optical character recognition. Since the outcomes of these activities are unknown to western scale-space researchers, we give an overview of the contribution to the development of linear scale-space theories and analyses. In particular, we review four Japanese axiomatic approaches that substantiate linear scale-space theories proposed between 1959 and 1981. By juxtaposing them to ten American or European axiomatics, we present an overview of the state-of-the-art in Gaussian scale-space axiomatics. Furthermore, we show that many techniques for analysing linear scale-space have also been pioneered by Japanese researchers.

2.1 Introduction

A rapidly increasing number of publications, workshops and conferences which are devoted to scale-space ideas confirms the impression that the scale-space paradigm belongs to the challenging new topics in computer vision. This is also reflected in a number of recent books on this subject [127, 382, 381, 230, 369, 412].

In scale-space theory one embeds an image $f : \mathbb{R}^2 \to \mathbb{R}$ into a continuous family $\{T_t f \mid t \geq 0\}$ of gradually smoother versions of it. The original image corresponds to the scale $t = 0$ and increasing the scale should simplify the image without creating spurious structures. Since a scale-space introduces a hierarchy of the image features, it constitutes an important step from a pixel-related image description to a semantical image description.

Usually a 1983 paper by Witkin [436] or an unpublished 1980 report by Stansfield [375] are regarded as the first references to the scale-space idea. Witkin obtained a scale-space representation by convolution of the original image with Gaussians of increasing width. Koenderink [213] pointed out that this *Gaussian scale-space* is equivalent to calculating $(T_t f)(x)$ as the solution $u(x, t)$ of the linear diffusion process

$$\partial_t u = \sum_i \partial_{x_i x_i} u =: \Delta u, \tag{2.1}$$

$$u(x, 0) = f(x). \tag{2.2}$$

Soon this linear filtering became very popular in image processing, and many results have been obtained with respect to the axiomatization, differential geometry, deep structure, and applications of linear diffusion filtering. An excellent overview of all these aspects can be found in the recent book edited by Sporring, Nielsen, Florack and Johansen [369].

Perona and Malik [303] pioneered the field of nonlinear diffusion processes, where the diffusivity is adapted to the underlying image structure. Many regularized variants of the Perona–Malik filter are well-posed and possess scale-space properties [412]. There exist also generalizations of the latter theories to anisotropic diffusion processes with a diffusion tensor [412] and to dynamic scale-space paradigms which can be analysed in terms of algebraic invariance theory and differential and integral geometry [328]. Other important classes of nonlinear scale-spaces are related to morphology. Some of them are continuous-scale versions of classical morphological processes such as dilation or erosion [390, 57, 195], others can be described as intrinsic evolutions of level curves [12, 207, 286, 333]. These geometric scale-spaces are generated by nonlinear partial differential equations (PDEs) which are designed to have Euclidean [390, 57, 207, 195], affine [12, 333] or projective invariances [119, 59, 286, 108]. Overviews of these nonlinear theories can be found in [104, 382, 412].

Besides the above mentioned continuous scale-space concepts there has also been research on how to construct adequate discrete scale-space frameworks, both for linear [25, 229, 230, 329] and nonlinear diffusion filtering [407, 412]. Since a more detailed discussion of these algorithmically important frameworks would be beyond the scope of this paper, we shall focus on continuous theories in the sequel.

This diversity of scale-space approaches has triggered people to investigate which axiomatic principles are involved in these theories and how they differ [213, 442, 28, 229, 130, 390, 12, 286, 295, 272, 231, 128]. Apart from a few notable exceptions such as the morphological equivalent of Gaussian scale-space [390], affine morphological scale-space [12] and intrinsic geometric heat flows [286], most of these axiomatics use (explicitly or implicitly) one requirement: a linearity assumption (superposition principle). Within such a linear framework it was always possible to derive the Gaussian scale-space as the unique possibility. The fact that derivations such as [295, 272, 231, 128] have been found recently shows that linear scale-space axiomatics belong to the current research topics in computer vision.

However, since the linear diffusion equation is well-established in mathematics and physics since Fourier's pioneering works in 1822 [133], and image processing was already an active field in the fifties, one might wonder whether the concept of Gaussian scale-space is not much older as well. Koenderink [214] states in a very nice preface discussing how scale-related ideas can be traced back in literature, poetry, painting and cartography that *"the key ideas have been around for centuries and essentially everything important was around by the end of the nineteenth century."* The goal of the present paper is to supplement these statements by showing that not only the ingredients, but also their final mixture and application to image processing is much older than generally assumed. To this end we present four Japanese linear scale-space approaches, which are older than American and European ones. The first one of them dates back to 1959.

The outline of this paper is as follows: In Section 2.2 we describe the basic ideas of a one-dimensional axiomatic for Gaussian scale-space that has been discovered by Taizo Iijima in 1959 [174]. Section 2.3 studies a two-dimensional version of this axiomatic leading to affine Gaussian scale-space. It has been established in 1962. In 1971 Iijima presented a more physical two-dimensional axiomatics of affine Gaussian scale-space [179]. Its principles are sketched in Section 2.4. The next section is concerned with further scale-space research of Iijima and his students. Section 2.6 describes a two-dimensional axiomatic which has been found by Nobuyuki Otsu in 1981 [292]. In Section 2.7 we shall relate all these results to the well-known linear axiomatics that have been established since 1984. We conclude with a discussion in Section 2.8. In order to give the reader an impression of the spirit of these Japanese papers, we stick closely to the notations in the original work. The discussions are supplemented with remarks on the philosophical background, physical and biological motivations, results on the deep structure in scale-space and applications to

optical character recognition (OCR). A preliminary version of this paper focusing exclusively on the axiomatic aspects of two of these four Japanese frameworks has been presented in [420].

2.2 Iijima's One-Dimensional Axiomatic (1959)

This section presents first some remarks on the historical and philosophical roots of Japanese scale-space research, and afterwards it describes the axioms and resulting lemmas and theorems of the earliest scale-space approach.

2.2.1 Historical and Philosophical Background

Japanese scale-space research was initiated by Taizo Iijima. After graduating in electrical engineering and mathematics from Tokyo Institute of Technology in 1948, he joined the Electrotechnical Laboratory (ETL). In his thesis titled *'A fundamental study on electromagnetic radiation'* he derived the third analytical solution of the radiation equation. During these studies he acquired the mathematical and physical background for his later scale-space research.

At the ETL Iijima was involved in different research activities on speech and pattern recognition. Triggered by actual needs such as optical character recognition (OCR), but also voice typewriting, or medical diagnosis, he wanted to establish a general theoretical framework for extracting characteristic information of patterns. This framework should avoid extreme standpoints such as purely deterministic or purely stochastic classifications, and it should make use of the original physical or geometric characteristics of the patterns [178].

Besides this problem-driven background, there was also a philosophical motivation for Iijima's scale-space research. Its principles go back to Zen Buddhism, and they can be characterized by the sentence *"Anything is nothing, nothing is anything"*. Applied to the scale-space context this means to obtain the desired information, it is necessary to control the unwanted information. The blurred scale-space evolution may be interpreted as a kind of unwanted information which helps to understand the semantical content of the unblurred image. In this sense important information is existing among seemingly unimportant information.

2.2.2 Axioms

Iijima's first axiomatic formulation of the scale-space concept can be found in a technical paper from 1959 titled *'Basic theory of pattern observation'* [174]. A journal version of this paper has been published in 1962 under the title *'Basic theory on normalization of pattern (In case of typical one-dimensional pattern)'* [175]. Both papers are written in Japanese. The restriction to the one-dimensional

case is for simplicity reasons. Extensions to the two-dimensional case are discussed in Section 2.3.

Iijima imposes basic principles which are in accordance with requirements from observation theory: a robust object recognition should be invariant under changes in the reflected light intensity, parallel shifts in position, and expansions or contractions of the object. In addition to these three transformations he considers an observation transformation Φ which depends on an observation parameter σ and which transforms the original image $g(x)$ into a blurred version $f(x)$.

More precisely, he calls this class of blurring transformations *'boke'* (defocusing), and he assumes that it has the structure[1]

$$f(x) = \Phi[g(x'), x, \sigma] = \int_{-\infty}^{\infty} \phi\{g(x'), x, x', \sigma\}\, dx', \tag{2.3}$$

satisfying four conditions:

(I) *Linearity (with respect to multiplications):*
If the intensity of a pattern becomes A times its original intensity, then the same should happen to the observed pattern:

$$\Phi[Ag(x'), x, \sigma] \;=\; A\,\Phi[g(x'), x, \sigma]. \tag{2.4}$$

(II) *Translation invariance:*
Filtering a translated image is the same as translating the filtered image:

$$\Phi[g(x'-a), x, \sigma] \;=\; \Phi[g(x'), x-a, \sigma]. \tag{2.5}$$

(III) *Scale invariance:*
If a pattern is spatially enlarged by some factor λ, then there exists a $\sigma' = \sigma'(\sigma, \lambda)$ such that

$$\Phi[g(x'/\lambda), x, \sigma] \;=\; \Phi[g(x'), x/\lambda, \sigma']. \tag{2.6}$$

(IV) *(Generalized) semigroup property:*
If g is observed under a parameter σ_1 and this observation is observed under a parameter σ_2, then this is equivalent to observing g under a suitable parameter σ_3:

$$\Phi\Big[\Phi[g(x''), x', \sigma_1], x, \sigma_2\Big] \;=\; \Phi[g(x''), x, \sigma_3], \tag{2.7}$$

where $\sigma_3 = \sigma_3(\sigma_1, \sigma_2)$, but not necessarily $\sigma_3 = \sigma_1 + \sigma_2$.

Later on we shall see that – in order to determine the Gaussian uniquely – this axiomatic has to be supplemented with a fifth requirement: preservation of positivity.

[1] The variable x' serves as a dummy variable.

2.2.3 Consequences

In [175] Iijima establishes four lemmas that confine the class of integral operators (2.3) by consecutively imposing conditions (I) up to (IV).

(a) **Lemma 1:**
If Φ has the structure (2.3) and satisfies the linearity axiom (I), then it can be written as the integral

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi(x, x', \sigma)\, dx'. \tag{2.8}$$

(b) **Lemma 2:**
If Φ is given by (2.8) and satisfies the translation invariance axiom (II), then it can be written as a convolution operation:

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi(x - x', \sigma)\, dx'. \tag{2.9}$$

(c) **Lemma 3:**
If Φ is given by (2.9) and satisfies the scale invariance axiom (III), then it can be written as

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi(\nu(\sigma)(x - x'))\, \nu(\sigma)\, dx', \tag{2.10}$$

where $\nu(\sigma)$ is an arbitrary function of σ.

(d) **Lemma 4:**
If Φ is given by (2.10) and satisfies the semigroup axiom (IV), then the convolution kernel ϕ has the specific structure

$$\phi(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-k^{2m}\xi^{2m} + i\xi u\right) d\xi$$
$$(k \in \mathbb{R},\; m = 1, 2, ...), \tag{2.11}$$

or $\Phi[g(x'), x, \sigma] \equiv 0$.

The proofs of the first three lemmas are rather short and not very complicated, whereas the longer proof of Lemma 4 involves some more sophisticated reasonings

in the Fourier domain: in order to derive the natural numbers m in Lemma 4, he shows that $\Psi(\xi)$, the Fourier transform of $\phi(u)$, must satisfy

$$\Psi(\xi) = \exp\left(\frac{\Psi^{(2m)}(0)}{(2m)!}\,\xi^{2m}\right). \tag{2.12}$$

for some $m \in \mathbb{N}$. Establishing $\Psi^{(2m)}(0) = -(2m)!\,k^{(2m)}$ finally yields (2.11).

In a next step Iijima simplifies the result of Lemma 4. The case $\Phi[g(x'), x, \sigma] \equiv 0$ is of no scientific interest and is not considered any further. In equation (2.10) the function ν can be eliminated by rescaling σ via $\nu(\sigma) = k/\sigma$. Then the k-dependence in (2.11) vanishes by means of substitution of variables.

The preceding results immediately yield

Theorem 1 *If Φ satisfies (2.3) and the axioms (I)–(IV), then it is given by*

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\,\phi_m\left(\frac{x-x'}{\sigma}\right)\,\frac{dx'}{\sigma} \tag{2.13}$$

with

$$\phi_m(u) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-\xi^{2m}+i\xi u\right) d\xi$$
$$(m = 1, 2, ...). \tag{2.14}$$

For this family it follows that σ' in (III) becomes $\sigma' = \sigma/\lambda$, and σ_3 in (IV) satisfies

$$\sigma_3^{2m} = \sigma_1^{2m} + \sigma_2^{2m}. \tag{2.15}$$

For the special case $m = 1$ equation (2.14) becomes

$$\phi_1(u) \;=\; \frac{1}{2\sqrt{\pi}}\exp\left(-\frac{u^2}{4}\right), \tag{2.16}$$

which gives

$$\Phi[g(x'), x, \sigma] \;=\; \frac{1}{2\sqrt{\pi}\sigma}\int_{-\infty}^{\infty} g(x')\exp\left(\tfrac{-(x-x')^2}{4\sigma^2}\right) dx'. \tag{2.17}$$

Thus, $\Phi[g(x'), x, \sigma]$ is just the convolution between g and a Gaussian with standard deviation $\sigma\sqrt{2}$.

In a subsequent theorem Iijima establishes that the case $m = 1$ is of specific interest: if one requires that Φ is *positivity preserving*, i.e.

$$\Phi[f(x'), x, \sigma] > 0 \qquad \forall\, f(x) > 0, \quad \forall\, \sigma > 0, \tag{2.18}$$

then $m = 1$ arises by necessity. The proof presents an explicit example, where the positivity is not preserved for $m > 1$.

This concludes Iijimas axiomatic derivation of the Gaussian kernel by requiring linearity, translation invariance, scale invariance, semi-group property, and preservation of positivity.

Interestingly, Iijima gives also reasons why e.g. the visual perception by a human is carried out through a lens [178]: an optical lens has a Gaussian-like blurring profile that be regarded as a natural consequence from elementary observational principles.

Iijima considered linear scale-space as a first step in his theory of pattern recognition. In [178] one can find an overview of his ideas, the main one-dimensional results, their motivation from a viewpoint of observation theory and their theoretical foundation as a classification tool for OCR and other problems. This paper is written in English.

2.3 Iijima's Two-Dimensional Axiomatic (1962)

Having obtained these one-dimensional results, it was straightforward for Iijima to generalize them to a two-dimensional scale-space axiomatic. This was done in a Japanese technical paper from 1962 [176], followed by a journal paper in 1963 [177], which also contains an informative English abstract.

In [177] he considers a blurring transformation of type

$$f^*(x) \;=\; \Phi[f(x'), x, \Sigma] \;=\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi\{f(x'), x, x', \Sigma\}\, dx'_1 dx'_2, \tag{2.19}$$

where Σ is a symmetric positive definite 2×2 matrix. That transformation should satisfy four conditions:

(I) *Linearity (with respect to multiplications):*

$$\Phi[F f(x'), x, \Sigma] \;=\; F\, \Phi[f(x'), x, \Sigma] \quad \forall\, F \in \mathbb{R}. \tag{2.20}$$

(II) *Translation invariance:*

$$\Phi[f(x'-a), x, \Sigma] \;=\; \Phi[f(x'), x-a, \Sigma] \quad \forall\, a \in \mathbb{R}. \tag{2.21}$$

(III) *Scale invariance and closedness under affine transformations:*
If the pattern is transformed by an (invertible) matrix Λ, then there exists a $\Sigma' = \Sigma'(\Sigma, \Lambda)$ such that

$$\Phi[f(\Lambda^{-1}x'), x, \Sigma] \;=\; \Phi[f(x'), \Lambda^{-1}x, \Sigma']. \tag{2.22}$$

(IV) *(Generalized) semigroup property:*
For every Σ_1 and Σ_2 there exists a $\Sigma_3 = \Sigma_3(\Sigma_1, \Sigma_2)$ such that

$$\Phi\Big[\Phi[f(x''), x', \Sigma_1], x, \Sigma_2\Big] \;=\; \Phi[f(x''), x, \Sigma_3]. \tag{2.23}$$

These four axioms in combination with the requirement of positivity preservation are sufficient to derive that the blurring transformation Φ is given by the affine Gaussian scale-space

$$\Phi[f(x'), x, \Sigma] \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x'_1, x'_2)\, \phi_1(x_1 - x'_1, x_2 - x'_2, \Sigma)\, dx'_1 dx'_2 \tag{2.24}$$

with

$$\phi_1(u_1, u_2, \Sigma) \;=\; \frac{1}{4\pi\sigma^2} \exp\left(-\frac{\mu_{22}u_1^2 - 2\mu_{12}u_1u_2 + \mu_{11}u_2^2}{4\sigma^2}\right) \tag{2.25}$$

and

$$\Sigma \;=\; \sigma^2 \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{pmatrix}, \qquad \det \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{pmatrix} \;=\; 1. \tag{2.26}$$

In order to obtain the usual (isotropic) Gaussian scale-space kernel, one has to impose one more axiom, namely invariance under rotations. Adding this rotational invariance condition to axiom (III) yields the ordinary pure scale invariance condition.

2.4 Iijima's Two-Dimensional Axiomatic Based on Physical Principles (1971)

In 1971 Iijima decided to reconsider his scale-space and pattern recognition theory in order to arrive at a physically consistent reformulation and simplification of his ideas [179, 181, 182, 180, 183, 184, 185, 186].

As a first step in this theory, two-dimensional affine Gaussian scale-space has been derived from physical principles. This is treated in a paper that is available as a complete English translation in [179].

The goal of this paper is to obtain a generalized figure $f(r, \tau)$ from an original figure $f(r)$ in a way which is comparable with the defocusing of an optical system. Two principles are assumed to hold:

(I) *Conservation principle:*
The blurring transformation does not change the total light energy within the image. This means that the image satisfies the continuity equation

$$\frac{\partial f(r, \tau)}{\partial \tau} + \operatorname{div} I(r, \tau) = 0 \tag{2.27}$$

where the flux I denotes the figure flow, r is the location, τ serves as a blurring parameter, and div is the divergence operator in $\mathbb{R}^2$.

(II) *Principle of maximum loss of figure impression:*
The figure flow is determined such that

$$J(I) \;:=\; \frac{|I^T \nabla f|^2}{I^T R^{-1} I} \tag{2.28}$$

takes its maximum value. Here, $R(\tau)$ denotes a positive definite matrix which is the medium constant of the blurring process.

This expression was motivated as follows. Iijima regards $f(r,\tau)$ as an energy density distribution and $\omega(r,\tau) := \frac{1}{2} f^2(r,\tau)$ as a quantity of information which he coins figure impression. He derives that

$$-\frac{d}{d\tau} \int\limits_{\mathbb{R}^2} \omega(r,\tau)\, dr = \int\limits_{\mathbb{R}^2} q(r,\tau)\, dr \tag{2.29}$$

where $q(r,\tau) := -\nabla^T f(r,\tau)\, I(r,\tau)$ is called loss of figure impression.

He measures the flow intensity $\|I\|$ by defining a norm which relates it to $R(\tau)$ via

$$\|I\|^2 := I^T(r,\tau)\, R^{-1}(\tau)\, I(r,\tau). \tag{2.30}$$

Thus, (2.28) describes just the (squared) loss of figure impression normalized by the flow intensity.

Iijima shows that (2.28) is maximized for

$$I(r,\tau) \;=\; -R(\tau) \cdot \nabla f(r,\tau). \tag{2.31}$$

Together with the conservation principle this leads to the anisotropic linear diffusion equation

$$\frac{\partial f(r,\tau)}{\partial \tau} = \mathrm{div}\left(R(\tau) \cdot \nabla f(r,\tau)\right) \tag{2.32}$$

which is just the formulation of affine Gaussian scale-space from Section 2.3 as a partial differential equation. $R(\tau)$ is a diffusion tensor. Iijima calls this equation the *basic equation of figure*.

2.5 Further Scale-Space Results by Iijima and His Students

Iijima considered scale-space as a useful tool in the context of pattern analysis with application to character recognition, feature extraction and focus-of-attention. Below we shall sketch some key ideas of his recognition theory as well as its applications to OCR and deep structure scale-space analysis.

2.5.1 Iijima's Pattern Recognition Theory

In [179] Iijima shows that the basic equation of figure is essentially invariant under conformal transformations describing multiplication of grey values with a constant, addition of linear brightness gradients, translations, and affine transformations. Essentially invariant means that it may be transformed into another anisotropic linear diffusion equation.

The set of these conformal mappings form an algebraic group, while the set of affine Gaussian blurring transformations form a semigroup. Iijima studies compositions of a conformal mapping and a blurring transformation under the name observational transformations. These transformations are used to construct a theory of pattern classification: two figures are considered as equivalent if they result from the same original figure by observational transformations.

Iijima compares the invariance of the basic equation of figure under conformal transformations with the invariance of Newton's basic equations of motion to Galileo's transformation, and the invariance of the Maxwell equations to the Lorentz transformation. It seems that he was fully aware of the future importance of his discovery when he wrote in [179] that *"this paper provides a basis for exploring the recognition theory of visual patterns and solving mathematically the various problems in visual physiology"*.

This subsequent recognition theory is documented in a series of Japanese papers which are available as full English translations [181, 182, 180, 183, 184, 185, 186]. Iijima regards a Gaussian-blurred figure as an element in a Hilbert space which can be expanded in an orthonormal function system given in terms of Hermite polynomials. The similarity between two patterns f and g is a function of the scalar product $(.,.)$ in this Hilbert space:

$$S(f,g) := \frac{(f,g)}{\|f\|\,\|g\|}, \tag{2.33}$$

where $\|f\| := \sqrt{(f,f)}$ denotes the induced norm.

Gaussian blurring plays a central role in this theory. It should be performed at the very beginning, because it makes the algorithms insensitive to noise, it reduces effects of positional deviation, and it allows a coarser sampling. In order to overcome the problem that blurred patterns become more similar, he devised a specific canonical transformation. Incorporating all these features has lead Iijima to a scale-space based subspace technique for matching two patterns which he called *simple similarity method*. In order to determine the similarity between an observed pattern and a whole equivalence class of reference patterns (e.g. different realizations of one letter of the alphabet), Iijima proposed an extension named *multiple similarity method* [190]. It is essentially a weighted mean square of the simple similarity measures between the observed pattern matched with each of the reference patterns.

In 1973 Iijima condensed his whole scale-space and pattern classification theory to a Japanese textbook [187]. It can be regarded as one of the first monographs on linear scale-space theory.

2.5.2 Applications to Character Recognition

Applications of Iijima's theory to OCR have been presented in English proceedings papers at the First USA–Japan Computer Conference in 1972 [189], and at the First International Joint Conference on Pattern Recognition in 1973 [190]. In [189] it is described how Iijima, Genchi and Mori have realized the multiple similarity method in hardware in the optical character reader ASPET/71. Gaussian blurring was performed by optical defocusing. ASPET/71 was capable of reading 2000 alphanumeric characters per second, and the scale-space part has been regarded as the reason for its reliability and robustness. Others stated about ASPET/71 that *"it has been proved to have better performance than any similar conventional method"* [279]. It is now exhibited at the National Museum of Sciences in Tokyo. The ASPET/71 was an analog machine, but its commercial variant OCR-V100 by Toshiba used digital technology fully. Iijima's multiple similarity method has also become the main algorithm of Toshiba's later OCR systems [254].

2.5.3 Deep Structure Analysis

In the eighties Iijima addressed together with Nan-yuan Zhao, a Ph.D. student of him, the problem of deep structure analysis in scale-space [447]; see also the discussion in [192]. For a solution $f(x,t)$ of the isotropic linear diffusion equation, they constructed a curve which comprises the stationary points, i.e. locations (x,t) with $\nabla f(x,t) = 0$. This so-called *stationary curve* $r(t)$ obeys the equation

$$\mathrm{Hess}(f)\,\frac{dr(t)}{dt} = -\Delta\left(\nabla f(r,t)\right). \tag{2.34}$$

They stated criteria for identifying *stable viewpoints* on the stationary curve, for instance by requiring that $\frac{dr(t)}{dt}$ vanishes there without vanishing in a neighbourhood. Afterwards they linked these stable viewpoints to a topological scale-space tree [446], which introduces a scale hierarchy. To each stable point (x_i, t_i) they assign a region of interest which is given by a disk with center x_i and radius t_i. Applied to an image of Zhao himself, this focus-of-attention method extracted eyes, nostrils and the mouth as regions of interest [445].

Some of this work on scale-space trees was further pursued by the image processing group of Makoto Sato, another former Ph.D. student of Iijima. Sato's group established linear scale-space results ranging from deep structure analysis [335, 336, 401, 83] to the filtering of periodic or spherical patterns [400, 211]. All cited Sato papers are written in English.

Iijima continued his research on linear scale-space techniques till the nineties [22]. After 1972 he held professorships at Tokyo Institute of Technology, Tokyo Engineering University, and the Advanced Institute of Science and Technology. In spring 1997 he retired at the age of 72. An English bibliography can be found in [22].

2.6 Otsu's Two-Dimensional Axiomatic (1981)

Iijima's scale-space research also inspired other people than his direct students. We shall illustrate this below by discussing the work of Nobuyuki Otsu on scale-space axiomatics and deblurring.

2.6.1 Derivation of the Gaussian

In 1981 another Japanese scale-space axiomatic has been established in the Ph.D. thesis of Nobuyuki Otsu [292]. This thesis, which was mainly concerned with the extraction of invariants for pattern recognition, was written at the ETL, where Iijima was working in the sixties. Otsu derived two-dimensional Gaussian scale-space in an axiomatic way by modifying the axioms described in Section 2.2. Section 4.1 of his thesis is titled *'Axiomatic derivation of the scale transformation'*. There he considers some transformation of an image f into an image $\tilde{f}$, for which the following holds:

(I) *Representation as a linear integral operator:*
There exists a function $W : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ such that

$$\tilde{f}(r) \;=\; \int_{\mathbb{R}^2} W(r, r')\, f(r')\, dr' \qquad \forall\, r \in \mathbb{R}^2. \tag{2.35}$$

(II) *Translation invariance:*
For all $r \in \mathbb{R}^2$ and for all $a \in \mathbb{R}^2$ it is required that

$$\tilde{f}(r-a) \;=\; \int_{\mathbb{R}^2} W(r, r')\, f(r'-a)\, dr'. \tag{2.36}$$

Since this is just $\int_{\mathbb{R}^2} W(r, r'+a)\, f(r')\, dr'$, and (I) states that $\tilde{f}(r-a) = \int_{\mathbb{R}^2} W(r-a, r')\, f(r')\, dr'$, it follows that the integral kernel is symmetric,

$$W(r, r'+a) \;=\; W(r-a, r'), \tag{2.37}$$

and, thus, it is a convolution kernel:

$$W(r, r') \;=\; W(r - r'). \tag{2.38}$$

(III) *Rotation invariance (of the kernel):*
For all rotation matrices T_Θ and for all $r = (x, y)^T \in \mathbb{R}^2$ it is assumed that

$$W(T_\Theta r) \;=\; W(r). \tag{2.39}$$

Hence, W depends only on $|r|$: $W(r) = W(x^2 + y^2)$.

(IV) *Separability:*
There exists a function $u : \mathbb{R} \to \mathbb{R}$ such that

$$W(r) \;=\; u(x)\, u(y). \tag{2.40}$$

Combining this with (III) implies after elementary manipulations that

$$W(r) \;=\; k\, \exp\left[c\,(x^2{+}y^2)\right]$$

with some parameters $k, c \in \mathbb{R}$. In order to arrive at $k > 0$ and $c < 0$, however, additional constraints are needed.

(V) His next requirement which he names "Normalization of energy" actually consists of two parts:
Preservation of nonnegativity,

$$\tilde{f}(r) \geq 0 \qquad \forall\, f(r) \geq 0, \tag{2.41}$$

and *average grey level invariance,*

$$\int\limits_{\mathbb{R}^2} \tilde{f}(r)\, dr \;=\; \int\limits_{\mathbb{R}^2} f(r)\, dr. \tag{2.42}$$

This leads to $W(r) \geq 0$ and $\int_{\mathbb{R}^2} W(r)\, dr = 1$, respectively.

Combining these results gives $k = \frac{1}{2\pi\sigma^2}$ and $c = -\frac{1}{2\sigma^2}$. This yields the Gaussian kernel

$$W(r) \;=\; \frac{1}{2\pi\sigma^2}\, \exp\left(-\frac{x^2{+}y^2}{2\sigma^2}\right) \tag{2.43}$$

and concludes the axiomatic derivation of two-dimensional linear scale-space.

2.6.2 Further Results

Section 4.2 of Otsu's thesis is titled *'Representation of scale-space transformation and semigroup'*. It is devoted to the N-dimensional Gaussian scale-space. With $\rho := \sigma^2/2$ he defines

$$T(\rho) f(r) := \frac{1}{(4\pi\rho)^{N/2}}\, \exp\left(-\frac{|r|^2}{4\rho}\right) * f(r). \tag{2.44}$$

Using Fourier techniques he shows that the generator of the scale-space transformation is the Laplacean:

$$\tilde{f}(r,\rho) \;=\; T(\rho)f(r) \;=\; \exp(\rho\Delta)\, f(r). \tag{2.45}$$

This gives

$$\frac{\partial \tilde{f}(r,\rho)}{\partial \rho} \;=\; \Delta\Big(\exp(\rho\Delta)f(r)\Big) \;=\; \Delta\,\tilde{f}(r,\rho).$$

Thus, $\tilde{f}$ satisfies the isotropic linear diffusion equation.

The formal inversion of the scale-space transformation by means of (2.45) is

$$\begin{aligned} f(r) &= [T(\rho)]^{-1}\tilde{f}(r,\rho) \\ &= \exp(-\rho\Delta)\,\tilde{f}(r,\rho) \\ &= \Big(I - \rho\Delta + \frac{\rho^2}{2}\Delta^2 - ...\Big)\,\tilde{f}(r,\rho). \end{aligned}$$

For the case that ρ or $\Delta^2\tilde{f}$ is small, Otsu proposes to approximate $[T(\rho)]^{-1}$ by $[I-\rho\Delta]$ and to use it for recovering the original image from a blurred one[2].

2.7 Relation to Other Work

Having sketched the basic ideas of these Japanese axiomatics, it is natural to ask about similarities and differences to other approaches. Table 2.1 gives an overview of the current axiomatics for the continuous Gaussian scale-space. These axioms and some of their relations can be explained as follows[3]:

- **Convolution kernel:**
  There exists a family of functions $\{k_t : \mathbb{R} \to \mathbb{R} \mid t \geq 0\}$ such that

  $$(T_t f)(x) = \int_{\mathbb{R}^N} k_t(x-x')\, f(x')\, dx'.$$

  In Section 2.6.1 we have already seen that this property can be derived from the two assumptions:

  - **Linear integral operator:**
    There exists a family of kernels $\{k_t \mid t \geq 0\}$ with

    $$(T_t f)(x) = \int_{\mathbb{R}^N} k_t(x,x')\, f(x')\, dx'.$$

[2] This is an ill-posed problem which may lead to unstable results.

[3] Of course, such a table can only give a "flavour" of the different approaches, and the precise description of each axiom may slightly vary from paper to paper. Several relations between the presented axioms are discussed in [12, 231, 295].

Table 2.1: Overview of continuous Gaussian scale-space axiomatics (I1 = Iijima [174, 175], I2 = Iijima [176, 177], I3 = Iijima [179], O = Otsu [292], K = Koenderink [213], Y = Yuille/Poggio [442], B = Babaud et al. [28], L1 = Lindeberg [229], F1 = Florack et al. [130], A = Alvarez et al. [12], P = Pauwels et al. [295], N = Nielsen et al. [272], L2 = Lindeberg [231], F2 = Florack [128]).

	I1	I2	I3	O	K	Y	B	L1	F1	A	P	N	L2	F2
convolution kernel	•	•		•		•	•	•	•	•	•		•	•
semigroup property	•	•						•	•	•	•	•	•	•
locality										•				
regularity						•	•	•	•	•	•		•	•
infinites. generator											•			
max. loss principle			•											
causality					•	•	•	•					•	
nonnegativity	•	•		•						•	•			•
Tikhonov regular.												•		
aver. grey level invar.			•	•			•	•		•	•			
flat kernel for $t \to \infty$						•			•					
isometry invariance		•		•		•	•	•	•	•	•	•	•	•
homogen. & isotropy					•									
separability				•					•					
scale invariance	•	•				•	•		•		•	•		•
valid for dimension	1	2	2	2	1,2	1,2	1	1	> 1	N	1,2	N	N	N

Since every continuous linear functional can be written as an integral operator, it follows that Florack's topological duality paradigm [128] can also be interpreted as requiring the existence of a linear integral operator[4].

– **Translation invariance:**
Let a translation τ_a be defined by $(\tau_a f)(x) := f(x-a)$. Then,

$$\tau_a T_t = T_t \tau_a \qquad \forall\, a \in \mathbb{R}^N, \quad \forall\, t > 0.$$

Since usually linearity and translation invariance are imposed in conjunction, we have summarized them under the term "convolution kernel".

- **Semigroup property:**

$$T_{t+s} f = T_t (T_s f) \qquad \forall\, t, s \geq 0, \quad \forall\, f.$$

This property ensures that one can implement the scale-space process as a cascade smoothing.

[4]Dirac point distributions and their derivatives are admitted as "functions under the integral".

- **Locality:**
  For small t the value of $T_t f$ at any point x is determined by its vicinity:

  $$\lim_{t \to 0^+} (T_t f - T_t g)(x) = \mathrm{o}(x)$$

  for all $f, g \in \mathrm{C}^\infty$ whose derivatives of order ≥ 0 are identical.

- **Regularity:**
  A precise definition of the smoothness requirements for the scale-space operator depends on the author:

  - Since the original image creates the scale-space, it is natural to assume that it is continuously embedded, i.e. $\lim_{t \to 0^+} T_t = I$. In the linear convolution case, this means that $k_t(x)$ tends to Dirac's delta distribution [442] and its Fourier transform becomes 1 everywhere [130].
  - Babaud et al. [28] and Florack [128] consider infinitely times differentiable convolution kernels which are rapidly decreasing functions in x, i.e. they are vanishing at ∞ faster than any inverse of polynomials.
  - Lindeberg uses kernels k_t which are Borel measurable in t [229], or kernels which converge for $t \to 0^+$ in the L^1 norm to the Dirac distribution [231].
  - Alvarez et al. [12] require that

    $$\|T_t(f + hg) - (T_t(f) + hg)\|_\infty \leq Cht$$

    for all $h, t \in [0, 1]$, and for all smooth f, g, where C may depend on f and g.
  - Pauwels et al. [295] assume that the convolution kernel $k_t(x)$ is separately continuous in x and in t.

- **Infinitesimal generator:**
  The existence of

  $$\lim_{t \to 0^+} \frac{T_t f - f}{t} =: A[f]$$

  guarantees that the semigroup can be represented by the evolution equation

  $$\partial_t u = A[u].$$

  From the mathematical literature it is well-known that the existence of an infinitesimal generator follows from the semigroup property when being combined with regularity assumptions [169].

- **Principle of maximum loss of figure impression:**
  See Section 2.4.

- **Causality:**
  The scale-space evolution should not create new level curves when increasing the scale parameter. If this is satisfied, iso-intensity linking through the scales is possible and a structure at a coarse scale can (in principle) be traced back to the original image.

  For this reason, Koenderink [213] required that at spatial extrema (with non-vanishing determinant of the Hessian) isophotes in scale-space are upwards convex; In two dimensions he showed that at these extrema the diffusion equation

  $$\partial_t u = \alpha(x,t)\,\Delta u \tag{2.46}$$

  has to be satisfied. Hereby, α denotes a positive-valued function.

  Hummel [173] established the equivalence between causality and a maximum principle for certain parabolic operators.

  We may also derive the causality equation (2.46) and its N-dimensional generalizations by requiring that local extrema with positive or negative definite Hessians are not enhanced [28, 231]: This assumption states that such an extremum in x_0 at scale t_0 satisfies

  $$\begin{aligned} \partial_t u > 0 \qquad & \text{if } x_0 \text{ is a minimum,} \\ \partial_t u < 0 \qquad & \text{if } x_0 \text{ is a maximum.} \end{aligned}$$

  This is just the causality requirement $\mathrm{sign}(\partial_t u) = \mathrm{sign}(\Delta u)$. Moreover, in one dimension, nonenhancement of local extrema is equivalent to the requirement that the number of local extrema does not increase [28, 229]. In higher dimensions, however, diffusion scale-spaces may create new extrema, see e.g. [442, 228, 98].

- **Nonnegativity:**
  If the nonnegativity of the convolution kernel,

  $$k_t(x) \geq 0 \qquad \forall\, x, \quad \forall\, t > 0,$$

  is violated, new level crossings may appear for $t > 0$, such that the causality property does not hold.

  Within a linear framework with spatially continuous convolution kernels, nonnegativity is equivalent to the monotony requirement [12]

  $$\begin{aligned} & f(x) \leq g(x) \quad \forall\, x \\ \Longrightarrow \quad & (T_t f)(x) \leq (T_t g)(x) \quad \forall\, x,\ \forall\, t > 0 \end{aligned}$$

and the preservation of nonnegativity:

$$\begin{aligned} & f(x) \geq 0 \quad \forall\, x \\ \Longrightarrow \quad & (T_t f)(x) \geq 0 \quad \forall\, x,\ \forall\, t > 0. \end{aligned}$$

- **Tikhonov regularization:**
  In the one-dimensional case, u is called a Tikhonov regularization of $f \in \mathrm{L}^2(\mathbb{R})$, if it minimizes the energy functional

$$E_f[u] = \int_{\mathbb{R}} \left[(f-u)^2 + \sum_{i=1}^{\infty} \lambda_i \left(\frac{d^i u}{dx^i} \right)^2 \right] dx.$$

  This concept and an N-dimensional generalization of it has been used by Nielsen, Florack and Deriche [272]. The first term under the integral ensures that u remains close to f, while the second one is responsible for the smoothness of u: the positive coefficients λ_i serve as weights which guarantee smoothness of the i-th derivative.

- **Average grey level invariance:**
  The average grey level invariance

$$\int_{\mathbb{R}^N} T_t f \, dx = \int_{\mathbb{R}^N} f \, dx \qquad \forall\, t > 0$$

  can be achieved by means of the continuity equation (2.27) in connection with reflecting or periodic boundary conditions. It boils down to the normalization condition

$$\int_{\mathbb{R}^N} k_t(x) \, dx = 1,$$

  if we consider linear convolution kernels.

  In this case normalization is also equivalent to grey level shift invariance [12]:

$$\begin{aligned} T_t(0) &= 0, \\ T_t(f + C) &= T_t(f) + C \end{aligned}$$

  for all images f and for all constants C.

- **Flat kernel for $t \to \infty$:**
  For $t \to \infty$, one expects that the kernel spreads the information uniformly over the image. Therefore, if the integral over the kernel should remain finite, it follows that the kernel has to become entirely flat: $\lim_{t \to \infty} k_t(x) = 0$.

- **Isometry invariance:**
  Let $R \in \mathbb{R}^N$ be an orthogonal transformation (i.e. $\det R = \pm 1$) and define $(Rf)(x) := f(Rx)$. Then,

  $$T_t(Rf) = R(T_t f) \qquad \forall f, \quad \forall t > 0.$$

  In the one-dimensional case with a linear convolution kernel this invariance under rotation and mirroring comes down to the symmetry condition $k_t(x) = k_t(-x)$.

- **Homogeneity and isotropy:**
  Koenderink [213] required that the scale-space treats all spatial points equally. He assumed that the diffusion equation (2.46), which results at extrema from the causality requirement, should be the same at each spatial position (regardless whether there is an extremum or not) and for all scales. He named these requirements homogeneity and isotropy[5].

- **Separability:**
  The convolution kernel $k_t(x)$ with $x = (x_1, ..., x_N)^T \in \mathbb{R}^N$ may be split into N factors, each acting along one coordinate axis:

  $$k_t(x) = k_{1,t}(x_1) \cdots k_{N,t}(x_N).$$

- **Scale invariance:**
  Let $(S_\lambda f)(x) := f(\lambda x)$. Then there exists some $t'(\lambda, t)$ with

  $$S_\lambda T_{t'} = T_t S_\lambda.$$

  One may achieve this by requiring that, in the N-dimensional case, the convolution kernel k_t has the structure

  $$k_t(x) = \frac{1}{\Psi^N(t)} \Phi\left(\frac{x}{\Psi(t)}\right)$$

  with a continuous, strictly increasing rescaling function Ψ. This means that all kernels can be derived by stretching a parent kernel such that its area remains constant [295]. It is evident that this is related to the normalization condition.

  Scale invariance follows also from the semigroup property when being combined with isometry invariance and causality [231]. Moreover, scale invariance, translation invariance and isometry invariance result from the more general assumption of invariance under the spacetime symmetry group; see [128] for more details.

[5] In our terminology, homogeneity and isotropy are much stricter requirements than translation and isometry invariance. They enable Koenderink to derive Gaussian scale-space under only one additional assumption (causality).

We observe that – despite the fact that all presented axiomatics use many similar requirements – not two of them are identical. Each of the 14 axiomatics confirms and enhances the evidence that the others give: that Gaussian scale-space is unique within a linear framework. This theoretical foundation is the backbone of a lot of successful applications of linear scale-space theory.

Nevertheless, apart from their historical merits, the early Japanese approaches differ from the well-known axiomatics after 1984 in several aspects:

Firstly, it is interesting to note that all Japanese axiomatics require only quite a few axioms in order to derive Gaussian scale-space. Even recent approaches which intend to use a minimal set of first principles do not utilize less axioms.

Iijima's one- and two-dimensional frameworks from 1959 and 1962, respectively, do not only belong to the most systematic derivations of Gaussian scale-space, they appear also rather modern: principles such as nonnegativity and the semigroup property are typical for axiomatizations after 1990, and also the importance of scale invariance has been emphasized mainly in recent years. Also affine Gaussian scale-space which he pioneered in 1962 has been further pursued only recently [230].

Iijima's physical motivation for affine Gaussian scale-space from 1971 uses only two principles which reduce the essential features of linear diffusion filtering to a minimum. In this sense it is written in a similar spirit as Koenderink's derivation [213], which also uses two highly condensed requirements (although of very different nature). Concepts such as the principle of maximum loss of figure impression may remind some of the readers of properties of nonlinear scale-spaces like the Euclidean and affine shortening flow [12, 286, 333]: they shrink the Euclidean or affine perimeter of a closed curve as fast as possible. Moreover, the group-theoretical studies in [179] prove that Iijima has also pioneered modern scale-space analysis such as in [12, 286, 328]. The canonical figures used in [179] to set up similarity measures can be regarded as predecessors of the canonical frames that have become popular in computer vision [258]. Last but not least, Iijima's anisotropic diffusion equation (2.32) may be viewed as a linear predecessor of various nonlinear anisotropic diffusion scale-spaces as considered in [412].

Otsu's two-dimensional axiomatic is very appealing due to its simplicity: in contrast to many other approaches it does not require advanced mathematical techniques like Fourier analysis, complex integrals, or functional analysis in order to derive the uniqueness of the Gaussian kernel. It is therefore a well-suited approach even for undergraduate courses in image processing.

2.8 Discussion

In this paper we have analysed Japanese axiomatics for the linear diffusion scale-space that have been unknown in the American and European image processing community. They reveal many interesting qualities which should trigger everyone who is interested in scale-space theory to have a closer look at them.

The discussed results demonstrate that an entire world of linear scale-space theory has evolved in Japan ranging from axiomatics for isotropic and affine Gaussian scale-space over group theory and deep structure analysis to hardware implementations for OCR. The Japanese scale-space paradigm was well-embedded into a general framework for pattern recognition and object classification [178, 187, 292], and many results have been established earlier than in America and Europe.

It is surprising that eastern and western scale-space theory have evolved with basically no interaction: To the best of our knowledge, the first citation of Iijima's work by non-Japanese scale-space researchers was made in 1996 [428]. Conversely, also Japanese work after 1983 was not always aware of American and European scale-space results. Some (English) papers by Makoto Sato's group [335, 336, 401, 400] cite both Iijima and Witkin. His paper with the probably most explicit reference to Iijima's work has been presented at the ICASSP '87 [335], where Sato and Wada cite the original Japanese versions of [179, 181] and state: *"The notion same as scale-space filtering was also proposed by T. Iijima in the field of pattern recognition. He derived a partial differential equation, called basic equation, from the continuity of the light energy in the waveform observation"*.

In 1992 several direct hints to Iijima's scale-space research can be found in the widespread journal *Proceedings of the IEEE* [254]. In an invited historical review of OCR methods written by Mori, Suen and Yanamoto, 10 out of the 193 key references were papers by Iijima. Concerning his contributions, the authors state that *"the concept of blurring was first introduced into pattern recognition by his work, whereas it was widely attributed to Marr in the West. Iijima's idea was derived from his study on modeling the vision observation system. (...) Setting reasonable conditions for the observation system, he proved that the mathematical form must be a convolution of a signal $f(x')$ with a Gaussian kernel"*. They also give a recommendation to the non-Japanese audience: *"Iijima's theory is not so easy to understand, but his recent book [188] is readable, although it is written in Japanese"*.

One can only speculate why nobody paid attention to these passages. Maybe, because none of the authors referred to one of Iijima's *English* scale-space papers. The present paper contains 12 references to English publications by Iijima. They can be found in many libraries in America and Europe, and a short look at papers such as [178, 179] should convince everybody that there remains no justification to deny Iijima's pioneering role in linear scale-space theory because of language

reasons.

Another reason might be that Iijima's work came too early to be appreciated: His theory was mathematically much more demanding than other methods at this time. At a stage where pattern recognition was still in its infancy and experimenting with very simple methods, it was not easy to make techniques popular, which are based on advanced mathematics. Also computing facilities were more restricted in the sixties and seventies than they are today. Despite very remarkable developments such as the scale-space based optical character readers, it was certainly more difficult to attract people by presenting computed results that demonstrate the advantages of a conceptually clean scale-space technique over ad-hoc strategies.

When scale-space became popular in America and Europe in the eighties, the situation was different: Computing power was much higher, and the pattern recognition and computer vision community had become mature for better founded techniques which take advantage of centuries of research in mathematics and physics. Today it is possible to establish an international conference solely devoted to scale-space ideas which attracts people from many countries and disciplines [381]. Would this have been possible 30 years ago? Certainly not.

Unfortunately, it seems that Iijima was not the only one who has pioneered the field of partial differential equations in image processing far ahead of his time, so that his work fell into oblivion for decades. Another example is the fact that already in 1965 the Nobel prize winner Dennis Gabor – the inventor of optical holography and the so-called Gabor functions – proposed a deblurring algorithm based on combining mean curvature flow with backward smoothing along flowlines [138, 233]. This long-time forgotten method is similar to modern PDE techniques for image enhancement.

Maybe this review helps a little bit that the pioneering work of these people receives the acknowledgement that it deserves. It would also be nice if it encourages the mutual interest between the Japanese and the western scale-space community. The fact that the same theory has been developed twice in two very different cultures shows that this theory is natural and that it is worthwhile to study all of its various aspects.

Acknowledgements. We thank Alfons Salden, Luc Florack, Peter Johansen and Jon Sporring for helpful comments and stimulating discussions. Parts of this work have been carried out while Seiji Ishikawa and Joachim Weickert were with the Image Sciences Institute of Utrecht University, The Netherlands. It has also been supported by the Danish National Science Research Council under grant no. 9502164.

Chapter 3

Relations Between Regularization and Diffusion Filtering

Otmar Scherzer and Joachim Weickert.
Journal of Mathematical Imaging and Vision,
Vol. 12, No. 1, 43–63, February 2000.

Abstract

Regularization may be regarded as diffusion filtering with an implicit time discretization where one single step is used. Thus, iterated regularization with small regularization parameters approximates a diffusion process. The goal of this paper is to analyse relations between noniterated and iterated regularization and diffusion filtering in image processing. In the linear regularization framework, we show that with iterated Tikhonov regularization noise can be better handled than with noniterated. In the nonlinear framework, two filtering strategies are considered: the total variation regularization technique and the diffusion filter technique of Perona and Malik. It is shown that the Perona-Malik equation decreases the total variation during its evolution. While noniterated and iterated total variation regularization is well-posed, one cannot expect to find a minimizing sequence which converges to a minimizer of the corresponding energy functional for the Perona–Malik filter. To overcome this shortcoming, a novel regularization technique of the Perona–Malik process is presented that allows to construct a weakly lower semi-continuous energy functional. In analogy to recently derived results for a well-posed class of regularized Perona–Malik filters, we introduce Lyapunov functionals and convergence results for regularization methods. Experiments on real-world images illustrate that iterated linear regularization performs better than noniterated, while no significant differences between noniterated and iterated total variation regularization have been observed.

3.1 Introduction

Image restoration is among other topics such as optic flow, stereo, and shape-from-shading one of the classical inverse problems in image processing and computer vision [40]. The inverse problem of image restoration consists in recovering information about the original image from incomplete or degraded data. Diffusion filtering has become a popular and well-founded tool for restoration in the image processing community [382, 412], while mathematicians have unified most techniques to treat inverse problems under the theory of regularization methods [114, 156, 255, 387]. Therefore it is natural to investigate relations between both approaches, as this may lead to a deeper understanding and a synthesis of these techniques. This is the goal of the present paper.

We can base our research on several previous results. In the linear setting, Torre and Poggio [388] emphasized that differentiation is ill-posed in the sense of Hadamard, and applying suitable regularization strategies approximates linear diffusion filtering or – equivalently – Gaussian convolution. Much of the linear scale-space literature is based on the regularization properties of convolutions with Gaussians. In particular, differential geometric image analysis is performed by replacing derivatives by Gaussian-smoothed derivatives; see e.g. [127, 230, 369] and the references therein. In a very nice work, Nielsen et al. [273] derived linear diffusion filtering axiomatically from Tikhonov regularization, where the stabilizer consists of a sum of squared derivatives up to infinite order.

In the nonlinear diffusion framework, natural relations between biased diffusion and regularization theory exist via the Euler equation for the regularization functional. This Euler equation can be regarded as the steady-state of a suitable nonlinear diffusion process with a bias term [281, 345, 80]. The regularization parameter and the diffusion time can be identified if one regards regularization as time-discrete diffusion filtering with a single implicit time step [253, 378, 339]. A popular specific energy functional arises from unconstrained total variation denoising [2, 76, 77]. Constrained total variation also leads to a nonlinear diffusion process with a bias term using a time-dependent Lagrange multiplier [326].

In spite of these numerous relations, several topics have not been addressed so far in the literature:

- *A comparison of the restoration properties of both approaches:* Since regularization corresponds to time-discrete diffusion filtering with a single time step, it follows that iterated regularization with a small regularization parameter gives a better approximation to diffusion filtering. An investigation whether iterated regularization is better than noniterated leads therefore to a comparison between regularization and diffusion filtering.

- *Energy formulations for stabilized Perona–Malik processes:* The Perona–Malik

filter is the oldest nonlinear diffusion filter [303]. Its ill-posedness has triggered many researchers to introduce regularizations which have shown their use for image restoration. However, no regularization has been found which can be linked to the minimization of an appropriate energy functional.

- *Lyapunov functionals for regularization:* The smoothing and information-reducing properties of diffusion filters can be described by Lyapunov functionals such as decreasing L^p norms, decreasing even central moments, or increasing entropy [412]. They constitute important properties for regarding diffusion filters as scale-spaces. A corresponding scale-space interpretation of regularization methods where the regularization parameter serves as scale parameter has been missing so far.

These topics will be discussed in the present paper. It is organized as follows. Section 3.2 explains the relations between variational formulations of diffusion processes and regularization strategies. In Section 3.3 we first discuss the noise propagation for noniterated and iterated Tikhonov regularization for linear problems. In the nonlinear framework, well-posedness results for total variation regularization are reviewed and it is explained why one cannot expect to establish well-posedness for the Perona–Malik filter. We will argue that, if the Perona–Malik filter admits a smooth solution, however, then it will be total variation reducing. A novel regularization will be introduced which allows to construct a corresponding energy functional. Section 3.4 establishes Lyapunov functionals for regularization methods that are in accordance with those for diffusion filtering. This leads to a scale-space interpretation for linear and nonlinear regularization. In Section 3.5 we shall present some experiments with noisy real-world images, that compare the restoration properties of noniterated and iterated regularization in the linear setting and in the nonlinear total variation framework. Moreover, the novel Perona–Malik regularization is juxtaposed to the regularization by Catté et al. [74]. Some preliminary results have been reported in a conference volume [310].

3.2 Variational Formulations of Diffusion Processes and the Connection to Regularization Methods

We consider a general *diffusion process* of the form

$$\begin{cases} \partial_t u = \nabla . (g(|\nabla u|^2)\nabla u) & \text{on } \Omega \times (0,\infty) \\ \partial_n u = 0 & \text{on } \Gamma \times (0,\infty) \\ u(x,0) = f_\delta(x) & \text{on } \Omega \, . \end{cases} \tag{3.1}$$

Here g is a smooth function satisfying certain properties which will be explained in the course of the paper; $\Omega \subseteq \mathbb{R}^d$ is a bounded domain with piecewise Lipschitzian boundary Γ with unit normal vector n, and f_δ is a degraded version of the original image $f := f_0 : \Omega \to \mathbb{R}$. The operator ∇ denotes the nabla operator in $\mathbb{R}^d$, i.e. $\nabla u = (\partial_{x_1} u, ..., \partial_{x_d} u)^T$, and $\nabla\cdot$ is its corresponding divergence operator.

For the numerical solution of (3.1) one can use explicit or implicit or semi-implicit difference schemes with respect to t.

The implicit scheme reads as follows.

$$\begin{cases} \frac{u(x,t+h)-u(x,t)}{h} &= \nabla.\,(g(|\nabla u|^2)\nabla u)\,(x,t+h) \\ u(x,0) &= f_\delta(x)\,. \end{cases} \tag{3.2}$$

Here $h > 0$ denotes the step-size in t–direction of the implicit difference scheme.

In the following we assume that g is measurable on $[0,\infty[$ and there exists a differentiable function $\hat{g}$ on $[0,\infty[$ which satisfies $\hat{g}' = g$. Then the minimizer of the functional (for some fixed given $u(x,t)$)

$$T(u) \;:=\; \|u - u(x,t)\|^2_{\mathrm{L}^2(\Omega)} + h \int_\Omega \hat{g}(|\nabla u|^2)\,dx, \tag{3.3}$$

satisfies $(T'(u), v) = 0$ for all v, where $(T'(u), v)$ is the formal Gateaux derivative of T in direction v:

$$\begin{aligned} (T'(u), v) &= \lim_{t\to 0^+} \frac{T(u+tv) - T(u)}{t} \\ &= \int_\Omega 2(u-f)v\,dx + h\int_\Omega 2g(|\nabla u|^2)\nabla u \nabla v\,dx. \end{aligned}$$

We observe that such a minimizer satisfies (3.2) at time $t+h$. If the functional T is convex, then a minimizer of T is uniquely characterized by the solution of the equation (3.2) with homogeneous Neumann boundary conditions.

The regularization functional $T(u)$ consists of the approximation functional $\|u-u(.,t)\|^2_{\mathrm{L}^2(\Omega)}$ and the stabilizing functional $\int_\Omega \hat{g}(|\nabla u|^2)\,dx$. The weight h is called *regularization parameter*. The case $\hat{g}(x) = x$ is called *Tikhonov regularization*.

In the next section we summarize some results on regularization and diffusion filtering and compare the theoretical results developed in both theories.

3.3 A Survey on Diffusion Filtering and Regularization

We have seen that each time step for the solution of the diffusion process (3.1) with an implicit, t-discrete scheme is equivalent to the calculation of the minimizer of the regularization functional (3.3). The numerical solution of the diffusion process with an implicit, t-discrete iteration scheme is therefore equivalent to *iterated regularization* where on has to minimize iteratively the set of functionals

$$T_n(u) := \|u - u_{n-1}\|^2_{\mathrm{L}^2(\Omega)} + h_n \int_\Omega \hat{g}(|\nabla u|^2)\, dx. \tag{3.4}$$

Here u_n is a minimizer of the functional $T_n, n = 1, 2, ...,$ and $u_0 := f_\delta$. If the functionals T_n are convex, then the minimizer of (3.4) denoted by u_n is the approximation of the solution of the diffusion process with an implicit, t-discrete method at time $t_1, ..., t_n$ where $t_k = \sum_{j=1}^k h_j$.

In the following we refer to iterated regularization if $h_n = h$ for all n. That corresponds to the solution of the diffusion process with an implicit, t-discrete method using a fixed time step size $h = h_n$.

If the regularization parameters h_n are adaptively chosen (this corresponds to the situation that the time discretization in the diffusion process is changed adaptively), then the method is called *nonstationary regularization*. For some recent results on nonstationary Tikhonov regularization we refer to Hanke and Groetsch [160]; however, their results do not fit directly into the framework of this paper. They deal with regularization methods for the stable solution of operator equations

$$Iu = y\,, \tag{3.5}$$

where I is a linear bounded operator from a Hilbert space X into a Hilbert space Y, and they use nonstationary Tikhonov regularization

$$\min(\|Iu - y\|^2_Y + h_n\|u - u_{n-1}\|^2_X)$$

for the stable solution of the operator equation (3.5), where $\|.\|_X$ and $\|.\|_Y$ denote the induced norms in the Hilbert spaces X and Y, respectively.

3.3.1 Error Propagation of Tikhonov Regularization with Linear Unbounded Operators

In this subsection we consider the problem of computing values of an unbounded operator L. We will always denote by $L : \mathcal{D}(L) \subseteq H_1 \to H_2$ a closed, densely

defined unbounded linear operator between two Hilbert spaces H_1 and H_2. A typical example is $Lu = \nabla u$. The problem of computing values $y = Lf_0$, for $f_0 \in D(L)$ is then ill–posed in the sense that small perturbations in f_0 may lead to data f_δ satisfying

$$\|f_0 - f_\delta\|_{\mathrm{L}^2(\Omega)} \leq \delta, \tag{3.6}$$

but $f_\delta \notin \mathcal{D}(L)$, or even if $f_\delta \in \mathcal{D}(L)$, it may happen that $Lf_\delta \not\to Lf_0$ as $\delta \to 0$, since the operator L is unbounded. Morozov has studied a stable method for approximating the value Lf_0 when only approximate data f_δ is available [255]. This method takes as an approximation to $y = Lf_0$ the vector $y_h^\delta = Lu_h^\delta$, where u_h^δ minimizes the functional

$$T_{\mathrm{TIK}}(u) := \|u - f_\delta\|^2_{\mathrm{L}^2(\Omega)} + h\|Lu\|^2_{\mathrm{L}^2(\Omega)} \quad (h > 0) \tag{3.7}$$

over $D(L)$.

The functional is strictly convex and therefore, if $D(L)$ is nonempty and convex, there exists a unique minimizer of the functional $T_{\mathrm{TIK}}(u)$. Thus the method is well–defined. For more background on the stable evaluation of unbounded operators we refer to [157].

Let $u_0 := f_\delta$. If $L = \nabla$, then the sequence $\{u_n\}_{n\geq 1}$ of minimizers of the family of optimization problems

$$T^n_{\mathrm{TIK}} := \|u - u_{n-1}\|^2_{\mathrm{L}^2(\Omega)} + h\|\nabla u\|^2_{\mathrm{L}^2(\Omega)}, \quad n \geq 1 \tag{3.8}$$

are identical to the semi-discrete approximations of the differential equation (3.1) at time nh ($n \geq 1$) where $g(x) = x$.

This shows:

> Methods for evaluating unbounded operators can be used for diffusion filtering and vice versa. However the motivations differ: For evaluating unbounded operators we solve the optimization and evaluate in a further step the unbounded operator. In diffusion filtering we "only" have to solve the optimization problem.

In the following we compare the error propagation in Tikhonov regularization with regularization parameter h and the error propagation in iterated Tikhonov regularization of order N with regularization parameter h/N. This corresponds to making an implicit, t-discrete ansatz for a diffusion process with one step h and an implicit, t-discrete ansatz with N steps of step h/N, respectively.

Tikhonov regularization with regularization parameter h reads as follows:

$$\hat{u} = (I + hL^*L)^{-1} f_\delta$$

where L^* is the adjoint operator to L (see e.g. [431] for more details). Tikhonov regularization of order N with regularization parameter h/N reads as follows:

$$u_N = \left(I + \frac{h}{N} L^*L\right)^{-N} f_\delta \; .$$

Let L^*L be an unbounded operator with spectral values

$$0 < \lambda_1 \leq ..... \leq \lambda_n ....$$

such that $\lambda_n \to \infty$ as $n \to \infty$. Then

$$\begin{aligned} u_N &= \left(I + \frac{h}{N} L^*L\right)^{-N} f_\delta \\ &= \left(I + \frac{h}{N} L^*L\right)^{-N} (f_\delta - f_0) + \left(I + \frac{h}{N} L^*L\right)^{-N} f_0 \; . \end{aligned}$$

$\left(I + \frac{h}{N} L^*L\right)^{-N} (f_\delta - f_0)$ denotes the propagated error of the initial data f_δ, which remains in u_N – this corresponds to the error propagation in diffusion filtering with an implicit, t-discrete method.

Let E_λ be the spectral family according to the operator L^*L. Then it follows that [431]

$$\left(I + \frac{h}{N} L^*L\right)^{-N} (f_\delta - f_0) \; = \; \int_0^\infty \left(1 + \frac{h}{N}\lambda\right)^{-N} dE_\lambda (f_\delta - f_0) \; .$$

Using

$$\left(1 + \frac{h}{N}\lambda\right)^N \; = \; 1 + \frac{h}{N}\lambda + \binom{N}{2} \left(\frac{h}{N}\right)^2 \lambda^2 + ... + \left(\frac{h}{N}\right)^N \lambda^N$$

we get that

$$\begin{aligned} &\left\| \left(I + \tfrac{h}{N} L^*L\right)^{-N} (f_\delta - f_0) \right\|^2_{\mathrm{L}^2(\Omega)} \; = \\ &\int_0^\infty \left(1 + \tfrac{h}{N}\lambda + \tbinom{N}{2} \left(\tfrac{h}{N}\right)^2 \lambda^2 + ... + + \left(\tfrac{h}{N}\right)^N \lambda^N\right)^{-2} d\|E_\lambda (f_\delta - f_0)\|^2_{\mathrm{L}^2(\Omega)} \; . \end{aligned} \tag{3.9}$$

In noniterated Tikhonov regularization the error propagation is

$$\left\| (I + hL^*L)^{-1} (f_\delta - f_0) \right\|^2_{\mathrm{L}^2(\Omega)} \; = \; \int_0^\infty (1 + h\lambda)^{-2} d\|E_\lambda (f_\delta - f_0)\|^2_{\mathrm{L}^2(\Omega)} \; . \tag{3.10}$$

For large values of λ (i.e., for highly oscillating noise) the term $(1 + h\lambda)^{-2}$ in (3.10) is significantly larger than the term $\left(1+\frac{h}{N}\lambda+\binom{N}{2}\left(\frac{h}{N}\right)^2\lambda^2+\ldots+\left(\frac{h}{N}\right)^N\lambda^N\right)^{-2}$ in (3.9). This shows that noise propagation is handled more efficiently by iterated Tikhonov regularization than by Tikhonov regularization.

Above we analysed the error of the (iterated) Tikhonov regularized solutions and not the error in evaluating L at the Tikhonov regularized solutions. We emphasize that the less noise is contained in a data set the better the operator L can be evaluated. Therefore we conclude that the operator L can be evaluated more accurately with the method of iterated Tikhonov regularization than with noniterated Tikhonov regularization. This will be confirmed by the experiments in Section 3.5.

3.3.2 Well-Posedness of Regularization with Nonlinear Unbounded Operators

In this subsection we discuss some theoretical results on regularization with nonlinear unbounded operators.

Well-Posedness and Convergence for Total Variation Regularization

Total variation regularization goes back to Rudin, Osher and Fatemi [326] and has been further analysed by many others, e.g. [2, 78, 77, 76, 109, 110, 193, 227, 378, 340, 399]. In the unconstrained formulation of this method the data f_0 is approximated by the minimizer of the functional over $\mathrm{TV}(\Omega)$, the space of all functions with finite total variation norm

$$T_{\mathrm{TV}}(u) := \|u - f_\delta\|^2_{\mathrm{L}^2(\Omega)} + h\,\mathrm{TV}(u)\,, \tag{3.11}$$

where $\mathrm{TV}(u) := \int_\Omega |\nabla u|\,dx$ and

$$\int_\Omega |\nabla u|\,dx \;:=\; \sup\left\{-\int_\Omega u\nabla.\rho\,dx : \rho \in C_0^1(\Omega,\mathbb{R}^d)\,,|\rho| \le 1\right\}\,.$$

This expression extends the usual definition of the total variation for smooth functions to functions with jumps [147].

TV-regularization can be also considered as a constrained optimization problem to find an approximation of f_δ in the class of functions of finite bounded variation, where the regularization parameter h serves as a trade-off parameter between approximation quality and desired smoothness. Using similar arguments it is easy to see that standard Tikhonov regularization is useful to calculate approximations

of f_δ in the class of differentiable functions. Clearly this method is not suitable to recover discontinuities in images.

It is easy to see that a smooth minimizer of the functional T_{TV} satisfies

$$u - f_\delta = \frac{h}{2}\nabla . \left(\frac{1}{|\nabla u|} \nabla u \right) . \tag{3.12}$$

Acar and Vogel [2] proved the following results concerning existence of a minimizer of (3.11) and concerning stability and convergence of the minimizers:

Theorem 2 (Existence of a minimizer)
Let $f_\delta \in L^2(\Omega)$, then for fixed $h > 0$ a minimizer $u_h \in TV(\Omega)$ of (3.11) exists and is unique.

Theorem 3 (Stability)
Let $f_\delta \in L^2(\Omega)$ and $f_0 \in TV(\Omega)$. Then for $\delta \to 0$,

$$u_h(f_\delta) \to u_h(f_0)$$

with respect to the L^p-norm ($1 \leq p < \frac{d}{d-1}$). Here $u_h(f_\delta)$ is the minimizer of (3.11) and $u_h(f_0)$ is the minimizer of (3.11) where f_δ is replaced by f_0.

Theorem 4 (Convergence)
Let $f_\delta \in L^2(\Omega)$ and $f_0 \in TV(\Omega)$ satisfy $\|f_\delta - f_0\|_{L^2(\Omega)} \leq \delta$.
Then for $h := h(\delta)$ satisfying $\frac{\delta^2}{h} \to 0$ as $\delta \to 0$,

$$u_h \to f_0$$

with respect to the L^p-norm ($1 \leq p < \frac{d}{d-1}$).

It is evident that analogous results to Theorem 2, Theorem 3 and 4 also hold for the minimizers of the *iterated total variation regularization* which consists of minimizing a sequence of functionals

$$T_{\mathrm{TV}}^N(u) := \|u - u_{n-1}\|^2_{\mathrm{L}^2(\Omega)} + h \int_\Omega |\nabla u| \, dx, \tag{3.13}$$

where u_{N-1} denotes the minimizer of the functional T_{TV}^{N-1} and $u_0 = f_\delta$.

This regularization technique corresponds to the implicit, t-discrete approximation of the diffusion process (3.1) with $\hat{g}(x) = \sqrt{x}$.

The Perona–Malik Filter.

In the Perona–Malik filter [303] we have $g(s) = \frac{1}{1+s}$ and $\hat{g}(s) = \ln(1+s)$. Iterated Perona–Malik regularization minimizes the family of functionals

$$T^n_{\mathrm{PM}}(u) \;:=\; \|u - u_{n-1}(x)\|^2_{\mathrm{L}^2(\Omega)} + h \int_\Omega \ln(1 + |\nabla u|^2)\, dx. \tag{3.14}$$

The functionals T^n_{PM} are not convex and therefore we cannot conclude that the minimizer of (3.14) (it it exists) satisfies the first order optimality condition

$$\frac{1}{h}(u - u_{n-1})(x) = \nabla. \left(\frac{\nabla u}{1 + |\nabla u|^2} \right), \tag{3.15}$$

with homogeneous Neumann boundary data.

In the following we comment on some aspects of the Perona–Malik regularization technique. For the definitions of the Sobolev spaces $W^{l,p}$ and the notion of weak lower semi-continuity we refer to [5].

1. *Neumann boundary conditions:* Let Ω be a domain with smooth boundary $\partial\Omega$. Using trace theorems (see e.g. [5]) it follows that the Neumann boundary data are well–defined in $L^2(\partial\Omega)$ for any function in $W^{\frac{3}{2},2}(\Omega)$. Suppose we could prove that there exists a minimizer of the functional T^n_{PM}, then this minimizer must satisfy

$$\int_\Omega \ln(1 + |\nabla u|^2)\, ds < \infty\,. \tag{3.16}$$

   Elementary calculations show that any function $u \in W^{1,p}(\Omega)$ $(p > 1)$ satisfies (3.16). Therefore we cannot deduce from (3.16) that the minimizer is in any Sobolev space $W^{1,p}(\Omega)(p > 1)$. Consequently, there exists no theoretical result that the Neumann boundary conditions are well-defined.

2. *Existence of a minimizer of the functional T^n_{PM}:* The function $\ln(1 + s^2)$ is not convex, and therefore the functional $T^n_{\mathrm{PM}}(u)$ is not weakly lower semi-continuous on $W^{1,p}(\Omega)$ for any $1 < p < \infty$ (see [96, p. 66] and also [95])). Therefore, there exists a sequence $u_k \in W^{1,p}(\Omega)$ with $u_k \rightharpoonup u$ in $W^{1,p}(\Omega)$, but

$$\liminf T^n_{\mathrm{PM}}(u_k) < T^n_{\mathrm{PM}}(u)\,.$$

   Consequently, we cannot expect that a minimizing sequence converges (in $W^{1,p}(\Omega)$) to a minimizer of the functional T^n_{PM}. Thus the solution of the Perona–Malik regularization technique is ill–posed on $W^{1,p}(\Omega)$!

The diffusion process associated with the Perona–Malik regularization technique is

$$\partial_t u = \nabla . \left(\frac{\nabla u}{1 + |\nabla u|^2} \right) . \tag{3.17}$$

The Perona–Malik diffusion filtering technique can be split up in a natural way into a forward and a backward diffusion process:

$$\begin{aligned} \partial_t u &= \nabla . \left(\frac{1}{1 + |\nabla u|^2} \nabla u \right) = \frac{(1 - |\nabla u|^2) \Delta u}{(1 + |\nabla u|^2)^2} \\ &=: (a(|\nabla u|) - b(|\nabla u|)) \, \Delta u \, . \end{aligned}$$

Here

$$\begin{aligned} a(|\nabla u|) &:= \frac{1}{(1 + |\nabla u|^2)^2} \, , \\ b(|\nabla u|) &:= \frac{|\nabla u|^2}{(1 + |\nabla u|^2)^2} \, . \end{aligned}$$

Both functions a and b are non-negative. In general the solution of a backward diffusion equation is severely ill–posed (see e.g. [114]). We argue below that this nonlinear backward diffusion is well–posed with respect to appropriate norms. In fact we argue that the backward diffusion equation

$$v_t = -b(|\nabla v|) \Delta v \tag{3.18}$$

satisfies

$$TV(v(.,t)) = TV(v(.,0)) =: TV(v_0(.)) \, . \tag{3.19}$$

The intuitive reason for the validity of this is the following: Let $v \in C^2(\Omega \times [0,T])$. Then

$$\partial_t |\nabla v| \approx \partial_t \sqrt{|\nabla v|^2 + \beta^2} = \frac{\nabla v \nabla \partial_t v}{\sqrt{|\nabla v|^2 + \beta^2}} \, . \tag{3.20}$$

Using (3.20), (3.18), and integration by parts it follows that

$$\begin{aligned} \int_\Omega \partial_t |\nabla v| \, dx &\approx \int_\Omega \frac{\nabla v}{\sqrt{|\nabla v|^2 + \beta^2}} \nabla \partial_t v \, dx \\ &= \int_\Omega \nabla . \left(\frac{\nabla v}{\sqrt{|\nabla v|^2 + \beta^2}} \right) \frac{|\nabla v|^2}{(1 + |\nabla v|^2)^2} \Delta v \, dx \\ &= \beta^2 \int_\Omega \frac{|\nabla v|^2}{|\nabla v|^2 + \beta^2} \frac{|\Delta v|^2}{(1 + |\nabla v|^2)^2} \, dx \, . \end{aligned}$$

If $v \in C^2(\Omega \times [0,T])$, then the right hand side tends to zero as $\beta \to 0$. These arguments indicate that

$$\partial_t \int_\Omega |\nabla v| \, dx = 0 \, .$$

Consequently the total variation of $v(.,t)$ does not change in the course of the evolutionary process (3.20). Indeed, (3.18) may be regarded as a total variation preserving shock filter in the sense of Osher and Rudin [290].

The diffusion process

$$\partial_t u = a(|\nabla u|)\Delta u$$

is a forward diffusion process which decreases the total variation during the evolution. In summary we have argued that the Perona–Malik diffusion equation decreases the total variation during the evolutionary process.

A Regularized Perona–Malik Filter

Although the ill-posedness of the Perona–Malik filter can be handled by applying regularizing finite difference discretizations [418], it would be desirable to have a regularization which does not depend on discretization effects. In this subsection we study a regularized Perona–Malik filter

$$T^n_{\text{R-PM}}(u) \;:=\; \|u - u_{n-1}(x)\|^2 + h \int_\Omega \ln(1 + |\nabla L_\gamma u|^2)\, dx, \tag{3.21}$$

where L_γ is linear and compact from $L^2(\Omega)$ into $C^1(\overline{\Omega})$. The applications which we have in mind include the case that L_γ is a convolution operator with a smooth kernel.

In the following we prove that the functional $T^n_{\text{R-PM}}$ attains a minimum:

Theorem 5 *The functional $T^n_{R\text{-}PM}$ is weakly lower semi continuous on $L^2(\Omega)$.*

Proof:
Let $\{u_s : s \in \mathbb{N}\}$ be a sequence in $L^2(\Omega)$ which satisfies

$$T^n_{\text{R-PM}}(u_s) \to \min\{T^n_{\text{R-PM}}(u) : u \in L^2(\Omega)\}\ .$$

Then $\{u_s\}$ has a weakly convergent subsequence (which is again denoted by $\{u_s\}$) with weak limit u. Since L_γ is compact from $L^2(\Omega)$ into $C^1(\overline{\Omega})$ the sequence $\ln(1 + |\nabla L_\gamma u_s|^2)$ converges uniformly to $\ln(1 + |\nabla L_\gamma u|^2)$. In particular, we have

$$\int_\Omega \ln\left(1 + |\nabla L_\gamma u_s|^2\right)\, dx \to \int_\Omega \ln\left(1 + |\nabla L_\gamma u|^2\right)\, dx\ .$$

Using the weak lower semi-continuity of the norm $\|.\|_{L^2(\Omega)}$ it follows that the functional $T^n_{\text{R-PM}}$ is weakly lower semi-continuous. □

The minimizer of the regularized Perona–Malik functional satisfies

$$u - u_{n-1}(x) \;=\; h\, L_\gamma^* \nabla . \left(\frac{\nabla L_\gamma u}{1 + |\nabla L_\gamma u|^2} \right) . \tag{3.22}$$

The corresponding nonlinear diffusion process associated with this regularization technique is

$$\partial_t u \;=\; L_\gamma^* \nabla . \left(\frac{\nabla L_\gamma u}{1 + |\nabla L_\gamma u|^2} \right) . \tag{3.23}$$

Regularized Perona–Malik filters have been considered in the literature before [34, 74, 280, 404, 412]. Catté et al. [74] for instance investigated the nonlinear diffusion process

$$\partial_t u = \nabla . \left(\frac{\nabla u}{1 + |\nabla L_\gamma u|^2} \right) . \tag{3.24}$$

This technique (as well as other previous regularizations) does not have a corresponding formulation as an optimization problem. The differences between (3.23) and (3.24) will be explained in Section 3.5.

3.4 Lyapunov Functionals for Regularization Methods

Lyapunov functionals play an important role in continuous diffusion filtering (see [411, 412]). In order to introduce Lyapunov functionals of regularization methods, we first give a survey on Lyapunov functionals in diffusion filtering. We consider the diffusion process (here and in the following Ω will be a domain with piecewise smooth boundary)

$$\begin{array}{lll} \partial_t u = \nabla . \left(g(L_\gamma(|\nabla u|^2)) \nabla u \right) & \text{on} & \Omega \times (0,T) \\ u(x,0) = f(x) & \text{on} & \Omega \\ \partial_n u = 0 & \text{on} & \Gamma \times (0,T) \end{array} \tag{3.25}$$

We assume that the following assumptions hold:

1. $f \in L^\infty(\Omega)$, with $a := \text{ess inf}_{x \in \Omega} f$ and $b := \text{ess sup}_{x \in \Omega} f$.

2. L_γ is a compact operator from $L^2(\Omega)$ into $C^p(\overline{\Omega})$ for any $p \in \mathbb{N}$.

3. $T > 0$.

4. For all $w \in L^\infty(\Omega, \mathbb{R}^d)$ with $|w(x)| \leq K$ on $\overline{\Omega}$, there exists a positive lower bound $\nu(K)$ for g.

The regularizing operator L_γ may be skipped in (3.25), if one assumes that $\hat{g}(|.|^2)$ is convex from $\mathbb{R}^d$ to $\mathbb{R}$. Moreover, it is also possible to generalize (3.25) to the anisotropic case where the diffusivity g is replaced by a diffusion tensor [412].

Under the preceding assumptions it can be shown that (3.25) is well-posed (see [74, 412]):

Theorem 6 *The equation (3.25) has a unique solution $u(x,t)$ which satisfies*

$$u \in C([0,T]; L^2(\Omega)) \cap L^2([0,T]; H^1(\Omega)) \tag{3.26}$$

$$\partial_t u \in L^2([0,T]; H^1(\Omega)). \tag{3.27}$$

Moreover,

$$u \in C^\infty(\overline{\Omega} \times [0,T]) \ .$$

The solution fulfills the extremum principle

$$a \le u(x,t) \le b \text{ on } \Omega \times (0,T]. \tag{3.28}$$

For fixed t the solution depends continuously on f with respect to $\|.\|_{L^2(\Omega)}$.

This diffusion process leads to the following class of Lyapunov functionals [412]:

Theorem 7 *Suppose that u is a solution of (3.25) and that assumptions 1 – 4 are satisfied. Then the following properties hold.*

(a) (Lyapunov functionals) For all $r \in C^2[a,b]$ with $r'' \ge 0$ on $[a,b]$, the function

$$V(t) := \phi(u(t)) := \int_\Omega r(u(x,t))\, dx \tag{3.29}$$

is a Lyapunov functional:

1. *$\phi(u(t)) \ge \phi(Mf)$ for all $t \ge 0$; here*

$$Mf := \frac{1}{|\Omega|} \int_\Omega f(x)\, dx.$$

2. *$V \in C[0,\infty) \cap C^1(0,\infty)$ and $V'(t) \le 0$ for all $t > 0$.*

Moreover, if $r'' > 0$ on $[a,b]$, then $V(t) = \phi(u(t))$ is a strict Lyapunov functional:

3. *$\phi(u(t)) = \phi(Mf)$ if and only if $u(t) = Mf$ on $\overline{\Omega}$ for $t > 0$ and $u(t) = Mf$ a.e. on $\overline{\Omega}$ for $t = 0$.*
4. *If $t > 0$, then $V'(t) = 0$ if and only if $u(t) = Mf$ on $\overline{\Omega}$.*

5. *$V(0) = V(T)$ for $T > 0$ if and only if $f = Mf$ a.e. on Ω and $u(t) = Mf$ a.e. on $\overline{\Omega} \times (0, T]$.*

(b) (Convergence)

1. *$\lim_{t\to\infty} \|u(t) - Mf\|_{L^p(\Omega)} = 0$ for $p \in [1, \infty)$.*
2. *If $\Omega \subseteq \mathbb{R}$, then the convergence*

$$\lim_{t\to\infty} u(x,t) = Mf$$

is uniform.

In the sequel we introduce Lyapunov functionals of regularization methods. In the beginning of this section we discuss existence and uniqueness of the minimizer of the following regularization functional in $H^1(\Omega)$:

$$I(u) := \|u - f_\delta\|^2_{L^2(\Omega)} + h \int_\Omega \hat{g}(|\nabla u|^2)\, dx\,. \tag{3.30}$$

Lemma 1 *Let $\Omega \subseteq \mathbb{R}^d$, $d \geq 1$. Moreover, let $\hat{g}$ satisfy:*

$$\begin{aligned} &\hat{g} \in C^0(K) \text{ for any compact } K \subseteq [0, \infty[, \\ &\text{and } \hat{g}(0) = \min\{\hat{g}(x) : x \in [0, \infty[\}\,. \end{aligned} \tag{3.31}$$

$$\hat{g}(|.|^2) \text{ is convex from } \mathbb{R}^d \text{ to } \mathbb{R}\,. \tag{3.32}$$

Moreover, we assume that there exists a constant $c > 0$ such that

$$\hat{g}(s) \geq cs + \hat{g}(0) \tag{3.33}$$

Then the minimizer of (3.30) exists and is unique in $H^1(\Omega)$.

Proof:
By virtue of (3.33) it follows that

$$\|u - f_\delta\|^2_{L^2(\Omega)} + h \int_\Omega \hat{g}(|\nabla u|^2)\, dx \;\geq\; \|u - f_\delta\|^2_{L^2(\Omega)} + h \int_\Omega c|\nabla u|^2\, dx\,. \tag{3.34}$$

Suppose now that u_n is a sequence such that $I(u_n)$ converges to the minimum of the functional $I(.)$ in $H^1(\Omega)$. From (3.34) it follows that u_n has a weakly convergent subsequence in $H^1(\Omega)$, which we also denote by u_n; the weak limit will be denoted by u_*. Since $\hat{g}(|.|^2)$ is convex, the functional $\int_\Omega \hat{g}(|\nabla u|^2)\, dx$ is weakly lower semi-continuous in $H^1(\Omega)$ (see [96, 95]), and thus

$$\int_\Omega \hat{g}(|\nabla u_*|^2)\, dx \;\leq\; \liminf_{n\in\mathbb{N}} \int_\Omega \hat{g}(|\nabla u_n|^2)\, dx\,.$$

Thanks to the the Sobolev embedding theorem (see [5]) it follows that the functional $\|u - f_\delta\|^2_{L^2(\Omega)}$ is weakly lower semi continuous on $H^1(\Omega)$. Consequently

$$I(u_*) \leq \liminf_{n \in \mathbb{N}} I(u_n)$$

and thus u_* is a minimizer of I in $H^1(\Omega)$. Suppose now that u_1 and u_2 are two minimizers of the functional I. Then, from the optimality condition it follows that

$$\begin{aligned} \langle u_1 - f_\delta, u_2 - u_1 \rangle + h \langle g(|\nabla u_1|^2) \nabla u_1, \nabla(u_2 - u_1) \rangle &= 0 \qquad (3.35) \\ \langle u_2 - f_\delta, u_2 - u_1 \rangle + h \langle g(|\nabla u_2|^2) \nabla u_2, \nabla(u_2 - u_1) \rangle &= 0. \qquad (3.36) \end{aligned}$$

Consequently,

$$0 = \|u_2 - u_1\|^2 + h \langle g(|\nabla u_2|^2) \nabla u_2 - g(|\nabla u_1|^2) \nabla u_1, \nabla(u_2 - u_1) \rangle , \qquad (3.37)$$

and, thus, the minimizer of I is unique. □

The minimizer of (3.30) will be denoted by u_h in the remaining of this paper.

In the following we establish the average grey level invariance of regularization methods.

Theorem 8 *Let (3.31), (3.32), (3.33) hold. Then for different values of h the minimizers of (3.30) are grey-level invariant, i.e., for $h > 0$*

$$\int_\Omega u_h \, dx = \int_\Omega f_\delta \, dx.$$

Proof:
Elementary calculations show that the minimizer of (3.30) satisfies, for all $v \in H^1(\Omega)$,

$$\langle u_h - f_\delta, v \rangle + h \left\langle g(|\nabla u_h|^2) \nabla u_h, \nabla v \right\rangle = 0 . \qquad (3.38)$$

Taking $v = 1$ the second term vanishes and the assertion follows. □

Next we establish some basic results on regularization techniques. As we will show the proofs of the following results can be carried out following the ideas of the corresponding results in the book of Morozov [255]. However Morozov's results can not be applied directly since they are only applicable in the case that $\hat{g}(|x|^2) = |x|^2$, which is not sufficient for the presentation of this paper. Later these results are used to establish a family of Lyapunov functionals for regularization methods.

Lemma 2 *Let (3.31), (3.32), (3.33) hold. Then, for any $h > 0$,*

$$\|u_{h+t} - u_h\|_{L^2(\Omega)} \to 0 \qquad \text{for } t \to 0$$

and for $h = 0$,

$$\|u_t - f_\delta\|_{L^2(\Omega)} \to 0 \qquad \text{for } t \to 0^+ \ .$$

Proof:
If $\hat{g}(|.|^2)$ is convex, then $g(|s|^2)s$ is monotone (see e.g. [96]), i.e., for all $s, t \in \mathbb{R}^d$

$$\langle g(|s|^2)s - g(|t|^2)t, s - t\rangle_{\mathbb{R}^d} \geq 0.$$

1. First we consider the case $h > 0$: from (3.38) it follows by using the notation

$$h_1 := h, \ \ h_2 := h + t, \ \ u_1 := u_h, \ \ u_2 := u_{h+t}$$

that

$$\begin{aligned} \langle u_1 - f_\delta, u_2 - u_1\rangle + \ h_1 \langle g(|\nabla u_1|^2)\nabla u_1, \nabla(u_2 - u_1)\rangle &= 0, \\ \langle u_2 - f_\delta, u_2 - u_1\rangle + \ h_2 \langle g(|\nabla u_2|^2)\nabla u_2, \nabla(u_2 - u_1)\rangle &= 0. \end{aligned}$$

Consequently,

$$\begin{aligned} &\|u_2 - u_1\|^2_{L^2(\Omega)} + h_1 \langle g(|\nabla u_2|^2)\nabla u_2 - g(|\nabla u_1|^2)\nabla u_1, \nabla(u_2 - u_1)\rangle \\ &= (h_1 - h_2) \cdot \langle g(|\nabla u_2|^2)\nabla u_2, \nabla(u_2 - u_1)\rangle. \end{aligned} \tag{3.39}$$

Thus, using the Cauchy–Schwarz inequality and the identity (3.39) it follows that

$$\|u_2 - u_1\|^2_{L^2(\Omega)} \ \leq \ |h_2 - h_1| \cdot \frac{\|u_2 - f_\delta\|_{L^2(\Omega)} \, \|u_2 - u_1\|_{L^2(\Omega)}}{h_2}$$

which shows the continuity of u_h.

2. If $h = 0$: There exists a sequence $f_n \in H^1(\Omega)$ with $f_n \to f_\delta$ in $L^2(\Omega)$. Consequently, for any $h > 0$, it follows from the definition of a minimum of the Tikhonov-like functional that

$$\|u_h - f_\delta\|^2_{L^2(\Omega)} \ \leq \ \|f_n - f_\delta\|^2_{L^2(\Omega)} + h \int_\Omega \hat{g}(|\nabla f_n|^2) \, dx \ .$$

By taking the limit $h \to 0$, it follows that for any $n \in \mathbb{N}$,

$$\lim_{h \to 0} \|u_h - f_\delta\|^2_{L^2(\Omega)} \ \leq \ \|f_n - f_\delta\|^2_{L^2(\Omega)} \, ,$$

which shows the assertion.

□

In the following we present some monotonicity results for the regularized solutions.

Lemma 2 implies that we can set $u_0 = f_\delta$ without causing any confusion.

Lemma 3 *Let (3.31), (3.32), (3.33) hold. Then* $\int_\Omega \hat{g}(|\nabla u_h|^2)\,dx$ *is monotonically decreasing in* h *and* $\|u_h - f_\delta\|^2$ *is monotonically increasing in* h.

Proof:
Using the definition of the regularized solution it follows that

$$\begin{aligned}
&\|u_h - f_\delta\|^2_{L^2(\Omega)} + h \int_\Omega \hat{g}(|\nabla u_h|^2)\,dx \\
&\leq \|u_{h+t} - f_\delta\|^2_{L^2(\Omega)} + (h+t) \int_\Omega \hat{g}(|\nabla u_{h+t}|^2)\,dx - t \int_\Omega \hat{g}(|\nabla u_{h+t}|^2)\,dx \\
&\leq \|u_h - f_\delta\|^2_{L^2(\Omega)} + h \int_\Omega \hat{g}(|\nabla u_h|^2)\,dx \\
&\quad + t \left(\int_\Omega \hat{g}(|\nabla u_h|^2)\,dx - \int_\Omega \hat{g}(|\nabla u_{h+t}|^2)\,dx \right)
\end{aligned}$$

and therefore, for $t > 0$,

$$\int_\Omega \hat{g}(|\nabla u_{h+t}|^2)\,dx - \int_\Omega \hat{g}(|\nabla u_h|^2)\,dx \;\leq\; 0.$$

This shows the monotonicity of the functional $\int_\Omega \hat{g}(|\nabla u_h|^2)\,dx$. Using very similar arguments it can be shown that $\|u_h - f_\delta\|^2_{L^2(\Omega)}$ is monotonically increasing in h. □

In the following we analyse the behaviour of the functionals $\int_\Omega \hat{g}(|\nabla u_h|^2)\,dx$ and $\|u_h - f_\delta\|^2_{L^2(\Omega)}$ for $h \to \infty$.

Lemma 4 *Let (3.31), (3.32), (3.33) hold. Then, for* $h \to \infty$ *the regularized solution converges (with respect to the* L^2*-norm) to the solution of the optimization problem*

$$\|u - f_\delta\|^2_{L^2(\Omega)} = \min$$

under the constraint

$$\int_\Omega \hat{g}(|\nabla u|^2)\,dx = \hat{g}(0)\; meas(\Omega)\;.$$

Proof:
The proof is similar to the proof in the book of Morozov [255, page 35] and thus omitted. □

In the following lemma we establish the boundedness of the regularized solution. For the proof of this result we utilize Stampacchia's Lemma (see [146]).

Lemma 5 *Let B be an open domain, u a function in $H^1(B)$ and a a real number. Then $u^a = \max\{a, u\} \in H^1(B)$ and*

$$\int_B |\nabla u^a|^2 \, dx \leq \int_B |\nabla u|^2 \, dx \; .$$

We are using this result to prove that each regularized solution lies between the minimal and maximal value of the data f.

Lemma 6 *Let (3.31), (3.32), (3.33) hold. Moreover, let*

$$\hat{g} \text{ be monotone in } [0, \infty[. \tag{3.40}$$

If $f \in L^\infty(\Omega)$, then for any $h > 0$ the regularized solution satisfies

$$\begin{aligned} &a := \operatorname{ess\,inf}\{f(x) : x \in \Omega\} \leq u_h \leq \\ &\operatorname{ess\,sup}\{f(x) : x \in \Omega\} =: b \; . \end{aligned} \tag{3.41}$$

Proof:
We verify that the maximum of u_h is less than b. The corresponding assertion for the minimum values can be proven analogously. Let $u_h^b = \min\{b, u_h\}$, then from Lemma 5 and the assumption (3.40) it follows that

$$\int_\Omega \hat{g}(|\nabla u_h|^2) \, dx \geq \int_\Omega \hat{g}(|\nabla u_h^b|^2) \, dx \; .$$

Since

$$\|u_h - f_\delta\|^2_{L^2(\Omega)} \geq \|u_h^b - f_\delta\|^2_{L^2(\Omega)}$$

it follows from the definition of a regularized solution that $u_h(x) \leq b$. □

Next we establish the announced family of Lyapunov functionals.

Theorem 9 *Let $\Omega \subseteq \mathbb{R}^d$, $d = 1, 2, 3$ and let a, b be as in (3.41). Moreover, let (3.31), (3.32), (3.33), and (3.40) be satisfied. Suppose that u_h is a solution of (3.30). Then the following properties hold.*

(a) (Lyapunov functionals for regularization methods) For all $r \in C^2[a,b]$ *with* $r'' \geq 0$*, the function*

$$V(h) := \phi(u_h) := \int_\Omega r(u_h(x))\,dx \tag{3.42}$$

is a Lyapunov functional for a regularization method: Let

$$Mf_\delta := \frac{1}{|\Omega|}\int_\Omega f_\delta\,dx\ .$$

Then

1. $\phi(u_h) \geq \phi(Mf_\delta)$ *for all* $h \geq 0$.
2. $V \in C[0,\infty)$, $DV(h) := \int_\Omega r'(u_h)(u_h - u_0)\,dx \leq 0$, $V(h) - V(0) \leq 0$ *for all* $h \geq 0$.

Moreover, if $r'' > 0$ *on* $[a,b]$*, then* $V(h) = \phi(u_h)$ *is a strict Lyapunov functional:*

3. $\phi(u_h) = \phi(Mf_\delta)$ *if and only if* $u_h = Mf_\delta$ *on* $\overline{\Omega}$ *for* $h > 0$ *and* $u_0 = Mf_\delta$ *a.e. on* $\overline{\Omega}$.
4. *if* $h > 0$*, then* $DV(h) = 0$ *if and only if* $u_h = Mf_\delta$ *on* $\overline{\Omega}$.
5. $V(H) = V(0)$ *for* $H > 0$ *if and only if* $f = Mf$ *a.e. on* Ω *and* $u_h = Mf$ *on* $\overline{\Omega} \times (0,H]$.

(b) (Convergence)

d=1: u_h *converges uniformly to* Mf_δ *for* $h \to \infty$

d=2: $\lim\limits_{h\to\infty} \|u_h - Mf_\delta\|_{L^p(\Omega)} = 0$ *for any* $1 \leq p < \infty$

d=3: $\lim\limits_{h\to\infty} \|u_h - Mf_\delta\|_{L^p(\Omega)} = 0$ *for any* $1 \leq p \leq 6$

Proof:

(a) 1. Since $r \in C^2[a,b]$ with $r'' \geq 0$ on $[a,b]$, we know that r is convex on $[a,b]$. Using the grey level invariance and Jensen's inequality it follows

$$\begin{aligned}
\phi(Mf_\delta) &= \int_\Omega r\left(\frac{1}{|\Omega|}\int_\Omega u_h(x)\,dx\right)dy \\
&\leq \int_\Omega \frac{1}{|\Omega|}\left(r\int_\Omega u_h(x)\,dx\right)dy \\
&= \int_\Omega r(u_h(x))\,dx \\
&= \phi(u_h)\ .
\end{aligned} \tag{3.43}$$

2. From Lemma 2 it follows that $V \in C[0,\infty[$. Setting $v = r'(u_h)$ it follows from (3.38) and (3.32) that

$$\langle u_h - u_0, r'(u_h)\rangle \;=\; -h\left\langle g(|\nabla u_h|^2)\nabla u_h, r''(u_h)\nabla u_h\right\rangle . \tag{3.44}$$

The right hand side is negative since r is convex.
We represent $V(h) - V(0)$ in the following way.

$$\begin{aligned}
V(h) - V(0) \;&=\; \phi(u_h) - \phi(u_0)\\
&= \int_\Omega r(u_h(x)) - r(u_0(x))\,dx\\
&= \int_\Omega \int_0^1 r'(u_0(x) + t(u_h(x) - u_0(x)))\,dt\;(u_h(x) - u_0(x))\,dx\\
&= \int_\Omega r'(u_h(x))(u_h(x) - u_0(x))\,dx\\
&\quad + \int_\Omega \int_0^1 (r'(u_0(x) + t(u_h(x) - u_0(x))) - \; r'(u_h(x)))\,dt\cdot\\
&\quad \cdot (u_h(x) - u_0(x))\,dx\\
&= \int_\Omega r'(u_h(x))(u_h(x) - u_0(x))\,dx\\
&\quad - \int_\Omega \int_0^1 \int_0^1 r''(u_h(x) - \tau(1-t)(u_h(x) - u_0(x))\,d\tau\cdot\\
&\quad \cdot (1-t)\,(u_h(x) - u_0(x))^2\,dt\,dx\;.
\end{aligned}$$

From (3.44) and the convexity of r it follows that the last two terms in the above chain of inequalities are negative. Thus the assertion is proved.

3. Let $\phi(u_h) = \phi(Mf_\delta)$. Let us now show that the estimate (3.43) implies that $u_h = \text{const}$ on $\overline{\Omega}$. Suppose that $u_h \neq c$ Since $u_h \in H^1(\Omega)$, there exists a partition $\Omega = \Omega_1 \cup \Omega_2$ with $|\Omega_1|, |\Omega_2| \in (0, |\Omega|)$ and

$$\alpha := \frac{1}{|\Omega_1|}\int_{\Omega_1} u_h\,dx \neq \frac{1}{|\Omega_2|}\int_{\Omega_2} u_h\,dx =: \beta\;.$$

This assertion follows from the Poincare inequality for functions in Sobolev spaces [117]. From the strict convexity of r it follows that

$$\begin{aligned}
r\left(\frac{1}{|\Omega|}\int_\Omega u_h\,dx\right) \;&=\; r\left(\frac{|\Omega_1|}{|\Omega|}\alpha + \frac{|\Omega_2|}{|\Omega|}\beta\right)\\
&<\; \frac{|\Omega_1|}{|\Omega|}r(\alpha) + \frac{|\Omega_2|}{|\Omega|}r(\beta)\\
&\leq\; \frac{1}{|\Omega|}\int_{\Omega_1} r(u_h)\;dx \;+\; \frac{1}{|\Omega|}\int_{\Omega_2} r(u_h)\;dx\\
&=\; \frac{1}{|\Omega|}\int_\Omega r(u_h)\;dx
\end{aligned}$$

If we utilize this result in (3.43) we observe that for $h > 0$ $\phi(u_h) = \phi(Mf_\delta)$ implies that $u_h = \text{const}$ on $\overline{\Omega}$. Thanks to the average grey value invariance we finally obtain $u_h = Mf_\delta$ on $\overline{\Omega}$.

We turn to the case $h = 0$. From (1.) and (2.) it follows that

$$\phi(Mf_\delta) \leq \phi(u_h) \leq \phi(u_0) \; .$$

If $\phi(u_0) = \phi(Mf_\delta)$, then for all $\theta > 0$

$$\phi(Mf_\delta) = \phi(u_\theta) \; .$$

Thus we have that for all $\theta > 0$ $u_\theta = Mf$. Using the continuity of u_θ with respect to $\theta \in [0, \infty[$ (cf. Lemma 2) the assertion follows.

4. The proof is analogous to the proof of the (iv)-assertion in Theorem 3 in [412].

5. Suppose that $V(H) = V(0)$, then from (2.) it follows that

$$V(h) = \text{const on } [0, H] \; .$$

Let $\epsilon > 0$. Then for any $h \in [\epsilon, H]$ it follows from (4.) that $u_h = Mf_\delta$. Using the continuity of u_h with respect to $h \in [0, \infty[$ (cf. Lemma 2) the assertion follows. The converse direction is obvious.

(b) From Lemma 4 and assumption 3.33 it follows that

$$\int_\Omega |\nabla u_h|^2 \, dx \to 0 \text{ and } \|u_h - Mf_\delta\|^2_{L^2(\Omega)} \to 0 \; .$$

This shows that

$$\|u_h - Mf_\delta\|_{H^1(\Omega)} \to 0 \; .$$

From the Sobolev embedding theorem it follows that, for $h \to \infty$,

$d = 1$: u_h converges uniformly to Mf_δ

$d = 2$: $\|u_h - Mf_\delta\|^2_{L^p(\Omega)} \to 0$ for any $1 \leq p < \infty$
(note that we assumed that Ω is bounded domain)

$d = 3$: $\|u_h - Mf_\delta\|^2_{L^p(\Omega)} \to 0$ for any $1 \leq p \leq 6$
(note that we assumed that Ω is bounded domain).

□

In Theorem 9 we obtained similar results as for Lyapunov functional of diffusion operators (see [412]). In (2.) of Theorem 9 the difference between Lyapunov functionals for diffusion processes and regularization methods becomes evident. For

Lyapunov functionals in diffusion processes we have $V'(t) \leq 0$, and in regularization processes we have $DV(h) \leq 0$. $DV(h)$ is obtained from $V'(t)$ by making a time discrete ansatz at time 0. We note that this is exactly the way we compared diffusion filtering and regularization techniques in the whole paper. It is therefore natural that the role of the time derivative in diffusion filtering is replaced by the time discrete approximation around 0.

Example. In this example we study different regularization techniques which have been used for denoising of images:

1. *Tikhonov regularization:* Here we have $\hat{g}(|u|^2) = |u|^2$. In this case the assumptions (3.31), (3.32), (3.33) and (3.40) are satisfied.

2. *Total variation regularization:* Here we have $\hat{g}(|u|^2) = \sqrt{|u|^2})$. In this case the assumption (3.33) is not satisfied.

   However, for the modified versions, proposed by Ito and Kunisch [193], where the functional is replaced by

   $$\hat{g}(|u|^2) = \sqrt{|u|^2} + \alpha|u|^2, \text{ with } \alpha > 0$$

   (3.31), (3.32), (3.33), and (3.40) are satisfied.

   For the functional [2, 80]

   $$\hat{g}(|u|^2) = \sqrt{|u|^2 + \beta^2}$$

   the assumption (3.33) is not satisfied. For the modified version

   $$\hat{g}(|u|^2) = \sqrt{|u|^2 + \beta^2} + \alpha|u|^2$$

   studied in [268], the assumptions (3.31), (3.32), (3.33), and (3.40) are satisfied.

   For the functional

   $$\hat{g}(|s|^2) = \begin{cases} \frac{1}{2\epsilon}|s|^2 & (|s| \leq \epsilon) \\ |s| - \frac{\epsilon}{2} & (\epsilon \leq |s| \leq \frac{1}{\epsilon}) \\ \frac{\epsilon}{2}|s|^2 + \frac{1}{2}\left(\frac{1}{\epsilon} - \epsilon\right) & (|s| > \frac{1}{\epsilon}) \end{cases}$$

   the assumptions (3.31), (3.32), (3.33), and (3.40) are satisfied. This method has been proposed by Geman and Yang [142] and was studied extensively by Chambolle and Lions [76] (see also [268]).

3. *Convex nonquadratic regularizations:* The functional used by Schnörr [345]

$$\hat{g}(|s|^2) = \begin{cases} \lambda_k^2 |s|^2 & (|s| \le c_\rho) \\ \lambda_l^2 |s|^2 + (\lambda_k^2 - \lambda_l^2) c_\rho (2|s| - c_\rho) & (|s| > c_\rho) \end{cases}$$

satisfies (3.31), (3.32), (3.33), and (3.40), whereas the Green functional [154]

$$\hat{g}(|s|^2) = \ln(\cosh(|s|)^2)$$

violates the assumption (3.33).

3.5 Experiments

In this section we illustrate some of the previous regularization strategies by applying them to noisy real-world images.

Regularization was implemented by using central finite differences. In the linear case this leads to a linear system of equations with a positive definite system matrix. It was solved iteratively by a Gauß–Seidel algorithm. It is not difficult to establish error bounds for its solution, since the residue can be calculated and the condition number of the matrix may be estimated using Gerschgorin's theorem. The Gauß-Seidel iterations were stopped when the relative error in the Euclidean norm was smaller than 0.0001.

Discretizing stabilized total variation regularization with

$$\hat{g}(x) = \sqrt{\beta^2 + x}$$

leads to a nonlinear system of equations. It was numerically solved for $\beta = 0.1$ by combining convergent fixed point iterations as outer iterations [110] with inner iterations using the Gauß-Seidel algorithm for solving the linear system of equations. The fixed point iteration turned out to converge quite rapidly, such that not more than 20 iterations were necessary.

Figure 3.1 shows three common test images and a noisy variant of each of them: an outdoor scene with a camera, a magnetic resonance (MR) image of a human head, and an indoor scene. Gaussian noise with zero mean has been added. Its variance was chosen to be a quarter, equal and four times the image variance, respectively.

The goal of our evaluation was to find out which regularization leads to restorations which are closest to the original images. We applied linear and total variation regularization to the three noisy test images, used 1, 4, and 16 regularization steps and varied the regularization parameter until the optimal restoration was found. The distance between some restored image u and the ground truth f_0 was computed using the signal-to-noise ratio

$$SNR(u/f_0) := 10 \log_{10} \left(\frac{\sigma^2(u)}{\sigma^2(f_0 - u)} \right) \tag{3.45}$$

where σ^2 denotes the variance. The results are shown in Table 3.1, as well as in Figs. 3.2 and 3.3. This gives rise to the following conclusions:

- In all cases, total variation regularization performed better than Tikhonov regularization. As expected, total variation regularization leads to visually sharper edges. The TV-restored images consist of piecewise almost constant patches.
- In the linear case, iterated Tikhonov regularization produced better restorations than noniterated. Visually, noniterated regularization resulted in images with more high-frequent fluctuations. This is in complete agreement with the theoretical considerations in our paper. Improvements caused by iterating the regularization were mainly seen between 1 and 4 iterations. Increasing the iteration number to 16 did hardly lead to further improvements.
- It appears that the theoretical and experimental results in the linear setting do not necessarily carry over to the nonlinear case with total variation regularization. For the slightly degraded camera image, iterated regularization performed worse than noniterated regularization. For the MR image, the differences are negligible, and the highly degraded office scene allows better restoration results with iterated regularization.

In those cases where noniterated and iterated regularization performs equally well, one should give the preference to the faster method. In our case iterated regularization was more efficient, since it led to matrices with smaller condition numbers and the Gauß–Seidel algorithm converged faster. Using for instance multigrid methods that solve the linear systems with a constant effort for all condition numbers would make noniterated total variation regularization favourable.

In a final experiment we juxtapose the regularizations (3.23) and (3.24) of the Perona–Malik filter. Both processes have been implemented using an explicit finite difference scheme. The results using the MR image from Figure 3.1(c) are shown in Figure 3.4, where different values for γ, the standard deviation of the Gaussian, have been used. For small values of γ, both filters produce rather similar results, while larger values lead to a completely different behaviour. For (3.23), the regularization smoothes the diffusive flux, so that it becomes close to 0 everywhere, and the image remains unaltered. The regularization in (3.24), however, creates a diffusivity which gets closer to 1 for all image locations, so that the filter creates blurry results resembling linear diffusion filtering.

3.6 Summary

The goal of this paper was to investigate connections between regularization theory and the framework of diffusion filtering. The regularization methods we considered

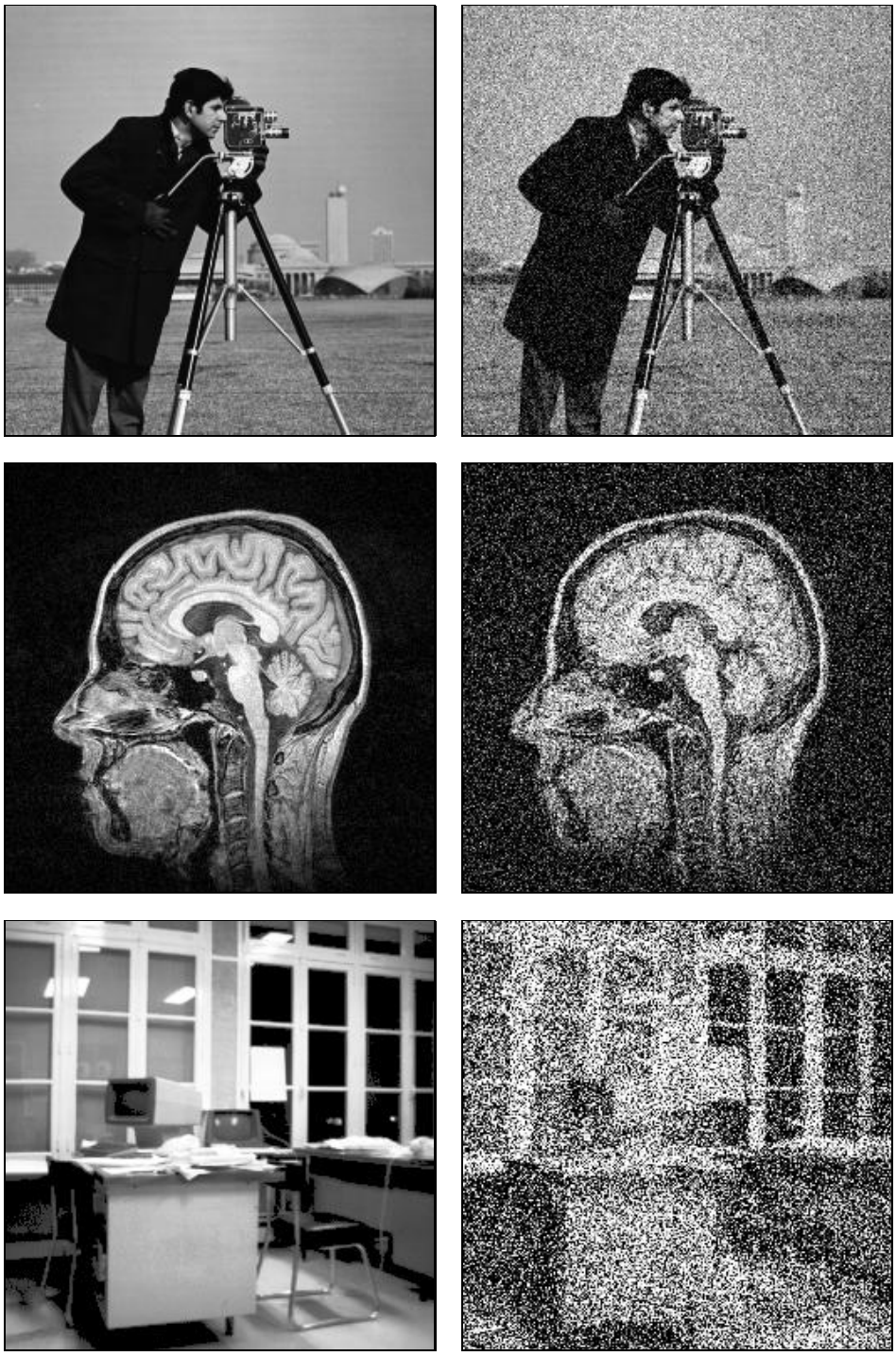


Figure 3.1: Test images, $\Omega = (0, 256)^2$. (a) TOP LEFT: Camera scene. (b) TOP RIGHT: Gaussian noise added. (c) MIDDLE LEFT: Magnetic resonance image. (d) MIDDLE RIGHT: Gaussian noise added. (e) BOTTOM LEFT: Office scene. (f) BOTTOM RIGHT: Gaussian noise added.

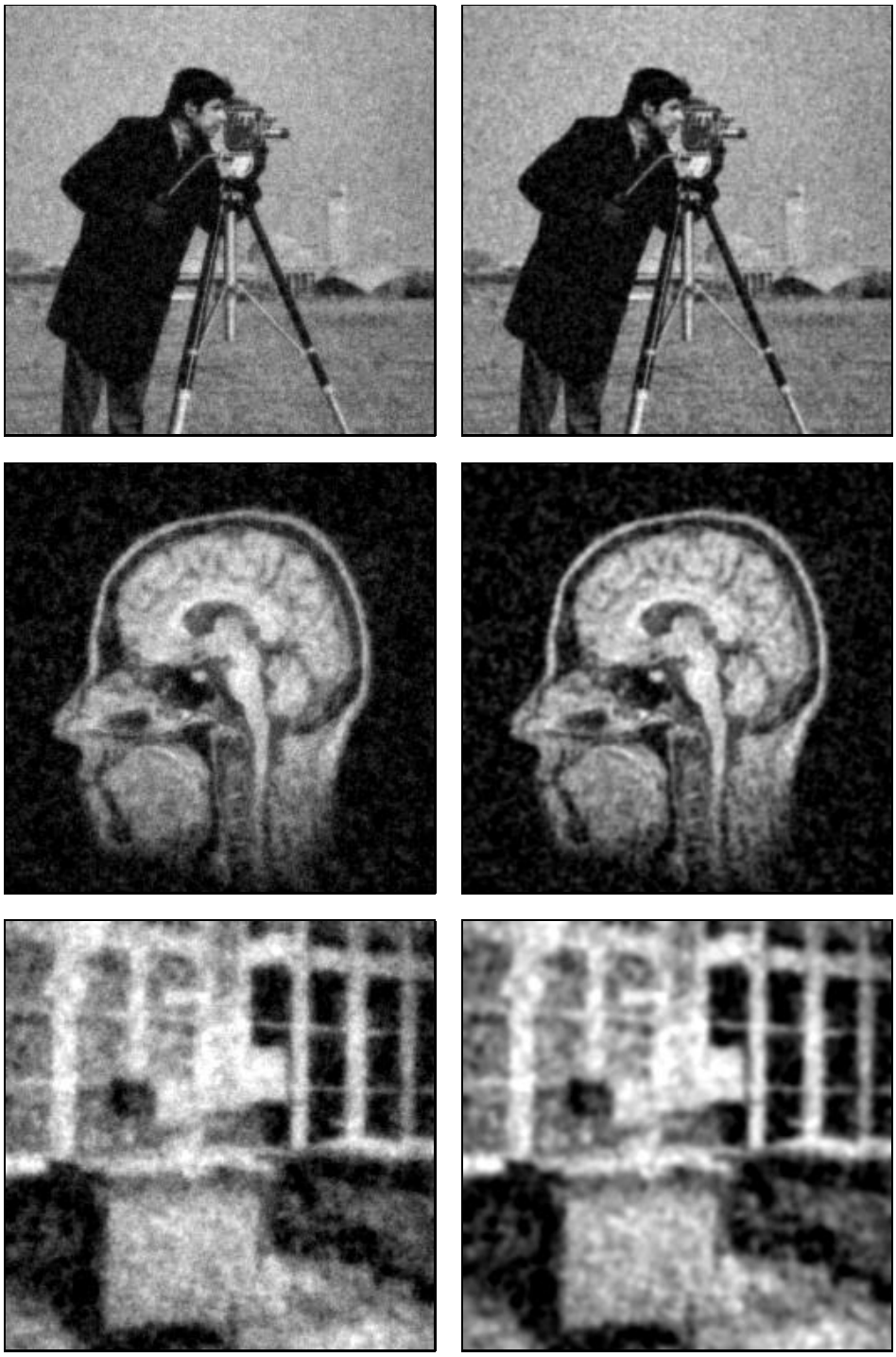


Figure 3.2: Optimal restoration results for Tikhonov regularization. (a) Top Left: Camera, 1 iteration. (b) Top Right: Camera, 16 iterations. (c) Middle Left: MR image, 1 iteration. (d) Middle Right: MR image, 16 iterations. (e) Bottom Left: Office, 1 iteration. (f) Bottom Right: Office, 16 iterations.

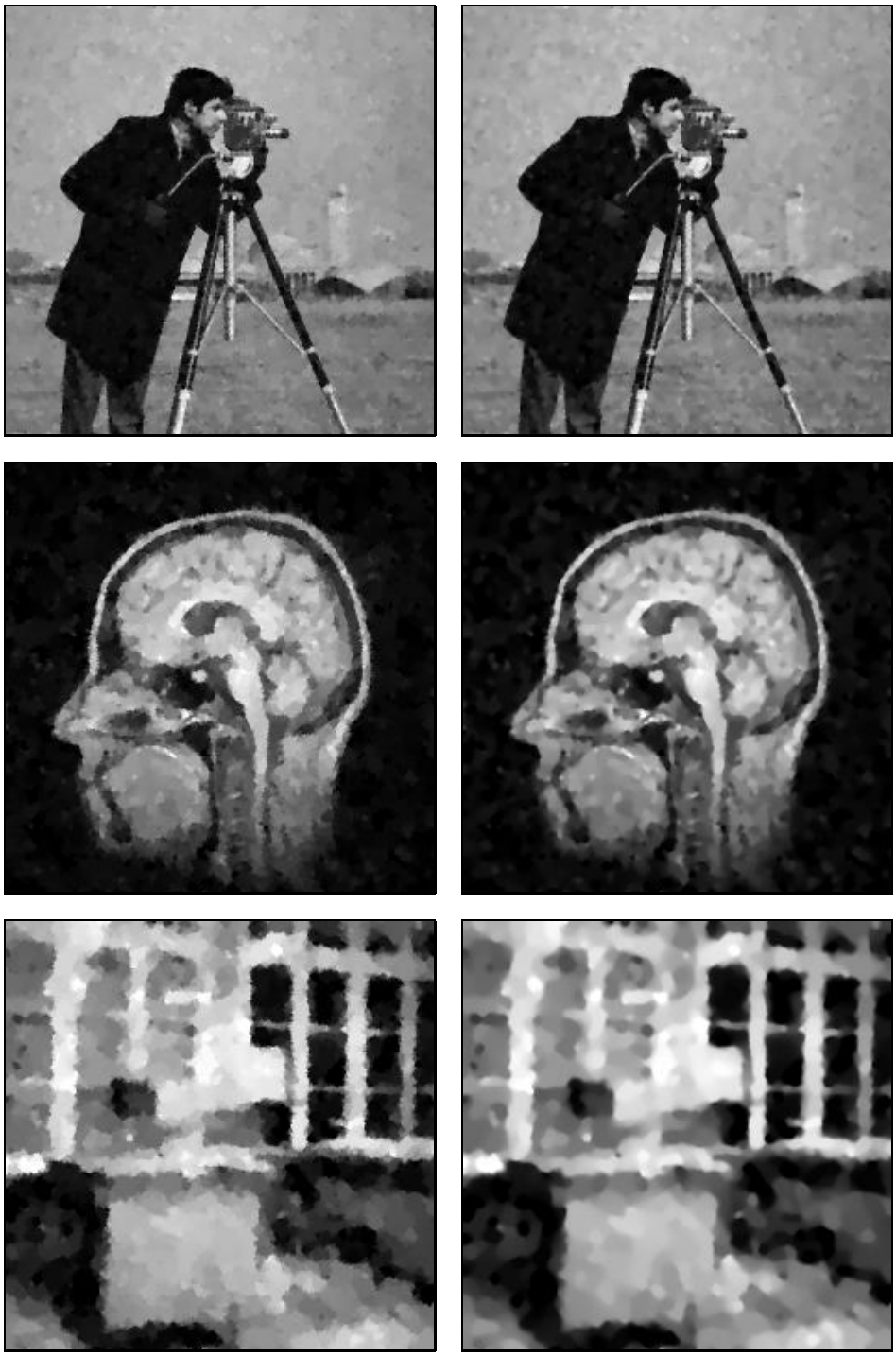


Figure 3.3: Optimal restoration results for total variation regularization. (a) Top Left: Camera, 1 iteration. (b) Top Right: Camera, 16 iterations. (c) Middle Left: MR image, 1 iteration. (d) Middle Right: MR image, 16 iterations. (e) Bottom Left: Office, 1 iteration. (f) Bottom Right: Office, 16 iterations.

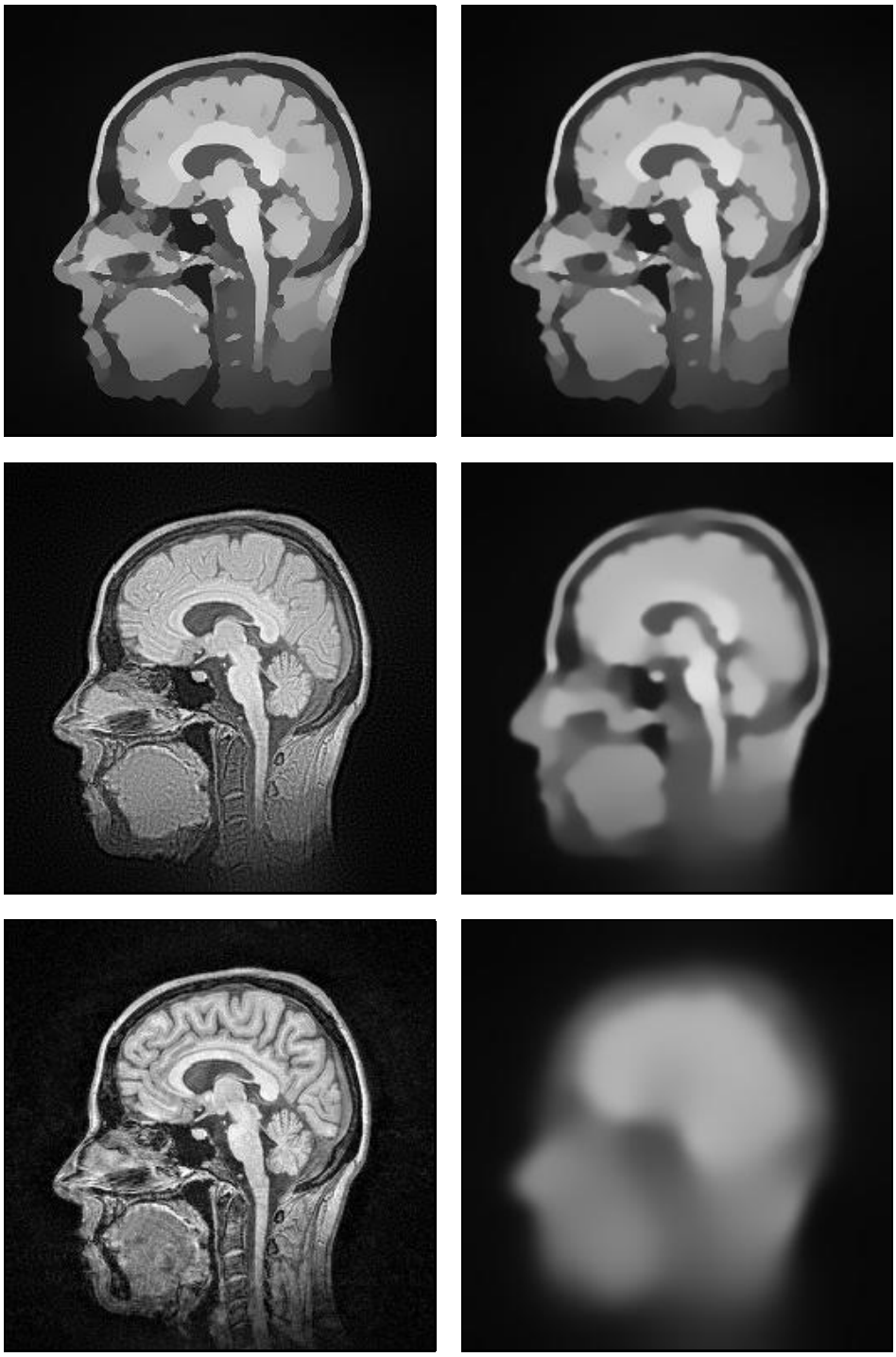


Figure 3.4: Comparison of two regularizations of the Perona–Malik filter ($t = 250$). (a) Top Left: Filter (3.23), $\gamma = 0.5$. (b) Top Right: Filter (3.24), $\gamma = 0.5$. (c) Middle Left: Filter (3.23), $\gamma = 2$. (d) Middle Right: Filter (3.24), $\gamma = 2$. (e) Bottom Left: Filter (3.23), $\gamma = 8$. (f) Bottom Right: Filter (3.24), $\gamma = 8$.

Table 3.1: Best restoration results for the different methods and images. The total regularization parameter for N iterations with parameter h is denoted $t = Nh$, and SNR describes the signal–to-noise ratio.

image	regularization	t	SNR
camera	linear, 1 iteration	0.82	11.80
camera	linear, 4 iterations	0.54	12.04
camera	linear, 16 iterations	0.48	12.09
MR	linear, 1 iteration	2.05	9.61
MR	linear, 4 iterations	1.16	9.91
MR	linear, 16 iterations	1.02	9.93
office	linear, 1 iteration	5.7	7.05
office	linear, 4 iterations	3.3	7.64
office	linear, 16 iterations	2.9	7.71
camera	TV, 1 iteration	13.2	14.11
camera	TV, 4 iterations	12.8	13.98
camera	TV, 16 iterations	12.4	13.93
MR	TV, 1 iteration	33.75	10.67
MR	TV, 4 iterations	33.5	10.64
MR	TV, 16 iterations	33	10.61
office	TV, 1 iteration	102	8.11
office	TV, 4 iterations	104	8.41
office	TV, 16 iterations	106	8.42

were Tikhonov regularization, total variation regularization, and we focused on linear diffusion filters as well as regularizations of the nonlinear diffusion filter of Perona and Malik. We have established the following results:

- We analysed the restoration properties of iterated and noniterated regularization both theoretically and experimentally. While linear regularization can be improved by iteration, there is no clear evidence that this is also the case in the nonlinear setting.

- We introduced an alternative regularization of the Perona–Malik filter. In contrast to previous regularization, it allows a formulation as a minimizer of a suitable energy functional.

- We have established Lyapunov functionals and convergence results for regularization methods using a similar theory as for nonlinear diffusion filtering.

These results can be regarded as contributions towards a deeper understanding as well as a better justification of both paradigms. It appears interesting to investigate

the following topics in the future:

- Regularization scale-spaces. So far, scale-space theory was mainly expressed in terms of parabolic and hyperbolic partial differential equations. Since scale-space methods have contributed to various interesting computer vision applications, it seems promising to investigate similar applications for regularization methods.

- Fully implicit methods for nonlinear diffusion filters using a single time step. This is equivalent to regularization and may be highly useful, if fast numerical techniques for solving the arising nonlinear systems of equations are applied.

Acknowledgements. The work of Otmar Scherzer is supported by the Fonds zur Förderung der Wissenschaftlichen Forschung (Austria), SFB F1310, as well as the Fonds zur Förderung der Gewerblichen Forschung (FFF), Project 200354. The present work has been carried out while Joachim Weickert was with the Department of Computer Science at the University of Copenhagen. His research was supported by the EU-TMR project VIRGO.

Chapter 4

Scale-Space Properties of Nonstationary Iterative Regularization Methods

Esther Radmoser, Otmar Scherzer, and Joachim Weickert.
Journal of Visual Communication and Image Representation,
Vol. 11, No. 2, 96–114, June 2000.

Abstract

Most scale-space concepts have been expressed as parabolic or hyperbolic partial differential equations (PDEs). In this paper we extend our work on scale-space properties of elliptic PDEs arising from regularization methods: we study linear and nonlinear regularization methods that are applied iteratively and with different regularization parameters. For these so-called nonstationary iterative regularization techniques we clarify their relations to both isotropic diffusion filters with a scalar-valued diffusivity and anisotropic diffusion filters with a diffusion tensor. We establish scale-space properties for iterative regularization methods that are in complete accordance with those for diffusion filtering. In particular, we show that nonstationary iterative regularization satisfies a causality property in terms of a maximum–minimum principle, possesses a large class of Lyapunov functionals, and converges to a constant image as the regularization parameters tend to infinity. We also establish continuous dependence of the result with respect to the sequence of regularization parameters. Numerical experiments in two and three space dimensions are presented that illustrate the scale-space behaviour of regularization methods.

4.1 Introduction

Decades after Iijima's pioneering axiomatic work in the sixties [175, 421], scale-spaces have become widely-used tools in image processing and computer vision [381, 274]. Alvarez *et al.* [12] have shown that imposing a reasonable set of architectural, invariance and simplification properties automatically leads to scale-spaces that can be described in terms of partial differential equations.

Partial differential equations may be classified into three main types: parabolic equations behaving in a diffusion-like manner, hyperbolic processes with wave-like character, and elliptic PDEs that can be related to variational problems. For more details on PDEs we refer to Colton [87] and Evans [116].

Examples for PDE-based scale-spaces incude parabolic PDEs such as linear and nonlinear diffusion scale-spaces [175, 303, 412], but also curvature scale-spaces like mean-curvature motion [13, 206] and affine morphological scale space [12, 333]. Hyperbolic PDEs with scale-space properties are given by the dilation and erosion equations arising from continuous-scale morphology [12, 23, 390, 57, 195].

Recently, Scherzer and Weickert [341] showed that a large class of regularization methods reveals the same scale-space properties as diffusion filtering, if one regards the regularization parameter of these elliptic PDEs as a scale parameter. This class includes the linear Tikhonov regularization as well as many nonlinear regularization methods that can be regarded as modified total variation (TV) denoising strategies.

The goal of the present paper is to extend this theory to regularization methods that are applied iteratively and with different regularization parameters. This framework is important since one can show that iterating regularization methods improves the restoration results in some cases [341], and varying the regularization parameter can be useful for accelerating the filtering procedure. We also extend our work by clarifying relations between diffusion filtering with nonmonotone fluxes or anisotropic diffusion filtering with a diffusion tensor on one hand, and iterated convex regularization methods on the other hand.

Our paper is organized as follows: In Section 4.2 and 4.3 we survey scale-space properties of diffusion filtering and noniterated regularization, respectively. Afterwards, this framework is extended to iterated nonstationary regularization in Section 4.4, where detailed proofs are presented. Section 4.5 gives an interpretation of diffusion filtering with nonmonotone fluxes or diffusion tensors in terms of iterated convex regularization methods. In Section 4.6 our theory is illustrated by experiments with 2D MR images and 3D ultrasound data.

Related work. Often there have been fruitful interactions between linear scale-space techniques and regularization methods. Torre and Poggio [388] emphasized that differentiation is ill-posed in the sense of Hadamard, and applying suitable regularization strategies approximates linear diffusion filtering or – equivalently

– Gaussian convolution. Much of the linear scale-space literature is based on the regularization properties of convolutions with Gaussians. In particular, differential geometric image analysis is performed by replacing derivatives by Gaussian-smoothed derivatives; see e.g. [127, 230, 369] and the references therein. In a very interesting work, Nielsen *et al.* [273] derived linear diffusion filtering axiomatically from Tikhonov regularization, where the stabilizer consists of a sum of squared derivatives up to infinite order.

Nonlinear diffusion filtering can be regarded both as a restoration method and a scale-space technique [303, 412]. When considering the restoration properties, natural relations between *biased* diffusion and regularization theory exist via the Euler equation for the regularization functional. This Euler equation can be regarded as the steady-state of a suitable nonlinear diffusion process with a bias term [80, 281, 345]. A popular specific energy functional arises from unconstrained total variation denoising [2, 76, 77]. Constrained total variation also leads to a nonlinear diffusion process with a bias term using a time-dependent Lagrange multiplier [326].

Strong and Chan [378] proposed to regard the regularization parameter of total variation denoising as a scale parameter. The present paper extends and completes our recent work on scale-space properties for noniterated regularization [341]. Following [253, 339, 378] we interpret the regularization parameter as a diffusion time by considering regularization as time-discrete diffusion filtering with a single implicit time step. Numerical implications of this relation are discussed in [419], and a shorter preliminary version of the present manuscript has been presented at the *Second International Conference on Scale-Space Theories in Computer Vision* [310].

4.2 Diffusion Filtering

In this section we review essential scale-space properties of nonlinear diffusion filtering. The presented results can also be extended to a broader class of methods including regularized filters with nonmonotone flux functions and anisotropic filters with a diffusion tensor. More details and proofs can be found in [412].

We consider a *diffusion process* of the form[1]

$$\begin{cases} \partial_t u(x,t) &= \nabla . \left(g(|\nabla u|^2)\nabla u\right)(x,t) & \text{on } \Omega \times [0,\infty) \\ \partial_n u(x,t) &= 0 & \text{on } \Gamma \times [0,\infty) \\ u(x,0) &= f(x) & \text{on } \Omega. \end{cases} \tag{4.1}$$

The image domain $\Omega \subseteq \mathbb{R}^d$ is assumed to be bounded with piecewise Lipschitzian boundary Γ with unit normal vector n, and $f \in L^\infty(\Omega)$ is a degraded

[1]We denote by (a,b) the open interval with startpoint a and endpoint b, $(a,b]$ denotes the interval which is open at a and closed at b, and $[a,b]$ denotes the closed interval.

original image with $a := \operatorname{ess\,inf}_\Omega f$ and $b := \operatorname{ess\,sup}_\Omega f$.

The diffusivity g satisfies the following properties:

1. Smoothness: $g \in C^\infty([0,\infty))$
2. The flux $g(s^2)s$ is monotonically increasing in s.
3. Positivity: $g(s) > 0$ for all $s \geq 0$.

Under these assumptions there exists a unique solution $u(x,t)$ of (4.1), such that $\|u(t)\|_{L^2(\Omega)}$ is continuous for $t \geq 0$. Here and in the following we use the abbreviation $u(t)$ for $u(.,t)$. It should be noted that this continuity property is necessary for relating structures over scales and for retrieving the original image for $t \to 0$. It is one of the fundamental architectural ingredients of scale-space theory. Furthermore, it is possible to show that $u(x,t) \in C^\infty(\bar{\Omega} \times (0,\infty))$.

Diffusion processes with reflecting boundary conditions preserve the average grey level:

$$\frac{1}{|\Omega|}\int_\Omega u(x,t)dx = Mf \quad \text{for all } t > 0\,,$$

with

$$Mf := \frac{1}{|\Omega|}\int_\Omega f(x)\,dx.$$

A constant average grey level is essential for scale-space segmentation algorithms such as the *hyperstack* [277]. It is also a desirable quality in medical imaging where grey values measure physical quantities of the depicted object, for instance proton densities in MR images.

The unique solution of (4.1) fulfills the extremum principle

$$a \leq u(x,t) \leq b \text{ on } \Omega \times (0,T]. \tag{4.2}$$

The extremum principle is an equivalent formulation of Koenderink's causality requirement [173]. Together with the continuity it ensures that level sets can be traced back in scale.

Another important simplification property can be expressed in terms of Lyapunov functionals. For all $r \in C^2[a,b]$ with $r'' \geq 0$ on $[a,b]$, the function

$$V(t) := \phi(u(t)) := \int_\Omega r(u(x,t))\,dx \tag{4.3}$$

is a Lyapunov functional:

1. It is bounded from below: $\phi(u(t)) \geq \phi(Mf)$ for all $t \geq 0$.
2. It is smoothly decreasing:

(a) $V \in C[0,\infty) \cap C^1(0,\infty)$

(b) $V'(t) \leq 0$ for all $t > 0$.

Lyapunov functionals show that diffusion filters create simplifying transformations: the special choices $r(s) := |s|^p$, $r(s) := (s - Mf)^{2n}$ and $r(s) = s\ln(s)$, respectively, imply that all L^p norms with $p \geq 2$ are decreasing, all even central moments are decreasing, and the entropy $S[u(t)] := -\int_\Omega u(x,t)\ln u(x,t)\,dx$, a measure of uncertainty and missing information, is increasing with respect to t. Lyapunov functionals have been used for scale-selection and texture analysis [373], for the synchronization of different diffusion scale-spaces [277], and for the automatic determination of stopping times [430]. Moreover, they allow to prove that the filtered image converges to a constant image as t tends to ∞: $\lim_{t\to\infty} \|u(t) - Mf\|_{L^p(\Omega)} = 0$ for $p \in [1,\infty)$. For $d = 1$ we have even uniform convergence.

4.3 Regularization

An interesting relation between nonlinear diffusion filtering and regularization methods becomes evident when considering an implicit time discretization [253, 339, 378]. The first step of an implicit scheme with step-size h in t–direction reads as follows.

$$\left\{ \begin{array}{rcl} \frac{u(x,h)-u(x,0)}{h} & = & \nabla\cdot\left(g(|\nabla u|^2)\nabla u\right)(x,h) \\ \partial_n u(x,h) & = & 0 \\ u(x,0) & = & f(x)\,. \end{array} \right. \tag{4.4}$$

In the following we assume the existence of a differentiable function $\hat{g}$ on $[0,\infty)$ which satisfies $\hat{g}' = g$. Then the minimizer of the functional

$$T(u) := \|u - f\|^2_{L^2(\Omega)} + h\int_\Omega \hat{g}(|\nabla u|^2)\,dx \tag{4.5}$$

satisfies (4.4). This can be seen by calculating the formal Gateaux derivative of T in direction v, i.e.

$$(T'(u), v) = \lim_{t\to 0^+} \frac{T(u+tv) - T(u)}{t} = \int_\Omega 2(u-f)v\,dx + h\int_\Omega 2g(|\nabla u|^2)\nabla u\nabla v\,dx.$$

We remark that for the numerical solution of parabolic differential equations several numerical schemes rely on implicit time discretizations, since they are unconditionally stable, i.e., for any choice of the time discretization the solution is stable with respect to data perturbations. In our context the unconditional stability of time implicit numerical schemes for solving the parabolic differential equation could as well be derived from regularization theory.

Since a minimizer of (4.5) satisfies $(T'(u), v) = 0$ for all v, we can conclude that the minimizer satisfies the differential equation (4.4). If the functional T is convex, then a minimizer of T is uniquely characterized by the solution of equation (4.4).

$T(u)$ is a typical regularization functional consisting of the approximation functional $\|u - f\|^2_{L^2(\Omega)}$ and the stabilizing functional $\int_\Omega \hat{g}(|\nabla u|^2)\, dx$. The weight h is called *regularization parameter*. An extensive discussion of regularization methods can be found in [114].

Now we sketch our scale-space theory for a broad class of regularization methods. For proofs and full details we refer to [341]. Let $\hat{g}$ satisfy the following properties.

I. $\hat{g}(.)$ is continuous for any compact $K \subseteq [0, \infty)$.

II. $\hat{g}(0) = \min\{\hat{g}(x) : x \in [0, \infty)\} \geq 0$.

III. $\hat{g}(|.|^2)$ is convex from $\mathbb{R}^d$ to $\mathbb{R}$.

IV. There exists a constant $c > 0$ such that $\hat{g}(s) - \hat{g}(0) \geq cs$.

V. $\hat{g}$ is monotone in $[0, \infty)$.

These assumptions guarantee existence and uniqueness of a minimizer u_h for the regularization functional (4.5) in the Sobolev space[2] $H^1(\Omega)$.

III. implies that $g(|.|^2)., $ where $g = \hat{g}'$ is monotone, i.e., for all $s_1, s_2 \in \mathbb{R}^d$

$$\left\langle g(|s_1|^2)s_1 - g(|s_2|^2)s_2, s_1 - s_2 \right\rangle \geq 0.$$

Assumptions I.-V. are satisfied for the following regularization techniques:

1. Tikhonov regularization:
$$\hat{g}(|s|^2) = |s|^2.$$

2. The modified total variation regularization of Ito and Kunisch [193]:
$$\hat{g}(|s|^2) = \sqrt{|s|^2} + \alpha|s|^2, \text{ with } \alpha > 0.$$

3. The modified total variation regularization of Nashed and Scherzer [268]:
$$\hat{g}(|s|^2) = \sqrt{|s|^2 + \beta^2} + \alpha|s|^2.$$

4. The regularization of Geman and Yang [142] and Chambolle and Lions [76]:
$$\hat{g}(|s|^2) = \begin{cases} \frac{1}{2\epsilon}|s|^2 & (|s| \leq \epsilon) \\ |s| - \frac{\epsilon}{2} & (\epsilon \leq |s| \leq \frac{1}{\epsilon}) \\ \frac{\epsilon}{2}|s|^2 + \frac{1}{2}\left(\frac{1}{\epsilon} - \epsilon\right) & (|s| > \frac{1}{\epsilon}). \end{cases}$$

[2] A function f belongs to the Sobolev space $H^m(\Omega)$ if f and all its derivatives up to order m belong to $L^2(\Omega)$. For more details on Sobolev spaces we refer to Adams [5].

5. Schnörr's [345] convex nonquadratic regularization:

$$\hat{g}(|s|^2) = \begin{cases} \lambda_h^2 |s|^2 & (|s| \le c_\rho) \\ \lambda_l^2 |s|^2 + (\lambda_h^2 - \lambda_l^2) c_\rho (2|s| - c_\rho) & (|s| > c_\rho). \end{cases}$$

Assumption IV. on $\hat{g}$ is violated for the total variation regularization in its original formulation by Rudin *et al.* [326]. Note that for TV-regularization $|\nabla u|$ only exists as a measure (see [117]). Therefore, we cannot set $\hat{g}(\sqrt{\cdot}) = \sqrt{\cdot}$ to obtain TV-regularization. Consequently, we cannot derive an equivalent optimality condition for a minimizer of (4.5). In this case our mathematical framework cannot guarantee existence of a minimizer of (4.5) in $H^1(\Omega)$, and in turn we have no existence theory for the partial differential equation (4.4). However, this does not mean that it is impossible to establish similar results by using other mathematical tools in the proofs; see e.g. the recent existence and uniqueness results by Andreu *et al.* [21].

The functional $\|u_h\|_{L^2(\Omega)}$ can also be shown to be continuous in $h \ge 0$. Regarding spatial smoothness, the solution belongs to $H^2(\Omega)$. This result is weaker than for the diffusion case where we have C^∞ results.

In analogy to diffusion filtering, the average grey level invariance

$$\int_\Omega u_h \, dx = \int_\Omega f \, dx \quad \text{for all } h \ge 0$$

and the extremum principle

$$a \le u_h \le b \quad \text{for all } h \ge 0$$

can be established.

Moreover, Lyapunov functionals for regularization methods can be constructed in a similar way. For all $r \in C^2[a, b]$ with $r'' \ge 0$, the function

$$V(h) := \phi(u_h) := \int_\Omega r(u_h(x)) \, dx \tag{4.6}$$

is a Lyapunov functional:

1. It is bounded from below: $\phi(u_h) \ge \phi(Mf)$ for all $h \ge 0$.

2. It is continuous and decreasing with respect to the original image:

   (a) $V \in C[0, \infty)$,

   (b) $DV(h) := \int_\Omega r'(u_h)(u_h - u_0) \le 0$, for all $h \ge 0$.

   (c) $V(h) - V(0) \le 0$ for all $h \ge 0$.

Here, a difference between Lyapunov functionals for diffusion processes and regularization methods becomes evident. For Lyapunov functionals in diffusion processes we have $V'(t) \leq 0$, and in regularization processes we have $DV(h) \leq 0$. $DV(h)$ is obtained from $V'(t)$ by making a time discrete ansatz at time 0. We note that this is exactly the way we compared diffusion filtering and regularization techniques. It is therefore natural that the role of the time derivative in diffusion filtering is replaced by the time discrete approximation around 0.

Again, these Lyapunov functionals allow to prove convergence of the filtered images to a constant image as $h \to \infty$. For $d = 3$, however, the convergence result is slightly weaker than in the diffusion case.

$$d = 1: \; u_h \text{ converges uniformly to } Mf \text{ for } h \to \infty$$

$$d = 2: \; \lim_{h \to \infty} \|u_h - Mf\|_{L^p(\Omega)} = 0 \text{ for any } 1 \leq p < \infty$$

$$d = 3: \; \lim_{h \to \infty} \|u_h - Mf\|_{L^p(\Omega)} = 0 \text{ for any } 1 \leq p \leq 6$$

4.4 Iterated Regularization

Regularization can be applied iteratively where the regularized solution of the previous step serves as initial image for the next iteration. For small regularization parameters, iterated regularization becomes therefore a good approximation to a nonlinear diffusion filter.

Let us consider an iterative regularization process with a sequence of positive regularization parameters $\mathcal{H} := (h_k)_{k \in \mathbb{N}}$. With $\mathcal{T} := (t_k)_{k \in \mathbb{N}}$ we denote the sequence of corresponding "diffusion times", i.e., $t_k := \sum_{i=1}^{k} h_i$. Note that $t_k - t_{k-1} = h_k$.

The n-th iteration of the nonstationary iterative regularization method reads as follows:

$$\begin{aligned} \frac{u^{\mathcal{H}}(x,t) - u^{\mathcal{H}}(x,t_{n-1})}{t - t_{n-1}} &= \nabla . \left(g(|\nabla u^{\mathcal{H}}|^2) \nabla u^{\mathcal{H}} \right)(x,t) && t \in (t_{n-1}, t_n], \; x \in \Omega \\ \partial_n u^{\mathcal{H}}(x,t) &= 0 && x \in \Gamma \\ u^{\mathcal{H}}(x,0) &= f(x) && x \in \Omega \end{aligned} \tag{4.7}$$

where now $t - t_{n-1}$ serves as the regularization parameter in the interval $(t_{n-1}, t_n]$. The superscript $\mathcal{H}$ at u refers to the fact that u is dependent on the discretization time in t direction. In the following we establish a scale-space theory for nonstationary iterated regularization. The terminology "iterative" refers to the fact that Tikhonov regularization is implemented iteratively. The terminology "nonstationary" refers to the fact that the parameters h_k may vary during the iterative process.

A minimizer $u \in H^1(\Omega)$ of the functional

$$T_t^{\mathcal{H}}(u) := \|u - u^{\mathcal{H}}(t_{n-1})\|^2_{L^2(\Omega)} + (t - t_{n-1}) \int_\Omega \hat{g}(|\nabla u|^2)\, dx, \tag{4.8}$$

satisfies (4.7) at time $t \in (t_{n-1}, t_n]$. If the functional $T_t^{\mathcal{H}}$ is strictly convex, then a minimizer of $T_t^{\mathcal{H}}$ is uniquely characterized by the solution of equation (4.7). Under these assumptions the minimizer of (4.8) exists and is unique in $H^1(\Omega)$ (cf. [341]).

Moreover, the spatial smoothness increases in each iteration step: a more detailed analysis using techniques from [437] shows that after n iterations the solution belongs to the Sobolev space $H^{2n}(\Omega)$ for fixed $t \in (t_{n-1}, t_n]$ (provided the diffusivity g is sufficiently smooth). This suggests that, if one uses the regularized solution for calculating derivatives of order $2n$, one should perform at least n iterations.

As for noniterated regularization, the average grey level invariance, temporal continuity in the L^2-norm and a maximum principle hold, if the function $\hat{g}$ satisfies I. – V. This can be seen, by noting that in the interval $(t_{n-1}, t_n]$ iterated regularization is noniterated regularization with initial data $u^{\mathcal{H}}(t_{n-1})$ and regularization parameter $t - t_{n-1}$. The results in [341] imply that in each interval $[t_{n-1}, t_n]$ the average grey level invariance holds, the function $u^{\mathcal{H}}(t)$ is bounded by the maximal and minimal values of $u^{\mathcal{H}}(t_{n-1})$, and the function is continuous with respect to t. The rest of the assertion follows by an inductive argument.

Using these properties we are able to establish a Ljapunov theory for iterated regularization.

Theorem 10 (Lyapunov functionals for nonstationary regularization) *Let $\Omega \subseteq \mathbb{R}^d$, $d = 1, 2, 3$ with smooth boundary. Let $f \in L^\infty(\Omega)$ with essential minimal value a and essential maximal value b. Moreover, let $\mathcal{H}$ be a sequence of positive numbers h_k, satisfying $\lim_{k\to\infty} h_k = \infty$. Then the following properties hold:*

(a) For all $r \in C^2[a, b]$ with $r'' \geq 0$, the function

$$V^{\mathcal{H}}(t) := \phi(u^{\mathcal{H}}(t)) := \int_\Omega r(u^{\mathcal{H}}(x, t))\, dx \tag{4.9}$$

is a Lyapunov functional for iterative regularization:

1. *It is bounded from below: $\phi(u^{\mathcal{H}}(t)) \geq \phi(Mf)$ for all $t \geq 0$,*
2. *It is continuous and decreasing with respect to the original image:*
   (a) $V^{\mathcal{H}} \in C[0, \infty)$,
   (b) $DV^{\mathcal{H}}(t) := \int_\Omega r'(u^{\mathcal{H}}(x, t)\,(u^{\mathcal{H}}(x, t) - u^{\mathcal{H}}(x, t_{n-1}))\, dx \leq 0$, for all $t \in (t_{n-1}, t_n]$,
   (c) $V^{\mathcal{H}}(t) - V^{\mathcal{H}}(t_{n-1}) \leq 0$ for all $t \in (t_{n-1}, t_n]$.

Moreover, if $r'' > 0$ on $[a,b]$, then $V^{\mathcal{H}}(t)$ is a strict Lyapunov functional:

3. *$\phi(u^{\mathcal{H}}(t)) = \phi(Mf)$ for all $t \in [0,\infty)$ if and only if $u^{\mathcal{H}}(t) = Mf$ on $\overline{\Omega}$ for $t > 0$, and $u^{\mathcal{H}}(.,0) = Mf$ almost everywhere on $\overline{\Omega}$.*
4. *If $t > 0$, then $DV^{\mathcal{H}}(t) = 0$ if and only if $u^{\mathcal{H}}(t) = Mf$ on $\overline{\Omega}$.*
5. *$V^{\mathcal{H}}(T) = V^{\mathcal{H}}(0)$ for $T > 0$ if and only if $f = Mf$ almost everywhere on Ω, and $u^{\mathcal{H}}(t) = Mf$ on $\overline{\Omega} \times (0,T]$.*

(b) (Convergence)

d=1: $u^{\mathcal{H}}(t_n)$ converges uniformly to Mf for $n \to \infty$

d=2: $\lim_{n\to\infty} \|u^{\mathcal{H}}(t_n) - Mf\|_{L^p(\Omega)} = 0$ for any $1 \le p < \infty$

d=3: $\lim_{n\to\infty} \|u^{\mathcal{H}}(t_n) - Mf\|_{L^p(\Omega)} = 0$ for any $1 \le p \le 6$

Proof:
Using the general result in [341] it follows that the assertions claimed in part (a) hold on the subintervals $(t_{n-1}, t_n]$. Moreover, from the results in [341] it follows that continuity of $u(.,t)$, with respect to t, also holds on $[t_{n-1}, t_n]$. By induction with respect to n the assertions of part (a) follow.

We turn to a verification of the assertions of part (b): since $u^{\mathcal{H}}(t_n)$ satisfies the first order optimality condition for a minimum of the functional $T^{\mathcal{H}}_{t_n}$ we get

$$\begin{aligned} &\left\langle u^{\mathcal{H}}(t_n), v \right\rangle_{L^2(\Omega)} + (t_n - t_{n-1}) \left\langle g(|\nabla u^{\mathcal{H}}(t_n)|^2) \nabla u^{\mathcal{H}}(t_n), \nabla v \right\rangle_{L^2(\Omega)} \\ &= \left\langle u^{\mathcal{H}}(t_{n-1}), v \right\rangle_{L^2(\Omega)} \end{aligned}$$

and taking $v = u^{\mathcal{H}}(t_n)$ shows that

$$\begin{aligned} &\|u^{\mathcal{H}}(t_n)\|^2_{L^2(\Omega)} + (t_n - t_{n-1}) \left\langle g(|\nabla u^{\mathcal{H}}(t_n)|^2) \nabla u^{\mathcal{H}}(t_n), u^{\mathcal{H}}(t_n) \right\rangle_{L^2(\Omega)} \\ &= \left\langle u^{\mathcal{H}}(t_{n-1}), u^{\mathcal{H}}(t_n) \right\rangle_{L^2(\Omega)}. \end{aligned}$$

Since $\left\langle g(|\nabla u^{\mathcal{H}}(t_n)|^2) \nabla u^{\mathcal{H}}(t_n), \nabla u^{\mathcal{H}}(t_n) \right\rangle_{L^2(\Omega)}$ is positive, which follows from the convexity of $\hat{g}$, we find that $u^{\mathcal{H}}(t_n)$ is uniformly bounded in $L^2(\Omega)$.

Since $u^{\mathcal{H}}(t_n)$ minimizes the functional $T^{\mathcal{H}}_{t_n}$ we immediately get that

$$\begin{aligned} &\|u^{\mathcal{H}}(t_n) - u^{\mathcal{H}}(t_{n-1})\|^2_{L^2(\Omega)} + (t_n - t_{n-1}) \int_\Omega \hat{g}(|\nabla u^{\mathcal{H}}(x,t_n)|^2)\, dx \\ &\le (t_n - t_{n-1}) \int_\Omega \hat{g}(|\nabla u^{\mathcal{H}}(x,t_{n-1})|^2)\, dx. \end{aligned} \tag{4.10}$$

Thus, the sequence $\int_\Omega \hat{g}(|\nabla u^{\mathcal{H}}(t_n)|^2)\, dx$ is monotonically decreasing in n.
Now we show that

$$\int_\Omega \hat{g}(|\nabla u^{\mathcal{H}}(x,t_n)|^2)\, dx \to \int_\Omega \hat{g}(0)\, dx \tag{4.11}$$

for $n \to \infty$. Since $u^{\mathcal{H}}(t_n)$ is the minimizing element of (4.8) (for $t = t_n$), we have

$$(t_n - t_{n-1}) \int_\Omega \hat{g}(|\nabla u^{\mathcal{H}}(x, t_n)|^2)\, dx \leq \|Mf - u^{\mathcal{H}}(t_{n-1})\|^2_{L^2(\Omega)} + (t_n - t_{n-1}) \int_\Omega \hat{g}(0)\, dx.$$

Dividing the inequality by $h_n = t_n - t_{n-1}$ and noting that $h_n \to \infty$ shows that

$$\lim_{n\to\infty} \int_\Omega \hat{g}(|\nabla u^{\mathcal{H}}(t_n)|^2)\, dx \;\leq\; \int_\Omega \hat{g}(0)\, dx.$$

Together with II. we get

$$\lim_{n\to\infty} \int_\Omega \hat{g}(|\nabla u^{\mathcal{H}}(t_n)|^2)\, dx \;=\; \int_\Omega \hat{g}(0)\, dx. \tag{4.12}$$

Since $u^{\mathcal{H}}(t_n)$ is uniformly bounded in $L^2(\Omega)$, it has a weakly convergent subsequence. Using IV. it follows that the weak limit is a constant function. Since iterative regularization is grey level invariant, we find that the limit is the constant function Mf. Moreover, from (4.12) it follows that the H^1-seminorm is strongly convergent to 0. Thus, the sequence $\{u^{\mathcal{H}}(t_n)\}_{n\in\mathbb{N}}$ itself is strongly convergent, i.e.

$$\|u^{\mathcal{H}}(t_n) - Mf\|_{H^1(\Omega)} \to 0.$$

By virtue of the Sobolev embedding theorem [5] it follows in particular that we obtain the following convergence results for $n \to \infty$.

d=1: $u^{\mathcal{H}}(t_n)$ converges uniformly to Mf

d=2: $\|u^{\mathcal{H}}(t_n) - Mf\|^2_{L^p(\Omega)} \to 0$ for any $1 \leq p < \infty$

d=3: $\|u^{\mathcal{H}}(t_n) - Mf\|^2_{L^p(\Omega)} \to 0$ for any $1 \leq p \leq 6$

This concludes the proof. □

The assumption that the sequence of regularization parameters h_k tends to infinity does not restrict practical applications. It actually suggest that in numerical simulations a monotonically increasing step size in time is very appropriate. Such an adaptation strategy would use small time steps in the beginning when much is happening, and afterwards, when the diffusion process slows down, the time step size becomes larger.

The previous result holds independently of the sequence $\mathcal{H} = \{h_k\}_{k\in\mathbb{N}}$. For numerical realizations of nonstationary iterated regularization it is important to verify continuous dependence of $u^{\mathcal{H}}(t)$ with respect $\mathcal{H}$. In order to prove this result we first show that the functional $u^{\mathcal{H}}(t)$ is Lipschitz continuous. Let $\mathcal{H}$ be a sequence of positive regularization parameters and let $\mathcal{T}$ be the according sequence of diffusion

times. Moreover, let τ_1 and τ_2 be two positive numbers in the interval $(t_{k-1}, t_k]$. Then

$$\langle u(\tau_1) - u(t_{k-1}), v\rangle_{L^2(\Omega)} + (\tau_1 - t_{k-1}) \left\langle g(|u(\tau_1)|^2)\nabla u(\tau_1), \nabla v\right\rangle_{L^2(\Omega)} = 0$$

and

$$\langle u(\tau_2) - u(t_{k-1}), v\rangle_{L^2(\Omega)} + (\tau_2 - t_{k-1}) \left\langle g(|u(\tau_2)|^2)\nabla u(\tau_2), \nabla v\right\rangle_{L^2(\Omega)} = 0.$$

Taking the difference of both equations gives

$$\begin{aligned} &\langle u(\tau_1) - u(\tau_2), v\rangle_{L^2(\Omega)} \\ &\quad + \quad (\tau_1 - t_{k-1}) \left\langle g(|u(\tau_1)|^2)\nabla u(\tau_1) - g(|u(\tau_2)|^2)\nabla u(\tau_2), \nabla v\right\rangle_{L^2(\Omega)} \\ &\quad - \quad (\tau_2 - \tau_1) \left\langle g(|\nabla u(\tau_2)|^2)\nabla u(\tau_2), \nabla v\right\rangle_{L^2(\Omega)} = 0. \end{aligned}$$

Taking $v = u(\tau_1) - u(\tau_2)$ and using the monotonicity of $g(|.|^2)$. gives:

$$\|u(\tau_1) - u(\tau_2)\|_{L^2(\Omega)} \leq \frac{|\tau_2 - \tau_1|}{\tau_2 - t_{k-1}} \|u(\tau_2) - u(t_{k-1})\|_{L^2(\Omega)}. \tag{4.13}$$

In particular for τ_1 and τ_2 both greater than $t_{k-1} + \epsilon$, with $\epsilon > 0$,

$$\|u(\tau_1) - u(\tau_2)\|_{L^2(\Omega)} \leq C_\epsilon |\tau_2 - \tau_1|\,,$$

where C_ϵ is independent of the particular choice of $\mathcal{H}$ as long as t_{k-1} is an element of $\mathcal{H}$.

With this Lipschitz continuity we are able to prove continuous dependence of $u^{\mathcal{H}}(t)$ on $\mathcal{H}$.

Lemma 7 *Let $\mathcal{H}^n$, $n \in \mathbb{N}$ and $\mathcal{H}$ be sequences of positive regularization parameters, where each sequence converges to infinity. Let $\mathcal{T}^n$, $n \in \mathbb{N}$ and $\mathcal{T}$ be the according sequences of diffusion times. Let*

$$t_k^n \to t_k \text{ for } n \to \infty, \text{ uniformly in } k.$$

Then

$$\|u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t)\|_{L^2(\Omega)} \to 0 \text{ for any } t \in [0, \infty).$$

Proof:
By means of the assumption on uniform convergence of t_k^n to t_k we get that, for any $t \in (t_{k-1}, t_k)$, there exists a sufficiently large index $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ also $t \in (t_{k-1}^n, t_k^n)$. Using that $u^{\mathcal{H}^n}(t)$ minimizes the functional $T_t^{\mathcal{H}^n}$ and $u^{\mathcal{H}}(t)$ minimizes the functional $T_t^{\mathcal{H}}$ we get

$$\begin{aligned} 0 &= \langle u^{\mathcal{H}^n}(t) - u^{\mathcal{H}^n}(t^n_{k-1}), v\rangle_{L^2(\Omega)} \\ &+ (t - t^n_{k-1})\langle g(|\nabla u^{\mathcal{H}^n}(t)|^2)\nabla u^{\mathcal{H}^n}(t), \nabla v\rangle_{L^2(\Omega)} \end{aligned}$$

and

$$\begin{aligned} 0 &= \langle u^{\mathcal{H}}(t) - u^{\mathcal{H}}(t_{k-1}), v\rangle_{L^2(\Omega)} \\ &+ (t - t_{k-1})\langle g(|\nabla u^{\mathcal{H}}(t)|^2)\nabla u^{\mathcal{H}}(t), \nabla v\rangle_{L^2(\Omega)}. \end{aligned} \tag{4.14}$$

Choosing $v := u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t)$ and subtracting these two equations gives

$$\begin{aligned} &\|u^{\mathcal{H}^n}(t) - u^{\mathcal{H}}(t)\|^2_{L^2(\Omega)} + \langle u^{\mathcal{H}^n}(t^n_{k-1}) - u^{\mathcal{H}}(t_{k-1}), u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t)\rangle_{L^2(\Omega)} \\ &+ (t - t^n_{k-1})\,\langle g(|\nabla u^{\mathcal{H}}(t)|^2)\nabla u^{\mathcal{H}}(t) - g(|\nabla u^{\mathcal{H}^n}(t)|^2)\nabla u^{\mathcal{H}^n}(t), \\ &\quad \nabla(u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t))\rangle_{L^2(\Omega)} \\ &- (t_{k-1} - t^n_{k-1})\,\langle g(|\nabla u^{\mathcal{H}}(t)|^2)\nabla u^{\mathcal{H}}(t), \nabla(u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t))\rangle_{L^2(\Omega)} = 0. \end{aligned}$$

Using that $g(|.|^2).$ is monotone, it follows from (4.14) that

$$\begin{aligned} \|u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t)\|_{L^2(\Omega)} &\leq \|u^{\mathcal{H}}(t_{k-1}) - u^{\mathcal{H}^n}(t^n_{k-1})\|_{L^2(\Omega)} \\ &+ \frac{|t_{k-1} - t^n_{k-1}|}{|t - t_{k-1}|}\|u^{\mathcal{H}}(t) - u^{\mathcal{H}}(t_{k-1})\|_{L^2(\Omega)}. \end{aligned} \tag{4.15}$$

Finally, we apply an inductive argument with respect to k. Let

$$\|u^{\mathcal{H}}(t_{k-1}) - u^{\mathcal{H}^n}(t^n_{k-1})\|_{L^2(\Omega)} \to 0 \text{ for } n \to \infty.$$

For $k = 0$ this is trivially satisfied, since $u^{\mathcal{H}}(.,0) = u^{\mathcal{H}^n}(.,0) = f$. Then it follows from (4.15) that, for any $t \in (t_{k-1}, t_k)$,

$$\|u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t)\|_{L^2(\Omega)} \to 0 \text{ for } n \to \infty.$$

Now, let $t = t_k$.

- If $t^n_k > t_k$, then we have $t = t_k \in (t^n_{k-1}, t^n_k)$ and analogously as above one can show that (4.15) holds with t replaced by t_k. Repeating the above arguments we find that
$$\|u^{\mathcal{H}}(t_k) - u^{\mathcal{H}^n}(t_k)\|_{L^2(\Omega)} \to 0 \text{ for } n \to \infty.$$

- If $t^n_k \leq t_k$, then let $t_0 \leq t^n_k$ fixed. Then it follows from triangle inequality that
$$\begin{aligned} \|u^{\mathcal{H}}(t_k) - u^{\mathcal{H}^n}(t^n_k)\|_{L^2(\Omega)} \;\leq\; &\|u^{\mathcal{H}}(t_0) - u^{\mathcal{H}^n}(t_0)\|_{L^2(\Omega)} + \\ &\|u^{\mathcal{H}}(t_k) - u^{\mathcal{H}}(t_0)\|_{L^2(\Omega)} + \|u^{\mathcal{H}^n}(t_0) - u^{\mathcal{H}^n}(t^n_k)\|_{L^2(\Omega)}. \end{aligned}$$

The second and third term on the right hand side of the last inequality are of the order $\max\{|t_0 - t_k|, |t_0 - t_k^n|\}$ (cf. (4.13)) (independent of n) - note that t_k^n converges uniformly to t_k. Since $t_k^n \to t_k$ we can choose t_0 in such a way that $\max\{|t_0 - t_k|, |t_0 - t_k^n|\}$ becomes arbitrarily small. Moreover, the first term tends to zero, since t_0 is an interior point of (t_{k-1}, t_k). These arguments show that

$$\|u^{\mathcal{H}}(t_k) - u^{\mathcal{H}^n}(t_k^n)\|_{L^2(\Omega)} \to 0.$$

Hence, the lemma is proved. □

We can use this lemma to show continuous dependence of $V^{\mathcal{H}}(t)$ on $\mathcal{H}$:

Theorem 11 *Let the assumptions of Lemma 7 hold. Then for, any $t \in [0, \infty)$,*

$$|V^{\mathcal{H}}(t) - V^{\mathcal{H}^n}(t)| \to 0.$$

Proof:
Using the Cauchy-Schwarz inequality it follows that

$$\begin{aligned} \left|V^{\mathcal{H}}(t) - V^{\mathcal{H}^n}(t)\right| \;\leq\; & \left(\int_\Omega \left(\frac{r(u^{\mathcal{H}}(x,t)) - r(u^{\mathcal{H}^n}(x,t))}{u^{\mathcal{H}}(x,t) - u^{\mathcal{H}^n}(x,t)}\right)^2 dx\right)^{\frac{1}{2}} \cdot \\ & \cdot \left\|u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t)\right\|_{L^2(\Omega)}. \end{aligned}$$

By virtue of Lemma 7 it follows that $\|u^{\mathcal{H}}(t) - u^{\mathcal{H}^n}(t)\|_{L^2(\Omega)}$ tends to zero. Since r is continuously differentiable, the first expression on the right hand side is bounded and the proof is accomplished. □

4.5 Extensions to the Nonconvex Case and to Anisotropic Filters

The previous sections analyse relations between regularization methods and diffusion filters for the case that $\hat{g}(|s|^2)$ is convex in s. This implies that the diffusive flux is monotonously increasing in s in the sense that

$$\langle g(|s|)s - g(|t|)t, s - t\rangle \;\geq\; 0 \quad \text{for all} \quad s, t \in \mathbb{R}^d.$$

In the context of diffusion filtering, however, nonmonotone fluxes leading to forward–backward diffusion processes are used frequently. While the earliest representative of this class, the Perona–Malik filter [303] is ill-posed, several well-posed forward–backward diffusion filters have been proposed afterwards; see e.g. [74, 412]. They offer the interesting property that they can enhance features like edges or flow-line

structures without renouncing smoothing properties in terms of Lyapunov functionals. It is also possible to replace the scalar-valued diffusivity g by a diffusion tensor D allowing true anisotropic behaviour.

These extensions are covered by the diffusion model

$$\begin{cases} \partial_t u(x,t) &= \nabla . \left(D(J_\rho(\nabla u_\sigma))\nabla u\right)(x,t) & \text{on } \Omega \times [0,\infty) \\ \langle D\nabla u, n\rangle &= 0 & \text{on } \Gamma \times [0,\infty) \\ u(x,0) &= f(x) & \text{on } \Omega \end{cases} \tag{4.16}$$

where $u_\sigma := K_\sigma * u$ denotes the convolution of u with a Gaussian K_σ of standard deviation σ, and J_ρ is the so-called structure tensor [132]

$$J_\rho(\nabla u_\sigma) = K_\rho * (\nabla u_\sigma \nabla u_\sigma^T),$$

a very useful matrix for the analysis of edges, corners and coherent structures. This model formulation comprises the regularized diffusion filter of Catté *et al.* [74] as well as edge-enhancing anisotropic diffusion filtering [408] and coherence-enhancing anisotropic diffusion filtering [415].

By assuming that the diffusion tensor D is a symmetric matrix-valued C^∞ function of J_ρ that remains uniformly positive definite, one can prove that all theoretical results from Section 4.2 carry over [412]. Well-posedness is achieved in the nonconvex case by the Gaussian smoothing in u_σ; see also [74].

Because of the Gaussian convolutions there is no straightforward way to derive a diffusion filter of this type as a minimizer of some energy functional. It is, however, instructive to study a semi-implicit time discretization of such a filter: it approximates the diffusion tensor D at the old time level and the remainder of the divergence expression at the new level. Such a discretization gives

$$\frac{u^{\mathcal{H}}(x,t) - u^{\mathcal{H}}(x,t_{n-1})}{t - t_{n-1}} = \nabla . \left(D(J_\rho(\nabla u^{\mathcal{H}}(x,t_{n-1})))\nabla u^{\mathcal{H}}(x,t)\right). \tag{4.17}$$

It can be regarded as an iterative regularization scheme where

$$\begin{aligned} T_t^{\mathcal{H}}(u) &:= \|u - u^{\mathcal{H}}(t_{n-1})\|^2_{L^2(\Omega)} \\ &+ (t - t_{n-1}) \int_\Omega (\nabla u)^T D(J_\rho(\nabla u_\sigma^{\mathcal{H}}(t_{n-1})))\nabla u \, dx \end{aligned} \tag{4.18}$$

is minimized. Now we are approximating a possibly nonconvex smoothing problem by a sequence of quadratic (and hence convex) regularization functionals.[3] As a consequence, the theoretical results for iterated regularization that we derived in Section 4.4 may also be extended to this case.

[3]This convexification by freezing the nonlinear part also relates our method to the adaptive linearization technique of Geman and Reynolds [141, 81] and the so-called Kačanov method from elasticity theory [137, 165, 419].

4.6 Experiments

The numerical experiments are performed using the software package DIFFPACK from the University of Oslo / Numerical Objects [97]. We have implemented the diffusion equation with $\hat{g}(|\nabla u|^2) = \sqrt{|\nabla u|^2 + \beta^2} + \alpha|\nabla u|^2$ which is a modified total variation regularization. For this diffusion filtering our theoretical results are applicable. The term $\alpha|\nabla u|^2$ is only of theoretical interest; in numerical realizations, the discretized version of the gradient is bounded, and there is no visible difference between using very small values of α (in which the theoretical results are applicable) and $\alpha = 0$ (where our theoretical results do not hold).

Our experiments were carried out for different sequences of time-steps and various smoothing parameters β. The influence of the parameter settings is as follows.

The impact of β on the numerical reconstruction is hardly viewable in the range from $\beta = 10^{-2}$ to 10^{-4}. Even the convergence rate is, although slower for smaller β, hardly affected.

For small values of regularization parameters h (up to approximately 5.0), there is no visible difference between iterated and noniterated regularization. The effect can only be seen for larger values of h. This is illustrated in Figure 4.2. It shows the result of noniterated and iterated regularization applied to the 2D MR image from Figure 4.1(a). The results are depicted at times $t = 10$, 30, and 100, respectively. For noniterated regularization this is achieved in one step, and for iterated regularization the regularization parameter $h = 1$ was chosen and 10, 30, or 100 iterations were performed. We observe that differences between the two methods are very small. They only become evident when subtracting one image from the other. This also indicates that even the semi-group property of regularization methods is well approximated in practice. It should be noted that the semi-group property is an ideal continuous concept which can only be approximated in time-discrete algorithms for partial differential equations.

As can be seen from the previous sections, the scale-space framework for noniterated and iterated regularization methods carries over to higher space dimensions. In the next figure we present results from a three-dimensional ultrasound data set of a fetus with $80 \times 80 \times 80$ voxels. Also in this case the differences between noniterated and iterated regularization are very small and iterated regularization appears to give slightly smoother results. This is in complete accordance with the theory in Section 4.4.

4.7 Conclusions

The novelty of our paper consists of establishing sequences of parameter dependent elliptic boundary value problems, namely nonstationary iterated regularization

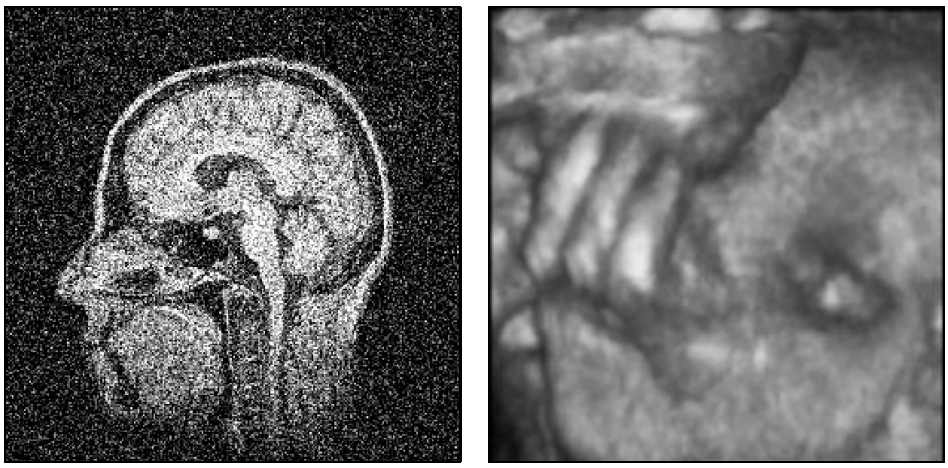
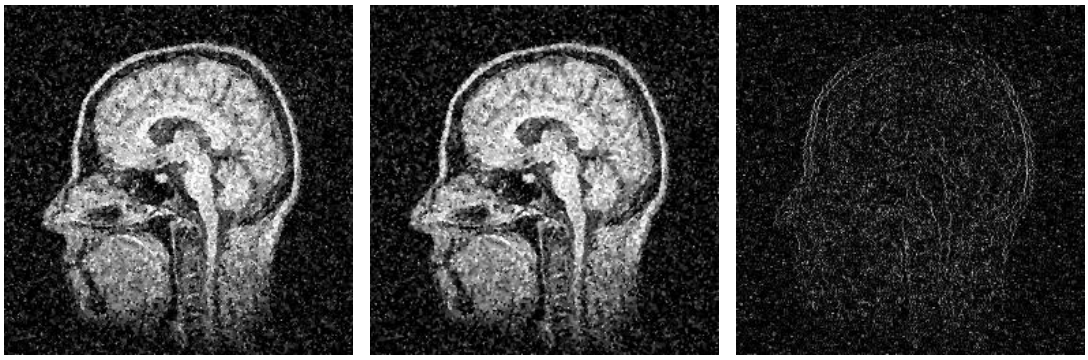


Figure 4.1: Test images. **(a) Left:** MR image with additive Gaussian noise ($SNR = 1$). **(b) Right:** Rendering of a three-dimensional ultrasound data set of a human fetus.

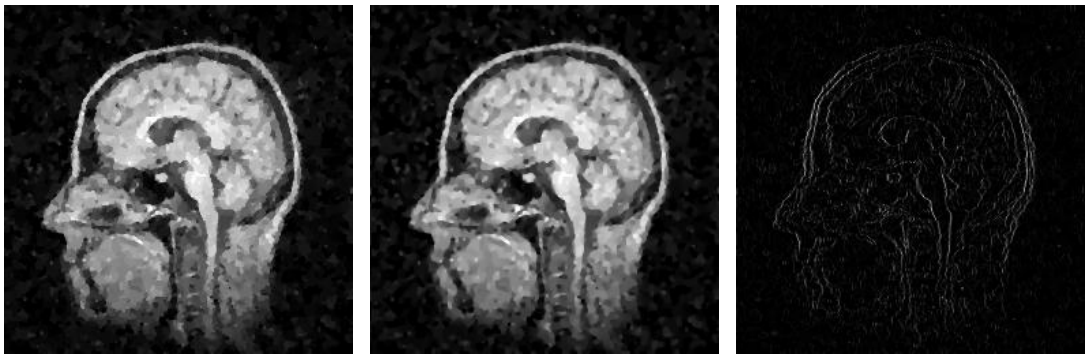
methods, as scale-space techniques. They satisfy the same scale-space properties as nonlinear diffusion filtering. The key ingredient for understanding this relation is the interpretation of iterated regularization methods as time-implicit or time-semi-implicit approximations to diffusion processes. In this sense, the scale-space theory of regularization methods is also a novel semi-discrete theory to diffusion filtering. This time-discrete framework completes the theory of diffusion scale-spaces where up to now only results for the continuous, the space-discrete and the fully discrete setting have been formulated [412].

The synthesis of regularization techniques and diffusion methods may lead to a deeper understanding of both fields, and it is likely that many more results can be transferred from one of these areas to the other. It would e.g. be interesting to study how results for optimal parameter selection in regularization methods can be used for diffusion filtering, or to further investigate the use of the iterated anisotropic functional (4.18) in the context of regularization theory. It is also promising to analyse and juxtapose efficient numerical techniques developed in both frameworks. First steps in this direction are reported in [419].

Acknowledgements. The work of E.R. and O.S. is supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung, SFB 1310, and J.W. received financial support from the EU-TMR project VIRGO. The authors thank Firma Kretztechnik AG, Austria, for the permission to publish the 3-D ultrasound data.



one iteration, t=10 | 10 iterations, t=10 | difference



one iteration, t=30 | 30 iterations, t=30 | difference

one iteration, t=100 | 100 iterations, t=100 | difference

Figure 4.2: Results for the MR image from Figure 4.1(a) with noniterated and iterated regularization ($\beta = 0.001$). The left column shows the results for noniterated, the middle column for iterated regularization. The images in the right column depict the modulus of the differences between the results for the iterated and noniterated method.

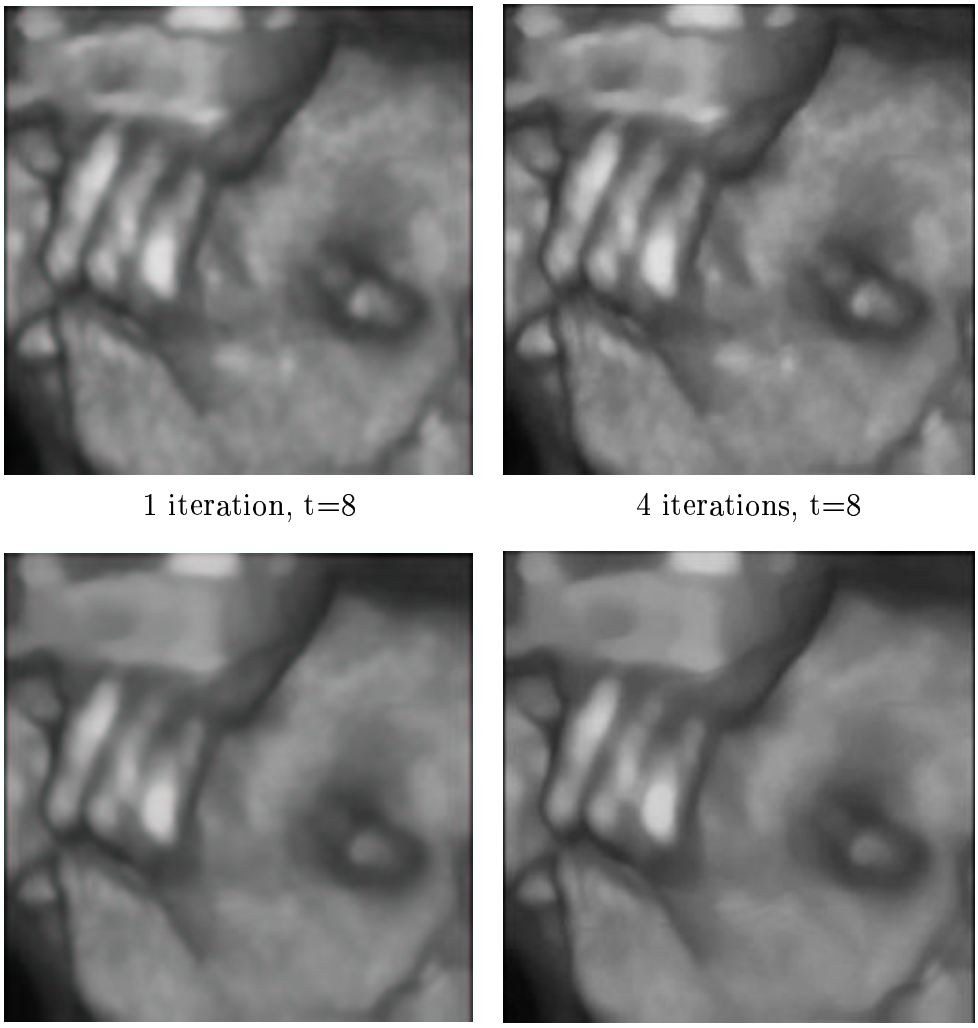


Figure 4.3: Results for the three-dimensional ultrasound data from Figure 4.1(b) with $\beta = 0.001$. The left column shows the renderings for noniterated, the right column for iterated regularization. The regularization parameter for iterated regularization was $h = 2$.

Chapter 5

Information Measures in Scale-Spaces

Jon Sporring and Joachim Weickert.
IEEE Transactions on Information Theory,
Vol. 45, No. 3, 1051–1058, April 1999.

Abstract

This paper investigates Rényi's generalized entropies under linear and nonlinear scale-space evolutions of images. Scale-spaces are useful computer vision concepts for both scale analysis and image restoration. We regard images as densities and prove monotony and smoothness properties for the generalized entropies. The scale-space extended generalized entropies are applied to global scale selection and size estimations. Finally, we introduce an entropy-based fingerprint description for textures.

5.1 Introduction

In recent years multiscale techniques have gained a lot of attention in the image processing community. Typical examples are pyramid and wavelet decompositions. They represent images at a small number of scales and have proven their use for many image processing tasks. Another important class of multiscale techniques consists of so-called *scale-space representations* [175, 420, 436, 213]. They embed an original image into a continuous family of subsequently simpler versions. Many scale-spaces can be formulated as the evolution of the initial image under a suitable linear or nonlinear diffusion process. Such an image evolution is useful for tasks such as feature extraction, scale selection, and segmentation, see [230, 382, 369] and the references therein.

Information theoretical concepts such as the Shannon–Wiener entropy [356, 434], Rényi's generalized entropies [316, 314, 315], and the Kullback–Leibler distance [220] have made contributions to image analysis; for instance Brink and Pendock [55], Brink [54], and Sahoo *et al.* [327] have used them for local image thresholding, and Vehel *et al.* [396] and Chaudhuri and Sarkar [82] study images in a multifractal setting. It is not difficult to see that the generalized entropies, the multifractal spectrum, the grey-value moments and the grey-value histogram itself are equivalent representations: they can be transformed into each other by one-to-one mappings. More details can be found in Appendix 5.6.

Since scale-spaces simplify images, it is only natural to investigate their simplification properties in terms of information measures. Already in 1949, Shannon mentioned that the Shannon–Wiener entropy decreases under averaging transformations [356, p. 52]. In 1993 Illner and Neunzert [191] studied a biased diffusion process, where the original image evolves towards a background image b along a path where its Kullback–Leibler distance with respect to b increases monotonically. Jägersand [196] used the Kullback–Leibler distance in linear scale-space for focus-of-attention. Oomes and Snoeren [287] used the entropy relative to a background measure to estimate the size of objects in images. Sporring [368] applied the Shannon–Wiener entropy in linear scale-space to perform scale selection in textures and showed the monotone behaviour using concepts from thermodynamics. Weickert [412] proved monotony of the Shannon–Wiener entropy in linear and nonlinear diffusion scale-spaces by regarding it as a Lyapunov functional. Lyaponov functionals have been used for scale-space synchronization [277] and for a uniform sampling of the scale axis with respect to its information content [430]. The fractal dimension in scale-spaces has been investigated by Peleg *et al.* [298], Barth *et al.* [36], and Pei *et al.* [297]. Relations between Shannon–Wiener entropy and multiscale concepts in terms of wavelets have been established by Krim and Brooks [217], where inequality theory was applied to propose optimal measures for feature-directed segmentation.

The present paper extends previous work in this field by studying both theoretical aspects and the practical potential of generalized entropies in a linear and nonlinear multiscale setting. Generalized entropies are complete in the sense that they allow for a reconstruction of the grey-value histogram (see Appendix 5.6). A scale-space extension is used to complement the entropies with spatial information. We prove monotony and smoothness properties with respect to the information order and the scale parameter. We use the scale-space behaviour of generalized entropies for scale selection and size estimation, and we introduce a fingerprint-like description for textures. The results indicate that our extensions broaden the potential use of entropy methods in image analysis. Some preliminary results in this paper have been presented at conferences [368, 372].

Throughout this paper we identify an image by its two-dimensional distribution of light on a rectangular image domain. It should be noted that this representation is invariant under multiplication with, but not under addition of, a constant. It is important to note that this two-dimensional distribution is *not* the grey-value histogram.

The outline of our paper is as follows. In Section 5.2 will be given a brief introduction to linear and nonlinear scale-spaces. Then in Section 5.3 we will investigate a scale-space extension of the generalized entropies. Finally in Section 5.4 we will describe some applications in image processing. A conclusion is given in Section 5.5.

5.2 A Short Introduction to Scale-Spaces

The images considered in this work are all discrete, but for simplicity we will in this section introduce two scale-spaces in the continuous setting. Discrete scale-space aspects are discussed by Lindeberg [230] for the linear framework, and by Weickert [412] for the nonlinear setting. Scale-spaces can be considered as an alternative to traditional smoothing methods from statistics [359].

In scale-space theory one embeds an image $p(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}$ into a continuous family $\{p(\mathbf{x}, t) \mid t \geq 0\}$ of gradually smoother versions of it. The original image corresponds to the scale $t = 0$, and increasing the scale should simplify the image without creating spurious structures. Since a scale-space creates a hierarchy of the image features, it constitutes an important step from a pixel-related image description to a semantical image description.

It has been shown that partial differential equations are the suitable framework for scale-spaces [12]. The oldest and best studied scale-space obtains a simplified version $p(\mathbf{x}, t)$ of $p(\mathbf{x})$ as the solution of the linear diffusion process with $p(\mathbf{x})$ as initial value.

$$\partial_t p \;=\; \partial_{x_1 x_1} p + \partial_{x_2 x_2} p, \tag{5.1}$$

$$p(\mathbf{x}, 0) \;=\; p(\mathbf{x}), \tag{5.2}$$

where $\mathbf{x} = (x_1, x_2)^T$. It is well-known from the mathematical literature that the solution $p(\mathbf{x}, t)$ can be calculated by convolving $p(\mathbf{x})$ with a Gaussian of standard deviation $\sigma = \sqrt{2t}$:

$$p(\mathbf{x}, t) \;=\; (G_t * p)(\mathbf{x}), \tag{5.3}$$

$$G_t(\mathbf{x}) \;:=\; \frac{1}{4\pi t} e^{\frac{-|\mathbf{x}|^2}{4t}}. \tag{5.4}$$

This process is called *Gaussian scale-space* or *linear scale-space*. It was first discovered by Iijima [175, 420] and became popular two decades later by the work of Witkin [436] and Koenderink [213]. A detailed treatment of the various aspects of Gaussian scale-space theory can be found in [230, 127, 369] and the references therein.

Unfortunately, Gaussian smoothing also blurs and dislocates semantically important features such as edges. This has triggered people to study nonlinear scale-spaces. Perona and Malik [303] proposed to replace the linear diffusion equation (5.1) by the nonlinear diffusion process

$$\partial_t p = \nabla \cdot (g(|\nabla p|)\, \nabla p), \tag{5.5}$$

where $\nabla = (\partial_x, \partial_y)^T$, and the diffusivity $g(|\nabla p|)$ is a decreasing function of $|\nabla p|$. The idea is to regard $|\nabla p|$ as an edge detector and to encourage smoothing within a region over smoothing across boundaries. Thus, locations where the gradient is large have a large likelihood of being an edge, and the diffusivity is reduced.

In our experiments we consider a nonlinear diffusion process where the diffusivity is given by [80]

$$g(|\nabla p|) := \frac{1}{\sqrt{1 + |\nabla p|^2/\lambda^2}} \qquad (\lambda > 0). \tag{5.6}$$

Such a choice guarantees that the nonlinear diffusion filter is well-posed.

This is one of the simplest representative of nonlinear scale-spaces. Overviews of other nonlinear scale-spaces can be found in [382, 412].

5.3 Generalized Entropies

Let us now consider a discrete image $\mathbf{p} = (p_1, \ldots, p_N)^T$, where $p_i > 0$ for all i. Note that a single index is used for the two-dimensional enumeration of pixels. Its family of generalized entropies is defined as,

$$S_\alpha(\mathbf{p}) := \frac{1}{1-\alpha} \log \sum_{i=1}^{N} p_i^\alpha \tag{5.7}$$

for $\alpha \neq 1$. The limit from left and right at $\alpha = 1$ is the Shannon–Wiener entropy,

$$S_1(\mathbf{p}) = -\sum_{i=1}^{N} p_i \log p_i, \tag{5.8}$$

and we might as well consider it as part of the continuum. The parameter α is called information order.

Let the vector-valued function $\mathbf{p}(t) = (p_1(t), \ldots, p_N(t))^T$ be the linear or non-linear scale-space extension, where the continuous parameter t denotes scale. These scale-spaces can be obtained by a spatial discretization of Equation 5.1 or 5.5 with reflecting boundary conditions.

We will now discuss some details of the mathematical structure of generalized entropies.

Proposition 1 *The generalized entropies are decreasing in α.*

Proof: Follows immediately from [316, 167]. □

Proposition 2 *The generalized entropies $S_\alpha(\mathbf{p}(t))$ are increasing in t for $\alpha > 0$, constant for $\alpha = 0$, and decreasing for $\alpha < 0$. For $t \to \infty$, they converge to S_0.*

Proof: The proof is based on a result from [412, Theorem 5]: For a discrete image $\mathbf{p}(t)$, which is obtained from a spatially discrete diffusion scale-space, the following holds. The expression

$$\Phi(\mathbf{p}(t)) := \sum_{i=1}^{N} r(p_i(t)) \tag{5.9}$$

is decreasing in t for every smooth convex function r. Moreover, $\lim_{t\to\infty} p_i(t) = 1/N$ for all i.

Using this we first prove the monotony of S_α with respect to t. Let $\alpha > 1$ and $s > 0$. Since $r(s) = s^\alpha$ satisfies

$$r''(s) = \alpha(\alpha - 1)s^{\alpha-2} > 0, \tag{5.10}$$

it follows that r is convex. Thus,

$$\Phi(\mathbf{p}(t)) = \sum_{i=1}^{N} r(p_i(t)) = \sum_{i=1}^{N} p_i^\alpha(t) \tag{5.11}$$

is decreasing in t and

$$S_\alpha(\mathbf{p}(t)) = \frac{1}{1-\alpha} \log \Phi(\mathbf{p}(t)) \tag{5.12}$$

is increasing in t.
Similar reasonings can be applied to establish monotony for the cases $0 < \alpha < 1$ and $\alpha < 0$.

For $\alpha = 1$ we obtain the Shannon–Wiener entropy for which monotony has already been shown in [412].

Let $\alpha = 0$. Then

$$S_0(\mathbf{p}(t)) = \log \sum_{i=1}^{N} p_i^0(t) = \log N = \text{const.} \quad \forall\, t. \tag{5.13}$$

To verify the asymptotic behaviour of the generalized entropies we utilize

$$\lim_{t\to\infty} p_i(t) = \frac{1}{N}. \tag{5.14}$$

For $\alpha \neq 1$ this gives

$$\lim_{t\to\infty} S_\alpha(\mathbf{p}(t)) = \frac{1}{1-\alpha} \log \sum_{i=1}^{N} \frac{1}{N^\alpha} = \log N = S_0, \tag{5.15}$$

and $\alpha = 1$ yields

$$\lim_{t\to\infty} S_1(\mathbf{p}(t)) = -\sum_{i=1}^{N} \frac{1}{N} \log \frac{1}{N} = \log N = S_0. \tag{5.16}$$

This completes the proof. □

The following smoothness results constitute the basis for studying derivatives of generalized entropies as will be done in Section 5.4.

Proposition 3 *The generalized entropies are* C^∞ *for* $\alpha \neq 1$ *and at least* C^1 *in* $\alpha = 1$. *For linear scale-space they are* C^∞ *in* t, *and for the nonlinear scale-space they are* C^1 *in* t.

Proof: In order to prove smoothness with respect to α, we first consider the case $\alpha \neq 1$. Then S_α is the product of the two C^∞ functions $\frac{1}{1-\alpha}$ and $\log \sum_{i=1}^{N} p_i^\alpha$, and thus also C^∞ in α.

The smoothness in $\alpha = 1$ is verified by applying l'Hôpital's rule. Straightforward calculations show that

$$\lim_{\alpha\to 1} \frac{\partial S_\alpha}{\partial \alpha} = \frac{\sum_{i=1}^{N} p_i (\log p_i)^2 - (\sum_{i=1}^{N} p_i \log p_i)^2}{2}. \tag{5.17}$$

Thus, $\frac{\partial S_\alpha}{\partial \alpha}$ exists and S_α is in C^1.

For linear scale-space, C^∞ in t follows directly from the fact that $G_t(x)$ is in C^∞ with respect to t. In the nonlinear case, C^1 in t is a consequence of the fact that the solution $\mathbf{p}(t)$ is in C^1 with respect to t. This is proven in [412, Theorem 4]. □

Figure 5.1 illustrates the monotony of the generalized entropies both in scale and order for both scale-spaces.

The figures have been created by finite difference algorithms which preserve the monotony properties established in this section [429].

5.4 Experiments

We will in this section demonstrate some applications for the generalized entropies in image processing. We will consider the change of entropies by logarithmic scale,

$$c_\alpha(\mathbf{p}(t)) := \frac{\partial S_\alpha(\mathbf{p}(t))}{\partial(\log t)}, \tag{5.18}$$

since this appears to be the natural parameter (at least for linear scale-space) [213],[130],[230, section 8.7.1],[372]. We emphasize that the generalized entropies are global measures and are thus best suited for images with homogeneous textures.

5.4.1 Shannon–Wiener Entropy and Zooming

This section analyses the zooming behaviour of the Shannon–Wiener entropy in linear scale-space.

Figure 5.2 (top left and right) shows images from a laboratory experiment. The camera is placed fronto-parallel to a plane with a simple texture: pieces of paper with discs arranged in a regular manner. A sequence is produced as a series of increasing zoom values.

In Figure 5.2 (bottom) we plot the scale $\sigma = \sqrt{2t}$ of the point of maximum entropy change against the mean size of the discs. As can be seen the relation is close to linear. It appears that in linear scale-space the point of maximal entropy change by logarithmic scale corresponds to the size of the dominating image structures.

5.4.2 Spatial Extent of Structures

In this section we show that the scaling behaviour in linear scale-space carries over to the generalized entropies, and that they can be used to simultaneously measure the size of light and dark structures. We shall also see that the latter cannot be done with the Shannon–Wiener entropy.

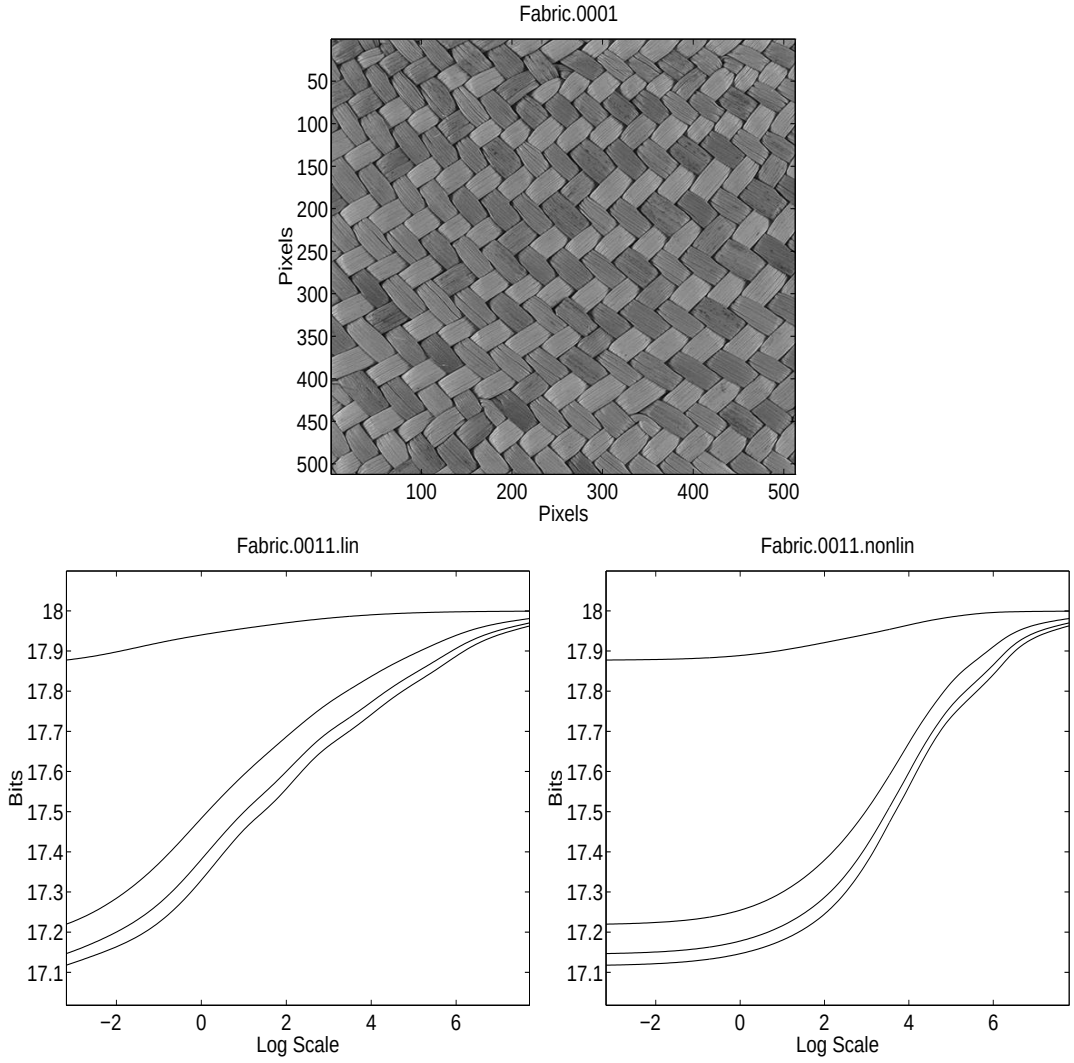


Figure 5.1: Examples of some generalized entropies. (a) TOP: A 512×512 grey-valued image. (b) BOTTOM LEFT: Generalized entropies in linear scale-space. From top to bottom: $\alpha = 1, 34, 67, 100$. (c) BOTTOM RIGHT: Ditto for nonlinear scale-space.

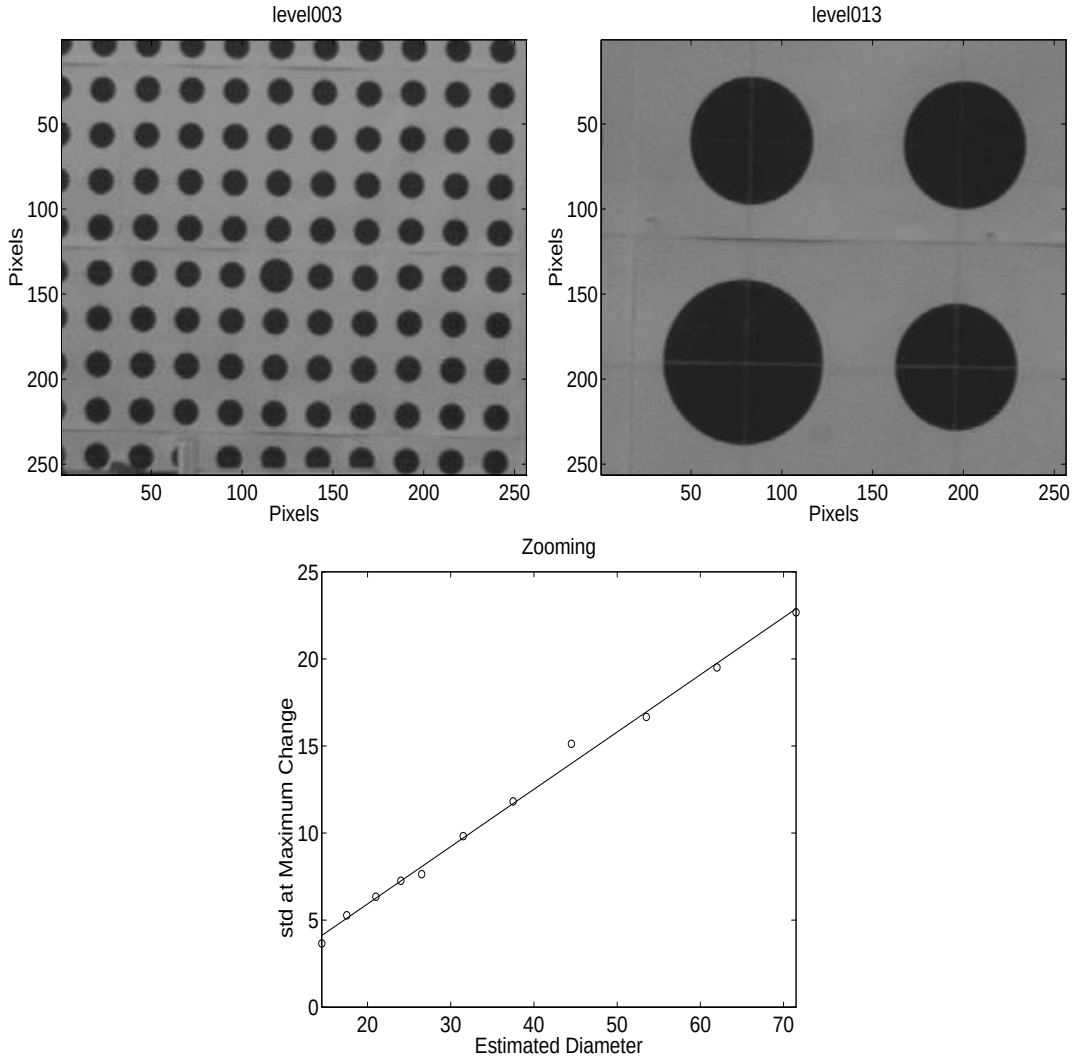


Figure 5.2: A zooming sequence. (a) TOP: First and last image. (b) BOTTOM: The $\sigma = \sqrt{t/2}$ values maximizing $c_1(\mathbf{p}(t))$ versus the estimated disc sizes.

The idea is as follows: The definition of the generalized entropies implies that entropies for large positive α focus on high grey-values (white areas), while for large negative value they analyse low grey-values (dark areas).

We expect that $c_\alpha(\mathbf{p}(t))$ is especially high for structures of diameter d, when the variance $\sigma^2 = t/2$ of the Gaussian is close to the variance of the structures. Let us for simplicity consider a random variable with uniform probability density function whose support is a disc of diameter d. Its variance is,

$$\sigma^2 = \int_0^{2\pi} \int_0^{d/2} \frac{r^2}{\pi(d/2)^2} r \; dr \; d\phi = \frac{d^2}{8}. \tag{5.19}$$

Hence we expect a light (or dark) structure of diameter d to have a significant entropy change by logarithmic scale at time $\sigma^2/2 = d^2/16$. This size estimate remains qualitatively correct for non-disc structures. In this case, it gives the size of the largest minimal diameter.

Figure 5.3 shows the result of a performance analysis.

The size estimate (5.19) has been applied to a number of simple sinusoidal images with structures (half wavelengths) between 1 and 257 pixels. As can be seen in the bottom graph, for sufficiently large structures the estimated sizes are close to the true size. Although by definition, the generalized entropies are not symmetric in order, both have a similar scaling behaviour which is close to linear.

In Figure 5.4 we show an experiment on a texture with a more complicated periodicity.

This real image has been created by the Belousov–Zhabotinsky reaction [199]. From orders ± 20 we find dominating low intensity values corresponding to a diameter 7.2, while the dominating high intensity values suggest structures of diameter 3.5. From this we conclude that the distance between the light spiral arms in the mean is approximately 7.2 pixels, and the width of the spiral arms is approximately 3.5 pixels. In spite of the fact that the disc model (5.19) is not very appropriate for the line like structure, the size estimates are in the correct order of magnitude.

The Shannon–Wiener entropy cannot be used for size estimation since it is a mixture of information from both light and dark areas. Thus is does not allow for a distinction between fore- and background.

5.4.3 Fingerprints for Entropies in Scale-Space

Section 5.4.1 and 5.4.2 have shown that the scales of extremal entropy change carry significant information for selected information orders. Thus it would be interesting to introduce a compact description of the extremal changes for the continuum of information orders. In analogy with edge analysis in linear scale-space [442] we call such a description a fingerprint image. In Figure 5.5 fingerprint images for two texture are given, both in the linear and nonlinear scale-space.

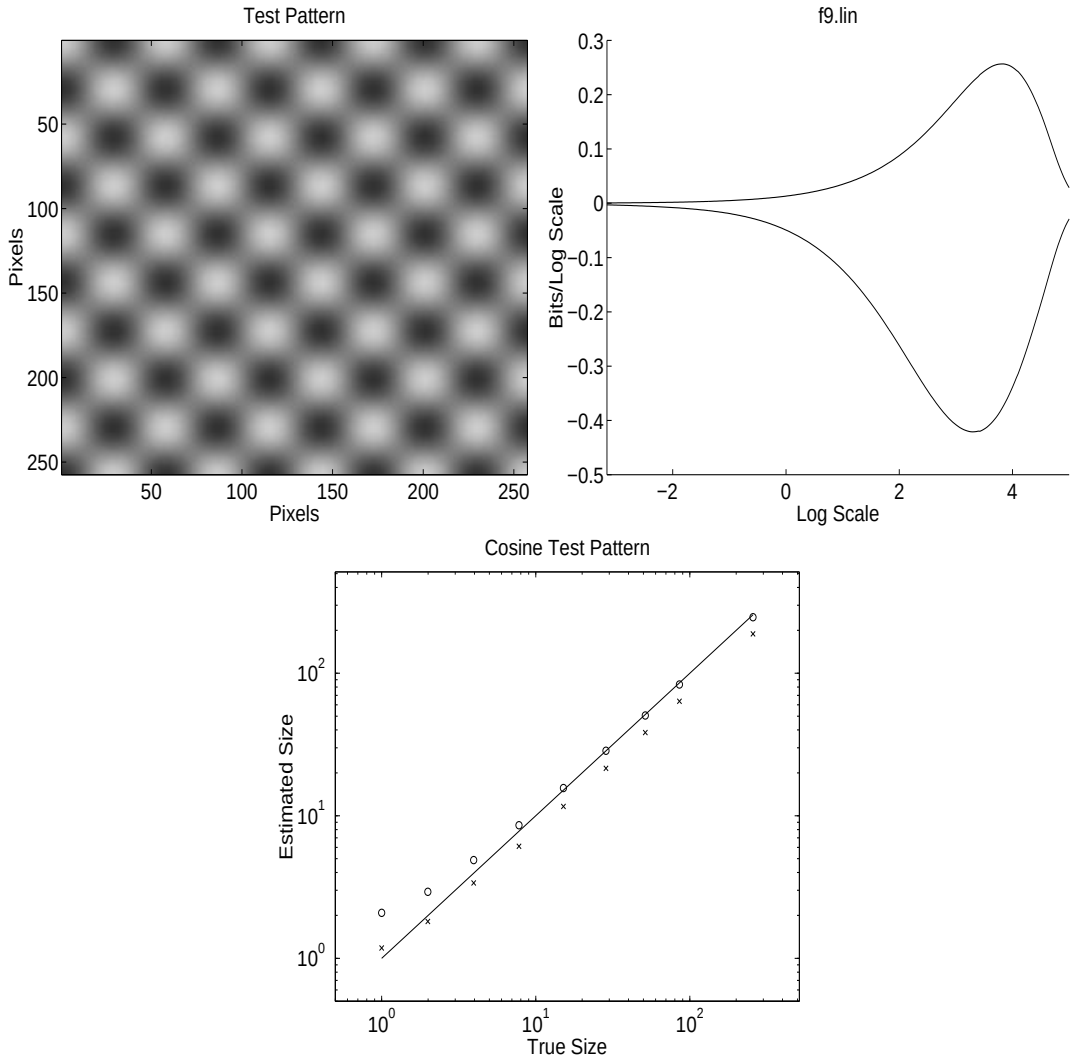


Figure 5.3: Scaling behaviour and size estimation with generalized entropies. (a) TOP LEFT: Test image generated by $257^{-2}(1 + 0.6\cos(\omega x_1)\cos(\omega x_2))$ with $\omega = 9\pi/257$. (b) TOP RIGHT: The corresponding $c_\alpha(\mathbf{p}(t))$ curves for $\alpha = \pm 100$. Top curve is for positive order and bottom curve for negative order. (c) BOTTOM: A double logarithmic plot of the true size versus the estimated size for various ω. The straight line depicts the truth, the circles the estimation from order 100, and the crosses for order -100.

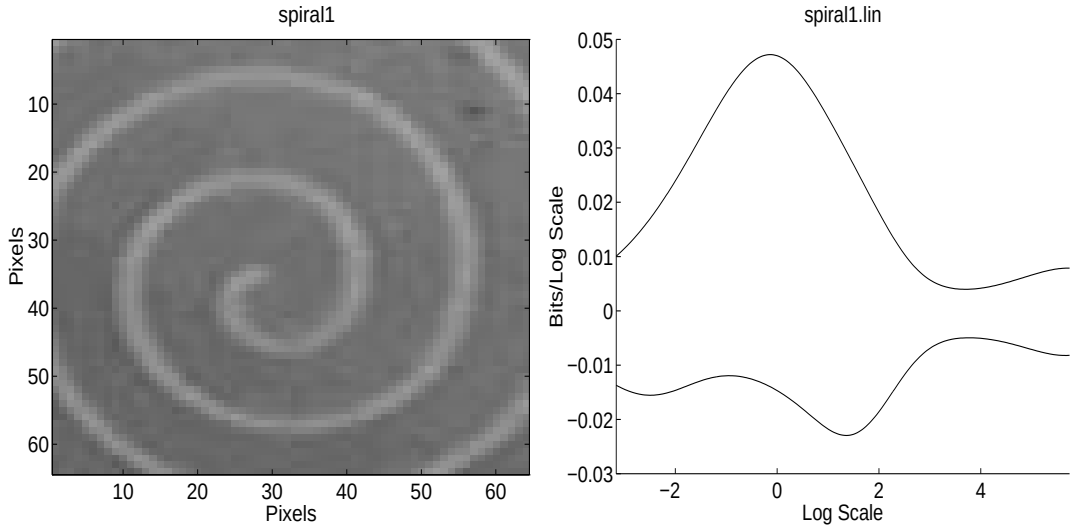


Figure 5.4: (a) LEFT: Spiral generated by a chemical reaction. (b) RIGHT: Entropy changes for orders 20 (top curve) and -20 (bottom curve).

The fingerprint lines are the extrema of $c_\alpha(\mathbf{p}(t))$ in t. Our monotony results immediately imply the following consequences: If there is only one fingerprint line for a given positive order, then it corresponds to a maximum (likewise to a minimum for negative orders); see also Figure 5.3. For almost all orders there will be an odd number of fingerprint lines, which correspond to alternating maxima and minima. This can be seen for instance in the middle right graph in Figure 5.5. For information order 60, the leftmost line is a maximum followed by alternating minima and maxima.

It appears that the location of the fingerprint lines is more stable over information orders for the nonlinear scale-space than for the linear one. Due to the reduced diffusivity of the nonlinear scale-space, the fingerprint lines are shifted towards higher scales.

5.5 Conclusions

In this paper we have investigated entropies as a means for extracting information from scale-spaces. This has lead to the following contributions.

- Monotony and smoothness properties for the Shannon–Wiener entropy and Rényi's generalized entropies have been proven for the linear and a nonlinear diffusion scale-space. The proofs hold also for all other nonlinear diffusion scale-spaces treated in [412].

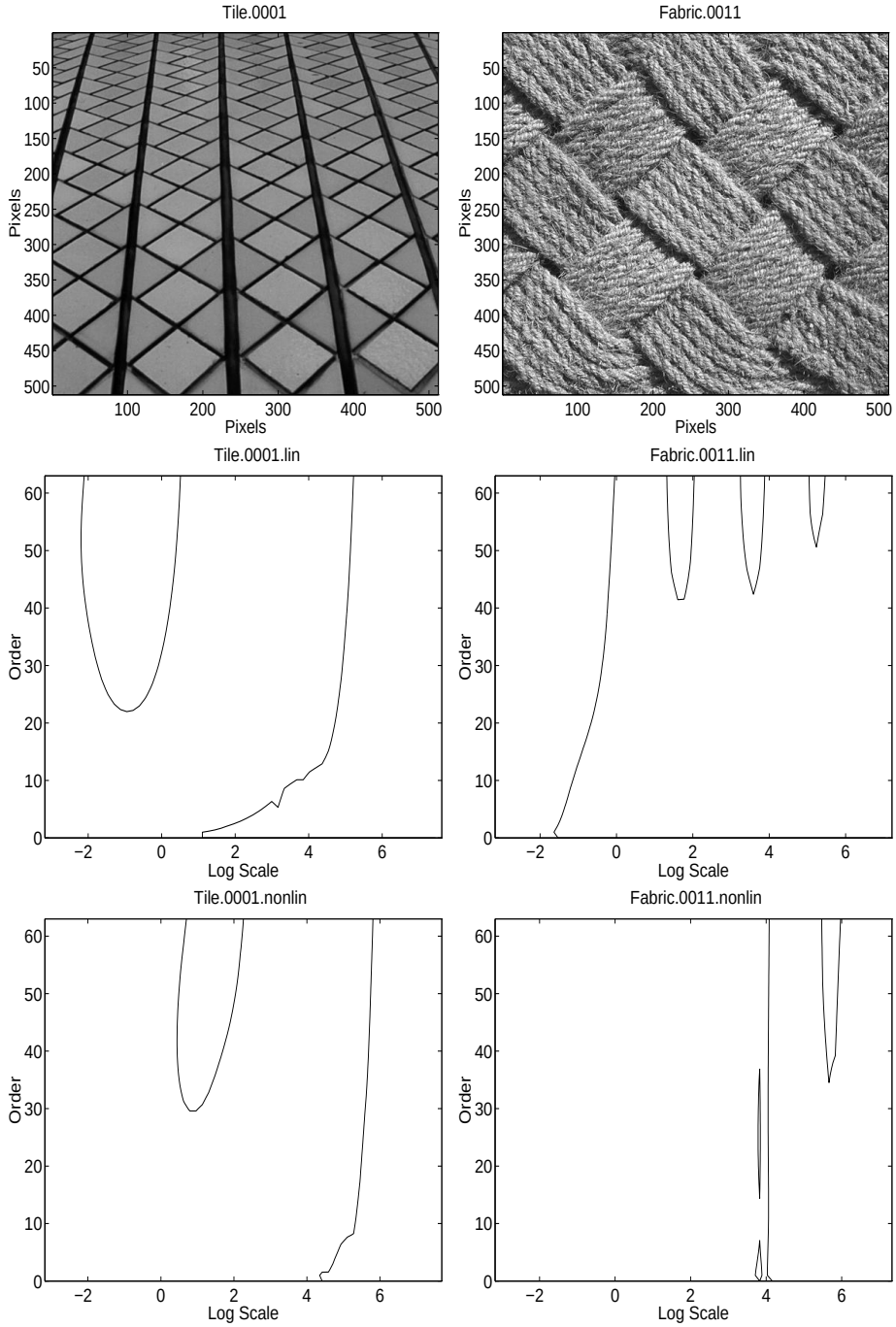


Figure 5.5: Fingerprints of generalized entropies. (a) Top Row: Two textures. (b) Middle Row: Fingerprints for linear scale-space. (c) Bottom Row: Ditto for nonlinear scale-space.

- We have illustrated that the generalized entropies can be used to perform size measurements for periodic textures. This is not possible with the Shannon–Wiener entropy. We have proceeded to define a fingerprint image for entropies in scale-space and analysed some of its basic properties. The localization of the fingerprint lines can be improved using nonlinear instead of linear scale-space.

The following topics appear promising for future work.

- In the context of texture analysis, it would be interesting to perform an in-depth study on the relation between the fingerprint topology and the structure of the texture.

- This paper has focused on the maximal entropy change by scale to estimate the size of image structures. The minimal change by scale, however, indicates especially stable scales with respect to evolution time. We expect these scales to be good candidates for stopping times in nonlinear diffusion scale-spaces.

- The entropies in this paper are global measures. For topics such as focus-of-attention it would be interesting to study local variants of them.

It should be emphasized that the analysis carried out in this paper is directly transferable to the analysis of multifractals, grey-value moments, and grey-value histograms.

Acknowledgments. This research has been funded in parts by the Real World Computing Partnership, the Danish National Research Council, and the EU-TMR Project VIRGO. We thank Peter Johansen, Mads Nielsen, Luc Florack, Ole Fogh Olsen, and Robert Maas for many discussions on this topic. Preben Graae Sørensen from the Department of Chemistry at the University of Copenhagen [199] has supplied the spiral image in Figure 5.4. The images in Figure 5.5 are taken from the 'VisTex' collection [304].

5.6 Appendix. Relations to Grey-Value Moments, Histograms, and Multifractal Spectra

The grey-value moments of an image are defined as,

$$m_\alpha(\mathbf{p}) = \sum_{i=1}^{N} p_i^\alpha. \tag{5.20}$$

From the definition of S_α in (5.7) it is clear that there is a one-to-one relation to m_α.

Let the image $(p_1, \ldots, p_N)^T$ consist of M distinct grey-values $v_1, \ldots, v_M$ occurring $f_1, \ldots, f_M$ times. We may use this grey-value histogram $\mathbf{f}$ to rewrite the moments as,

$$m_\alpha(\mathbf{p}) := \sum_{j=1}^{M} f_j v_j^\alpha. \tag{5.21}$$

Considering the moments $m_0, \ldots, m_{M-1}$ gives the relation:

$$\begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ \vdots \\ m_{M-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ v_1 & v_2 & \ldots & v_M \\ v_1^2 & v_2^2 & \ldots & v_M^2 \\ & & \ddots & \\ v_1^{M-1} & v_2^{M-1} & \ldots & v_M^{M-1} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_M \end{bmatrix} \tag{5.22}$$

The system matrix is a so-called Vandermonde matrix. By induction over M the determinant can be shown to be $\prod_{1 \le n < m \le M} (v_m - v_n)$. Since $v_j, j = 1, \ldots, M$ are distinct, the matrix is invertible (but ill-conditioned). Thus there is a one-to-one relation between the moments $m_0, \ldots, m_{M-1}$ and the histogram $f_1, \ldots, f_M$.

The equivalence of the multifractal spectrum and the generalized entropies is discussed in [163].

Part II

Coherence-Enhancing Anisotropic Diffusion

Chapter 6

Coherence-Enhancing Diffusion Filtering

Joachim Weickert.
International Journal of Computer Vision,
Vol. 31, No. 2/3, 111–127, April 1999.

Abstract

The completion of interrupted lines or the enhancement of flow-like structures is a challenging task in computer vision, human vision, and image processing. We address this problem by presenting a multiscale method in which a nonlinear diffusion filter is steered by the so-called interest operator (second-moment matrix, structure tensor). An m-dimensional formulation of this method is analysed with respect to its well-posedness and scale-space properties. An efficient scheme is presented which uses a stabilization by a semi-implicit additive operator splitting (AOS), and the scale-space behaviour of this method is illustrated by applying it to both 2-D and 3-D images.

6.1 Introduction

Oriented flow-like structures arise in many computer vision and image processing problems: Within the field of texture analysis they appear for instance in the automatic grading of fabrics or wood surfaces, but they are also of importance for fingerprint analysis in forensic applications. They are present in many scientific imaging problems ranging from fluid dynamics to meteorology, and last but not least in medical imaging, for instance in the analysis of trabecular structures in bones. Interestingly, related tasks such as gap completion or the completion of interrupted lines also play a role in human vision tasks such as perceptual grouping. Moreover, the success of paintings by Munch or van Gogh suggests that emphasizing flow-like structures may create effects which fascinate many people.

Analysing flow-like patterns is an active research topic for certainly more than one decade; see for instance Kass and Witkin [200]. Besides methods such as Gabor filters or steerable filters, many of these approaches are equivalent to the so-called structure tensor (interest operator, second moment matrix) [132, 313]. In the meantime this field has even entered textbooks [153], and it can be regarded as well-established.

Much less attention has been paid to the question how to *enhance* flow-like patterns. Poor quality of fingerprint or trabecular bone images is not unusual. In those cases it would be desirable to have a tool which improves the quality of flow-like structures without destroying for instance semantically important singularities like the minutiae in fingerprints.

For problems like the grading of fabrics or applications to fluid dynamics it is also useful to have a multiscale simplification of the original image by embedding it into a scale-space in order to obtain a subsequently coarser, more global impression of the main flow-like structures. Of course, such a scale-space should take into account the coherence of the structures by smoothing mainly along their preferred orientation instead of perpendicular to it.

Since flow-like structures can also be present in higher dimensional data sets, e.g. 3-D images of trabecular bones, it should be possible to generalize such a method to arbitrary dimensions.

The preceding problems will be addressed in this paper by presenting an m-dimensional scale-space for the enhancement of coherent structures. The underlying concept can be motivated from ideas of Perona and Malik for improving edge detection by creating a feedback loop of an edge detector and a nonlinear diffusion process [303]. In a similar way we embed a classical method for describing flow-like structures – namely the structure tensor approach – into a nonlinear diffusion process. This turns a method for *analysing* coherent pattern into a technique for *enhancing* and *simplifying* them. In contrast to most nonlinear diffusion filters, however, we use an approach where the process is steered by a diffusion tensor

instead of a scalar-valued diffusivity. This enables true anisotropic behaviour by adapting the diffusion process not only to the location, but allowing also different smoothing in different directions. We shall see that this filter belongs to a class of nonlinear diffusion methods for which many well-posedness and scale-space properties can be proved. This is also in contrast to the Perona–Malik filter, whose continuous formulation is only locally well-posed [202]. Details of the numerical implementation will be discussed as well. In particular, an efficient novel stabilization technique based on a semi-implicit additive operator splitting (AOS) is presented. It extends previous encouraging experiments with AOS schemes for nonlinear diffusion filtering with a scalar-valued diffusivity to the fully anisotropic case with a diffusion tensor. All theoretical and numerical discussions hold in the m-dimensional case. As examples, results from 2-D and 3-D implementations are presented.

The outline of the papers is as follows: Section 6.2 surveys the underlying structure tensor method for describing coherence in images. This method is used in Section 6.3 for constructing a nonlinear diffusion process which smoothes along coherent flow-like structures. This process is called coherence-enhancing diffusion (CED). Section 6.4 gives a detailed analysis of the theoretical properties of a more general class of diffusion filters comprising CED. In Section 6.5 numerical and algorithmical questions are addressed; in particular, a more efficient alternative to the explicit scheme is presented. Section 6.6 applies CED to 2-D and 3-D images from different application areas. The paper is concluded with a summary in Section 6.7. Some results in this paper have been presented earlier at conferences [406, 427].

Related work. Since the work of Perona and Malik [303] numerous nonlinear diffusion filters have been proposed; see e.g. [382, 411] for an overview. Nevertheless, most of them use a (spatially varying) scalar diffusivity, not a diffusion tensor. Thus, they act inhomogeneously (nonuniformly) on the image, but – in our terminology – they remain isotropic.

True anisotropic diffusion filtering is studied in the reaction-diffusion model of Cottet and Germain [91]. Its diffusion tensor uses the eigenvectors $v_1 \parallel \nabla u_\sigma$, $v_2 \perp \nabla u_\sigma$ and its eigenvalues are given by

$$\lambda_1 := 0, \tag{6.1}$$

$$\lambda_2 := \frac{\eta |\nabla u_\sigma|^2}{1 + (|\nabla u_\sigma|/\sigma)^2} \qquad (\eta > 0). \tag{6.2}$$

This choice is similar to our method it that sense that it diffuses mainly along strongly anisotropic structures. However, there are two important differences: First, we observe that this diffusion tensor cannot be treated within a scale-space framework using uniformly positive definite diffusion tensors. Moreover, the Cottet–Germain model uses an additional reaction term which leads to nontrivial steady-states and qualifies it as a pure restoration method without scale-space ambitions.

Second, the eigendirections of D are adapted to ∇u_σ, not to the eigendirections of the structure tensor. We shall see that the introduction of an integration scale in the structure tensor is an essential model feature in order to obtain reliable orientation estimates for flow-like structures. Recently Cottet and El Ayyadi have proposed a modified image restoration model which replaces the spatial regularization by a temporal regularization [90].

Nitzberg and Shiota [280] pioneered shape-adapted Gaussian smoothing, where the shape of an anisotropic Gaussian convolution kernel is a function of the structure tensor. Later on similar proposals have been made in [232, 440] and supplemented with scale-selection methods [6]. It should be noted that a common feature of all the preceding shape-adapted Gaussian smoothing methods is the combination of isotropic smoothing inside a region with anisotropic smoothing along edges. Space-variant Gaussian blurring is, however, not equivalent to an inhomogeneous diffusion process, and it does not preserve the average grey value. Anisotropic diffusion filters with isotropic diffusion within regions and anisotropic diffusion along edges can be found in [408] and the references therein. These methods are different from the present approach: CED is basically a pure 1-D diffusion, where a minimal amount of isotropic smoothing is added only for regularization purposes.

Other anisotropic partial differential equations for smoothing images rely on morphological methods such as the *mean-curvature motion (geometric heat equation, Euclidean shortening flow)* [207, 13]

$$\partial_t u \;=\; u_{\xi\xi} \;=\; |\nabla u| \operatorname{curv}(u) \tag{6.3}$$

with ξ being the direction perpendicular to ∇u.

Since mean-curvature motion propagates isophotes in inner normal direction with curvature-dependent speed, we should not expect that such completely local methods are capable of closing interrupted line-like structures. The same holds for the affine invariant version of this process [12, 333]. All these methods do not take into account semilocal information as can be gained from the structure tensor. Recent results by Carmona and Zhong [66] confirm the importance of semilocal estimates of the smoothing direction also for equations of mean-curvature type, if one is interested in specific goals such as enhancement of line-like structures.

Three-dimensional nonlinear diffusion filters have been investigated first by Gerig *et al.* [144] in the isotropic case, and by Rambaux and Garçon [312] in the edge-enhancing anisotropic case. A generalization of coherence-enhancing anisotropic diffusion to higher dimensions was first proposed in [427]. A recent three-dimensional PDE-based filter by Krissian *et al.* [218] and a 3-D reaction-diffusion process by Payot *et al.* [296] may be related to these anisotropic diffusion techniques.

The discussion above shows that the distinctive features of our approach is a

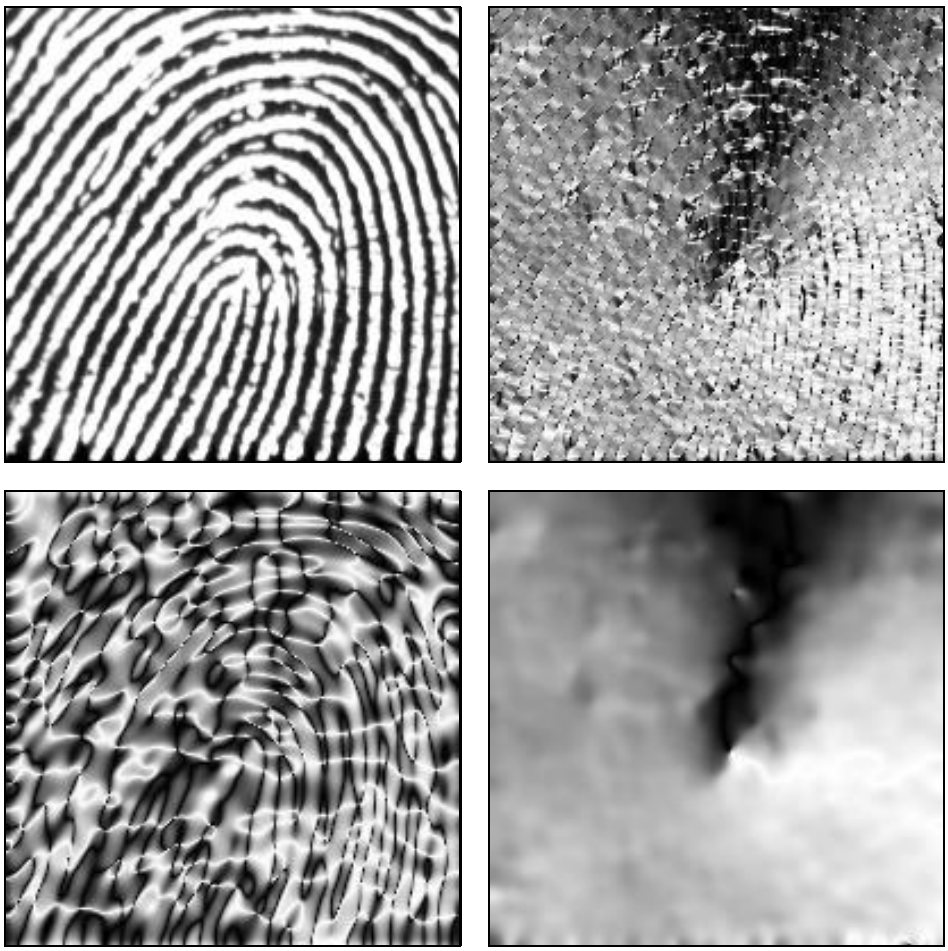


Figure 6.1: Local orientation in a fingerprint image. FROM TOP LEFT TO BOTTOM RIGHT: (a) Original fingerprint, $\Omega = (0, 200)^2$. (b) Orientation of smoothed gradient, $\sigma = 0.5$. (c) Orientation of smoothed gradient, $\sigma = 5$. (d) Structure tensor orientation, $\sigma = 0.5$, $\rho = 4$.

semi-local analysis by means of the structure tensor combined with 1-D diffusion along one of its eigenvectors.

6.2 The Structure Tensor

First we review a reliable tool for analysing coherent flow-like structures. Consider an m-dimensional image domain $\Omega := (0, a_1) \times \cdots \times (0, a_m)$, and let an image $u(x)$ be represented by a bounded mapping $u : \Omega \to \mathbb{R}$.

A very simple structure descriptor is given by ∇u_σ, the gradient of a Gaussian-

smoothed version of u:

$$K_\sigma(x) := \frac{1}{(2\pi\sigma^2)^{m/2}} \cdot \exp\left(-\frac{|x|^2}{2\sigma^2}\right), \tag{6.4}$$

$$u_\sigma(x,t) := (K_\sigma * u(.,t))(x) \qquad (\sigma > 0). \tag{6.5}$$

The standard deviation σ denotes the *noise scale*, since it makes the edge detector ignorant of details smaller than $O(\sigma)$. Convolution on a finite domain Ω can be defined as convolution between K_σ and $\tilde{u}$ on $\mathbb{R}^m$, where $\tilde{u}$ denotes an extension of u by mirroring. This results in applying no-flux boundary conditions.

Although ∇u_σ is useful for detecting edges, it is unsuited for finding parallel structures, as we can see from Figure 6.1. The left image shows an original fingerprint. Figure 6.1(b),(c) illustrates the gradient orientation using grey values: vertical gradients are depicted in black, and horizontal ones in white. We observe that for small σ high fluctuations remain, while larger σ lead to entirely useless results. This is due to the fact that for larger σ neighbouring gradients with the same orientation, but opposite sign cancel one another. Gradient smoothing averages directions instead of orientations[1]. To make the structure descriptor invariant under sign changes, we may replace ∇u_σ by its tensor product

$$J_0(\nabla u_\sigma) := \nabla u_\sigma \otimes \nabla u_\sigma := \nabla u_\sigma \nabla u_\sigma^T. \tag{6.6}$$

This matrix is symmetric and positive semidefinite, and its eigenvectors are parallel and orthogonal to ∇u_σ, respectively. The corresponding eigenvalues $|\nabla u_\sigma|^2$ and 0 describe just the contrast in the eigendirections. Now that we have replaced directions by orientations, we can average the orientations by applying a componentwise convolution with a Gaussian K_ρ:

$$J_\rho(\nabla u_\sigma) := K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma) \qquad (\rho \geq 0). \tag{6.7}$$

This matrix is named *structure tensor*, *interest operator* or *second-moment matrix*. It is useful for many different tasks, for instance for analysing flow-like textures [313], corners and T-junctions [132, 280], shape cues [232] and spatio–temporal image sequences [197]. Equivalent approaches have been discovered independently in [42, 200]. A book by Jähne [197] gives a nice overview of these methods and clarifies their mutual relations.

It is not hard to verify that the symmetric matrix J_ρ is positive semidefinite. Let its eigenvalues $\mu_1,...,\mu_m$ be ordered such that

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m \tag{6.8}$$

[1] In our terminology, gradients with opposite sign share the same orientation, but point in opposite directions.

and let $\{w_1, ..., w_m\}$ denote the corresponding orthonormal set of eigenvectors. Since the eigenvalues integrate the variation of the grey values within a neighbourhood of size $O(\rho)$, they describe the average contrast in the eigendirections w_1,...,w_m. Thus, the *integration scale* ρ should reflect the characteristic size of the texture. Usually, it is large in comparison to the noise scale σ. The eigenvector w_m corresponds to the smallest eigenvalue μ_m. It is the orientation with the lowest fluctuations, the so-called *coherence orientation*[2].

Figure 6.1(d) depicts the coherence orientation for the fingerprint image. We observe that it is exactly the desired average orientation of the lines. It should be noted how well the singularity corresponds to the singularity in the original fingerprint image.

The eigenvalues of J_ρ provide useful information on the coherence of a structure, i.e. the actual amount of anisotropy. As a measure for the coherence, one can define

$$\kappa := \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} (\mu_i - \mu_j)^2. \tag{6.9}$$

It becomes large for strongly differing eigenvalues, and it tends to zero for isotropic structures.

6.3 Coherence-Enhancing Anisotropic Diffusion in m Dimensions

Now that we know how to analyse coherent structures, we draw our attention to the question of how to enhance them. This can be done by embedding the structure tensor analysis into a nonlinear diffusion filter.

The principle of nonlinear diffusion filtering is as follows. One calculates a processed version $u(x,t)$ of $f(x)$ with a scale parameter $t \geq 0$ as the solution of a diffusion equation with f as initial condition and reflecting boundary conditions:

$$\partial_t u = \operatorname{div}(D\,\nabla u) \quad \text{on} \quad \Omega \times (0,\infty), \tag{6.10}$$

$$u(x,0) = f(x) \quad \text{on} \quad \Omega, \tag{6.11}$$

$$\langle D\nabla u,\, n\rangle = 0 \quad \text{on} \quad \partial\Omega \times (0,\infty). \tag{6.12}$$

Hereby, n denotes the outer normal and $\langle .,. \rangle$ the usual inner product. If one wants to adapt the diffusion process to the image itself one should choose the symmetric positive definite diffusion tensor $D = (d_{ij}) \in \mathbb{R}^{m\times m}$ as a function of the local image

[2]For the case that we have p identical minimal eigenvalues, one may regard every orientation within the span of w_{m-p+1},...,w_m as a coherence orientation. For natural images, this situation happens almost never and can be neglected.

structure. In the case of enhancing flow-like patterns, it is natural to adapt the diffusion tensor D to the structure tensor $J_\rho(\nabla u_\sigma)$.

How should this function look like? For enhancing coherence in images with flow-like structures, we need a smoothing process which acts mainly along the flow direction w_m (with the notations from Section 6.2) and the smoothing should increase with the strength of its orientation given by the coherence κ. This may be achieved in the following way:

We construct D such that it has the same eigenvectors as J_ρ and its eigenvalues are given by

$$\lambda_i \;:=\; \alpha \tag{6.13}$$

for $i = 1,...,m-1$, and by

$$\lambda_m \;:=\; \begin{cases} \alpha & \text{if } \kappa = 0, \\ \alpha + (1-\alpha)\exp\left(\frac{-C}{\kappa}\right) & \text{else.} \end{cases} \tag{6.14}$$

$C > 0$ serves as a threshold parameter: For $\kappa \gg C$ we get $\lambda_m \approx 1$, and $\kappa \ll C$ leads to $\lambda_m \approx \alpha$. The exponential function and the small positive parameter $\alpha \in (0, 1)$ were introduced mainly for two theoretical reasons: First, this guarantees that the smoothness of the structure tensor carries over to the diffusion tensor. The second reason is that the process never stops: Even if the structure becomes isotropic ($\kappa \to 0$), there remains some small linear diffusion with diffusivity $\alpha > 0$. Thus, α serves as a regularization parameter which keeps the diffusion tensor uniformly positive definite. In the next section we shall see that these are useful requirements in order to establish many theoretical properties for CED.

6.4 A General Well-Posedness and Scale-Space Framework

Coherence-enhancing anisotropic diffusion filtering can be regarded as a special case within a more general framework for nonlinear diffusion filtering, for which many well-posedness and scale-space properties can be found. This framework shall be analysed now.

6.4.1 Well-Posedness Properties

We study a diffusion filter with the following properties:

Assume that $f : \Omega \to \mathbb{R}$ is bounded, $\rho \geq 0$, and $\sigma > 0$.
Let $a := \inf_\Omega f$, $b := \sup_\Omega f$, and consider the problem

$$\partial_t u = \operatorname{div}\left(D(J_\rho(\nabla u_\sigma))\,\nabla u\right) \qquad \text{on} \qquad \Omega \times (0, \infty),$$

$$\begin{aligned} u(x,0) &= f(x) \qquad \text{on} \qquad \Omega, \\ \langle D(J_\rho(\nabla u_\sigma))\nabla u,\, n\rangle &= 0 \qquad \text{on} \qquad \partial\Omega \times (0,\infty), \end{aligned}$$

where the diffusion tensor $D = (d_{ij})$ satisfies the following properties:

(C1) Smoothness:
$D \in \mathrm{C}^\infty(\mathbb{R}^{m\times m}, \mathbb{R}^{m\times m})$.

(C2) Symmetry:
$d_{ij}(J) = d_{ji}(J)$ for all symmetric matrices $J \in \mathbb{R}^{m\times m}$.

(C3) Uniform positive definiteness:
If $w : \Omega \to \mathbb{R}^2$ satisfies $|w(x)| \leq K$ on $\bar{\Omega}$, there exists a positive lower bound $\nu(K)$ for the eigenvalues of $D(J_\rho(w))$.

This filter class is denoted by (P_c). Evidently, coherence-enhancing anisotropic diffusion satisfies the preceding requirements. Under the assumptions (P_c) the following theorem, which generalizes and extends results from [74, 408], can be proved.

Theorem 12 (Well-posedness properties, smoothness, and extremum principle)
The problem (P_c) has a unique solution $u(x,t)$ in the distributional sense, which is smooth for $t > 0$:

$$u \in \mathrm{C}^\infty(\bar{\Omega} \times (0,\infty)).$$

This solution depends continuously on f with respect to $\|\,.\,\|_{\mathrm{L}^2(\Omega)}$, and it fulfills the extremum principle

$$a \leq u(x,t) \leq b \quad \text{on} \quad \Omega \times (0,\infty). \tag{6.15}$$

Proof: See Appendix 6.8.1.

Some remarks are in order.

(a) We observe a strong smoothing effect which is characteristic for many diffusion processes: boundedness of the initial image is sufficient to obtain an infinitely often differentiable solution for arbitrary small positive times. Additional requirements – for instance that f should be uniformly continuous in order to apply the theory of viscosity solutions – are not necessary in our case.

(b) The continuous dependence of the solution on the initial image has significant practical impact as it ensures stability with respect to perturbations of the original image. This is of importance when considering stereo image pairs, spatio-temporal image sequences or slices from medical CT or MRI sequences, since we know that similar images remain similar after filtering.

(c) The extremum principle offers the practical advantage that, if we start for instance with an image within the range $[0, 255]$, we will never obtain results with grey value such as 257.

(d) Coherence-enhancing anisotropic diffusion in not the only filter which is covered by the theorem above. The well-posedness theory (as well as the subsequent scale-space framework) comprises the regularized Perona–Malik filter by Catté, Lions, Morel and Coll [74] as well as the edge-enhancing anisotropic diffusion filter described in [408]. It is also possible to establish a semidiscrete and fully discrete well-posedness and scale-space theory which proceeds in the same way; see [412] for more details.

6.4.2 Scale-Space Properties

Scale-space representations embed an image f into a family $\{T_t f \mid t \geq 0\}$ of gradually smoother, simplified versions of it. Long before this concept became popular in the western world by the work of Witkin [436] and Koenderink [213], it had been discovered and axiomatically justified in Japan by Iijima [175, 420]. Most axiomatic scale-space representations are devoted to the linear diffusion scale-space (Gaussian scale-space), but one can also create nonlinear scale-spaces which can be classified in a unique way [12, 390, 286]. All these approaches are uncommitted, since their unique classification does not give us the liberty to incorporate a-priori knowledge.

Anisotropic nonlinear diffusion filtering, however, does offer the possibility to incorporate knowledge into its evolution. So let us now study some of its scale-space properties.

It is evident that the filter class (P_c) satisfies typical scale-space properties such as the semi-group property, invariance under Euclidean image transformations, grey-level shifts, or contrast reversion. Moreover, since it can be written in divergence form and uses reflecting boundary conditions, it is also not hard to verify that the average grey level

$$\mu \;:=\; \frac{1}{|\Omega|} \int_{\Omega} f(x)\, dx \tag{6.16}$$

is not affected by nonlinear diffusion filtering [412, pp. 63–64]:

$$\frac{1}{|\Omega|} \int_{\Omega} T_t f \, dx \;=\; \mu \qquad \forall\, t > 0. \tag{6.17}$$

Average grey level invariance is a property in which diffusion scale-spaces differ from morphological scale-spaces. In general, the evolution PDEs of the latter ones

are not of divergence form and do not preserve the mean grey value. A constant average grey level is useful for scale-space based segmentation algorithms such as the hyperstack [277, 398]. It is also a desirable quality for applications in medical imaging where grey values measure physical qualities of the depicted object.

Causality in Terms of Nonenhancement of Local Extrema.

Let us now turn to the question in which sense an evolution equation of type (P_c) can be considered as a smoothing, information-reducing image transformation.

Koenderink [213] required that a scale-space evolution should not create new level curves when increasing the scale parameter. If this is satisfied, iso-intensity linking through the scales is possible and a structure at a coarse scale can (in principle) be traced back to the original image *(causality)*. For this reason, he required that at spatial extrema with nonvanishing determinant of the Hessian isophotes in scale-space are upwards convex. He showed that this constraint can be written as

$$\operatorname{sign}(\partial_t u) = \operatorname{sign}(\Delta u). \tag{6.18}$$

A sufficient condition for the causality equation (6.18) to hold is requiring that local extrema with positive or negative definite Hessians are not enhanced: an extremum in ξ at scale θ satisfies $\partial_t u > 0$ if ξ is a minimum, and $\partial_t u < 0$ if ξ is a maximum. This implication is easily seen: In the first case, for instance, the eigenvalues $\eta_1,...,\eta_m$ of $\operatorname{Hess}(u)$ are positive. Thus,

$$\Delta u = \operatorname{tr}(\operatorname{Hess}(u)) = \sum_{i=1}^{m} \eta_i > 0, \tag{6.19}$$

which gives just the causality requirement (6.18).

Nonenhancement of local extrema has first been used by Babaud *et al.* [28] in the context of linear diffusion filtering. However, it is also satisfied by nonlinear diffusion scale-spaces, as we shall see now.[3]

Theorem 13 (Nonenhancement of local extrema).
Let u be the unique solution of (P_c) and consider some $\theta > 0$. Suppose that $\xi \in \Omega$ is a local extremum of $u(.,\theta)$ with nonvanishing Hessian. Then,

$$\partial_t u(\xi,\theta) < 0, \quad \textit{if } \xi \textit{ is a local maximum,} \tag{6.20}$$

$$\partial_t u(\xi,\theta) > 0, \quad \textit{if } \xi \textit{ is a local minimum.} \tag{6.21}$$

[3] As in the linear diffusion case, nonenhancement of local extrema generally does not imply that their number is nonincreasing for dimensions $m \geq 2$; cf. [320].

Proof: See Appendix 6.8.2.

Nonenhancement of local extrema is just one possibility to end up with Koenderink's causality requirement. Another way to establish causality is via the extremum principle (6.15) following Hummel's reasoning; see [173] for more details.

Lyapunov Functionals and Behaviour for $t \to \infty$.

Interestingly, causality in terms of nonenhancement of local extrema can be regarded as a first representative of a much larger class of smoothing properties of nonlinear diffusion filtering. Other representatives result from studying the Lyapunov functionals of these processes. Lyapunov functionals are energy-like expressions which decrease during the evolution and which have a lower bound. They can be useful for finding attractors of a process for $t \to \infty$. The theorem below establishes a class of Lyapunov functionals and shows that processes of type (P_c) converge to a flat steady-state which is given by the average grey value of the original image. This is a desirable property for scale-spaces, since such a steady-state can be regarded as the coarsest, most global representation of the original image.

Theorem 14 (Lyapunov functionals and behaviour for $t \to \infty$).
Suppose that u is the solution of (P_c) and let a, b, μ be defined as in (P_c) and (6.16). Then the following properties are valid:

(a) (Lyapunov functionals)
For all $r \in \mathrm{C}^2[a,b]$ which are convex on $[a,b]$ the function

$$V(t) := \Phi(u(t)) := \int_\Omega r(u(x,t))\,dx \tag{6.22}$$

is a Lyapunov functional:

(i) It is bounded from below by the value of Φ for a constant image with the same average grey level:

$$\Phi(u(t)) \geq \Phi(Mf) \quad \forall \geq 0, \tag{6.23}$$

where $(Mf)(x) := \mu$.

(ii) $V(t)$ is continuous in 0, and decreasing for all $t > 0$.

(b) (Convergence)
$u(x,t)$ converges to a constant image with the same average grey value:

$$\lim_{t\to\infty} \|u(t) - Mf\|_{\mathrm{L}^p(\Omega)} = 0 \tag{6.24}$$

for all $p \in [1,\infty)$.

Proof: See Appendix 6.8.3.

What are interesting representatives of this family of Lyapunov functionals? Considering the Lyapunov functions associated with $r(s) := |s|^p$, $r(s) := (s-\mu)^{2n}$ and $r(s) := s \ln s$, respectively, the preceding theorem gives the following corollary.

Corollary 1 (Special Lyapunov functionals).
Let u be the solution of (P_c) and a and μ be defined as in (P_c) and (6.16). Then the following functions are decreasing for $t \in [0, \infty)$:

(a) $\|u(t)\|_{L^p(\Omega)}$ *for all* $p \geq 2$.

(b) $M_{2n}[u(t)] := \frac{1}{|\Omega|} \int_\Omega (u(x,t) - \mu)^{2n}\, dx$
for all $n \in \mathbb{N}$.

(c) $H[u(t)] := \int_\Omega u(x,t) \ln(u(x,t))\, dx,$
if $a > 0$.

Corollary 1 offers multiple possibilities of how to interpret nonlinear anisotropic diffusion filtering as a smoothing transformation.

As a special case of (a) it follows that the energy $\|u(t)\|^2_{L^2(\Omega)}$ is reduced by diffusion. Using Parseval's equality we know that a decreasing energy is also equivalent to a decreasing sum of the squared Fourier coefficients.

Part (b) gives a probabilistic interpretation of anisotropic diffusion filtering. Consider the intensity in an image f as a random variable Z_f with distribution $F_f(z)$, i.e. $F_f(z)$ is the probability that an arbitrary grey value Z_f of f does not exceed z. By the average grey level invariance, μ is equal to the expected value

$$EZ_{u(t)} := \int_{\mathbb{R}} z\, dF_{u(t)}(z), \tag{6.25}$$

and it follows that $M_{2n}[u(t)]$ is just the even central moment

$$\int_{\mathbb{R}} \left(z - EZ_{u(t)}\right)^{2n} dF_{u(t)}(z). \tag{6.26}$$

The second central moment (the variance) characterizes the spread of the intensity about its mean. It is a common tool for constructing measures for the relative smoothness of the intensity distribution. The fourth moment is frequently used to describe the relative flatness of the grey value distribution. Higher moments are more difficult to interpret, although they do provide important information

for tasks like texture discrimination [150, pp. 414–415]. All decreasing even moments demonstrate that the image becomes smoother during diffusion filtering. Hence, local enhancement effects, which object to increase central moments, are overcompensated by smoothing in other areas.

If we choose another probabilistic model of images, then part (c) characterizes the information-theoretical side of our scale-space. Provided the initial image f is strictly positive on Ω, we may regard it also as a two-dimensional density.[4] Then,

$$S[u(t)] \;:=\; -\int_{\Omega} u(x,t)\,\ln(u(x,t))\,dx \tag{6.27}$$

is called the *entropy* of $u(t)$, a measure of uncertainty and missing information [61]. Since anisotropic diffusion filters increase the entropy, the corresponding scale-space embeds the genuine image f into a family of subsequently likelier versions of it which contain less information. Moreover, for $t \to \infty$, the process reaches the state with the lowest possible information, namely a constant image. This information-reducing property indicates that anisotropic diffusion might be generally useful in the context of image compression. In particular, it helps to explain the success of nonlinear diffusion filtering as a preprocessing step for subsampling as observed in [131]. The interpretation of the entropy in terms of Lyapunov functionals carries also over to generalized entropies, see [372] for more details.

From all the previous considerations, we recognize that anisotropic diffusion does really simplify the original image in a steady way.

6.5 Numerical Aspects

To approximate CED numerically, we replace the derivatives by finite differences. Since continuous CED has the structure

$$\partial_t u = \sum_{i,j=1}^{m} \partial_{x_i}(d_{ij}\partial_{x_j} u), \tag{6.28}$$

its simplest discretization is given by the finite difference scheme

$$\frac{U^{k+1} - U^k}{\Delta t} = \sum_{i,j=1}^{m} L_{ij}^k U^k. \tag{6.29}$$

In this notation, U describes a vector containing the values at each pixel. The upper index denotes the time level and L_{ij} is a central difference approximation to the operator $\partial_{x_i}(d_{ij}\partial_{x_j})$.

[4] Without loss of generality we omit the normalization.

Since (6.29) can be rewritten as

$$U^{k+1} = \Big(I + \Delta t \sum_{i,j=1}^{m} L_{ij}^k\Big) U^k \tag{6.30}$$

we observe that U^{k+1} can be calculated explicitly from U^k without any matrix inversions. For this reason it is called an *explicit* scheme.

Unfortunately, such explicit schemes require very small time steps Δt in order to be stable. Therefore, it is desirable to replace (6.30) by an implicit scheme which has the same first-order Taylor expansion in Δt, but better stability properties. One possibility is the *AOS-stabilized scheme*

$$U^{k+1} \;=\; \frac{1}{m} \sum_{l=1}^{m} \Big(I - m\Delta t L_{ll}^k\Big)^{-1} \Big(I + \Delta t \sum_{i=1}^{m} \sum_{j \neq i} L_{ij}^k\Big)\, U^k. \tag{6.31}$$

This method achieves a stabilization through the nonnegative matrices of type $(I - m\Delta t L_{ll}^k)^{-1}$. They describe a semi-implicit discretization of the diffusion caused by the l-th diagonal entry of the diffusion tensor. The typically used step size $\Delta t = 2$ is about one order of magnitude larger than the ones for which an explicit scheme of type (6.29) is stable in 2-D or 3-D. For the standard approximations with central derivatives within a (3×3)-stencil, the matrix inversions in (6.31) come down to solving diagonally dominant tridiagonal systems of linear equations. This can be performed in linear complexity with a modified Gaussian algorithm (Thomas algorithm). It should be noted that (6.31) uses an additive operator splitting (AOS) instead of the usual multiplicative operator splitting from the mathematical literature such as [241]. This guarantees that all axes are treated in the same way. AOS schemes have been introduced in [429] as an efficient and reliable method for isotropic nonlinear diffusion filtering, which perform well on parallel computer architectures [430]. Equation (6.31) shows one way how to extend them to anisotropic processes with a diffusion tensor.

The final CED algorithm is as follows:

1. Calculation of the structure tensor in each pixel. This requires one convolution with a Gaussian K_σ, and $\sum_{k=1}^{m} k = \frac{1}{2} m(m+1)$ convolutions with K_ρ. The convolutions were implemented in the spatial domain exploiting the separability and symmetry of the Gaussian.

2. Principal axis transformation of the structure tensor in each pixel. In our case this was done by a cyclic Jacobi algorithm [352].

3. Calculation of the diffusion tensor in each pixel. Here a significant speed-up is possible by creating a look-up table for the function (6.14) at the beginning of the programme run.

4. Calculation of
$$V^k := \Big(I + \Delta t \sum_{i=1}^{m} \sum_{j \neq i} L_{ij}^k\Big) U^k.$$

5. For $l = 1, ..., m$: calculation of
$$W_l^{k+1} := \Big(I - m\Delta t L_{ll}^k\Big)^{-1} V^k$$
by means of the Thomas algorithm [429].

6. Calculation of
$$U^{k+1} := \frac{1}{m} \sum_{l=1}^{m} W_l^{k+1}.$$

This algorithm is computationally less complex than it looks at first glance: typical execution times for one iteration on a HP 9000/889 are 0.3 seconds for a 256×256 image and 9 seconds for a $64 \times 64 \times 64$ image.

6.6 Examples

Figure 6.2 illustrates the importance of choosing the appropriate smoothing direction in anisotropic processes. In Figure 6.2(b) mean curvature motion is applied to the fingerprint image. We observe that – although mean curvature motion is a completely anisotropic technique with many merits in computer vision – it cannot be used for closing interrupted line-like structures: topologically connected structures remain connected, but gaps become larger, since both ends move in opposite directions. MCM does not exploit any semilocal information.

Figure 6.2(c) shows the effect of coherence-enhancing anisotropic diffusion where α and the integration scale ρ are set to 0. This model is close in spirit of the Cottet–Germain filter [91]. For such a model it is difficult to find a useful noise scale σ for smoothing along flow-like structures: for $\sigma \to 0$, ∇u becomes eigenvector of D and the diffusion is halted, since the corresponding eigenvalue λ_1 is zero. If σ becomes larger, the effect illustrated in Figure 6.1(c) dominates: opposite gradients cancel one another leading to more or less random smoothing directions.

Coherence-enhancing diffusion filtering with a nonvanishing integration scale is depicted in Figure 6.2(d). We observe that diffusion along the coherence orientation w_2 is well-suited for closing interrupted lines in flow-like textures. Due to its reduced diffusivity at noncoherent structures, the location of the semantically important singularity in the fingerprint remains the same. This is an important prerequisite that any image processing method has to satisfy if it is to be applied to fingerprint analysis. In this and all subsequent CED images the parameters $C = 1$, and $\alpha = 0.001$ have been used.

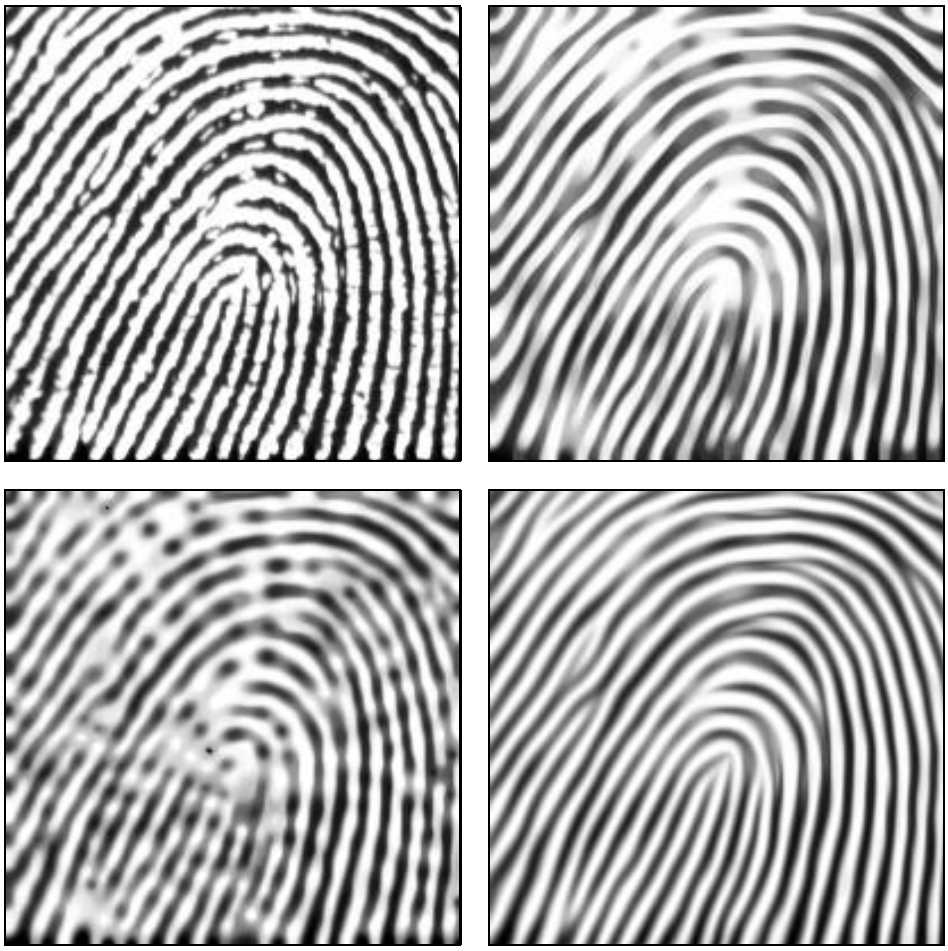


Figure 6.2: Anisotropic equations applied to the fingerprint image. (a) Top Left: Original image. (b) Top Right: Mean curvature motion, $t = 10$. (c) Bottom Left: Anisotropic diffusion similar to the Cottet–Germain model, $\alpha = 0$, $\sigma = 2$, $t = 20$. (d) Bottom Right: Coherence-enhancing anisotropic diffusion, $\sigma = 0.5$, $\rho = 4$, $t = 20$.

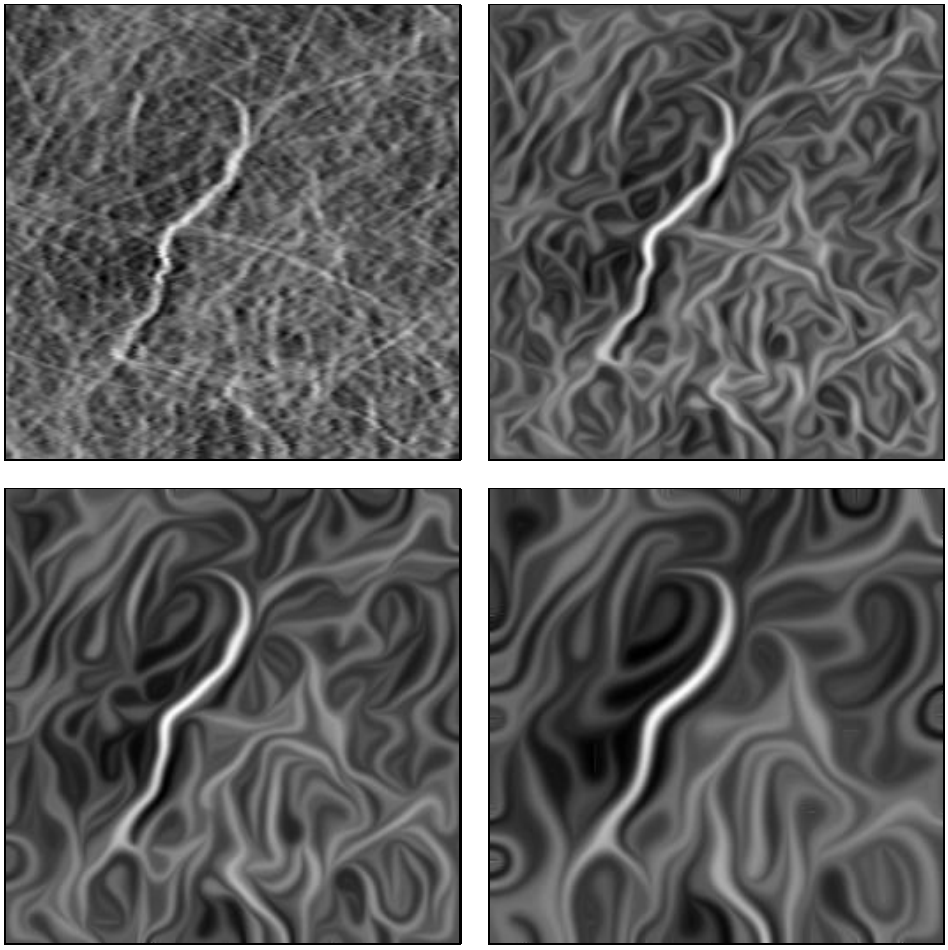


Figure 6.3: Scale-space behaviour of coherence-enhancing diffusion ($\sigma = 0.5$, $\rho = 2$). (a) Top Left: Original fabric image, $\Omega = (0, 257)^2$. (b) Top Right: $t = 20$. (c) Bottom Left: $t = 120$. (d) Bottom Right: $t = 640$.

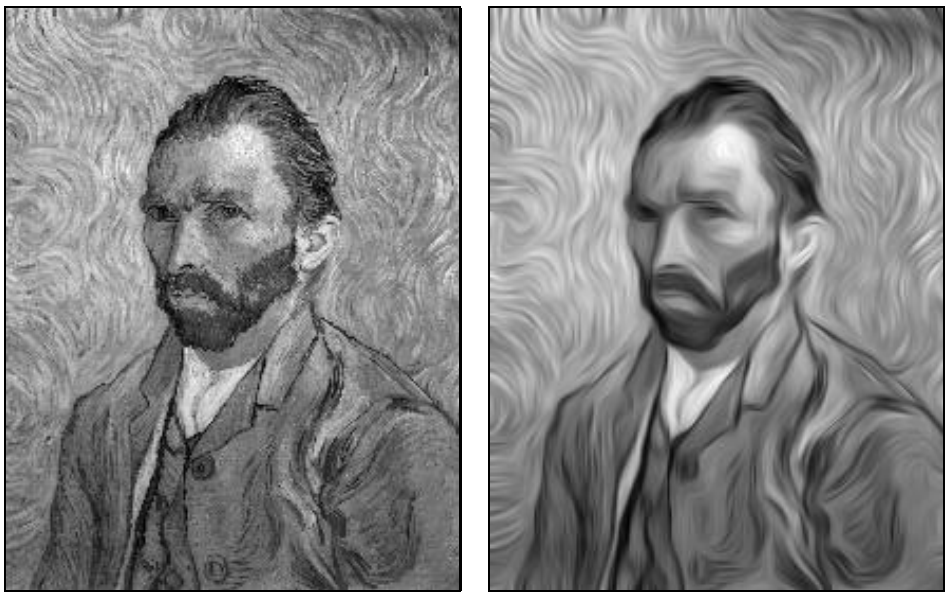


Figure 6.4: Image restoration using coherence-enhancing anisotropic diffusion. (a) LEFT: "Selfportrait" by van Gogh (Saint-Rémy, 1889; Paris, Museé d'Orsay), $\Omega = (0, 215) \times (0, 275)$. (b) RIGHT: Filtered, $\sigma = 0.5$, $\rho = 4$, $t = 6$.

Figure 6.3 depicts the scale-space behaviour of coherence-enhancing anisotropic diffusion applied to a fabric image arising in computer aided quality control. The temporal behaviour of this diffusion filter seems to be appropriate for visualizing coherent fibre agglomerations (stripes) at different scales, a difficult problem for the automatic grading of nonwovens.

Let us now investigate the impact of coherence-enhancing diffusion on images, which are not typical texture images, but still reveal a flow-like character. To this end, we shall process expressionistic paintings by Vincent van Gogh.

Fig. 6.4 shows the restoration properties of coherence-enhancing anisotropic diffusion when being applied to a selfportrait of the artist [391]. We observe that the diffusion filter can close interrupted lines and enhance the flow-like character which is typical for van Gogh paintings.

The next painting we are concerned with is called "Road with Cypress and Star" [393]. It is depicted in Fig. 6.5. In order to demonstrate the influence of the integration scale ρ, all filter parameters are fixed except for ρ. In Fig. 6.5(b) we observe that a value for ρ which is too small does not lead to the visually dominant coherence orientation and, thus, the filtered structures reveal a lot of undesired fluctuations. Increasing the value for ρ improves the image significantly (Fig. 6.5(c)). Interestingly, a further increasing of ρ does hardly alter this result (Fig. 6.5(d)), which indicates that this van Gogh painting possesses a uniform

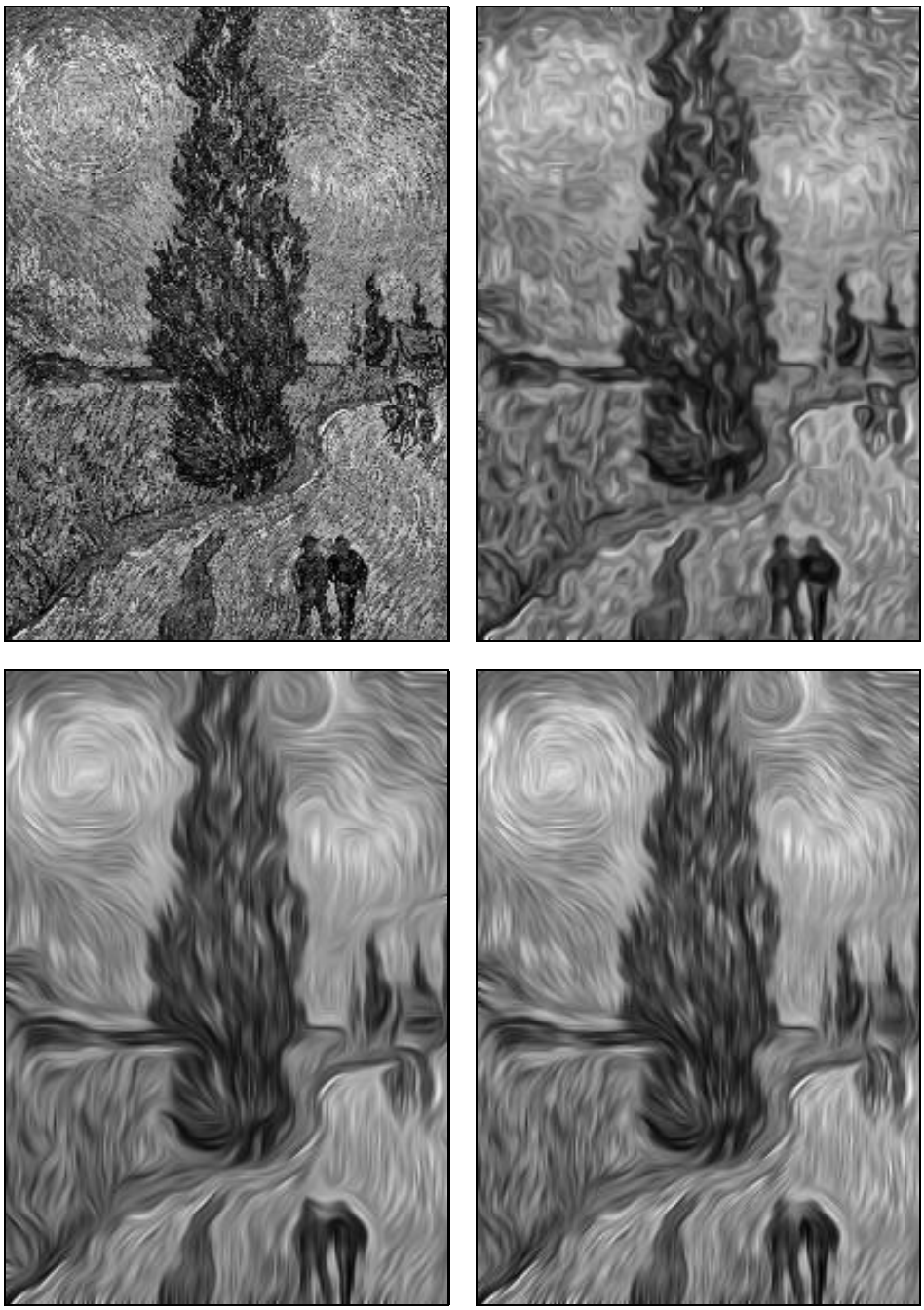


Figure 6.5: Impact of the integration scale on coherence-enhancing anisotropic diffusion ($\sigma = 0.5$, $t = 8$). (a) TOP LEFT: "Road with Cypress and Star" by van Gogh (Auvers-sur-Oise, 1890; Otterlo, Rijksmuseum Kröller–Müller), $\Omega = (0, 203) \times (0, 290)$. (b) TOP RIGHT: Filtered with $\rho = 1$. (c) BOTTOM LEFT: $\rho = 4$. (d) BOTTOM RIGHT: $\rho = 6$.

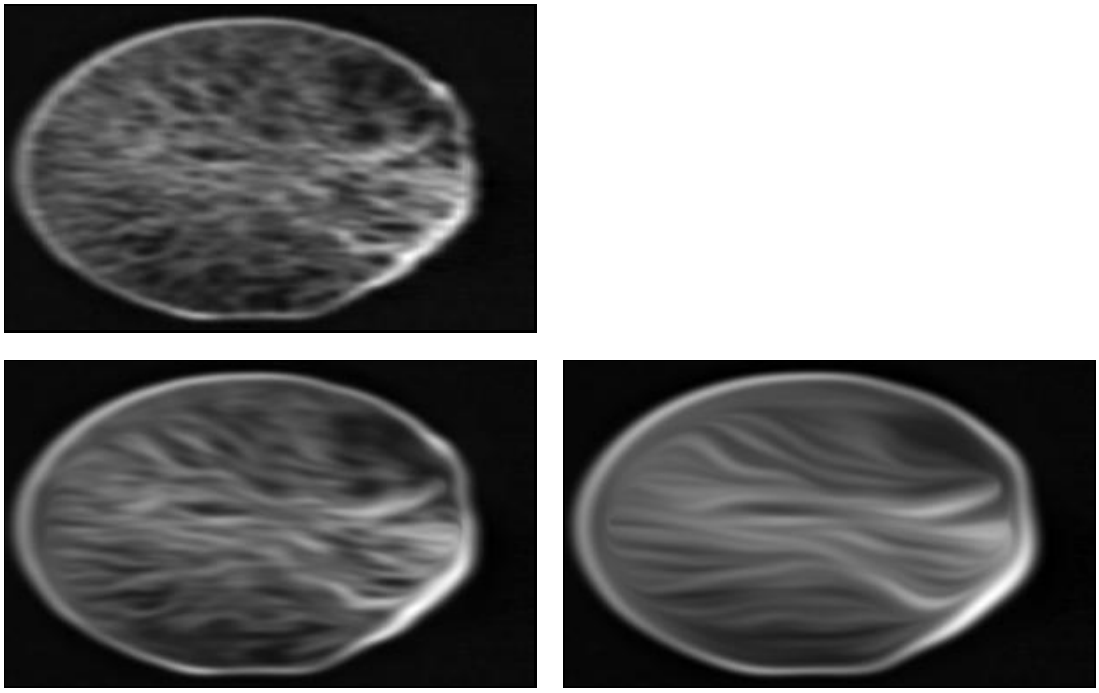


Figure 6.6: (a) TOP: High resolution slipring CT scan of a femural bone, showing the trabecular formation. Slice thickness 1 mm, field of view 6 cm. $\Omega = (0, 300) \times (0, 186)$. (b) BOTTOM LEFT: Filtered by coherence-enhancing anisotropic diffusion, $\sigma = 0.5$, $\rho = 6$, $t = 16$. (c) BOTTOM RIGHT: Ditto with $t = 128$.

"texture scale" reflecting the characteristic painting style of the artist.

Figure 6.6 illustrates the potential of CED for medical applications. It depicts a human bone. Its internal structure has a distinctive texture through the presence of tiny elongated bony structural elements, the *trabeculae*. There is evidence that the trabecular formation is for a great deal determined by the external load [438, 249, 29]. For this reason the trabecular structure constitute an important clinical parameter in orthopedics. Examples are the control of recovery after surgical procedures, such as the placement or removal of metal implants, quantifying the rate of progression of rheumatism and osteoporosis, the determination of left-right deviations of symmetry in the load or establishing optimal load corrections for physiotherapy. The high resolution of contemporary CT using slipring technology and MR scanners now enables in vivo analysis of detailed trabecular structure.

While in vivo analysis of the trabecular bone by means of a structure tensor is investigated in [384, 275], the present paper points out ways to enhance the coherence information in medical images, in order to ease such an analysis.

From Figure 6.6(b),(c) we observe that CED is indeed capable of closing interrupted lines. Parallel flow-like structures are enhanced and a subsequent coherence

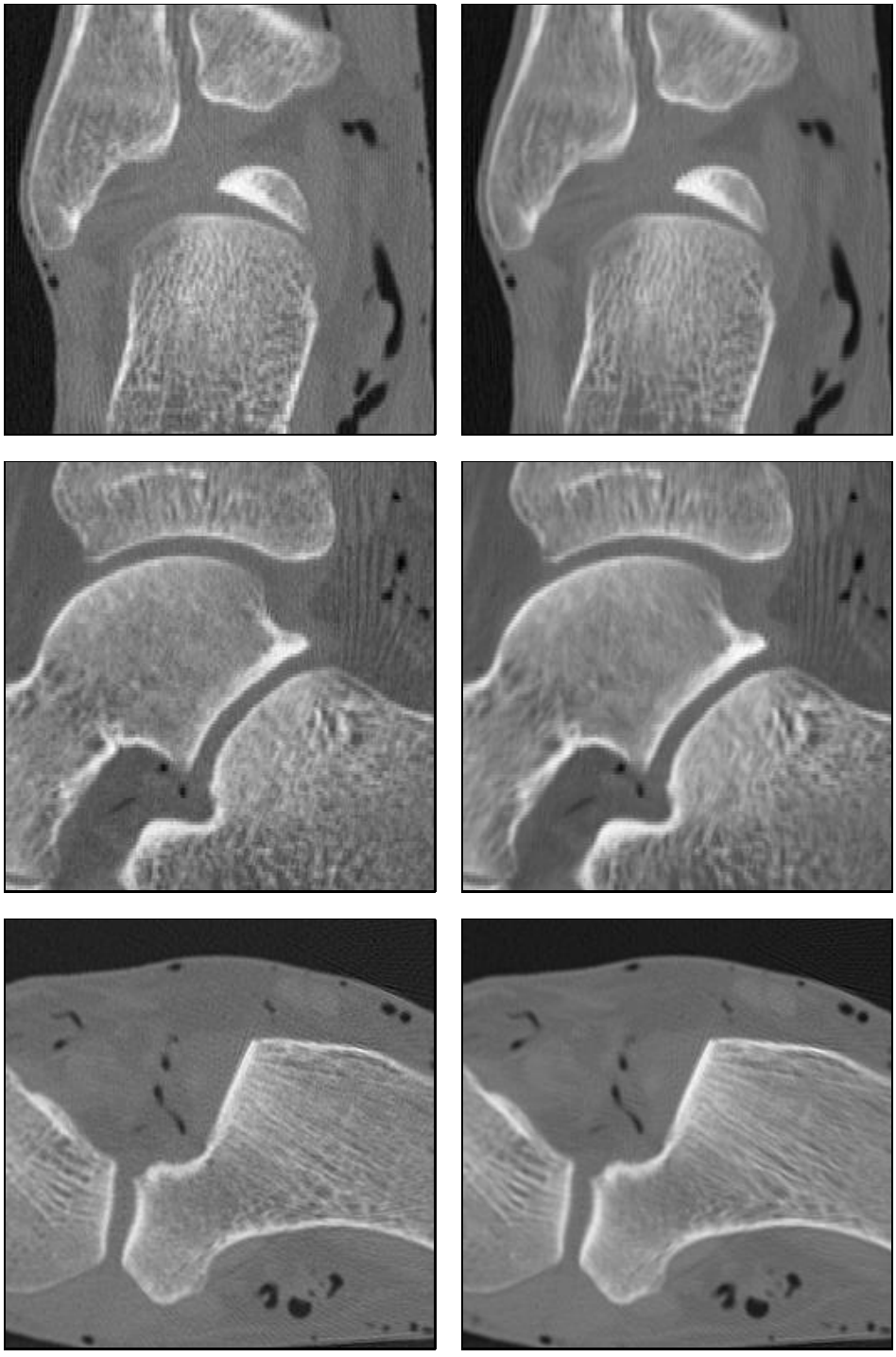


Figure 6.7: Three-dimensional CED applied to a CT data set of size $256 \times 256 \times 128$. LEFT COLUMN: 2-D sections depicting slice no. 180 in x, no. 128 in y, and no. 100 in z, respectively. RIGHT COLUMN: Filtered, $\sigma = 0.5$, $\rho = 6$, $t = 4$.

analysis becomes much easier. Increasing the time t gives a coarser representation of the coherence. The entire evolution creates a task-driven scale-space which is dedicated to flow-like patterns.

In medical applications physicians often insist in only very small amounts of smoothing, in order to reduce some noise without affecting clinically relevant structures too much. Simplifications as in Figure 6.6(c) would be considered as misleading. Figure 6.7 shows the results for two iterations of three-dimensional CED filtering. The data set reveals a size of $256 \times 256 \times 128$, and the voxel dimensions are $0.25 \times 0.25 \times 0.5\,\mathrm{mm}^3$. It depicts a CT scan of a foot area by means of a Philips SR8000 slipring CT at 140 kV. Even for 3-D data sets, such a slight preprocessing using only a few iterations can easily be performed within the acquisition time. The price of a suitable hardware would be a small fraction of the price for a CT scanner.

It should be noted that in some 3-D applications it might also be desirable to smooth along two directions instead of one as in CED. Diffusion methods of this type have been studied by Rambaux and Garçon [312]. They are 3-D versions of the edge-enhancing anisotropic diffusion filter from [408].

6.7 Summary and Conclusions

In the present paper we have treated the problem of enhancing flow-like patterns. For such tasks a reliable measurement of local orientation is needed. Our experiments demonstrate that the structure tensor satisfies this requirement. Unlike many other applications, we do not restrict its application to pure image *analysis*, we use it as a tool for *steering* a scale-space evolution.

To this end, we evolve the original image by means of a nonlinear anisotropic diffusion equation. Its diffusion tensor reflects the local image structure by using the same set of eigenvectors as the structure tensor. The eigenvalues are chosen in such a way that diffusion acts mainly along the direction with the highest coherence, and becomes stronger when the coherence increases. The resulting coherence-enhancing diffusion (CED) process gives a theoretically well-founded scale-space representation: proofs have been presented which show that its unique solution is stable under perturbations of the initial image, satisfies Koenderink's causality requirement, and creates a large family of Lyapunov functionals which ensure that the process is image simplifying and converges to a constant steady-state.

This theoretical framework is valid in any dimension. We have also presented an m-dimensional numerical scheme where a stabilization based on additive operator splitting (AOS) allows time steps which are one order of magnitude larger than in the explicit case.

Finally, the use of CED was illustrated by applying it to 2-D and 3-D data

sets. The results indicate a variety of possible application areas ranging from computer aided quality control over fingerprint enhancement to medical imaging. This, however, gives only one aspect of the filter concept. Coherence-enhancing diffusion allows also generalizations to vector-valued images. Details in this direction can be found in [416].

Acknowledgments. The author thanks Robert Maas and Stiliyan Kalitzin for useful comments on a draft version of this paper, and Wiro Niessen for providing the 3-D trabecular bone data set.

6.8 Appendix

6.8.1 Proof of Theorem 12

(a) **Existence, Uniqueness and Regularity**
Existence, uniqueness and regularity are straightforward anisotropic extensions of the proof for the isotropic case studied by Catté, Lions, Morel and Coll [74], and the extremum principle follows from Stampacchia's truncation method (cf. [52], p. 211); see [412] for more details.

(b) **Continuous Dependence on the Initial Image**
In order to discuss the continuous dependence on f, let us first define some useful notations. Let $\mathrm{H}^1(\Omega)$ be the Sobolev space of functions $u(x) \in \mathrm{L}^2(\Omega)$ with all distributional derivatives of first order being in $\mathrm{L}^2(\Omega)$. We equip $\mathrm{H}^1(\Omega)$ with the norm

$$\|u\|_{\mathrm{H}^1(\Omega)} := \Big(\|u\|^2_{\mathrm{L}^2(\Omega)} + \sum_{i=1}^{m} \|\partial_{x_i} u\|^2_{\mathrm{L}^2(\Omega)} \Big)^{1/2} \tag{6.32}$$

and identify it with its dual space. Let $\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))$ be the space of functions u, strongly measurable on $[0,T]$ with range in $\mathrm{H}^1(\Omega)$ (for the Lebesgue measure dt on $[0,T]$) such that

$$\|u\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))} := \Big(\int_0^T \|u(t)\|^2_{\mathrm{H}^1(\Omega)}\, dt \Big)^{1/2} < \infty. \tag{6.33}$$

The first part of the proof is similar to the uniqueness proof in [74]. Let $f, h \in \mathrm{L}^\infty(\Omega)$ be two initial values and u, w the corresponding solutions. Then one shows that

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\|u(t)-w(t)\|^2_{\mathrm{L}^2(\Omega)} \;\leq\; & \|\nabla u(t)-\nabla w(t)\|_{\mathrm{L}^2(\Omega)} \cdot (\|D(J_\rho(\nabla u_\sigma(t)))) \\ & - D(J_\rho(\nabla w_\sigma(t)))\|_\infty)_{\mathrm{L}^\infty(\Omega)} \cdot \|\nabla u(t)\|_{\mathrm{L}^2(\Omega)} \\ & - \nu \|\nabla u(t)-\nabla w(t)\|^2_{\mathrm{L}^2(\Omega)} \end{aligned} \tag{6.34}$$

where $\|B\|_\infty := \max_{1=1..m} \sum_{j=1}^{m} |b_{ij}|$ for $B = (b_{ij}) \in \mathbb{R}^{m\times m}$, and ν denotes the lower positive bound for the eigenvalues of D.

In order to apply the Gronwall–Bellman lemma, we have to estimate

$$(\|D(J_\rho(\nabla u_\sigma(t))) - D(J_\rho(\nabla w_\sigma(t)))\|_\infty)_{\mathrm{L}^\infty(\Omega)}$$

by $\|u(t)-w(t)\|_{\mathrm{L}^2(\Omega)}$. By the boundedness of ∇u_σ and ∇w_σ, and the smoothness property (C1) we know that there exists a Lipschitz constant L such that

$$\begin{aligned} &(\|D(J_\rho(\nabla u_\sigma(t))) - D(J_\rho(\nabla w_\sigma(t)))\|_\infty)_{\mathrm{L}^\infty(\Omega)} \\ &\leq\ L \operatorname*{ess\,sup}_{x\in\Omega} \|\nabla u(t) - \nabla w(t)\|_2 \end{aligned} \tag{6.35}$$

By iteratively reflecting Ω at its boundaries we get a partitioning $\mathbb{R}^m = \dot{\bigcup}_i \Omega_i$. Using Jensen's inequality, our definition of convolution on a finite domain Ω, and this partitioning we obtain

$$\begin{aligned} &(\|D(J_\rho(\nabla u_\sigma(t))) - D(J_\rho(\nabla w_\sigma(t)))\|_\infty)_{\mathrm{L}^\infty(\Omega)} \\ &\leq\ L \cdot \operatorname*{ess\,sup}_{x\in\Omega} \Big(\int_{\mathbb{R}^m} \|\nabla K_\sigma(y)\|_2^2 \, ((\tilde{u} - \tilde{w})(x-y))^2 \, dy \Big)^{1/2} \\ &\leq\ L \cdot \Big(\sum_i \max_{z\in\bar{\Omega}_i} \|\nabla K_\sigma(z)\|_2^2 \Big)^{1/2} \|u(t) - w(t)\|_{\mathrm{L}^2(\Omega)} \\ &=\ c \cdot \|u(t) - w(t)\|_{\mathrm{L}^2(\Omega)} \end{aligned} \tag{6.36}$$

with some finite constant c. Plugging (6.36) into (6.34) gives

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}\|u(t)-w(t)\|_{\mathrm{L}^2(\Omega)}^2 \\ &\leq \underbrace{\tfrac{c}{\sqrt{2\nu}}\|\nabla u(t)\|_{\mathrm{L}^2(\Omega)}\|u(t)-w(t)\|_{\mathrm{L}^2(\Omega)}}_{=:a} \underbrace{\sqrt{2\nu}\|\nabla u(t)-\nabla w(t)\|_{\mathrm{L}^2(\Omega)}}_{=:b} \\ &\quad -\nu\,\|\nabla u(t)-\nabla w(t)\|_{\mathrm{L}^2(\Omega)}^2. \end{aligned}$$

To get at the right hand side a quadratic expression in $\|\nabla u(t)-\nabla w(t)\|_{\mathrm{L}^2(\Omega)}$, we use $|ab| \leq \frac{a^2}{2} + \frac{b^2}{2}$ and end up with

$$\frac{d}{dt}\Big(\|u(t)-w(t)\|_{\mathrm{L}^2(\Omega)}^2\Big) \ \leq\ \frac{c^2}{2\nu} \cdot \|\nabla u(t)\|_{\mathrm{L}^2(\Omega)}^2 \cdot \|u(t)-w(t)\|_{\mathrm{L}^2(\Omega)}^2.$$

Applying the Gronwall–Bellman lemma [51, pp. 156–157] yields

$$\|u(t)-w(t)\|_{\mathrm{L}^2(\Omega)}^2 \ \leq\ \|f-h\|_{\mathrm{L}^2(\Omega)}^2 \cdot \exp\left(\frac{c^2}{2\nu} \cdot \int_0^t \|\nabla u(s)\|_{\mathrm{L}^2(\Omega)}^2 \, ds\right).$$

Next we show that $\int_0^t \|\nabla u(s)\|^2_{\mathrm{L}^2(\Omega)}\,ds$ is finite:

$$\begin{aligned}
\int_0^t \|\nabla u(s)\|^2_{\mathrm{L}^2(\Omega)}\,ds &\leq \int_0^T \|\nabla u(s)\|^2_{\mathrm{L}^2(\Omega)}\,ds \\
&\leq \frac{1}{\nu}\int_0^T \left| \int_\Omega \langle \nabla u(x,s), D\nabla u(x,s)\rangle\,dx \right| ds \\
&= \frac{1}{\nu}\int_0^T \left| \int_\Omega u(x,s)\cdot \operatorname{div}\Big(D\nabla u(x,s)\Big)\,dx \right| ds \\
&\leq \frac{1}{\nu}\int_0^T \|u(s)\|_{\mathrm{L}^2(\Omega)}\,\|\partial_t u(s)\|_{\mathrm{L}^2(\Omega)}\,ds \\
&\leq \frac{1}{\nu}\,\|u\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))}\,\|\partial_t u\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))}.
\end{aligned}$$

By virtue of the existence proof in [74], we know that the right-hand of this estimate exists. Now, let $\epsilon > 0$ and choose

$$\delta := \epsilon\cdot\exp\Big(\frac{-c^2}{2\nu^2}\,\|u\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))}\cdot\,\|\partial_t u\|_{\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))}\Big).$$

Then for $\|f-h\|_{\mathrm{L}^2(\Omega)} < \delta$, the preceding results imply

$$\|u(t)-w(t)\|_{\mathrm{L}^2(\Omega)} < \epsilon \qquad \forall\, t\in[0,T],$$

which proves the continuous dependence on the initial data. □

6.8.2 Proof of Theorem 13

Let $D(J_\rho(\nabla u_\sigma)) =: (d_{ij}(J_\rho(\nabla u_\sigma)))$. Then we have

$$\begin{aligned}
\partial_t u &= \sum_{i=1}^m\sum_{j=1}^m \Big(\partial_{x_i} d_{ij}(J_\rho(\nabla u_\sigma))\Big)\,\partial_{x_j} u \\
&+ \sum_{i=1}^m\sum_{j=1}^m d_{ij}(J_\rho(\nabla u_\sigma))\,\partial_{x_i x_j} u. \qquad (6.37)
\end{aligned}$$

Since $\nabla u(\xi,\theta) = 0$ and $\partial_{x_i} d_{ij}(J_\rho(\nabla u_\sigma(\xi,\theta)))$ is bounded, the first term of the right-hand side of (6.37) vanishes in (ξ,θ).

We know that the diffusion tensor $D := D(J_\rho(\nabla u_\sigma(\xi,\theta)))$ is positive definite. Hence, there exists an orthogonal transformation $S \in \mathbb{R}^{m\times m}$ such that

$$S^T D S \;=\; \mathrm{diag}(\lambda_1,...,\lambda_m) \;=:\; \Lambda$$

with λ_1,...,λ_m being the positive eigenvalues of D.

Now, let us assume that (ξ,θ) is a local maximum where the Hessian $H := \mathrm{Hess}(u(\xi,\theta))$ and, thus, $B := (b_{ij}) := S^T H S$ are negative definite. Then we have

$$b_{ii} < 0 \quad (i = 1,...,m),$$

and by the invariance of the trace with respect to orthogonal transformations it follows that

$$\begin{aligned}
\partial_t u(\xi,\theta) &= \mathrm{tr}\,(DH) \\
&= \mathrm{tr}\,(S^T D S\, S^T H S) \\
&= \mathrm{tr}\,(\Lambda B) \\
&= \sum_{i=1}^{m} \lambda_i\, b_{ii} \\
&< 0.
\end{aligned}$$

If ξ is a local minimum of $u(x,\theta)$, one proceeds in the same way utilizing the positive definiteness of the Hessian. □

6.8.3 Proof of Theorem 14

(a) (i) Let $r \in \mathrm{C}^2[a,b]$ be convex on $[a,b]$. Using the average grey level invariance and Jensen's inequality we obtain, for all $t \geq 0$,

$$\begin{aligned}
\Phi(Mf) &= \int_\Omega r\left(\frac{1}{|\Omega|}\int_\Omega u(x,t)\,dx\right) dy \\
&\leq \int_\Omega \left(\frac{1}{|\Omega|}\int_\Omega r(u(x,t))\,dx\right) dy \\
&= \int_\Omega r(u(x,t))\;dx \\
&= \Phi(u(t)). \qquad (6.38)
\end{aligned}$$

(ii) Let us start by proving the continuity of $V(t)$ in 0. Thanks to the maximum–minimum principle, we may choose a constant

$$L := \max_{s\in[a,b]} |r'(s)|$$

such that for all $t > 0$, the Lipschitz condition

$$|r(u(x,t)) - r(f(x))| \;\leq\; L\,|u(x,t) - f(x)|$$

is verified almost everywhere on Ω. From this and the Cauchy–Schwarz inequality, we get

$$\begin{aligned} |V(t) - V(0)| &\leq |\Omega|^{1/2}\,\|r(u(t)) - r(f)\|_{\mathrm{L}^2(\Omega)} \\ &\leq |\Omega|^{1/2}\,L\,\|u(t) - f\|_{\mathrm{L}^2(\Omega)}. \end{aligned}$$

From [74] we know that $u \in \mathrm{C}([0,T];\mathrm{L}^2(\Omega))$. Thus, the limit $t \to 0^+$ gives the announced continuity in 0.
By Theorem 12 and the boundedness of r' on $[a,b]$, we know that V is differentiable for $t > 0$ and $V'(t) = \int_\Omega r'(u)\,u_t\,dx$. Thus, the divergence theorem yields

$$\begin{aligned} V'(t) &= \int_\Omega r'(u)\,\mathrm{div}\,(D\nabla u)\,dx \\ &= \int_\Gamma r'(u)\,\underbrace{\langle D\nabla u, n\rangle}_{=0}\,dS \\ &\quad - \int_\Omega \underbrace{r''(u)}_{\geq 0}\,\underbrace{\langle \nabla u, D\nabla u\rangle}_{\geq 0}\,dx \\ &\leq 0. \end{aligned}$$

(b) By the grey level shift invariance we know that $v := u - Mf$ satisfies the diffusion equation as well. We multiply this equation by v, integrate, and use the divergence theorem to obtain

$$\int_\Omega v v_t\,dx \;=\; -\int_\Omega \langle \nabla v, D(J_\rho(\nabla v_\sigma))\nabla v\rangle\,dx.$$

Since ∇v_σ is bounded, we can find some $\nu > 0$ such that

$$\frac{1}{2}\frac{d}{dt}\left(\|v\|^2_{\mathrm{L}^2(\Omega)}\right) \;\leq\; -\nu\,\|\nabla v\|^2_{\mathrm{L}^2(\Omega)}.$$

For $t > 0$, there exists some x_0 with $v(x_0) = 0$. Therefore, we may apply Poincaré's inequality (cf. [7, p. 122]), which tells us that

$$\|v\|^2_{\mathrm{L}^2(\Omega)} \;\leq\; C_0\,\|\nabla v\|^2_{\mathrm{L}^2(\Omega)}$$

with some constant $C_0 = C_0(\Omega) > 0$. This yields

$$\frac{d}{dt}\,\|v\|^2_{\mathrm{L}^2(\Omega)} \;\leq\; -2\nu\,C_0\,\|v\|^2_{\mathrm{L}^2(\Omega)}$$

and hence the exponential decay of $\|v\|_{\mathrm{L}^2(\Omega)}$ to 0.

By the maximum principle, we know that $\|v(t)\|_{\mathrm{L}^\infty(\Omega)}$ is bounded by $\|f - Mf\|_{\mathrm{L}^\infty(\Omega)}$. Thus, for $q \in \mathbb{N}$, $q \geq 2$, we get

$$\|v(t)\|_{L^q(\Omega)}^q \;\leq\; \|f - Mf\|_{\mathrm{L}^\infty(\Omega)}^{q-2} \cdot \|v(t)\|_{\mathrm{L}^2(\Omega)}^2 \;\to\; 0,$$

and, for $1 \leq p < q < \infty$, Hölder's inequality gives

$$\|v(t)\|_{L^p(\Omega)} \;\leq\; |\Omega|^{(1/p)-(1/q)} \cdot \|v(t)\|_{L^q(\Omega)} \;\to\; 0.$$

This proves the assertion. □

Chapter 7

Coherence-Enhancing Diffusion of Colour Images

Joachim Weickert.
Image and Vision Computing,
Vol. 17, No. 3-4, 201–212, March 1999.

Abstract

Many image processing problems require the enhancement of coherent flow-like structures. This can be accomplished in a natural way by combining anisotropic diffusion filtering and texture analysis by means of the structure tensor (second-moment matrix, interest operator). In this paper an extension of these ideas to vector-valued images is presented. A structure tensor for vector-valued images is constructed as the mean of the structure tensors of each channel. This idea generalizes Di Zenzo's gradient for colour images by introducing an additional integration scale. The common structure tensor is used for steering the diffusion processes in each channel. The additional integration scale turns out to be crucial for orientation smoothing and it leads to significantly improved filter results. After analysing the role of all filter parameters, examples from different application areas are presented, and it its demonstrated that this type of diffusion filtering is highly robust under additive Gaussian noise.

7.1 Introduction

Flow-like structures appear in many image processing problems, for instance the automatic assessment of wood surfaces or fabrics, fingerprint analysis, or scientific image processing in oceanography [197]. Most humans consider flow-like structures as pleasant, harmonic, or at least interesting. Thus, it is not surprising that also artists like van Gogh or Munch have emphasized these features in their paintings.

Some images containing flow-like patterns are of poor quality, such that it becomes necessary to enhance them by closing interrupted lines[1]. Moreover, in many "natural" images, colour plays an important role and it would be desirable to make use of this additional information. Vector-valued image processing material with coherent structures, however, can also arise in meteorology and oceanography, if information is collected at different channels corresponding to different wavelengths.

In the present paper we shall address the problem of enhancing such flow-like structures in vector-valued images. To this end, we first have to *analyse* their coherence. This is done by generalizing a well-established tool from texture analysis, the structure tensor (second-moment matrix, interest operator) to vector-valued images. Its eigenvectors and eigenvalues provide us with all required information. This coherence descriptor enables us to construct a diffusion tensor which steers the diffusion process in each channel in such a way that diffusion is encouraged along the preferred structure orientation.

The paper is organized as follows. Section 7.2 gives a review of the structure tensor concept and it presents its generalization to vector-valued images. Then we discuss in Section 7.3 how this information is used for designing an appropriate diffusion process. Its parameters are analysed in Section 7.4, and its properties are illustrated by several examples in Section 7.5. We conclude with a summary in Section 7.6.

Related work. The work presented here makes a synthesis of two earlier ideas of the author: Using a common structure tensor for diffusing vector-valued images was proposed in [405], while scalar-valued coherence-enhancing anisotropic diffusion goes back to [406]. Its distinctive feature is the combination of colour texture processing based on structure tensor analysis with a nonvanishing integration scale.

This concept also generalizes work of Whitaker and Gerig [432] for isotropic diffusion of vector-valued images with a common scalar diffusivity to the anisotropic case with a common diffusion tensor. Recently, different techniques which are based on partial differential equations (PDEs) have been applied to colour images [50, 75, 210, 308, 332, 355]. Among them, the ones by Chambolle [75] and Sapiro and Ringach [332] reveal the most structural similarities to our approach. How-

[1] This problem can even be related to finding illusory contours.

ever, their goal is edge-preserving smoothing, while the method presented here achieves coherence-enhancement by looking at an additional, much larger scale: the integration scale. We shall see that this is a very essential model feature which guarantees stable orientation estimates. This additional integration scale is also the main difference to the anisotropic diffusion approach of Cottet and Germain [91] for scalar images.

Related anisotropic techniques which convolve with a Gaussian, whose shape is adapted to a second moment matrix, have been pioneered by Nitzberg and Shiota [280], and further pursued by Lindeberg and Garding [232] and Yang *et al.* [440]. Unlike our approach they do not utilize the difference of the eigenvalues of the structure tensor as a coherence measure, and they focus on scalar images. In this sense they are more similar to the methods presented in [408], but they are not equivalent to a diffusion process in divergence form, which is capable of preserving the average grey value.

PDE methods for texture smoothing or enhancement often create a feature vector image by means of Gabor techniques, apply a nonlinear vector-valued diffusion to this feature image, and assemble the filtered texture from the processed features. Techniques in this spirit have been proposed by Whitaker and Gerig [432], Rubner and Tomasi [324], and Kimmel *et al.* [210]. Although these methods make use of generalized gradients for vector-valued images in the sense of Kreyzsig [216] and Di Zenzo [107], they do not take into account the integration scale. Recently, a PDE-based smoothing method which utilizes Gabor filters and which works directly on the image (instead of the feature vector image) has been studied by Carmona and Zhong [66]. Most of the preceding texture processing methods were designed for grey-scale images, and techniques such as [324, 210, 66] do not necessarily lead to diffusion processes in divergence form.

Since our method works on colour textures and may serve to create visually more pleasant structures with increased coherence, it is not surprising that it gives perceptually similar results as the line integral convolution method by Cabral and Leedom [63], which has become a popular tool in computer graphics. A preliminary version of the present paper can be found in a proceedings volume [409].

7.2 Analysing Coherent Structures

For analysing coherent flow-like structures, we first focus on scalar images. Consider a rectangular image domain $\Omega := (0, a_1) \times (0, a_2)$, and let an image $u(x)$ be represented by a bounded mapping $u : \Omega \to \mathbb{R}$.

A very simple structure descriptor is given by ∇u_σ, the gradient of a Gaussian-smoothed version of u:

$$K_\sigma(x) \quad := \quad \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{|x|^2}{2\sigma^2}\right), \tag{7.1}$$

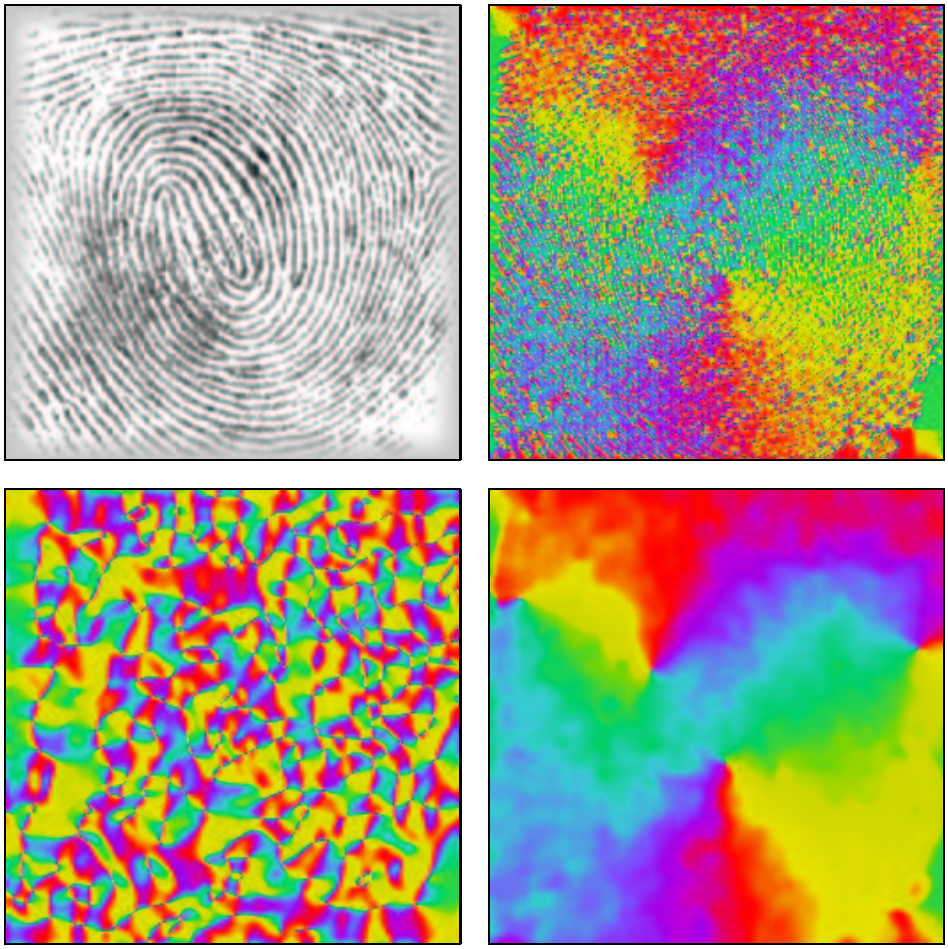


Figure 7.1: Local orientation in a fingerprint image. (a) TOP LEFT: Original fingerprint, $\Omega = (0, 256)^2$. (b) TOP RIGHT: Orientation of smoothed gradient, $\sigma = 0.5$. (c) BOTTOM LEFT: Orientation of smoothed gradient, $\sigma = 5$. (d) BOTTOM RIGHT: Structure tensor orientation, $\sigma = 0.5$, $\rho = 4$. Adapted from [406].

$$u_\sigma(x,t) \quad := \quad (K_\sigma * u(.,t))\,(x) \qquad (\sigma > 0). \tag{7.2}$$

The standard deviation σ denotes the *noise scale*, since it makes the edge detector ignorant of details smaller than $O(\sigma)$. Although ∇u_σ is useful for detecting edges, it is unsuited for finding parallel structures, as we can see from Figure 7.1.

The left image shows an original fingerprint. Figure 7.1(b),(c) depict the gradient orientation using colours: vertical gradients are depicted in red, horizontal ones in green etc. We observe that for small σ high fluctuations remain, while larger σ lead to entirely useless results. This is due to the fact that for larger σ neighbouring gradients with same orientation, but opposite sign cancel each

other. Gradient smoothing averages directions instead of orientations[2]. To make the structure descriptor invariant under sign changes, we may replace ∇u_σ by its tensor product

$$J_0(\nabla u_\sigma) := \nabla u_\sigma \nabla u_\sigma^T. \tag{7.3}$$

This matrix is symmetric and positive semidefinite, and its eigenvectors are parallel and orthogonal to ∇u_σ, respectively. The corresponding eigenvalues $|\nabla u_\sigma|^2$ and 0 describe the contrast in the eigendirections. Now that we have replaced directions by orientations, we can average the orientations by applying a componentwise convolution with a Gaussian K_ρ:

$$J_\rho(\nabla u_\sigma) := K_\rho * (\nabla u_\sigma \nabla u_\sigma^T) \qquad (\rho \geq 0). \tag{7.4}$$

This matrix is named *structure tensor*, *interest operator* or *second-moment matrix.* It is useful for many different tasks, for instance for analysing flow-like textures [313], corners and T-junctions [132, 161, 280], shape cues [230] and spatio–temporal image sequences [197]. Equivalent approaches may also be found in [200, 42]. A book by Jähne [197] gives a nice overview of these methods and clarifies their equivalence.

It is not hard to verify that the symmetric matrix $J_\rho = \left(\begin{smallmatrix} j_{11} & j_{12} \\ j_{12} & j_{22} \end{smallmatrix}\right)$ is positive semidefinite. Its eigenvalues can be calculated as

$$\mu_{1,2} = \frac{1}{2}\left(\mathrm{tr}(J_\rho) \pm \sqrt{\mathrm{tr}^2(J_\rho) - 4\det(J_\rho)}\right), \tag{7.5}$$

where $\mathrm{tr}(J_\rho) = j_{11} + j_{22}$, $\det(J_\rho) = j_{11}j_{22} - j_{12}^2$, and $\mu_1 \geq \mu_2$. The corresponding orthonormal set of eigenvectors $\{w_1, w_2\}$ is given by $w_1 = (\cos\phi, \sin\phi)^T$, where ϕ satisfies

$$\tan(2\phi) = \frac{2j_{12}}{j_{11} - j_{22}}. \tag{7.6}$$

The eigenvalues integrate the variation of the grey values within a neighbourhood of size $O(\rho)$. They describe the average contrast in the eigendirections. Thus, the *integration scale* ρ should reflect the characteristic size of the texture. Usually, it is large in comparison to the noise scale σ. The eigenvector w_2 corresponds to the smaller eigenvalue μ_2. It is the orientation with the lowest fluctuations, the so-called *coherence orientation.*

Figure 7.1(d) depicts this direction. We observe that it is exactly the desired average orientation. It should also be noted how well the singularities correspond to the singularities in the original fingerprint image.

[2]In our terminology, gradients with opposite sign share the same orientation, but point in opposite directions.

Not only the eigenvectors, but also the eigenvalues provide useful information. Constant areas are characterized by $\mu_1 = \mu_2 = 0$, straight edges give $\mu_1 \gg \mu_2 = 0$, and corners yield $\mu_1 \geq \mu_2 \gg 0$. The expression

$$\kappa \;:=\; (\mu_1 - \mu_2)^2 \;=\; (j_{11} - j_{22})^2 + 4j_{12}^2 \tag{7.7}$$

becomes large for anisotropic structures. It measures the *coherence* within a window of scale ρ, and its value can range from 0 to ∞.

The reason why we prefer this coherence measure over the popular *normalized coherence measure*

$$\tilde{\kappa} \;:=\; \frac{(\mu_1 - \mu_2)^2}{(\mu_1 + \mu_2)^2} \;=\; \frac{(j_{11} - j_{22})^2 + 4j_{12}^2}{(j_{11} + j_{22})^2}, \tag{7.8}$$

which attains values between 0 and 1, is that the normalized coherence measure is discontinuous for $\mu_1 = \mu_2 = 0$. This may lead to unreliable coherence estimates in flat image regions. A simple example can illustrate the problem: if $\mu_1 = \mu_2 \downarrow 0$, then $\tilde{\kappa} \to 0$, while $\mu_2 = 0$ and $\mu_1 \downarrow 0$ lead to $\tilde{\kappa} \to 1$. The latter limit also contradicts the intuition that a homogeneous region should reveal no anisotropy.

Now that we have analysed the structure tensor for scalar-valued images, we can draw our attention to the vector-valued case. We denote a vector-valued image by $\vec{u} : \Omega \to \mathbb{R}^m$ and its channels by u_i, $i = 1,...,m$. Di Zenzo [107] proposed an edge detector for vector images by considering the eigenvalues and eigenvectors of

$$\sum_{i=1}^{m} \nabla u_i (\nabla u_i)^T. \tag{7.9}$$

If we extend his idea by introducing a noise and integration scale, this comes down to averaging the structure tensors of each channel in a common structure tensor

$$J_\rho(\nabla \vec{u}_\sigma) \;:=\; \sum_{i=1}^{m} w_i J_\rho(\nabla u_{i,\sigma}) \tag{7.10}$$

with $\sum w_i = 1$ and $w_i > 0$ for all i. An interpretation of this structure tensor for vector-valued images in terms of eigenvalues and eigenvectors carries immediately over from the scalar case.

In absence of specific *a-priori* knowledge or if one has a colour model where all channels have a similar meaning, range and reliability, one usually chooses equal weights:

$$w_i = \frac{1}{m} \qquad i = 1, ..., m. \tag{7.11}$$

In those cases where measurements in some channels are less reliable or more noisy, one may choose the weights as a function of the noise variance σ_i^2 such that [60]:

$$w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{m} 1/\sigma_j^2} \qquad i = 1, ..., m. \tag{7.12}$$

If the noise variance turns out to be signal dependent, these weights can be adapted locally resulting in an inhomogeneous treatment of the pixels [60].

7.3 Diffusing Colour Images

Now that we have a tool for *analysing* coherence in vector images, we can try to *enhance* it.

This may be achieved by using anisotropic diffusion filtering. In the scalar case the idea is as follows: One obtains a processed version $u(x,t)$ of an image $f(x)$ with a scale parameter $t \geq 0$ as the solution of a diffusion equation

$$\partial_t u = \operatorname{div}(D\,\nabla u), \tag{7.13}$$

with f as initial condition,

$$u(x,0) = f(x), \tag{7.14}$$

and reflecting boundary conditions:

$$\langle D\nabla u,\, \vec{n}\rangle = 0. \tag{7.15}$$

Hereby, $\vec{n}$ denotes the outer normal and $\langle .,. \rangle$ the usual Euclidean scalar product.

The diffusion tensor D is a positive definite 2×2 matrix, which steers the diffusion process: its eigenvalues determine the diffusivities in the directions of the eigenvectors. In the nonlinear case one adapts the diffusion tensor to the evolving image, for instance to reduce undesired smoothing across edges; see [412] and the references therein. The use of a diffusion tensor allows a more flexible, orientation-dependent filter design than early nonlinear diffusion filters [303, 74] which use only scalar-valued diffusivities.

The simplest idea to process vector images would be to diffuse each channel separately. If the diffusion tensor depends on the local image structure, however, this causes a risk that a structure (e.g. an edge) evolves at different locations for different channels. Thus, it is plausible to synchronize the evolution by a common diffusion tensor for all channels. In this case the vector-valued diffusion filter has the following structure ($i = 1, ..., m$):

$$\partial_t u_i = \operatorname{div}(D\,\nabla u_i), \tag{7.16}$$

$$u_i(x,0) = f_i(x), \tag{7.17}$$

$$\langle D\nabla u_i,\, \vec{n}\rangle = 0. \tag{7.18}$$

Since D should take into account information from all channels, a natural choice would be to make it a function of $J_\rho(\nabla \vec{u}_\sigma)$, the structure tensor for vector images. How should this function look like? For enhancing coherence in vector images,

we need a smoothing process which acts mainly along the coherence direction w_2 (with the notations from Section 7.2) and the smoothing should increase with the coherence $(\mu_1-\mu_2)^2$. This may be achieved in the following way:

We require that D should possess the same eigenvectors w_1, w_2 as the structure tensor $J_\rho(\nabla \vec{u}_\sigma)$. The eigenvalues of D are chosen as

$$\lambda_1 := \alpha \tag{7.19}$$

$$\lambda_2 := \begin{cases} \alpha & \text{if } \mu_1 = \mu_2, \\ \alpha + (1-\alpha)\exp\left(\frac{-C}{(\mu_1-\mu_2)^2}\right) & \text{else} \end{cases} \tag{7.20}$$

with $C > 0$ and a small parameter $\alpha \in (0,1)$.

We observe that λ_2 is an increasing function with respect to the coherence $(\mu_1-\mu_2)^2$. Since the corresponding eigenvector w_2 describes the coherence direction, we have constructed a diffusion process acting preferably along coherent structures.

The exponential function and the positive parameter α were introduced mainly for two theoretical reasons: First, the exponential function guarantees that the smoothness of the structure tensor carries over to the diffusion tensor, and that λ_2 does not exceed 1. This bound is a typical scaling convention in nonlinear diffusion filtering. Second, the positivity of α guarantees that the process never stops: Even if the structure becomes isotropic ($(\mu_1-\mu_2)^2 \to 0$), there remains some small linear diffusion with diffusivity $\alpha > 0$. Thus, the diffusion tensor is uniformly positive definite.

Exploiting the smoothness and uniform positive definiteness properties, we may find a well-founded scale-space interpretation in a similar way as for the scalar-valued anisotropic diffusion filters from [405]. This scale-space representation simplifies the vector image with respect to many aspects: within each channel maxima decrease, minima increase, all L^p-norms ($2 \le p \le \infty$) decrease, even central moments are diminished, and the entropy increases. Moreover, the solution depends continuously on the original image. For $t \to \infty$, all channels tend to a constant image. The average value within each channel remains unaltered during the whole evolution. Existence and uniqueness results for this initial boundary value problem can be obtained in a similar way as in [74]. For a more detailed treatment of the well-posedness and scale-space theory for scalar-valued anisotropic diffusion filters in the continuous and discrete setting, the reader is referred to [412].

7.4 Parameter Selection

The preceding model contains several parameters which have to be specified in practical situations. The goal of this section is to clarify their meaning and to present empirical guidelines for their selection.

We have already seen that the regularization parameter α was introduced to ensure a small amount of isotropic diffusion and to limit the spectral condition number of the diffusion tensor to $1/\alpha$. This parameter is mainly important for theoretical reasons. In practice, it can be fixed to a small value, and no adaptation to the actual image material is required.

The parameter C is a threshold parameter. Strucures with coherence measures $(\mu_1 - \mu_2)^2 \ll C$ are regarded as almost isotropic, and the diffusion along the coherence direction w_2 tends to α. For $(\mu_1 - \mu_2)^2 \gg C$, the diffusion along the coherence direction w_2 tends to its maximal value, which is limited by 1. One possibility to determine a good practical value for C is to calculate a cumulate histogram for $(\mu_1 - \mu_2)^2$ evaluated for the initial image f, and to set C to a certain quantile of this histogram. For instance, if one estimates that 95 % of the image locations have strongly preferred one-dimensional structures, one may set C to the 95 % quantile of the process.

Since the time t is an inherent parameter in each continuous diffusion process, it has nothing to do with its discretization. The common tradition in image analysis, however, is to assign unit length to a pixel. In this case, a different discretization has to be regarded as a rescaling of the image domain. The scaling behaviour of diffusion processes implies that a spatial rescaling which replaces x by βx, has to replace t by $\beta^2 t$. This means for instance that a subsampling in each image direction by a factor 2 results in a four times faster image evolution. Moreover, typical finite difference implementations reveal a computational effort which is proportional to the pixel number. This gives another speed-up by a factor 4, such that the whole calculation becomes 16 times faster.

There remains another question to be addressed: what is a suitable stopping time t of the process?[3] Let us first address this question for scalar-valued images. In a classic linear scale-space representation based on the diffusion process $\partial_t u = \Delta u$, the time t corresponds to a convolution with a Gaussian of standard deviation $\sigma = \sqrt{2t}$. Thus, specifying at spatial smoothing radius σ immediately determines the stopping time t.

In the nonlinear diffusion case, the smoothing is nonuniform and the time t is not directly related to a spatial scale. Other intuitive measures like counting the number of extrema are also problematic for diffusion filters, since it is well-known that for linear and nonlinear diffusion filters in dimensions ≥ 2, the number of local extrema does not necessarily decrease in a monotone way: creation of extrema is not an exception but an event which happens generically [320].

However, it is possible to define average measures for the globality of the representation which reveal monotone behaviour. By identifying it as a Lyapunov func-

[3]It should be observed that this question only appears when regarding the diffusion process as a *restoration method*. Considering it as a *scale-space* means that one is interested in the entire evolution.

tional of a large class of scalar-valued nonlinear diffusion filters, Weickert [405, 412] has shown that the variance $\eta^2(u(t))$ is monotonically decreasing. This reasoning immediately carries over to vector-valued images, where one may define the variance as a suitable convex combination of the variances in each channel:

$$\eta^2(\vec{u}(t)) := \sum_{i=1}^{m} \beta_i \, \eta^2(\vec{u_i}(t)) \tag{7.21}$$

with $\sum_i \beta_i = 1$ and $\beta_i > 0$ for all i. We also know that $\eta^2(\vec{u}(\infty)) = 0$, since every channel converges to a constant image. Therefore, the relative variance

$$s_{\vec{f}}(\vec{u}(t)) \;:=\; \frac{\eta^2(\vec{u}(t))}{\eta^2(\vec{f})} \tag{7.22}$$

decreases monotonically from 1 to 0. It gives the average locality of $\vec{u}(t)$ and its value can be used to measure the distance of $\vec{u}(t)$ from the initial state $\vec{f}$ and the final state $\vec{u}(\infty)$. Prescribing a certain value for $s_{\vec{f}}$ provides us with an a-posteriori criterion for the stopping time of the nonlinear diffusion process. Moreover, this strategy frees the users from any recalculations of the stopping time, if the image is resampled. Practical applications to the restoration of scalar valued medical images have demonstrated the usefulness and simplicity of this criterion [430, 278].

The relative variance can also be used as a heuristic guideline for restoring images with a known signal-to-noise ratio (SNR). Let us illustrate this for a simple example where a degraded image $\vec{g}$ is a noisy variant of some original image $\vec{f}$. By defining the SNR as the ratio between the variance of the original image and the noise variance, one knows that

$$\frac{\eta^2(\vec{f})}{\eta^2(\vec{g})} = \frac{1}{1 + \frac{1}{SNR}}. \tag{7.23}$$

An ideal diffusion filter which works optimally for a denoising task would first eliminate the noise before significantly affecting the signal. For such a filter, one should choose the stopping time T such that the relative variance satisfies

$$s_{\vec{g}}(\vec{u}(T)) = \frac{1}{1 + \frac{1}{SNR}}. \tag{7.24}$$

This stopping time is uniquely determined, as we know that $s_{\vec{g}}(\vec{u}(t))$ is monotone in t. In practice, such a stopping criterion may underestimate the optimal stopping time, since even a well-adapted diffusion filter cannot completely avoid influencing the signal to a certain amount while eliminating the noise. Nevertheless, in Section 7.5 we shall see that (7.24) can give results which are rather close to the optimal value.

The local scale σ and the integration scale ρ of the structure tensor should be adapted to the noise and the texture scale of the problem. In many cases it is not very difficult to get parameter estimates which work well over the whole image domain. In other situations, one may want to find local estimates by applying scale selection strategies as proposed by Lindeberg [230]. The next section will show an example which illustrates the impact of the integration scale.

7.5 Examples

Coherence-enhancing anisotropic diffusion can be implemented by means of standard finite-difference methods from the numerical literature; see [427] for some implementational details as well as extensions to higher dimensions. In the following test examples we used the RGB channels to regard colour images as vector images. Since the variances in all channels were comparable and no additional a priori knowledge was available, equal weights were assigned to all channels. The parameter α was set to 0.001.

Figure 7.2 shows the well-known mandrill test image and its processed version for different integration scales ρ. The case $\rho = 0$ is depicted in Figure 7.2(b). This choice is essentially a colour extension of the greyscale model by Cottet and Germain [91], and its smoothing direction w_2 is identical with the Chambolle and Sapiro–Ringach filters [75, 332]. As can be seen from Figure 7.1(b), the reason for the limited performance of this filter in the context of orientation smoothing lies in highly fluctuating local orientation estimates. The introduction of an additional integration scale stabilizes these estimates and leads to a significantly improved smoothing along coherent structures such as the mandrill's hair. This effect, depicted in Figure 7.2(c)–(f), shows that *the integration scale is an important feature for the success of any orientation diffusion scheme.* It should be equal to or larger than the texture scale. Overestimating it is significantly less critical than underestimations.

Figure 7.3 analyses the behaviour of coherence-enhancing diffusion for noisy input images. The mandrill image has been corrupted with additive Gaussian noise and the filtered results were depicted at those times when the l^2 distance to the uncorrupted mandrill image was minimal. The observable high robustness confirms the theoretically established stability of a broad class of nonlinear diffusion filters under perturbations of the original image [412]. Even for a SNR of only 0.25 and a noise scale of only $\sigma = 1$ pixel, it is possible to correctly restore many coherent structures which are hardly visible for a human observer.

Table 7.1 shows that in this case coherence-enhancing diffusion was able to reduce the l^2 distance to the original image by a factor 15.58. Table 7.2 and Figure 7.4 illustrate the denoising behaviour when using the stopping criterion (7.24). We

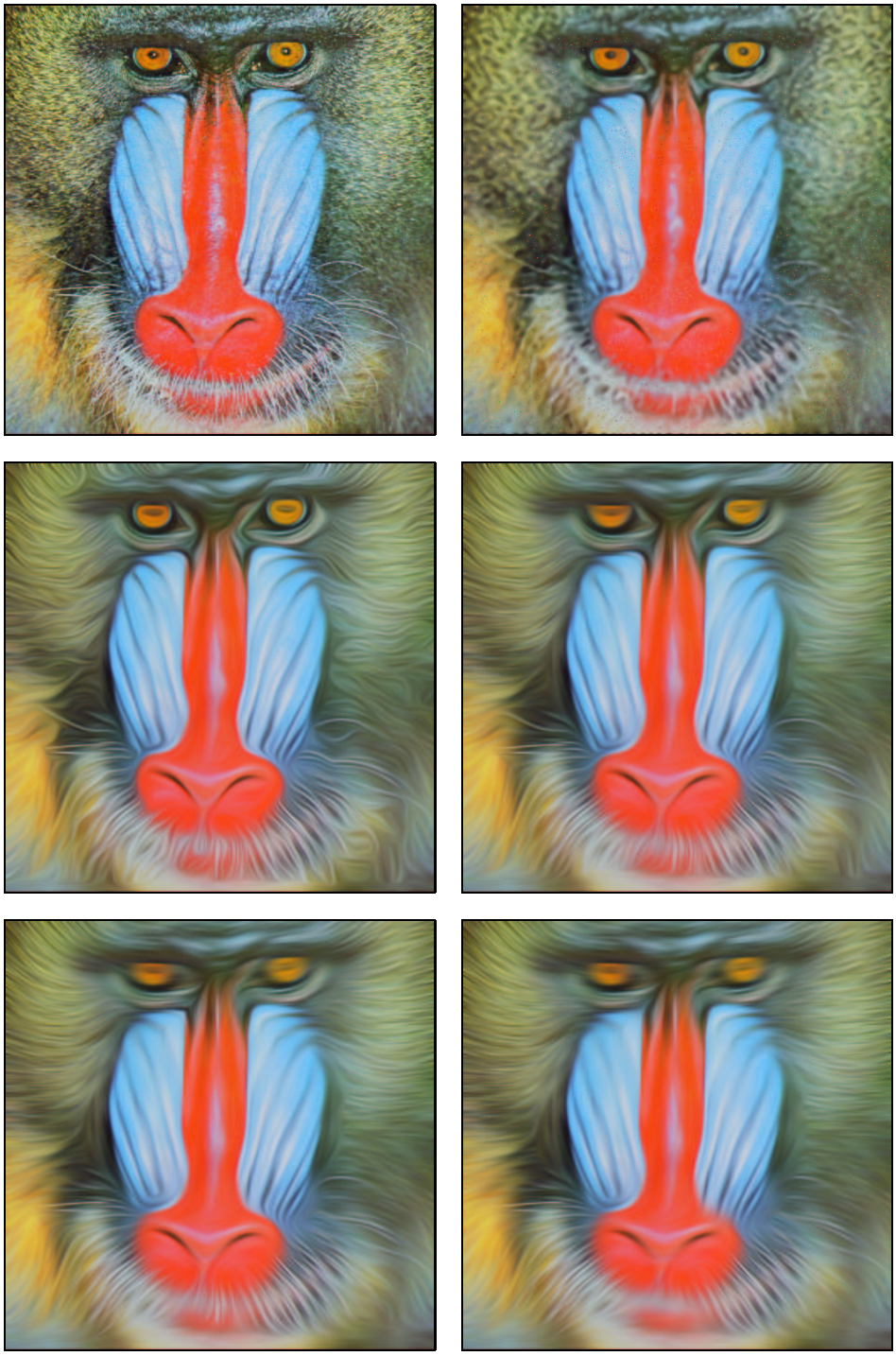


Figure 7.2: Coherence-enhancing diffusion of the mandrill test image with different integration scales ρ. C was set to the 99 % quantile. (a) Top Left: Original, $\Omega = (0, 512)^2$. (b) Top Right: Filtered with $\sigma = 1$, $\rho = 0$, $t = 40$ (c) Middle Left: Ditto with $\rho = 5$. (d) Middle Right: $\rho = 10$. (e) Bottom Left: $\rho = 15$. (f) Bottom Right: $\rho = 20$.

Figure 7.3: Robustness of coherence-enhancing diffusion under noise ($\sigma = 1$, $\rho = 12$, λ was set to the 99% quantile). (a) Top Left: Mandrill image with additive Gaussian noise. $\eta_n = 56.2$, $SNR = 1$. (b) Top Right: Optimal restoration of (a). $t = 6.4$. (c) Middle Left: $\sigma_n = 79.0$, $SNR = 0.5$. (d) Middle Right: Optimal restoration of (c). $t = 16.5$. (e) Bottom Left: $\sigma_n = 112.4$, $SNR = 0.25$. (f) Bottom Right: Optimal restoration of (e). $t = 46.0$.

Table 7.1: Restoration properties of coherence-enhancing diffusion applied to the mandrill image with additive Gaussian noise with zero mean. η_n denotes the standard deviation of the noise, and η^2 is the average image variance per channel. T is the optimal stopping time with respect to the minimization of $\delta(t)$, the l^2 distance between $\vec{u}(t)$ and the uncorrupted original image.

SNR	4	2	1.0	0.5	0.25
η_n	28.1	39.7	56.2	79.0	112.4
T	1.3	2.7	6.4	16.5	46.0
$\eta^2(t=0)$	3945.34	4730.24	6308.89	9384.54	15761.21
$\eta^2(t=T)$	2990.08	2917.97	2807.48	2669.16	2502.25
$\delta(t=0)$	787.50	1571.88	3149.99	6224.20	12599.97
$\delta(t=T)$	299.13	415.21	547.20	678.24	808.29
$\delta(0)/\delta(T)$	2.63	3.79	5.76	9.18	15.58

Table 7.2: Restoration properties of coherence-enhancing diffusion when applying the stopping time criterion instead of the optimal stopping time.

SNR	4	2	1.0	0.5	0.25
η_n	28.1	39.7	56.2	79.0	112.4
T	0.7	1.3	2.4	4.8	9.0
$\eta^2(t=0)$	3945.34	4730.24	6308.89	9384.54	15761.21
$\eta^2(t=T)$	3155.85	3144.92	3143.47	3119.03	3145.70
$\delta(t=0)$	787.50	1571.88	3149.99	6224.20	12599.97
$\delta(t=T)$	299.13	415.21	547.20	678.24	808.29
$\delta(0)/\delta(T)$	2.46	3.48	5.14	7.93	12.55

Figure 7.4: Optimal stopping time vs. automatic stopping time. (a) LEFT: Optimal restoration of mandrill image with additive Gaussian noise ($SNR = 1$). (b) RIGHT: Restoration according to the stopping criterion (7.24).

observe that the improvements in the l^2 distance are not much worse than in the optimal case.

After these discussions on the restoration properties of coherence-enhancing diffusion, let us now have a look at its scale-space qualities. Figure 7.5 depicts the evolution of a wood surface under three diffusion scale-spaces.

The first column shows the results when isotropic linear diffusion with the unit matrix as diffusion tensor is applied. Since this scale-space is designed to be uncommitted, it blurs all features in a uniform way. In the middle column evolution under nonlinear isotropic diffusion filtering can be seen. This method is based on an extension of the techniques from [303, 74] to vector images; see [432] for more details. It can be regarded as an edge-preserving smoothing where the (scalar-valued) diffusivity is low whenever $\sum_i |\nabla u_{i,\sigma}|^2$ is large. It tends to preserve high contrasts, but it is ignorant of the flow-like image character. The right column shows the effect of coherence-enhancing anisotropic diffusion. We observe that it gives a flow-like gradual simplification of the original image. It is our believe that an evolution of this type will be potentially useful for the automatic grading of wood surfaces. This would be in accordance with the good results of coherence-enhancing diffusion filtering for the grading of scalar-valued nonwoven fabric images that were reported in [406].

Other potential application fields where one might be interested in processing coherent structures are sketched in Figure 7.6 and 7.7. They describe problems from analysing sonograms and medical images from light microscopy, respectively.

Since coherence-enhancing anisotropic colour diffusion seems to create optically pleasant structures, it is tempting to apply it to van Gogh paintings which

Figure 7.5: Scale-space evolution of a wood surface under different diffusion processes TOP: Original image, $\Omega = (0, 256)^2$. LEFT COLUMN: Linear diffusion, top to bottom: $t = 0$, 2, 10, 50. MIDDLE COLUMN: Isotropic nonlinear diffusion ($\lambda = 10$, $\sigma = 1$), $t = 0$, 20, 200, 2000. RIGHT COLUMN: Coherence-enhancing nonlinear diffusion ($C = 17.6$, $\sigma = 1$, $\rho = 5$), $t = 0$, 20, 200, 2000.

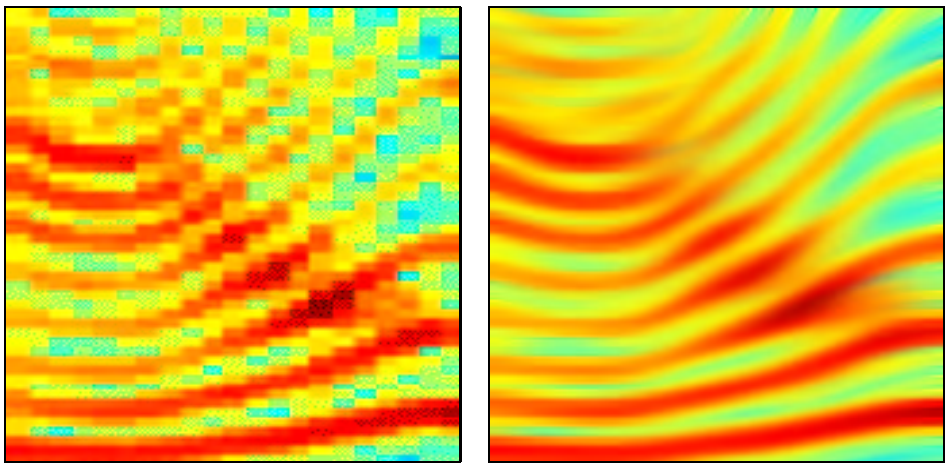


Figure 7.6: Coherence-enhancing diffusion of a sonogram. It is a local frequency analysis of the Danish word *hej*. The horizontal axis is the time axis, the vertical axis describes the logarithm of the frequency, and the colours display their intensities. One is interested in estimating the distance between the different lines, and in interpolating such that sampling artifacts become smaller. (a) LEFT: $\Omega = (0, 168) \times (0, 192)$. (b) RIGHT: Filtered ($C = 28.5$ (99 % quantile), $\sigma = 1$, $\rho = 10$, $t = 50$).

emphasize such coherent structures. Thus, in a last example the temporal evolution of the "Starry Night" painting is shown in Figure 7.8 [392]. While the image becomes gradually simpler, its flow-like character, which is typical for many van Gogh paintings, is maintained for a very long time. The coherence-enhancing effect even creates full moon.

7.6 Conclusions

In this paper a filter for enhancing coherent structures in vector-valued images has been presented. It is based on two ideas: a generalized structure tensor for vector images, and anisotropic nonlinear diffusion filtering with a diffusion tensor. We have shown that its main distinction from other colour diffusion models lies in an additional integration scale which gives a semilocal average over the preferred orientation. This integration leads to significantly improved smoothing orientations, which is of importance for the enhancement of one-dimensional structures. We have clarified the role of the necessary parameters and proposed heuristics for their selection. Examples have been presented which show that coherence-enhancing colour diffusion is very robust under noise and of potential interest in various application areas.

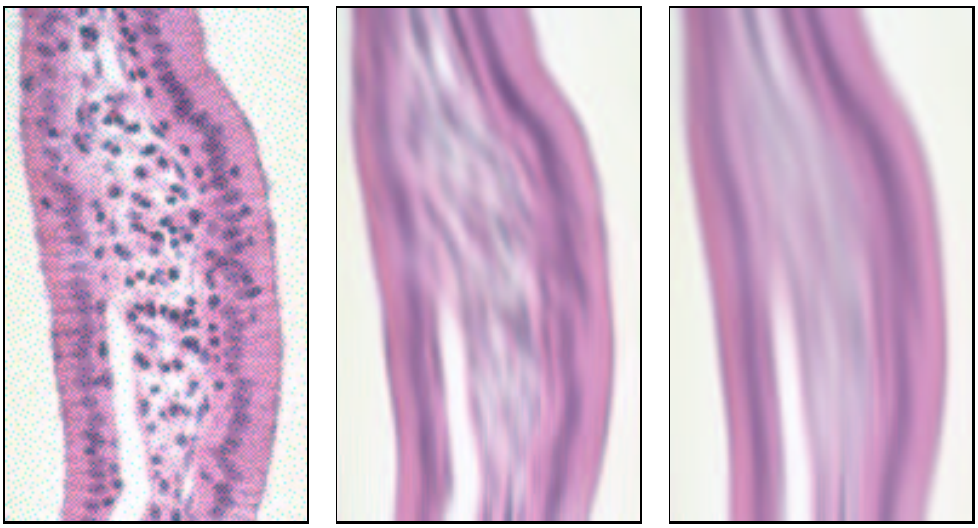


Figure 7.7: Coherence-enhancing colour diffusion for simplifying medical images. This is of potential interest for easing their automatic evaluation by subsequent image processing methods. (a) LEFT: Light microscopy image of a columnar epithelium of the gall bladder, $\Omega = (0, 125) \times (0, 212)$. (b) MIDDLE: Filtered ($\sigma = 1$, $\rho = 30$, $C = 1$, $t = 30$). (c) Ditto with $t = 100$.

Acknowledgments. This work has been funded by *Stiftung Innovation des Landes Rheinland–Pfalz*, the *Real World Computing Partnership*, the *Danish National Science Research Council*, and the *EU-TMR project VIRGO*. The sonogram has been kindly provided by Kristoffer Jensen (Department of Computer Science, University of Copenhagen).

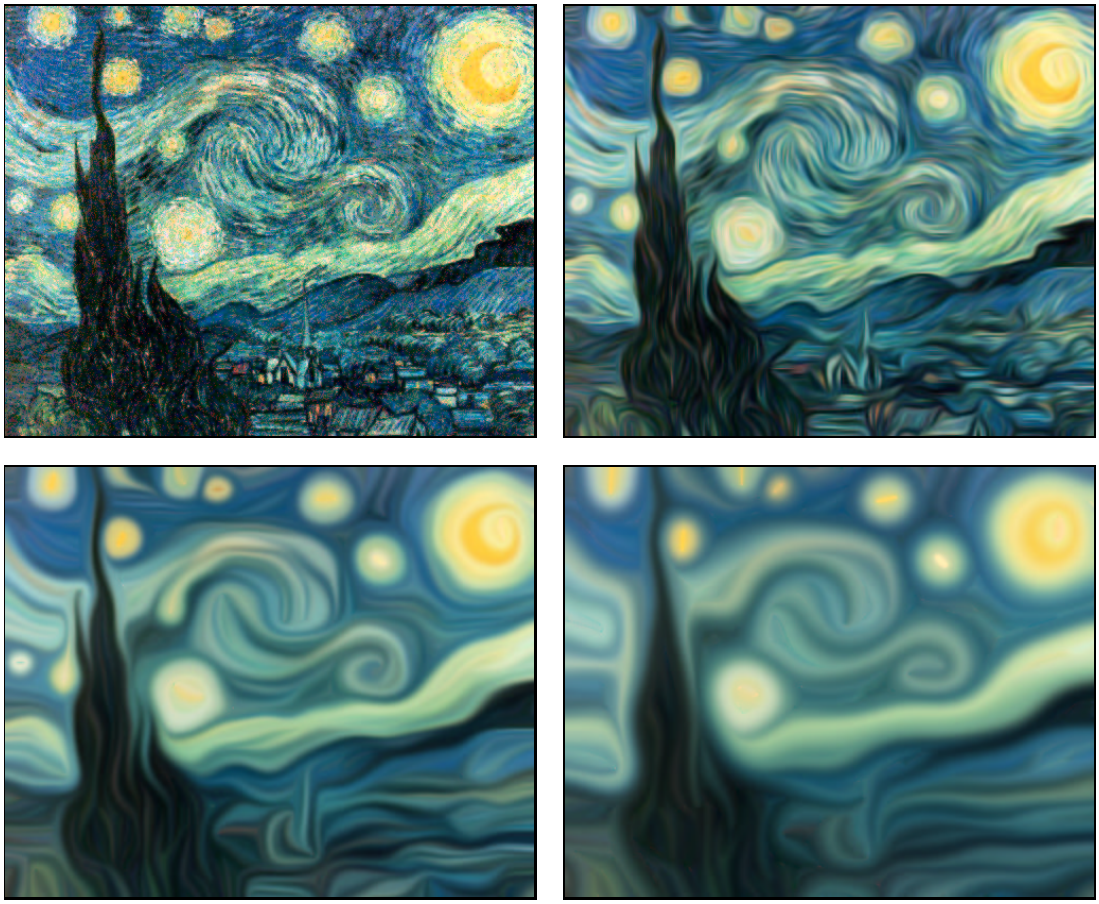


Figure 7.8: Coherence-enhancing anisotropic diffusion of an expressionistic painting ($C = 4.6$ (99 % quantile), $\sigma = 1$, $\rho = 4$). (a) Top Left: "Starry Night" by van Gogh (Saint-Rémy, 1889; New York, The Museum of Modern Art), $\Omega = (0, 500) \times (0, 406)$. (b) Top Right: $t = 15$. (c) Bottom Left: $t = 300$. (d) Bottom Right: $t = 3000$.

Part III

Efficient Algorithms

Chapter 8

Efficient and Reliable Schemes for Nonlinear Diffusion Filtering

Joachim Weickert, Bart M. ter Haar Romeny, and Max A. Viergever.
IEEE Transactions on Image Processing,
Vol. 8, No. 3, 398–410, March 1998.

Abstract

Nonlinear diffusion filtering is usually performed with explicit schemes. They are only stable for very small time steps, which leads to poor efficiency and limits their practical use. Based on a recent discrete nonlinear diffusion scale-space framework we present semi-implicit schemes which are stable for all time steps. These novel schemes use an additive operator splitting (AOS) which guarantees equal treatment of all coordinate axes. They can be implemented easily in arbitrary dimensions, have good rotational invariance and reveal a computational complexity and memory requirement which is linear in the number of pixels. Examples demonstrate that, under typical accuracy requirements, AOS schemes are at least ten times more efficient than the widely-used explicit schemes.

8.1 Introduction

Impressive results are the main reason for using nonlinear diffusion filtering in image processing: unlike linear diffusion filtering (which is equivalent to convolving with a Gaussian) edges remain well-localized and can even be enhanced. Spatial regularizations of this filter class have a solid mathematical foundation as well-posed scale-spaces [74, 408, 412], whose parameter influence is well-understood [37, 226].

Poor efficiency is the main reason for *not* using nonlinear diffusion filtering: most approaches are based on the simplest finite difference discretization by means of a so-called explicit or Euler-forward scheme. This scheme requires very small time steps in order to be stable. Hence the whole filtering procedure is rather time-consuming.

In the present paper we are going to address this problem. We present a novel type of separable schemes which do not suffer from any time step size restriction since all stability-relevant terms are discretized in an implicit manner. The backbone of these schemes is a Gaussian algorithm for solving a tridiagonal system of linear equations. It is fast, stable and requires only a few lines programming work. Its forward and backward substitution step can be regarded as a causal and anticausal filter of a recursive scheme. The presented algorithms are applicable in arbitrary dimensions and their computational and storage effort is linear in the image size. This shows their *efficiency*.

We prove the *reliability* of these schemes by verifying that they satisfy recently established criteria for nonlinear diffusion scale-spaces [407, 412]. This comes down to checking six simple criteria. If these requirements are fulfilled we can be sure that the scheme preserves the average grey value, satisfies a causality property in terms of a maximum–minimum–principle, reveals a large class of smoothing Lyapunov functionals, and converges to a constant steady-state as the time tends to infinity. It should be noted that the discrete maximum–minimum principle is a very restrictive stability criterion (more restrictive than the von Neumann stability), since it also takes into account the boundary conditions and guarantees that over- and undershoots cannot appear.

The goal of this paper is to guide the reader in a systematic way to these so-called additive operator splitting (AOS) schemes. Specific knowledge in numerical analysis is not necessary, as we shall refer to the required material in the literature whenever it is needed. However, the reader who is interested in a more detailed introduction to the matrix algebra which is useful for the present paper, may find this in Chapter 6 of Ortega's textbook [289]. As a prototype of a well-founded nonlinear diffusion filter we focus on a spatial regularization of the Perona–Malik filter [303] by Catté, Lions, Morel and Coll [74], and Whitaker and Pizer [433].

The paper is organized as follows:

Section 8.2 gives a brief survey on this diffusion model (which will call CLMC equation henceforth). In Section 8.3 we review the simplest scheme for the 1-D CLMC equation: the explicit (Euler forward) discretization in time. We analyse it by means of criteria for discrete nonlinear diffusion scale-spaces in order to explain why it requires rather prohibitive time step sizes. As a remedy we study a semi-implicit discretization for which we show that it satisfies all discrete scale-space criteria (including stability) even for arbitrary large time steps. It requires to solve a tridiagonal linear system of equations, which is easily and efficiently done by a special variant of the well-known Gaussian elimination algorithm. This so-called *Thomas algorithm* will be presented in detail, since it forms the core of the whole scheme.

In Section 8.4 we consider the higher-dimensional case. It is argued that the simple explicit scheme leads to even more restrictive stability conditions than in the 1-D case, while the semi-implicit scheme remains absolutely stable. However, solving the m-dimensional linear system becomes significantly less efficient for dimensions ≥ 2.

As as remedy we present an alternative scheme which is also semi-implicit, has the same approximation order, and is absolutely stable, but it can be separated into one-dimensional processes. Thus, the simple and efficient Thomas algorithm can be applied again. Unlike classical multiplicative splitting schemes from the mathematical literature, we consider an additive operator splitting (AOS). It ensures that all coordinate axes are treated equally, a very desirable symmetry property in the context of image processing. Furthermore, we shall check that the AOS schemes satisfy all criteria for discrete nonlinear scale-spaces.

The section is concluded by proposing a related method for the regularization step within the CLMC model. Since this regularization is based on a Gaussian convolution, it is natural to regard it as a linear diffusion filter for which one may also apply splitting techniques based on the Thomas algorithm.

Section 8.5 presents an m-dimensional algorithmic formulation of the AOS schemes and analyses its complexity.

In Section 8.6 we evaluate the results by checking the performance of AOS schemes with respect to rotational invariance and accuracy. This allows us to propose reasonable time step size and to analyse the accuracy and efficiency in comparison to the unsplit semi-implicit scheme and the widely-used explicit scheme.

We conclude the paper with a summary in Section 8.7. A shortened preliminary version of this paper can be found in [410].

Related work. Our work has been influenced by a number of related approaches which we would like to mention here.

Implicit splitting-based approaches for linear diffusion filtering have been proposed in [151, 64], and also in [8, 14, 444] where their realization as recursive filters is suggested. Impressive results on improved efficiency by means of recursive filtering

can be found in [102, 103], and the close relation between recursive filters and linear scale-space approaches has been clarified in [273]. Semidiscrete or fully discrete analogues of linear diffusion filtering have been proposed in [282, 229, 25, 329].
In the nonlinear diffusion field one can find several approaches which aim to be efficient alternatives to the conventional two-level explicit finite-difference scheme, for instance three-level methods [136], semi-implicit approaches [74], multiplicative splittings [403], multigrid methods [4], finite element techniques with adaptive mesh coarsening [33], numerical schemes with wavelets as trial functions [136], and pseudospectral methods [136]. Even hardware proposals for nonlinear diffusion filtering can be found in the literature [302, 145].
Schemes which inherit a large number of the properties of their continuous counterparts have also been proposed in the context of curvature-based nonlinear scale-spaces [58, 72, 73, 86]. Sophisticated algorithms for such processes comprise fast level set methods [353], high-order ENO schemes [357], and implicit algorithms for mean curvature motion [8, 271].

8.2 The Continuous Filter Process

In the m-dimensional case the filter of Catté, Lions, Morel and Coll [74] has the following structure:
Let $\Omega := (0, a_1) \times .... \times (0, a_m)$ be our image domain and consider a (scalar) image as a bounded mapping from Ω into the real numbers $\mathbb{R}$. Then the CLMC filter calculates a filtered image $u(x, t)$ of $f(x)$ as a solution of the diffusion equation

$$\partial_t u \;=\; \mathrm{div}\left(g(|\nabla u_\sigma|^2)\;\nabla u\right) \tag{8.1}$$

with the original image as initial state,

$$u(x, 0) \;=\; f(x), \tag{8.2}$$

and reflecting boundary conditions:

$$\partial_n u := 0 \quad \text{on} \quad \partial\Omega, \tag{8.3}$$

where n denotes the normal to the image boundary $\partial\Omega$.
The "time" t is a scale parameter: increasing t leads to simpler image representations. The whole embedding of the original image into such a one-parameter family of simplified images is called *scale-space*. The first representative of this very general and useful image processing concept, namely linear diffusion filtering, has been derived in an axiomatic way by Taizo Iijima more than 35 years ago [175, 420].

In order to reduce smoothing at edges, the diffusivity g is chosen as a decreasing function of the edge detector $|\nabla u_\sigma|$. Here, ∇u_σ is the gradient of a smoothed version of u which is obtained by convolving u with a Gaussian of standard deviation σ:

$$\nabla u_\sigma := \nabla(K_\sigma * u), \tag{8.4}$$

$$K_\sigma := \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \tag{8.5}$$

We use the following form for the diffusivity:

$$g(s) := \begin{cases} 1 & (s \le 0) \\ 1 - \exp\left(\frac{-3.315}{(s/\lambda)^4}\right) & (s > 0). \end{cases} \tag{8.6}$$

For such rapidly decreasing diffusivities smoothing on both sides of an edge is much stronger than smoothing across it. As a result, the gradient at edges may even be enhanced, see [303] for more details. λ plays the role of a contrast parameter: Structures with $|\nabla u_\sigma| > \lambda$ are regarded as edges, where the diffusivity is close to 0, while structures with $|\nabla u_\sigma| < \lambda$ are considered to belong to the interior of a region. Here the diffusivity is close to 1. In this sense, the CLMC model serves as a selective smoothing which prefers intraregional smoothing to interregional blurring. After some time it leads to segmentation-like results which are piecewise almost constant.

The parameter $\sigma > 0$ makes the filter insensitive to noise at scales smaller than σ. It is also a regularization parameter which guarantees well-posedness of the process: Catté *et al.* [74] have shown that their filter has a unique solution which is infinitely times differentiable for $t > 0$. Weickert [408, 412] has proved that it depends continuously on the original image, satisfies a maximum–minimum principle and reveals a large family of smoothing Lyapunov functionals which guarantee that the solution tends to a constant image for $t \to \infty$. During the whole evolution, the average grey value remains unaltered.

Equations of this type have been successfully applied to process medical images, see e.g. [222, 433, 235]. Nevertheless, they are only one representative of a large class of nonlinear scale-spaces. Overviews of other methods can be found in [104, 382, 412].

8.3 One-Dimensional Case

8.3.1 Explicit Scheme

The Scheme

The one-dimensional CLMC equation is given by

$$\partial_t u = \partial_x\Big(g(|\partial_x u_\sigma|^2)\, \partial_x u\Big). \tag{8.7}$$

Let us now consider the simplest discrete approximation of this process. A discrete image can be regarded as a vector $f \in \mathbb{R}^N$, whose components f_i, $i \in J := \{1, ..., N\}$ display the grey values at each pixel. Pixel i represents some location x_i, and h is the grid size. We consider discrete times $t_k := k\tau$, where $k \in \mathbb{N}_0$ and τ is the time step size. By u_i^k we denote approximations to $u(x_i, t_k)$.
The simplest discretization of (8.7) with reflecting boundary conditions is given by

$$\frac{u_i^{k+1} - u_i^k}{\tau} = \sum_{j \in \mathcal{N}(i)} \frac{g_j^k + g_i^k}{2h^2} (u_j^k - u_i^k), \tag{8.8}$$

where $\mathcal{N}(i)$ is the set of the two neighbours of pixel i (boundary pixels have only one neighbour).
The diffusivities g_i^k approximate $g(|\nabla u(x_i, t_k)|^2)$. They can be obtained as follows: In the spatially discrete case the convolution $u_\sigma = K_\sigma * u$ comes down to a multiplication of $u \in \mathbb{R}^N$ with a suitable matrix $H \in \mathbb{R}^{N \times N}$. In Section 8.4.3 we shall present an efficient way to achieve this in the spatial domain. A gradient approximation by central differences gives

$$g_i^k := g\left(\frac{1}{2} \sum_{p,q \in \mathcal{N}(i)} \left(\frac{u_p^k - u_q^k}{2h}\right)^2\right) \tag{8.9}$$

for some inner pixel i. This expression remains also valid at the boundary pixels, if we extend the image by reflecting it at the boundary.
We can write the explicit scheme in matrix–vector notation as

$$\frac{u^{k+1} - u^k}{\tau} = A(u^k)\, u^k \tag{8.10}$$

with $A(u^k) = (a_{ij}(u^k))$ and

$$a_{ij}(u^k) := \begin{cases} \frac{g_i^k + g_j^k}{2h^2} & (j \in \mathcal{N}(i)), \\ -\sum\limits_{n \in \mathcal{N}(i)} \frac{g_i^k + g_n^k}{2h^2} & (j = i), \\ 0 & (\text{else}). \end{cases} \tag{8.11}$$

This comes down to the iteration scheme

$$u^{k+1} = \left(I + \tau A(u^k)\right) u^k, \tag{8.12}$$

where $I \in \mathbb{R}^N$ is the unit matrix. This scheme is called *explicit*, since u^{k+1} can be directly calculated from u^k without solving a system of equations.
Such an explicit iteration step is computationally very cheap: its requires mainly to calculate the three nonvanishing matrix extries per row and to perform a matrix–vector multiplication. The computational and storage effort is linear in the pixel

number N. But does this explicit scheme also create a good discrete scale-space and how far can we come with one step? We can find an answer to these question by applying a framework for discrete nonlinear diffusion scale-spaces which we shall review next.

Criteria for Discrete Nonlinear Diffusion Scale-Spaces

Recently a scale-space interpretation for the continuous CLMC equation and its anisotropic generalizations has been established [408, 412]. In addition to invariances such as the preservation of the average grey value, it has been shown that – it spite of its contrast-enhancing potential – these equations create smoothing scale-spaces: the obey a maximum–minimum principle, have a large class of smoothing Lyapunov functionals, and converge to a constant steady-state.
It would be desirable to ensure that discrete approximations do also reveal these qualities *exactly*. Criteria have been identified under which one can guarantee that a discrete scheme of type

$$u^0 = f, \tag{8.13}$$

$$u^{k+1} = Q(u^k)\, u^k, \qquad \forall\, k \in \mathbb{N}_0, \tag{8.14}$$

possesses such properties [407, 412]. All one has to check are the following criteria for $Q(u^k) = (q_{ij}(u^k))$:

(D1) *Continuity in its argument:*

$$Q \in \mathrm{C}(\mathbb{R}^N, \mathbb{R}^{N\times N}) \tag{8.15}$$

(D2) *Symmetry:*

$$q_{ij} = q_{ji} \qquad \forall\, i,j \in J. \tag{8.16}$$

(D3) *Unit row sum:*

$$\sum_{j\in J} q_{ij} = 1 \qquad \forall\, i \in J. \tag{8.17}$$

(D4) *Nonnegativity:*

$$q_{ij} \geq 0 \qquad \forall\, i,j \in J. \tag{8.18}$$

(D4) *Positive diagonal:*

$$q_{ii} > 0 \qquad \forall\, i \in J. \tag{8.19}$$

(D6) *Irreducibility:*
We can connect any two pixels by a path with nonvanishing diffusivities. Formally: For any $i,j \in J$ there exist $k_0,\ldots,k_r \in J$ with $k_0 = i$ and $k_r = j$ such that $q_{k_p k_{p+1}} \neq 0$ for $p = 0,\ldots,r-1$.

Under these prerequisites the filtering process is well-posed and satisfies the following discrete scale-space properties [407, 412]:

(a) *Average grey level invariance:*
The average grey level $\mu := \frac{1}{N} \sum_{j \in J} f_j$ is not affected by the discrete diffusion filter:

$$\frac{1}{N} \sum_{j \in J} u_j^k = \mu \qquad \forall\, k \in \mathbb{N}_0. \tag{8.20}$$

This invariance is required in scale-space based segmentation algorithms such as the hyperstack [277].

(b) *Extremum principle:*

$$\min_{j \in J} f_j \leq u_i^k \leq \max_{j \in J} f_j \qquad \forall\, i \in J, \quad \forall\, k \in \mathbb{N}_0. \tag{8.21}$$

This property is much more than a stability result which forbids under- and overshoots. It also ensures that iso-intensity linking towards the original image is possible. Hence, it states an important causality property, cf. [173].

(c) *Smoothing Lyapunov sequences:*
The process is a simplifying, information-reducing transform with respect to many aspects:

(i) The p-norms

$$\|u^k\|_p \;:=\; \Big(\sum_{i=1}^{N} |u_i^k|^p \Big)^{1/p} \tag{8.22}$$

are decreasing in k for all $p \geq 1$.

(ii) All even central moments

$$M_{2n}[u^k] \;:=\; \frac{1}{N} \sum_{j=1}^{N} (u_j^k - \mu)^{2n} \quad (n \in \mathbb{N}) \tag{8.23}$$

are decreasing in k.

(iii) The entropy

$$S[u^k] \;:=\; -\sum_{j=1}^{N} u_j^k \ln u_j^k, \tag{8.24}$$

a measure of uncertainty and missing information, is increasing in k (if f_j is positive for all j).

(d) *Convergence to a constant steady-state:*

$$\lim_{k\to\infty} u_i^k = \mu \quad \forall\, i \in J. \tag{8.25}$$

Thus, the discrete scale-space evolution tends to the most global image representation that is possible: a constant image with the same average grey level as f.

Does the Explicit Scheme Create a Discrete Scale-Space?

Let us now investigate if the explicit scheme (8.12) satisfies the criteria (D1)–(D6) for discrete nonlinear scale-spaces. Let

$$Q(u^k) \;:=\; (q_{ij}(u^k)) := I + \tau A(u^k). \tag{8.26}$$

By virtue of (8.11) we observe that the continuity of Q with respect to its argument follows directly from the continuity of the diffusivity g.

The symmetry of Q follows from (8.11) and the symmetry of the neighbourhood relation ($i \in \mathcal{N}(j) \iff j \in \mathcal{N}(i)$).

By the construction of A it is also evident that the row sums of A vanish. Hence, all row sums of Q are 1, which proves (D3).

Thus, let us investigate the nonnegativity. From $a_{ij} \geq 0$ for $i \neq j$, we also have $q_{ij} \geq 0$ for $i \neq j$. Thus, we can focus on the diagonal entries. If they are all positive, both (D4) and (D5) are satisfied. Since

$$q_{ii} = 1 - \tau \sum_{j\neq i} a_{ij} \tag{8.27}$$

and $\sum_{j\neq i} a_{ij} > 0$, positive diagonal entries require that

$$\tau \;<\; \frac{1}{\max_i \sum_{j\neq i} a_{ij}(u^k)}. \tag{8.28}$$

In order to show that Q is irreducible, let us assume that τ satisfies this restriction and consider two arbitrary pixels s_1 and s_2. If $s_1 \leq s_2$ then the positivity of g implies that

$$q_{s_1,s_1+1} > 0, \quad q_{s_1+1,s_1+2} > 0, \quad \cdots, \quad q_{s_2-1,s_2} > 0. \tag{8.29}$$

If $s_1 > s_2$ then

$$q_{s_1,s_1-1} > 0, \quad q_{s_1-1,s_1-2} > 0, \quad \cdots, \quad q_{s_2+1,s_2} > 0. \tag{8.30}$$

This establishes (D6).

From these considerations we conclude that the explicit scheme creates a discrete scale-space provided that the time step size satisfies the restriction (8.28). In image processing one usually sets $h := 1$. Since the diffusivity g is bounded from above by 1, definition (8.11) allows us to guarantee (8.28) for $\tau < 1/2$.
In practice, this is often a very severe step size restriction. It means that the use of an explicit scheme is limited rather by its stability than its accuracy. For this reason it would be interesting to look for schemes with better stability properties. This shall be done next.

8.3.2 Semi-Implicit Scheme

The Scheme

We consider a slightly more complicated discretization of (8.7), namely

$$\frac{u^{k+1} - u^k}{\tau} \;=\; A(u^k)\, u^{k+1}, \tag{8.31}$$

which leads to the scheme

$$\Big(I - \tau A(u^k)\Big)\, u^{k+1} = u^k. \tag{8.32}$$

We observe that this scheme does not give the solution u^{k+1} directly (explicitly): It requires to solve a linear system first. For this reason it is called a *linear-implicit (semi-implicit)* scheme.

Remark. One may also be interested in studying the *(fully) implicit* scheme

$$\frac{u^{k+1} - u^k}{\tau} \;=\; A(u^{k+1})\, u^{k+1}, \tag{8.33}$$

leading to a nonlinear system of equations. This is more complicated to solve. Below we shall see, however, that such a high effort is not necessary, since already semi-implicitness is sufficient to guarantee absolute stability.

Does the Semi-Implicit Scheme Create a Discrete Scale-Space?

In order to establish the semi-implicit scheme (8.32) as a discrete scale-space we have to check (D1)–(D6) again.
First we have to show that

$$B(u^k) \;:=\; (b_{ij}(u^k)) \;:=\; (I - \tau A(u^k)) \tag{8.34}$$

is invertible. This is easily seen, because B is strictly diagonally dominant:

$$|b_{ii}| > \sum_{j \neq i} |b_{ij}| \qquad \forall\, i \in J. \tag{8.35}$$

It is well-known from linear algebra that strictly diagonally dominant matrices are invertible, see e.g. [289, p. 226]. Thus,

$$Q(u^k) \;:=\; (q_{ij}(u^k)) \;:=\; B^{-1}(u^k) \tag{8.36}$$

exists and the continuity of Q in its argument follows from the continuity of g. Moreover, the symmetry of A carries also over to B and Q, which establishes (D2). In order to prove (D3), consider $w := (1, ..., 1)^T \in \mathbb{R}^N$. Since B has unit row sum, we have $Bw = w$. This implies that

$$w = B^{-1}w = Qw. \tag{8.37}$$

Reading this componentwise shows that Q has also unit row sum.
(D4)–(D6) can be verified in one step. We already know that B is strictly diagonally dominant. It is also immediately seen that B is irreducible, $b_{ij} \leq 0$ for $i \neq j$, and $b_{ii} > 0$ for all i. Then a theorem by Varga [395, p. 85] tells us that $Q = B^{-1}$ satisfies

$$q_{ij} > 0 \qquad \forall\, i, j \in J. \tag{8.38}$$

Thus, Q is nonnegative, has positive diagonal and is irreducible.

From these considerations we observe that the semi-implicit scheme creates a discrete nonlinear diffusion scale-space for arbitrarily large time steps. In particular, it is unconditionally stable and does not suffer from any time step size restriction. Unlike the explicit scheme, it can be fully adapted to the desired accuracy without the need to choose small time steps for stability reasons.

Solving the Tridiagonal Linear System: The Thomas Algorithm

The semi-implicit scheme requires to solve a linear system, where the system matrix is tridiagonal and diagonally dominant. The most efficient way to do this is the so-called *Thomas algorithm*, a Gaussian elimination algorithm for tridiagonal systems. It can be found in many textbooks on numerical analysis, e.g. [352, pp. 43–45]. However, since it builds the backbone of our algorithms and since we want to keep this paper selfcontained, we survey its algorithmic features here.
The principle is as follows. Suppose we want to solve a tridiagonal linear system $Bu = d$ with

$$B = \begin{pmatrix} \alpha_1 & \beta_1 & & & \\ \gamma_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_{N-2} & \alpha_{N-1} & \beta_{N-1} \\ & & & \gamma_{N-1} & \alpha_N \end{pmatrix}. \tag{8.39}$$

Then the Thomas algorithm consists of three steps.

Step 1: LR decomposition

We decompose B into the product of a lower bidiagonal matrix

$$L = \begin{pmatrix} 1 & & & \\ l_1 & 1 & & \\ & \ddots & \ddots & \\ & & l_{N-1} & 1 \end{pmatrix} \tag{8.40}$$

and an upper bidiagonal matrix

$$R = \begin{pmatrix} m_1 & r_1 & & \\ & \ddots & \ddots & \\ & & m_{N-1} & r_{N-1} \\ & & & m_N \end{pmatrix} \tag{8.41}$$

Comparing the coefficients shows that $r_i = \beta_i$ for all i, and m_i and l_i can be obtained as follows:

$m_1 := \alpha_1$
`for` $i = 1, 2, ..., N-1$:
    $l_i := \gamma_i / m_i$
    $m_{i+1} := \alpha_{i+1} - l_i \beta_i$

Solving $LRu = d$ for u is done in two steps:

Step 2: Forward substitution

We solve $Ly = d$ for y. This gives

$y_1 := d_1$
`for` $i = 2, 3, ..., N$:
    $y_i := d_i - l_{i-1} y_{i-1}$

Step 3: Backward substitution

We solve $Ru = y$ for u. This leads to

$u_N := y_N / m_N$
`for` $i = N-1, N-2, ..., 1$:
    $u_i := (y_i - \beta_i u_{i+1}) / m_i$

This completes the Thomas algorithm. It is stable for every strictly diagonally dominant system matrix. One may also regard it as a recursive filtering: The LR decomposition determines the filter coefficients, Step 2 is a causal filter and Step 3 an anticausal one. The whole scheme is very efficient: it requires only

$$2(N-1) + (N-1) + 1 + 2(N-1) \;=\; 5N - 4 \tag{8.42}$$

multiplications/divisions, and

$$(N-1) + (N-1) + (N-1) \;=\; 3N-3 \tag{8.43}$$

subtractions. Hence the CPU effort is *linear* in N. The same is true for the storage effort.
Applying the Thomas algorithm to the semi-implicit scheme takes almost twice as long as one iteration of the explicit scheme, but we may use much larger time steps, since the scheme is absolutely stable.

8.4 Higher-Dimensional Case

8.4.1 Explicit and Semi-Implicit Schemes

The m-dimensional CLMC equation is given by

$$\partial_t u \;=\; \sum_{l=1}^{m} \partial_{x_l}\Big(g(|\nabla u_\sigma|^2)\;\partial_{x_l} u\Big). \tag{8.44}$$

We can discretize the m summands of the right hand side in the same manner as in the 1-D case. Using only one index for pixel numbering, we may represent the whole image of size $N_1 \times ... \times N_m$ as a vector of size $N := N_1 \cdots N_m$. In this vector–matrix notation we can write the m-dimensional explicit scheme as

$$u^{k+1} \;=\; \Big(I + \tau \sum_{l=1}^{m} A_l(u^k)\Big)\, u^k \tag{8.45}$$

and its semi-implicit counterpart as

$$u^{k+1} \;=\; \Big(I - \tau \sum_{l=1}^{m} A_l(u^k)\Big)^{-1} u^k. \tag{8.46}$$

In both cases the matrix $A_l = (a_{ijl})_{ij}$ corresponds to derivatives along the l-th coordinate axis. Let us also introduce

$$A(u^k) \;:=\; (a_{ij}(u^k)) \;:=\; \sum_{l=1}^{m} A_l(u^k). \tag{8.47}$$

What about the reliability of both schemes? Checking the discrete scale-space requirements (D1)–(D6) can be done in a similar way as in the 1-D case, see [412, Theorem 8] for more details. As in the 1-D case one obtains that the explicit scheme creates a discrete scale-space for

$$\tau \leq \frac{1}{\max\limits_i \sum\limits_{j \neq i} a_{ij}(u^k)}, \tag{8.48}$$

and that the semi-implicit scheme satisfies all requirements unconditionally.
What does this mean regarding efficiency? In the m-dimensional case each inner pixel i_0 has $2m$ neighbours with which it is connected via nonvanishing entries in the i_0-th row of A. From (8.11) we see that we can estimate

$$\sum_{j \neq i} a_{ij}(u^k) \leq \left(\frac{2}{h_1^2} + ... + \frac{2}{h_m^2} \right) \sup_s g(s) \tag{8.49}$$

where h_1,...,h_m denote the dimensions of an m-dimensional pixel. With $h_1 = ... = h_m = 1$ and $\sup\limits_s g(s) = 1$, restriction (8.48) may be replaced by

$$\tau < \frac{1}{2m}. \tag{8.50}$$

Thus, the allowed step size of the explicit scheme becomes even smaller for higher dimensions.
However, this does not necessarily imply that the semi-implicit scheme becomes superior. There appears a new problem as well: Although the actual structure of the matrix A depends on the pixel numbering, it is not possible anymore to order the pixels in such a way that in the i-th row all nonvanishing matix elements can be found within the positions $[i, i-m]$ to $[i, i+m]$: Usually, the matrix reveals a much larger bandwidth. Applying direct algorithms such as Gaussian elimination would destroy the zeros within the band and would lead to an immense storage and computation effort. Hence, iterative algorithms have to be applied. Classical methods like Gauß–Seidel or SOR do not need additional storage and convergence can be guaranteed for the special structure of A. This convergence, however, is rather slow. Faster iterative methods such as the SSOR-CG algorithm [250, pp. 154–161] need significantly more storage, which can become prohibitive for large images. A typical problem of iterative methods is also that their convergence becomes slower for larger τ, since this increases the condition number of the system matrix. Multigrid methods [53] appear to be one possibility to circumvent many of these problems, but their implementation is more complicated.
Recapitulating, we see that for dimensions ≥ 2 the semi-implicit scheme remains absolutely stable, but it is difficult to take full advantage of this because of the problems to solve the arising linear system as efficiently as it was possible in the 1-D case with the Thomas algorithm.

8.4.2 AOS Schemes

In order to address the abovementioned problem let us consider a modification of the semi-implicit scheme (8.46), namely the *additive operator splitting (AOS) scheme*

$$u^{k+1} \;=\; \frac{1}{m} \sum_{l=1}^{m} \Big(I - m\tau A_l(u^k) \Big)^{-1} u^k. \tag{8.51}$$

Several points should be noted:

- The explicit scheme (8.45), the semi-implicit scheme (8.46) and the AOS scheme (8.51) have the same first-order Taylor expansions in τ. It is easy to see that all schemes are $\mathcal{O}(\tau + h_1^2 + \ldots + h_m^2)$ approximations to the continuous equation. From this viewpoint, all schemes are consistent to the original equation. One should not make the mistake to regard the AOS scheme as an algebraically incorrect reformulation of the semi-implicit scheme: The explicit scheme is also different from the semi-implicit one, but it approximates the same continuous diffusion process.

- The operators

$$B_l(u^k) \;:=\; I - m\tau A_l(u^k) \tag{8.52}$$

  describe one-dimensional diffusion processes along the x_l axes. Thus, under a suitable pixel numbering they come down to strictly diagonally dominant tridiagonal matrices which can be inverted in an efficient and stable way by the Thomas algorithm from Section 8.3.2.

- Since it is an *additive* splitting, all coordinate axes are treated in exactly the same manner. This is in contrast to conventional splitting techniques from the literature [111, 241, 250, 439]. They are *multiplicative* splittings such as the *locally one-dimensional (LOD) scheme*

$$u^{k+1} \;=\; \prod_{l=1}^{m} \Big(I - \tau A_l(u^k) \Big)^{-1} u^k. \tag{8.53}$$

  Since in the general nonlinear case the split operators do not commute, the result of multiplicative splittings will depend on the order of the one-dimensional operators. This disadvantage will be discussed in Section 8.6 in more detail.

Does the AOS Scheme Create a Discrete Scale-Space?

The discussed properties suggest that the AOS scheme is an interesting candidate for an efficient discrete diffusion scale-space. Thus, let us now assess its reliability by checking the criteria (D1)–(D6).

Many reasonings carry over from the 1-D semi-implicit scheme: First we observe that $Q_l := (q_{ijl})_{ij} := B_l^{-1}$ exist, since B_l is strictly diagonally dominant. Also the continuity of

$$Q := \frac{1}{m} \sum_{l=1}^{m} Q_l \tag{8.54}$$

in its argument is a direct consequence of the continuous diffusivity g and the construction of A_l.

In the same way the symmetry of Q goes back to the symmetry of A_l. Note that the symmetry of A_l is independent of the pixel numbering: a permutation of their numbering transforms A_l into PA_lP^{-1} for some permutation matrix P. Since $P^{-1} = P^T$ and there exists a pixel numbering such that A_l is transformed into a symmetric tridiagonal matrix just as in the 1-D case, it is clear that A_l has to be symmetric.

With the same reasoning as in 8.3.2, we know that not only B_l, but also Q_l has row sum 1. Thus, Q has also unit row sum.

To verify (D4), we observe that $B_l = (b_{ijl})_{ij}$ is strictly diagonally dominant, $b_{iil} > 0$ for all i, and $b_{ijl} \leq 0$ for $i \neq j$. Under these circumstances we may conclude from [245, p. 192] that $Q_l = B_l^{-1}$ is nonnegative in all components. This implies the nonnegativity of Q.

Let us now check (D5) and (D6) in one step. Since B_l, $l = 1, ..., m$ represent one-dimensional diffusion operators, it follows that there exist permutation matrices P_l, $l = 1, ..., m$ such that $P_l B_l P_l^T$ is not only diagonally dominant, but also tridiagonal and block irreducible[1]. Within each irreducible matrix block, we have a positive diagonal and nonnegative off-diagonals. Applying again Varga's theorem [395, p. 85], we conclude that the inverse of each block contains only positive elements. From this it follows that $a_{i_0 j_0 l} \neq 0$ for some $i_0, j_0 \in J$ implies that $q_{i_0 j_0 l} > 0$. Thus, the irreducibility of $A = \sum_{l=1}^{m} A_l$ carries over to $Q = \frac{1}{m} \sum_{l=1}^{m} Q_l$, and (D6) is satisfied. In particular, since A_l is constructed such that $a_{iil} \neq 0$ for all $i \in J$, it is clear that Q contains only positive diagonal elements. Therefore, (D5) is verified as well.

These discussions show that the AOS scheme creates a discrete nonlinear diffusion scale-space for all time step sizes.

8.4.3 Regularization

This section describes a simple method for calculating the presmoothing $u_\sigma = K_\sigma * u$ in a way which is consistent with the ideas presented above.

[1] Each of the N/N_l blocks represents the pixels where all components except for the $l-$th are identical.

It is well-known that Gaussian convolution with standard deviation σ is equivalent to linear diffusion filtering ($g \equiv 1$) for some time $T = \sigma^2/2$. Thus we may use the (semi-)implicit[2] scheme again in order to obtain a stable algorithm. Several things make the situation even easier than in the nonlinear setting:

- Frequently, σ is in the order of the pixel size. In this case we may regularize in a single step by filtering once with a time step size $T = \sigma^2/2$.

- The linear diffusion process is separable. Therefore, the order of the one-dimensional approximations is not of importance and we may also use a multiplicative splitting:

$$Pu^k := \prod_{l=1}^{m} S_l u^k. \tag{8.55}$$

- The system in step l can be decomposed into N/N_l tridiagonal systems with the same system matrix. Thus, the LR decomposition needs to be done only once for an $N_l \times N_l$-matrix of type

$$\begin{pmatrix} 1+r_l & -r_l & & & \\ -r_l & 1+2r_l & -r_l & & \\ & \ddots & \ddots & \ddots & \\ & & -r_l & 1+2r_l & -r_l \\ & & & -r_l & 1+r_l \end{pmatrix} \tag{8.56}$$

  with $r_l := \tau/h_l^2$. Therefore, the main effort boils down to performing N/N_l times the same forward and backward substitution step from the Thomas algorithm. This requires only $3N-2$ multiplications/divisions and $2N-2$ subtractions. Such an effort is comparable with the recursive filters presented in [14, 8, 444], but unlike those Fourier-based methods, the algorithm presented here allows an adequate treatment of the reflecting boundary conditions and preserves the average grey value.

8.5 Algorithmic Structure

8.5.1 AOS Algorithm

We may summarize our considerations in the following algorithm for one AOS step in m dimensions.

[2] Semi-implicit and implicit are identical in the linear case.

input: $u = u^n$

regularization: $v := K_\sigma * u$
(according to 8.4.3)

calculate diffusivity $g(|\nabla v|^2)$
(approximate ∇v by central differences)
(use look-up table for evaluating g)

create copy: $f := u$
initialize sum: $u := 0$

for $l = 1, ..., m$:
    calculate $v := (mI - m^2 \tau A_l)^{-1} f$
    (solve N/N_l tridiagonal systems of
    size N_l with Thomas algorithm 8.3.2)
    update: $u := u + v$

output: $u = u^{n+1}$

8.5.2 Complexity

In order to assess the complexity of AOS algorithms, let us consider dimensions $m \geq 2$ and focus on terms of order N.
From the preceding algorithm we recognize that only the four vectors u, v, g, and f are required. Thus, since all calculations may be performed in single precision, the main storage effort is $4N \times 4$ Bytes. This is independent of the dimension m.
Table 8.1 summarizes the relevant computational requirements for each step of the AOS algorithm. We observe that the effort is proportional to the number N of pixels and the dimension m. The total effort is only $11mN$ multiplications or divisions, $(10m-1)N$ additions or subtractions, and N look-ups in a table. This is less than twice the typical effort needed for an explicit scheme, a rather low price for gaining absolute stability.

8.6 Evaluation

We have seen that AOS schemes with large time steps still reveal average grey value invariance, stability based on extremum principle, Lyapunov functionals, and convergence to a constant steady-state. Thus, they are legitimate when being considered as a pure discrete process which is not intended to approximate a continuous process.
But does this mean that it is recommendable to consider arbitrarily large time

Table 8.1: Main operations for one m-dimensional AOS step (M/D: multiplications or divisions; A/S: additions or subtractions; LUT: look-up operations in a table).

task	M/D	A/S	LUT
regularization	3mN	2mN	
calculate $\|\nabla v\|^2$	2mN	(2m-1)N	
calculate diffusivity			N
create system matrix	mN	3mN	
Thomas algorithms	5mN	3mN	
total	11mN	(10m-1)N	N

step sizes? In the extreme case: can one filter an image in one step?

In this case we should expect problems with those properties which a naturally linked to continuous ideas and which can only be satisfied approximately by discrete schemes: rotational invariance and accuracy.

A loss of rotational invariance becomes visible as a preference of certain directions, while a loss of accuracy becomes evident in those cases where filtering with time step $n\tau$ differs visually from n times filtering with τ. So let us now check these approximation effects by applying a 2-D AOS scheme to two test images.

First we check the rotational invariance. Since the AOS scheme is consistent to the original equation, we should expect good rotational invariance for small spatial and temporal steps.

Figure 8.1 is used as a test for rotational invariance. It depicts a Gaussian-like image and its filtered versions. For $\tau = 0.25$ both the explicit and AOS scheme are visually undistinguishable. This step size is also the stability limit for the explicit scheme, while the AOS scheme allows to increase τ further. We see that for $\tau \leq 5$ no significant changes appear. Thus, AOS may be used with 20 times larger time steps than the explicit scheme. On the other hand, even for $\tau = 20$ the deviations from a perfect circular structure are not very severe.

What about the accuracy? Figure 8.2 depicts the filtering of a brain image. The situation is similar as in Figure 8.1: For $\tau = 0.25$ the explicit and the AOS scheme are undistinguishable. The AOS scheme remains close to these results up to $\tau \approx 5$. For $\tau = 20$ we get more severe deviations: the filtering effect becomes weaker. This is a typical behaviour for implicit schemes with large time steps: implicit techniques always remain on the "save" side (by orienting the diffusion on the "smoother" future rather than on the "rougher" past), while their deviation from the true solution becomes larger with increasing step size. Thus, their filtering effect on the final image at a specified time decreases with increasing time step size. Again $\tau = 5$ is a good compromise between efficiency and accuracy.

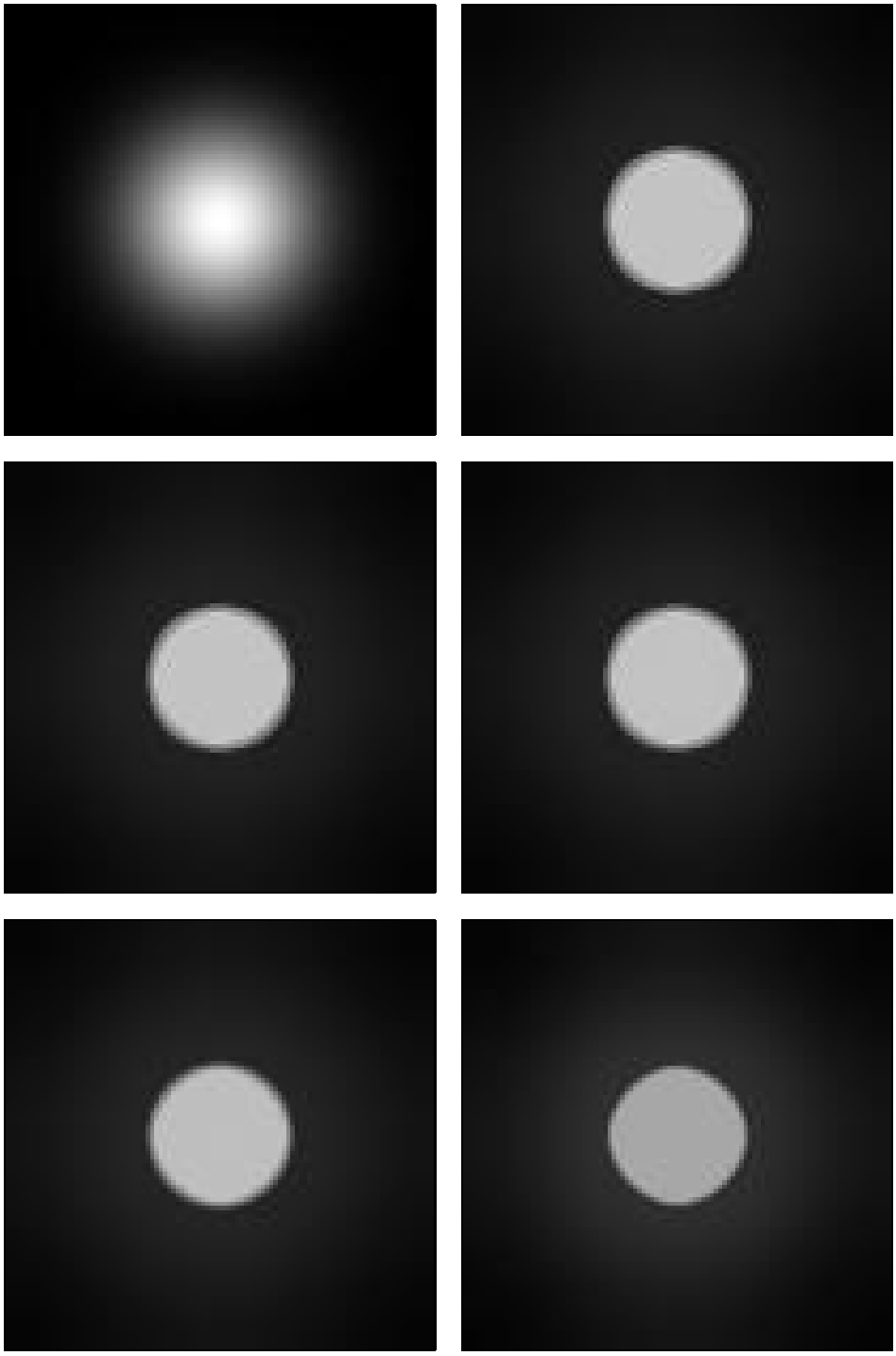


Figure 8.1: Nonlinear diffusion filtering of a Gaussian-like test image ($\lambda = 8$, $\sigma = 1.5$). (a) TOP LEFT: Original image, $\Omega = (0, 101)^2$. (b) TOP RIGHT: Explicit scheme, 800 iterations, $\tau = 0.25$. (c) MIDDLE LEFT: AOS scheme, 800 iterations, $\tau = 0.25$. (d) MIDDLE RIGHT: AOS scheme, 200 iterations, $\tau = 1$. (e) BOTTOM LEFT: AOS scheme, 40 iterations, $\tau = 5$. (f) BOTTOM RIGHT: AOS scheme, 10 iterations, $\tau = 20$.

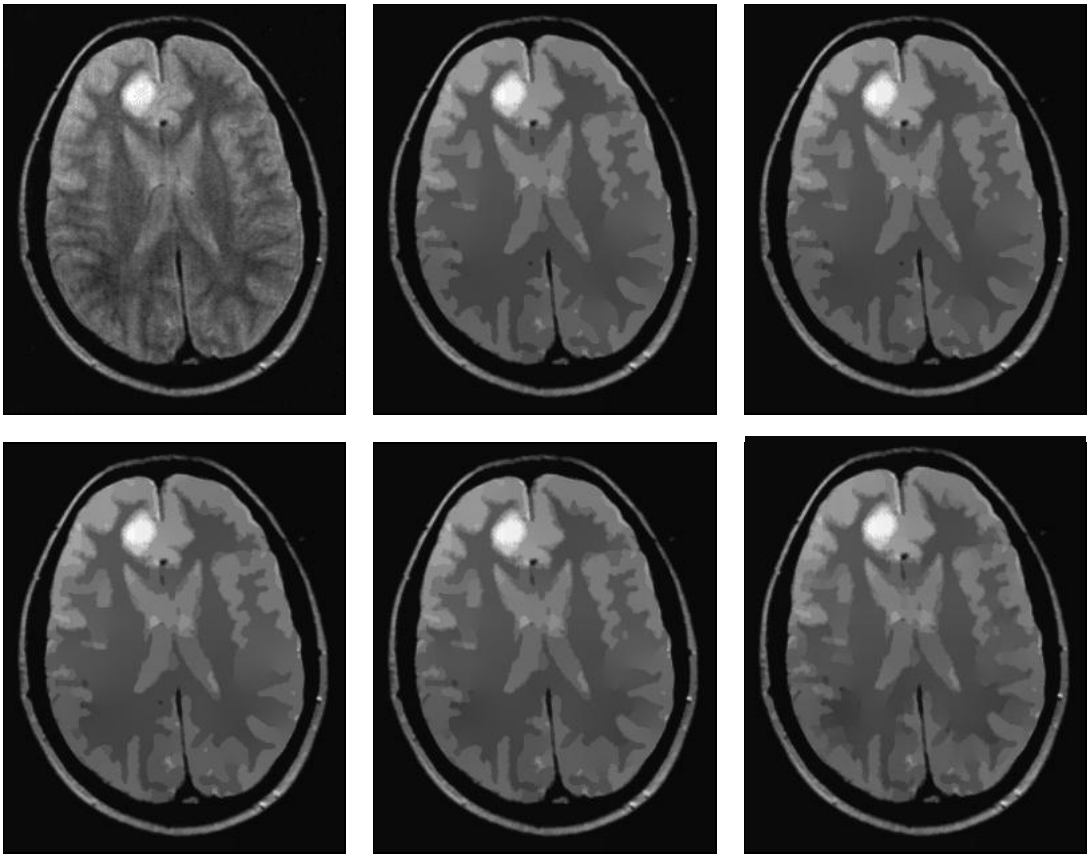


Figure 8.2: Nonlinear diffusion filtering of a medical image ($\lambda = 2$, $\sigma = 1$). (a) TOP LEFT: Original image, $\Omega = (0, 255) \times (0, 308)$. (b) TOP MIDDLE: Explicit scheme, 800 iterations, $\tau = 0.25$. (c) TOP RIGHT: AOS scheme, 800 iterations, $\tau = 0.25$. (d) BOTTOM LEFT: AOS scheme, 200 iterations, $\tau = 1$. (e) BOTTOM MIDDLE: AOS scheme, 40 iterations, $\tau = 5$. (f) BOTTOM RIGHT: AOS scheme, 10 iterations, $\tau = 20$.

After these visual inspections we shall investigate the accuracy more quantitatively. To this end we perform a comparison between the explicit scheme (8.45), the semi-implicit scheme (8.46) and the AOS scheme (8.51). Since no analytical solution to the CLMC equation is known, we have to use a good numerical approximation to a test example as a standard for comparison. In our case we took the explicit scheme with the small step size $\tau = 0.1$ and applied it 2000 times to the test image from Fig. 8.2.

The linear system of the 2-D semi-implicit scheme is solved by a *Gauß–Seidel algorithm*. Iterative methods of this type are quite popular for nonlinear PDEs in image processing [8, 271], since they are easy to implement and they do not require additional memory. Let the diffusion operator

$$B(u^k) \;:=\; I - \tau \sum_{l=1}^{m} A_l(u^k) \tag{8.57}$$

be decomposed into the strictly lower triangular matrix L, the diagonal matrix D, and the strictly upper triangular matrix U. Then the Gauß–Seidel method approximates the solution u^{k+1} of the semi-implicit scheme

$$B(u^k)\, u^{k+1} \;=\; u^k \tag{8.58}$$

by a sequence of vectors $y^{(n)}$ with

$$y^{(0)} \;=\; u^k, \tag{8.59}$$

$$y^{(n+1)} \;=\; (L+D)^{-1}\,(u^k - U y^{(n)}) \qquad (n \geq 0). \tag{8.60}$$

Every second step we calculate the residue

$$r^{(n)} := B y^{(n)} - u^k, \tag{8.61}$$

and we stop the iteration process if its l^2 norm $\|r^{(n)}\|_2 = (\sum_i |r_i^{(n)}|^2)^{1/2}$ satisfies

$$\|r^{(n)}\|_2 \;\leq\; \alpha \|r^{(0)}\| \tag{8.62}$$

with some accuracy parameter $\alpha = 0.01$ or $\alpha = 0.1$.

Table 8.2 compares the explicit scheme, the semi-implicit scheme with accuracies $\alpha = 0.01$ and $\alpha = 0.1$, respectively, and the AOS scheme.

If v denotes our reference solution (explicit scheme, $\tau = 0.1$), then we calculate the relative l^2 error of an approximation u as

$$\frac{\|u - v\|_2}{\|v\|_2}. \tag{8.63}$$

First we observe that the explicit scheme with $\tau = 0.25$ reveals a very small error, while the semi-implicit method with $\alpha = 0.01$ is not only less accurate, but also

Table 8.2: Comparison of nonlinear diffusion schemes.

scheme	τ	CPU time	rel. l^2 error
explicit	0.25	65.65 s	0.14 %
semi-implicit	0.25	145.65 s	0.27 %
($\alpha = 0.01$)	0.5	92.62 s	0.68 %
	1	62.97 s	1.05 %
	2	43.58 s	1.46 %
	5	27.62 s	1.97 %
	10	19.49 s	2.26 %
	20	12.45 s	2.74 %
	50	5.54 s	3.77 %
semi-implicit	0.25	106.43 s	0.83 %
($\alpha = 0.1$)	0.5	72.04 s	1.00 %
	1	36.70 s	1.40 %
	2	23.53 s	1.83 %
	5	12.30 s	2.51 %
	10	7.00 s	3.27 %
	20	3.53 s	4.19 %
	50	1.17 s	5.43 %
AOS	0.25	114.04 s	0.73 %
	0.5	56.55 s	1.32 %
	1	28.56 s	1.66 %
	2	14.25 s	1.83 %
	5	5.80 s	2.22 %
	10	2.95 s	2.73 %
	20	1.52 s	3.37 %
	50	0.67 s	4.29 %

slower for $\tau = 0.25$ and 0.5. For $\tau \geq 1$ the semi-implicit scheme becomes faster than the explicit one. On the other hand, the Gauß-Seidel algorithm slows down for larger τ, since this increases the condition number of the system matrix. Hence, the overall CPU time per semi-implicit step increases with increasing τ. If we relax the accuracy from $\alpha = 0.01$ to $\alpha = 0.1$, the semi-implicit scheme becomes faster, but the l^2 error also increases. For $\tau > 0.5$, the AOS scheme becomes the fastest method. Interestingly, for $\tau > 2$, it is also more accurate than the semi-implicit scheme with $\alpha = 0.1$.

It is worth noticing that there is a fundamental difference between errors in the AOS scheme and errors which are introduced by an insufficient number of Gauß–Seidel iterations: unlike AOS errors which are compatible with the discrete scale-space framework, Gauß–Seidel errors can violate these requirements. Thus, properties such as the average grey level invariance are no more satisfied in an exact manner. In order to avoid these difficulties, one would have to apply more Gauß–Seidel iterations, which will finally destroy all efficiency advantages compared to the explicit scheme; see also [271].

Figure 8.3 gives a graphical representation of Table 8.2, which allows us to find the most efficient schemes for a desired accuracy. We observe that for very high accuracy requirements the explicit scheme is most appropriate[3]. This is at the expense of a hight overall computational effort. On the other hand, even relaxing the accuracy requirements to a relative l^2 error of 1 % does not permit to find a more efficient technique. For errors between 1 % and 1.7 % the semi-implicit scheme with $\alpha = 0.1$ is fastest, and for errors larger than 1.7 % AOS schemes become rapidly superior. In our previous experiments we have observed that the accuracy of AOS with $\tau = 5$ appears to be tolerable for many applications. This corresponds to an error of about 2.2 %. In this case, AOS is almost 2.5 times more efficient than the semi-implicit scheme with $\alpha = 0.1$, more than 3.5 times faster than the semi-implicit scheme with $\alpha = 0.01$, and about 11 times more efficient than the explicit scheme. Although these relations have been illustrated by one example only, additional experiments have indicated that these basic relations between explicit, semi-implicit and AOS discretizations carry over to a large class of images: The accuracy requirements of many practical problems allow an efficiency gain by one order of magnitude. All one has to do is to replace the explicit scheme by an AOS scheme with 20 times larger time step sizes.

It should be noted that the AOS scheme calculate the average of operators of type $(I - m\tau A_l)^{-1}$. They describe 1-D diffusions with a step size $m\tau$. Since multiplicative splittings such as the LOD scheme (8.53) use operators of type $(I - \tau A_l)^{-1}$, one can expect that they give even better accuracy. However, multiplicative splittings

[3] One can achieve even higher accuracy by methods which are of second order in time, for instance predictor–corrector techniques [412]. Such a high accuracy, however, is rarely required in image processing.

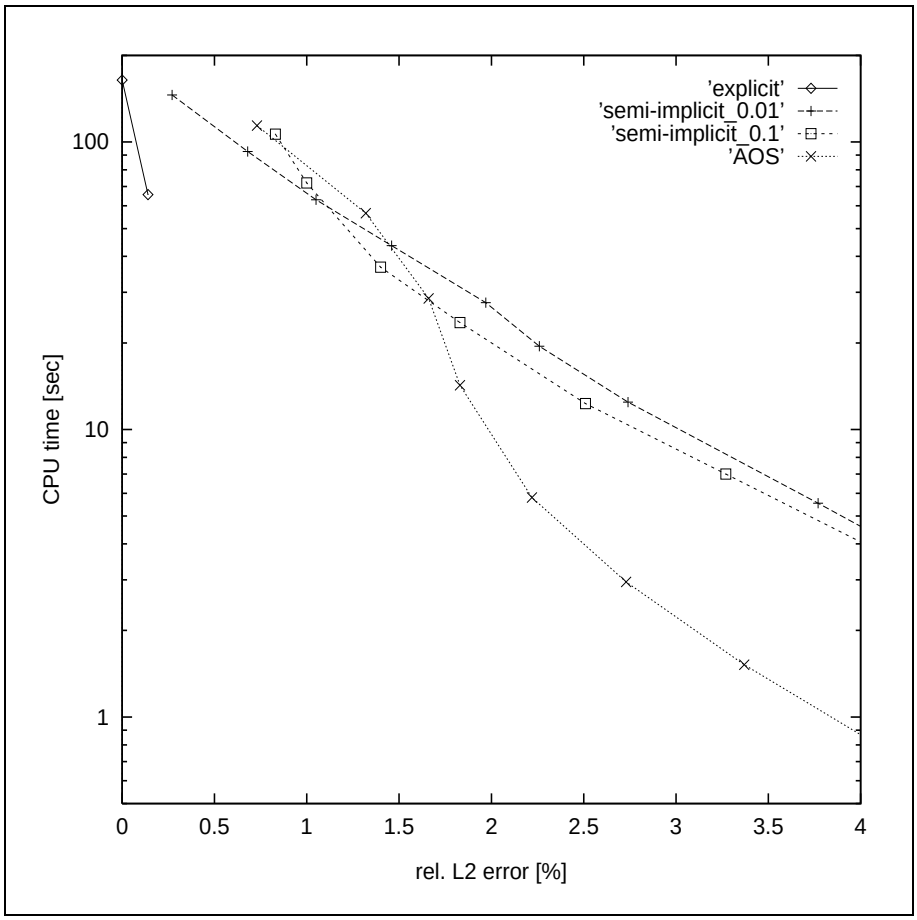


Figure 8.3: Tradeoff between efficiency and accuracy of nonlinear diffusion solvers. The data were calculated on the test image from Fig. 8.2, size $\Omega = (0, 255) \times (0, 308)$. Filter parameters: $\lambda = 2$, $\sigma = 1$. Stopping time: $T = 200$. Hardware: one R10000 processor on an SGI Challenge XL.

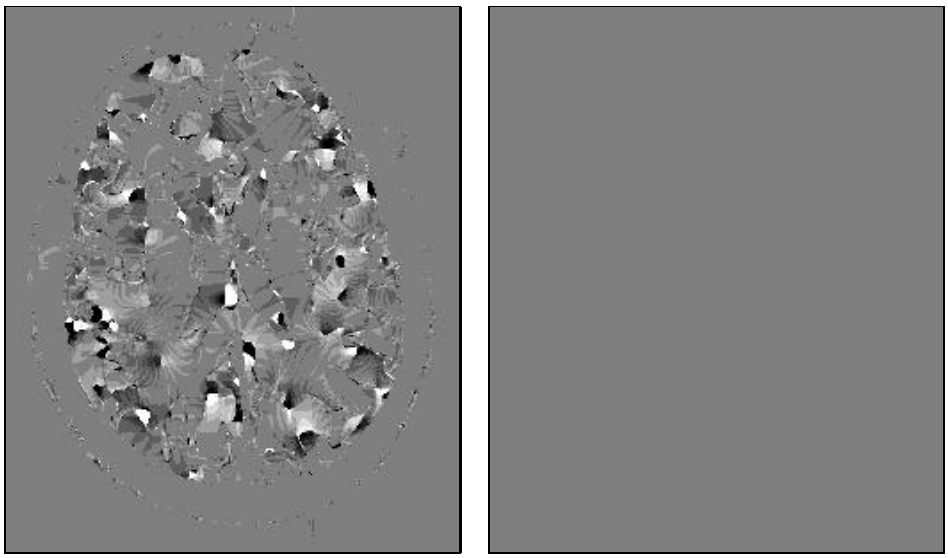


Figure 8.4: (Non-)Commutation of nonlinear diffusion operators. The difference between filtering prior to rotation by 90 degrees, and rotation prior to filtering is depicted. Test image: Figure 8.2 ($\lambda = 2$, $\sigma = 1$, $\tau = 20$, 10 iterations). (a) LEFT: A multiplicative splitting such as LOD treats x and y axes differently. (b) RIGHT: Additive operator splitting (AOS) treats all axes equally.

for nonlinear problems reveal one big disadvantage, which makes their use in many image processing applications problematic: In the general nonlinear case the split operators do not commute any longer. Thus, the result of multiplicative splittings depends on the order of the one-dimensional operators, and the grid axes are treated differently. In practice, this means that these schemes produce different results if the image is rotated by 90 degrees. Such an undesirable effect is illustrated in Fig. 8.4. Since AOS schemes apply the 1-D operators in parallel instead of sequentially, they do not suffer from this limitation.

Moreover, most multiplicative splittings lead to a nonsymmetric system matrix $Q(u^k)$. This violates criterion (D2) for discrete diffusion scale-spaces. For this reason, we have not considered these approaches in the present paper.

Finally we check the relation between the computational effort and the number of pixels. Table 8.3 shows the measured CPU times on a single R10000 processor of an SGI Challenge XL and on an HP 900-755, both for 2-D and 3-D images.

For small image sizes the computing times reveal good proportionality to the overall number of pixel. This is what we expect from theory. Because of Cache limitations, the CPU time per pixel becomes slightly higher for huger data sets: We also observe that this deviation from the linear scaling behaviour is machine dependent. The HP remains closer to the linear scaling behaviour than the SGI.

Table 8.3: Measured CPU times for one AOS iteration.

image size	SGI	HP
64×64	0.0086 s	0.0168 s
128×128	0.0324 s	0.0676 s
256×256	0.134 s	0.307 s
512×512	0.711 s	1.63 s
1024×1024	6.35 s	7.38 s
2048×2048	27.8 s	34.0 s
4096×4096	145 s	
8192×8192	724 s	
$16 \times 16 \times 16$	0.0159 s	0.0331 s
$32 \times 32 \times 32$	0.116 s	0.304 s
$64 \times 64 \times 64$	1.15 s	2.45 s
$128 \times 128 \times 128$	13.4 s	20.1 s
$256 \times 256 \times 256$	237 s	
$512 \times 512 \times 256$	1340 s	

On the other hand, with its CPU memory of 1 GByte the SGI permits even to process data sets of size 8192×8192 and $512 \times 512 \times 256$.

Three-dimensional data sets from medicine with typical sizes such as $256 \times 256 \times 64$ can be processed in less than one minute per AOS iteration (both on the HP and the SGI). In many practical applications less than 10 iterations are sufficient for the denoising of such data sets.

Recapping we have observed that – although the desired approximation quality is of course purpose dependent – under typical circumstances 20 times larger step sizes than the stability limit of the explicit scheme appear reasonable. They give an efficiency gain of a factor ten[4]. Especially for large data sets such as 3-D medical data this is often the difference between not applicable and applicable. We are currently testing our schemes for the filtering of 3-D ultrasound images and preprocessing 3-D MR data for segmentation. In both cases first results are encouraging.

[4]We have seen that an m-dimensional AOS scheme averages 1-D operators with an effective step size of $m\tau$. Thus, for higher dimensions m one should reduce the step size in order to have the same accuracy. However, since explicit schemes also have to decrease the step size for larger m in the same way, the factor 10 remains valid for every dimension.

8.7 Conclusions

We have presented absolutely stable additive operator splitting (AOS) schemes for the nonlinear diffusion filter of Catté *et al.* and Whitaker and Pizer. These schemes satisfy all criteria for discrete nonlinear diffusion scale-spaces and are easy to implement in any dimension. Both computational and storage effort is linear in the number of pixels. Experiments have shown that under realistic accuracy requirements one can gain an increase of efficiency by a factor 10. This makes this type of schemes attractive for applications such as medical 3D data sets.
Implementations of AOS schemes on parallel architectures are studied in [430]. These experiments demonstrate that it is possible to gain a speed-up by another order of magnitude by exploiting the intrinsic parallelism of AOS schemes. Last but not least, there are also ways to generalize AOS schemes to anisotropic diffusion filters with diffusion tensors; a first proposal in this direction can be found in [412, Section 4.4.2].

Acknowledgements. We would like to thank Robert Maas for careful proofreading and helpful hints.

Chapter 9

Fast Parallel Algorithms for a Broad Class of Nonlinear Variational Diffusion Approaches

Joachim Weickert, Josef Heers, Christoph Schnörr, Karel J. Zuiderveld, Otmar Scherzer, and H. Siegfried Stiehl.
Real-Time Imaging, in press.

Abstract

Variational segmentation and nonlinear diffusion approaches have been very active research areas in the fields of image processing and computer vision during the last years. In the present paper, we review recent advances in the development of efficient numerical algorithms for these approaches. The performance of parallel implementations of these algorithms on general–purpose hardware is assessed. A mathematically clear connection between variational models and nonlinear diffusion filters is presented that allows to interpret one approach as an approximation of the other, and vice versa. Extending this continuous connection to the fully discrete setting enables us to derive many structural similarities between efficient numerical algorithms for both frameworks. These results provide a perspective for uniform implementations of nonlinear variational models and diffusion filters on parallel architectures.

9.1 Introduction

Variational approaches for image processing and computer vision have been the subject of considerable interest during the last years. A growing community has focused on various research problems including the mathematical foundations of image segmentation [253], nonlinear regularization approaches to image restoration [326, 142], stochastic modeling of spatial context [225, 435], and in on wide range of corresponding applications [299, 300]. In each case, the approach is formulated in terms of an energy functional which precisely specifies the criteria being used for assessing the output of a processing stage. Often results from various branches of mathematics like functional analysis, convex optimization, or numerical mathematics, for example, can successfully be applied to clarify the properties of a variational approach. As a result, variational modeling has contributed much to the fields of image processing and computer vision during the last years.

Parallel to this development, the application of partial differential equations (PDEs) to image processing has been a very active field of research [382, 381, 412, 70, 274]. Research problems that motivated corresponding work include nonlinear extensions to the linear scale–space paradigm [12], corresponding curve evolutions [333], active contours and surfaces [67, 239], and sound mathematical models for image enhancement through local backward diffusion. For reviews, we refer to [411, 412].

In many cases, these approaches are superior to conventional existing image processing methods and thus provide new perspectives for various application areas like, for example, medical image analysis. Consequently, there is an increasing interest in efficient numerical algorithms enabling implementations of these computationally demanding approaches that work at acceptable computing speed. For example, a challenging task concerns the fast processing of 3D medical image data with a PDE–based approach such that interactivity becomes feasible.

In the present paper, we review recent advances in the development of efficient numerical algorithms for both variational modeling approaches [164, 165] and PDE–based approaches [429, 430] to nonlinear adaptive image processing. Numerical experiments with parallel implementations of these algorithms on general–purpose hardware are reported. Although the underlying mathematical models appear to be quite different (non–quadratic minimization problems vs. nonlinear evolution equations), we show that they are related and can lead to very similar algorithms. One important tool to understand these close relations is a connection which has recently been exploited in [341, 311] to establish scale-space properties for regularization methods. In the present paper we extend this close relation for the continuous case to discrete formulations by deriving various relations between efficient numerical algorithms for both frameworks. It is our hope that such a unified description of the structural similarities may also shed light on how the algorithmic advances reported may get more interrelated in future work.

The remainder of this paper is organized as follows. Section 9.2 reviews non-quadratic variational models for adaptive image processings and discusses a connection to approaches based on nonlinear parabolic evolution equations. Section 9.3 is devoted to efficient numerical approximations by means of finite elements and finite difference methods. It describes a linearization technique as well as corresponding algorithms suited to compute minimizing functions efficiently. Furthermore it is shown how similar ideas can be used for diffusion filtering, and how they can be modified towards an even more efficient additive splitting scheme. Numerical experiments with parallel implementations of our algorithms using general–purpose hardware are reported in Section 9.4. We conclude with pointing out possible directions of further research in Section 9.5.

9.2 Continuous Formulations

9.2.1 Variational Approaches

In this section, we sketch the mathematical formulation of a variational approach for adaptive image processing. For a more detailed account we refer to [348].
Let $g : x \in \Omega \to \mathbb{R}$ denote the grey-value function of a given image defined over an m-dimensional image domain $\Omega \subset \mathbb{R}^m$. The goal is to find a filtered version u of the original image g such that u minimizes a suitable energy functional which precisely specifies the optimization criterion of our processing step. We consider minimization problems of the following form:

$$u = \arg\inf_{v \in \mathcal{H}} J(v) \,, \quad J(v) = \frac{1}{2} \int_\Omega \Big\{ (v-g)^2 + \alpha \cdot \lambda(|\nabla v|) \Big\} dx \tag{9.1}$$

where $\mathcal{H}$ is some Hilbert space and $\alpha > 0$ serves as a weight factor. The function λ depends on the gradient of admissible functions v and is piecewise defined:

$$\lambda(s) = \begin{cases} \lambda_{low}(s) & (0 \le s \le c) \\ \lambda_{high}(s) & (s > c). \end{cases}$$

Typical definitions include $\lambda_{low}(s) \sim s^2$ and $\lambda_{high}(s) \sim s$ or $\lambda_{high}(s) \sim const.$ (cf., e.g., [48, 140, 47]). Accordingly, the functional (9.1) can be written in a form which is easier to interpret:

$$J(v) = \frac{1}{2} \Big[\int_\Omega (v-g)^2 \, dx + \alpha \int_{\{x:|\nabla v| \le c\}} \lambda_{low}(|\nabla v|) \, dx + \alpha \int_{\{x:|\nabla v| > c\}} \lambda_{high}(|\nabla v|) \, dx \Big]. \tag{9.2}$$

The first term of the right hand side of (9.2) measures the similarity between functions v and given image data g. The second term measures the smoothness of functions v within regions with a low gradient of v. The third term measures properties

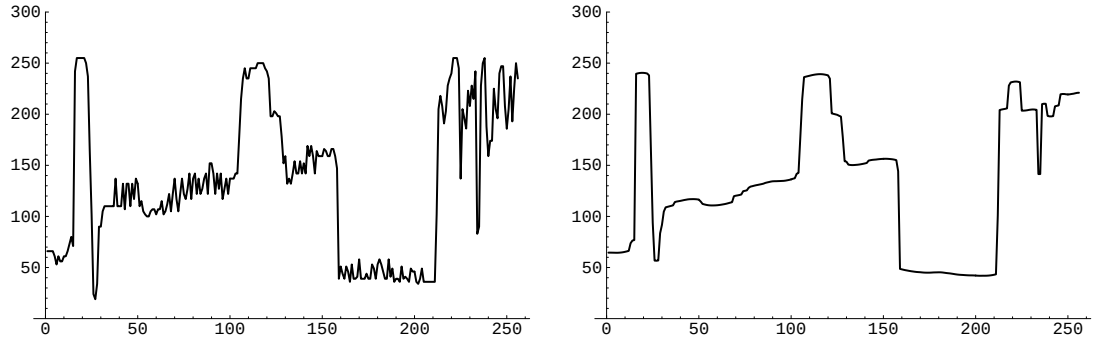


Figure 9.1: Nonlinear adaptive smoothing of a 1D-signal

of functions v where v refers to local features in terms of significant variations. For example, using $\lambda_{high}(s) \sim s$, this term measures the length of iso–contour lines of v summed up over the local contrast (cf. [347]). As a result, algorithms minimizing the functional (9.1) lead to an approximation of the given image function g with a piecewise smooth function u. Since the domains of integration in (9.2) depend on the solution u itself, the corresponding local smoothing process generating u from g is nonlinear and adaptive. Figures 9.1 and 9.2 show examples.
Vanishing of the first variation of the functional J in (9.1) yields a variational equation determining minimizing functions u:

$$L(u,v) := \int_\Omega \Big\{ (u-g)v + \alpha\, \rho(|\nabla u|) \nabla u \cdot \nabla v \Big\} dx = 0 \qquad \forall v \in \mathcal{H} \tag{9.3}$$

where the so–called diffusivity function $\rho(\cdot)$ is derived from $\lambda(\cdot)$ as

$$\rho(s) = \frac{\lambda'(s)}{2s}. \tag{9.4}$$

Partial integration in (9.3) shows that u is also a (weak) solution to the Euler equation

$$\alpha \,\mathrm{div}\, (\rho(|\nabla u|)\nabla u) - (u-g) = 0 \tag{9.5}$$

with reflecting (homogeneous Neumann) boundary conditions:

$$\frac{\partial u}{\partial n} \;=\; 0 \quad \text{on } \partial\Omega, \tag{9.6}$$

where n denotes the normal to the image boundary $\partial\Omega$.

9.2.2 Related Diffusion Filters

Diffusion Filters with Monotone Fluxes

Interestingly, the preceding variational approach has strong connections to nonlinear diffusion filtering. This shall be explained in the sequel.

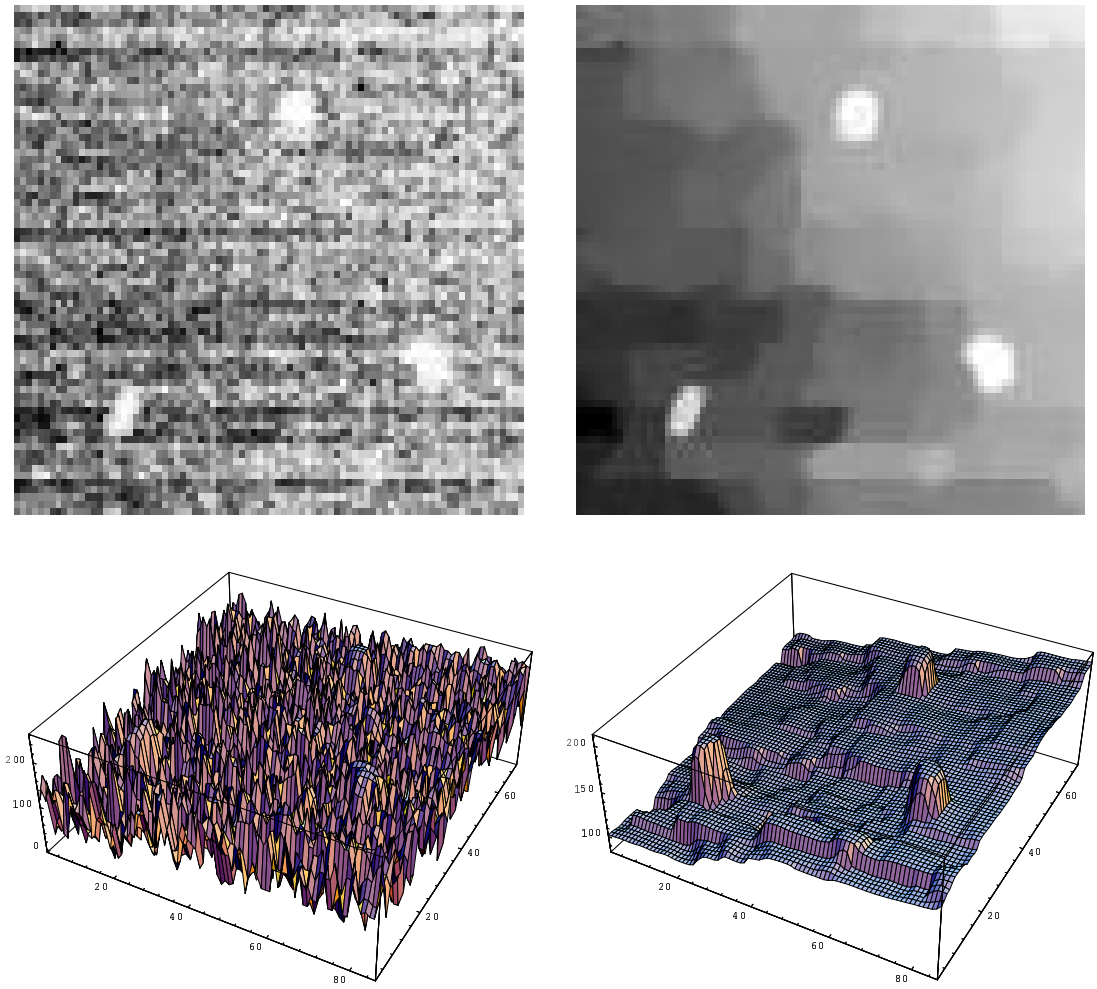


Figure 9.2: TOP: Variational restoration of a mammography. BOTTOM: The images from above as 3D–plots.

A simple nonlinear diffusion filter can be constructed as follows. A processed version $u(x,t)$ of the original image $g(x)$ is obtained by solving the diffusion equation

$$\partial_t u \;=\; \mathrm{div}\left(\rho(|\nabla u|)\,\nabla u\right) \tag{9.7}$$

with the original image as initial state,

$$u(x,0) \;=\; g(x), \tag{9.8}$$

and reflecting boundary conditions:

$$\frac{\partial u}{\partial n} \;=\; 0 \quad \text{on } \partial\Omega. \tag{9.9}$$

The "time" t is a scale parameter: larger values lead to "simpler" image representations. One can assure well-posedness of the diffusion filter if the diffusivity $\rho(s)$ creates a flux $\rho(s)\cdot s$ that is monotonically increasing for $s \geq 0$. Such diffusivities ρ arise from equation (9.4) when the smoothness potential λ is convex.
By writing the Euler equation (9.5) as

$$\frac{u-g}{\alpha} \;=\; \mathrm{div}\left(\rho(|\nabla u|)\nabla u\right) \tag{9.10}$$

it becomes clear that variational image restoration with regularization parameter α approximates a diffusion filter at time α: indeed, (9.10) is nothing more than an implicit time discretization of the diffusion equation (9.7). This relation has been exploited in [341, 311] for establishing various relations between variational image restoration and nonlinear diffusion filtering.
One would expect that this discretization is of higher accuracy for smaller α. However, Figure 9.3 shows that also for rather large α, the approximation remains surprisingly good. In this example we have used a diffusivity of type

$$\rho(s) = \frac{1}{\sqrt{1+s^2/\nu^2}} \tag{9.11}$$

with $\nu = 2$. It can be derived from the convex potential

$$\lambda(s) = 2\nu^2\sqrt{1+s^2/\nu^2}. \tag{9.12}$$

This example illustrates that *variational image restoration and nonlinear diffusion filtering are basically equivalent.* As a consequence, efficient numerical methods for one of these paradigms can also be used for the other.
We note that the diffusivity (9.11) is a decreasing function. As a consequence diffusion is small at those locations where the gradient is large which in turn reduces blurring at edges.

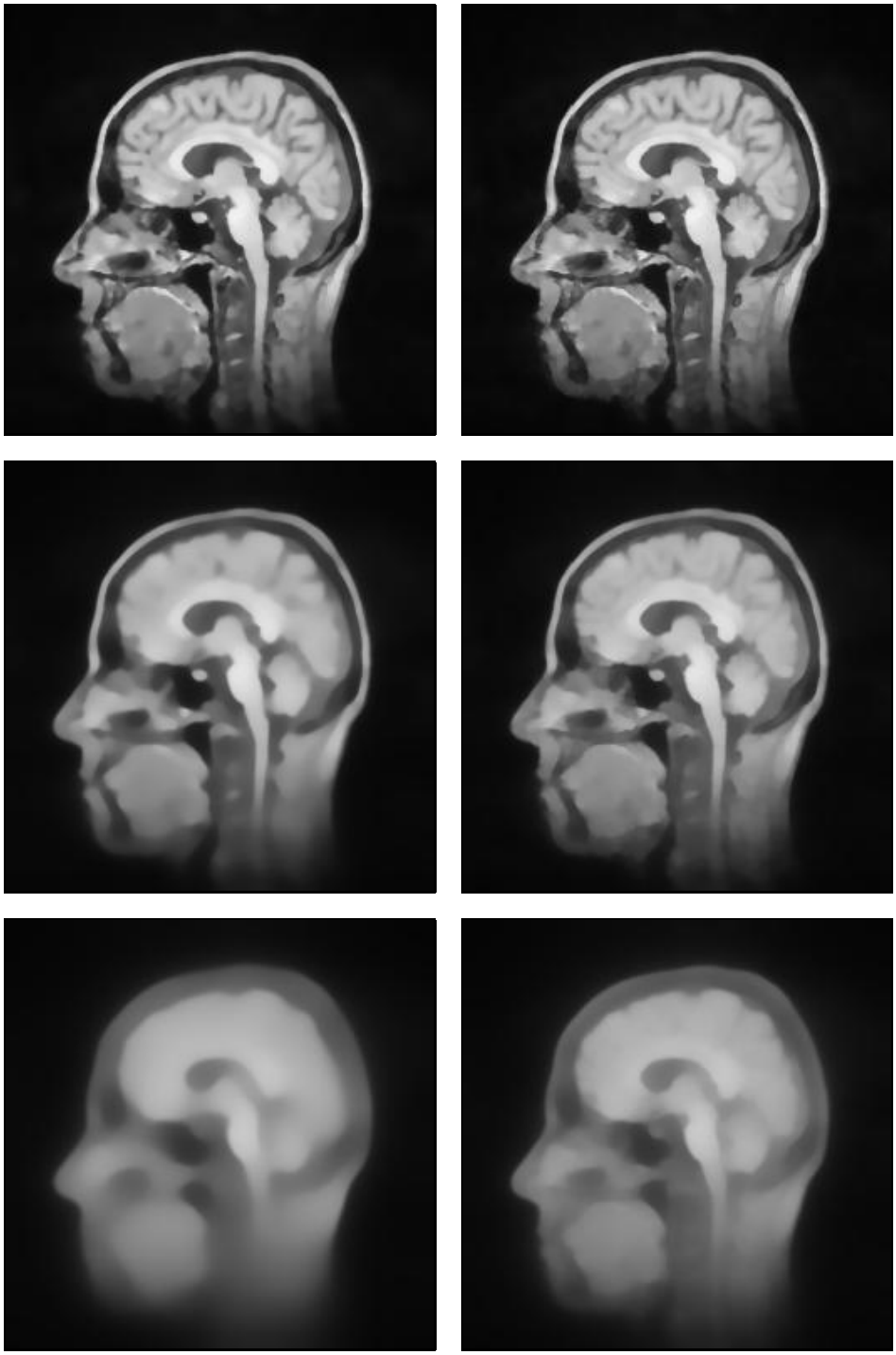


Figure 9.3: Comparison between the nonlinear diffusion filter (9.7) and the variational image restoration (9.10). (a) LEFT COLUMN: Temporal evolution of the diffusion filter, $t = 10$, 30, 100. (b) RIGHT COLUMN: Variational restoration with regularization parameters $\alpha = 10$, 30, 100.

Diffusion Filters with Nonmonotone Fluxes

Intuitively, one would expect that edges are better preserved if the diffusivity decreases more rapidly. For instance, a diffusivity of type

$$\rho(s) := \begin{cases} 1 & (s = 0) \\ 1 - \exp\left(\frac{-3.315}{(s/\nu)^8}\right) & (s > 0) \end{cases} \tag{9.13}$$

decreases more rapidly than (9.11).
However, the corresponding potential λ is only convex for gradient magnitudes that do not exceed ν: the factor 3.315 ensures that the flux $= s\rho(s)$ is increasing for $|s| \leq \nu$ and decreasing for $|s| > \nu$. Thus, ν is a contrast parameter separating low-contrast regions with (smoothing) forward diffusion from high-contrast locations where backward diffusion may enhance edges [303]. In this case, we should not expect the existence of a unique solution which is stable.
However, it is possible to regularize diffusion filters with nonmonotone fluxes in such a way that they become well-posed [74]. This can be achieved by replacing the edge detector $|\nabla u|$ in $\rho(|\nabla u|)$ by a Gaussian derivative based edge detector $|\nabla u_\sigma|$:

$$\nabla u_\sigma := \nabla(K_\sigma * u), \tag{9.14}$$

$$K_\sigma := \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \tag{9.15}$$

After some time this filter creates segmentation-like results which are piecewise almost constant. This is illustrated in Figure 9.4. For $t \to \infty$, however, the image converges to a constant grey-value function [412]. Well-posedness results for this filter can be found in [74, 412] and a scale-space interpretation is given in [412].
Because of the Gaussian convolution there is no straightforward way to express such regularized diffusion filters as minimizers of energy functionals. However, there are possibilities to approximate them by a sequence of quadratic variational approaches; see [311] for details. Regularized filters with nonmonotone fluxes offer additional perspectives in terms of contrast enhancement, and they can be treated with the same numerical techniques that are used for unregularized diffusion filters with monotone fluxes.

9.3 Discrete Formulations

9.3.1 Variational Approach

Linearization

The goal of this section is to replace the nonlinear equation (9.3) by a sequence of linear equations. Their solution methods are well understood and suited for parallel

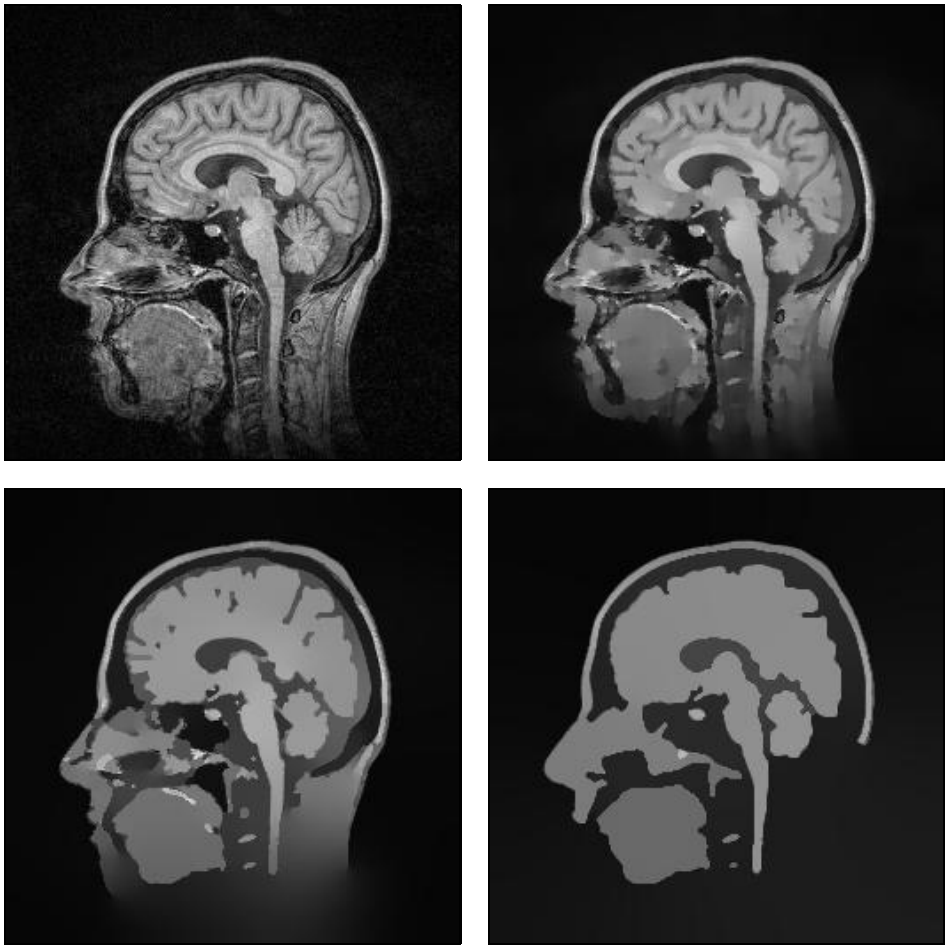


Figure 9.4: Temporal evolution of the regularized nonlinear diffusion filter with nonmonotone flux function.

implementations. In the following, we describe a particular linearization technique introduced to image processing by Geman and Reynolds [141] and further studied in [81, 165]. For alternative linearization techniques and a comparison within a more general framework we refer to [165].
The key idea is to replace the non-quadratic functional (9.1) – from which equation (9.3) is derived – by a sequence of quadratic functionals. Following Geman and Reynolds [141], we introduce an auxiliary function w and consider the problem of minimizing

$$\tilde{J}(v,w) = \frac{1}{2}\int_\Omega \Big\{(v-g)^2 + \alpha\cdot\big(w|\nabla v|^2 + \psi(w)\big)\Big\}dx \tag{9.16}$$

where the function $\psi(\cdot)$ depends on $\lambda(\cdot)$ in (9.1) and is chosen such that $\tilde{J}(v,w)$ is convex in w, and

$$J(v) = \inf_w \tilde{J}(v,w).$$

The minimization of (9.16) is accomplished by the following two–step procedure (k denotes the iteration index):

$$w^k = \arg\inf_w \tilde{J}(u^k, w), \tag{9.17}$$

$$u^{k+1} = \arg\inf_u \tilde{J}(u, w^k). \tag{9.18}$$

Variational calculus shows that equation (9.18) is equivalent to

$$\int_\Omega \Big\{(u^{k+1}-g)v + \alpha\, w^k \nabla u^{k+1}\cdot\nabla v dx\Big\}dx = 0. \tag{9.19}$$

After discretization, u^{k+1} can be computed as solution to a *linear* system of equations.
Making equation (9.17) more explicit, however, is not as straightforward as with equation (9.18), in the *general* case. Nevertheless, the computation of w^k is not difficult due to the convexity of $\tilde{J}(v,w)$ with respect to w. In the particular case of a *convex* original functional (9.1) equation (9.17) explicitly reads:

$$w^k = \rho(|\nabla u^k|)$$

with $\rho(\cdot)$ from (9.4). Note that in this case equation (9.19) amounts to "freeze" the nonlinear part of equation (9.3) for one iteration step. To our knowledge, this so-called *Kačanov method* has been introduced more than 30 years ago [201, 137]. Nevertheless, it turned out to be both efficient and competitive with respect to other techniques (see Section 9.4 and [165]).

Finite Element Discretization

In this section we explain briefly how a discrete version of the variational approach (9.3) can be obtained using the finite element method. Note that the linear system corresponding to (9.19) can easily be computed by replacing $\rho(|\nabla u|)$ in (9.3) with w^k. For a sound introduction to the finite element method we refer to, e.g., [84]. The finite element method can be extended easily to adaptive algorithms with a coarse resolution in homogeneous regions and a fine one near edges. For more details on such speed-up techniques we refer to [347] in the context of variational image restoration, and to [33, 305] in the context of nonlinear diffusion filtering.
The basic idea behind the finite element method is the restriction of optimization problems to finite–dimensional subspaces. Let $\{\phi_1, ..., \phi_N\}$ be basis functions of a finite–dimensional subspace $\mathcal{H}_h \subset \mathcal{H}$. Then, the restriction of (9.3) to $\mathcal{H}_h$ reads:

$$L(u_h, \phi_i) = 0 \,, \quad \forall i = 1, ..., N \,, \tag{9.20}$$

with minimizer $u_h \in \text{span}\{\phi_1, ..., \phi_N\}$. If we define the mappings

$$I : \mathbb{R}^N \to \mathcal{H}_h; \; \mathbf{u} \to \sum_j u_j \phi_j \,, \tag{9.21}$$

and

$$\mathbf{L}_i(\mathbf{u}) := L(I(\mathbf{u}), \phi_i) \,, \quad \mathbf{u} = (u_1, ..., u_N)^T \,, \tag{9.22}$$

then the solution of (9.20) is equivalent to the solution of the nonlinear system:

$$\mathbf{L}(\mathbf{u}) = \mathbf{0} \,, \quad \mathbf{L} = (\mathbf{L}_1, ..., \mathbf{L}_N)^T \,. \tag{9.23}$$

For the case of two–dimensional (2D) grey-value images we use piecewise linear basis functions as follows. The first step is to triangulate the underlying image domain, in this case the rectangular area $\Omega = [0, N_1 - 1] \times [0, N_2 - 1]$, as illustrated in Figure 9.5.
Next, we assign to each mesh node $P_{i,j}$ a basis function $\phi_{i,j}$ which is uniquely defined by the following conditions:

$$\begin{array}{ll} \phi_{i,j}(x) & \text{is linear within each triangle } d_k \,, \\ \phi_{i,j}(x) = 1 & \text{at node } P_{i,j} \,, \\ \phi_{i,j}(x) = 0 & \text{at every node } P_{k,l} \neq P_{i,j} \,. \end{array}$$

Given discrete grey-value images, and elements of the subspace $\mathcal{H}_h$ in general, images are represented by simply interpolating the values of corresponding nodal variables $v_{i,j}$ in a piecewise linear fashion:

$$I : \mathbb{R}^{N_1 \times N_2} \to \mathcal{H}_h \,, \quad \mathbf{v} \to \sum_{i,j} v_{i,j} \phi_{i,j}$$

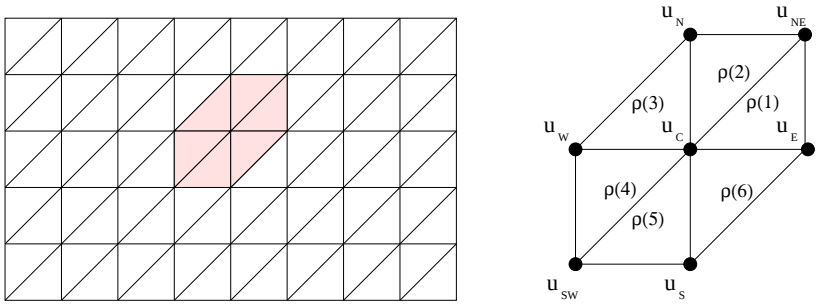


Figure 9.5: LEFT: Triangulation of the rectangular image domain Ω with mesh width 1. The nodes correspond to pixel positions. RIGHT: An interior node u_C with adjacent triangles d_i and the corresponding ρ-values (for piecewise linear basis functions $\rho(k) = \rho(|\nabla u|) = const.$ for each triangle d_k).

with grey-value $v_{i,j}$. From (9.21) and (9.22) we thus obtain:

$$\sum_{k,l}(u_{k,l} - g_{k,l}) \int_\Omega \phi_{k,l}\phi_{i,j}\, dx + \sum_{k,l} u_{k,l} \int_\Omega \rho(|\nabla u_h|)\nabla\phi_{k,l} \cdot \nabla\phi_{i,j}\, dx = 0. \qquad (9.24)$$

These integrals vanish for all pairs of nodes (i, j) and (k, l) which have no triangle in common. The remaining integrals can be computed analytically to obtain a sparse system of nonlinear equations in terms of the nodal variables of the solution image $\mathbf{u}$. Additional details and applications to different variational problems can be found in [348].

The expressions in (9.24) are weighted sums. Applying the linearization technique described in Section 9.3.1, that is "freezing" the nonlinear part as in equation (9.19), the nonlinear system (9.24) reduces to a linear system of type

$$\mathbf{B}^k\mathbf{u}^{k+1} = \mathbf{b}\,, \qquad (9.25)$$

where

$$\mathbf{B}^k = \mathbf{D} + \mathbf{R}(\mathbf{u}^k)\,, \qquad \mathbf{b} = \mathbf{D}\mathbf{g}\,.$$

The matrices $\mathbf{D}$ and $\mathbf{R}$ are sparse, since they describe convolutions within a 3×3 neighbourhood. For some inner node, the corresponding stencil notations are given by

0	$\frac{1}{12}$	$\frac{1}{12}$
$\frac{1}{12}$	$\frac{1}{2}$	$\frac{1}{12}$
$\frac{1}{12}$	$\frac{1}{12}$	0

(9.26)

for $\mathbf{D}$, and by

$$\begin{array}{|c|c|c|}
\hline
0 & \frac{1}{12} - \alpha \frac{\rho(2)+\rho(3)}{2} & \frac{1}{12} \\
\hline
\frac{1}{12} - \alpha \frac{\rho(3)+\rho(4)}{2} & \frac{1}{2} + \alpha \frac{\rho(2)+\rho(3)}{2} + \alpha \frac{\rho(3)+\rho(4)}{2} \\ +\alpha \frac{\rho(1)+\rho(6)}{2} + \alpha \frac{\rho(5)+\rho(6)}{2} & \frac{1}{12} - \alpha \frac{\rho(1)+\rho(6)}{2} \\
\hline
\frac{1}{12} & \frac{1}{12} - \alpha \frac{\rho(5)+\rho(6)}{2} & 0 \\
\hline
\end{array} \tag{9.27}$$

for the system matrix $\mathbf{B}^k$. It should be noted that the entries $\rho(.)$ are functions of $\mathbf{u}^k$; see also Figure 9.5. The necessary modifications of these stencils at boundary points are automatically obtained by taking into consideration the correct domain of integration in (9.24). It is not hard to see that $\mathbf{B}^k$ is symmetric and positive definite.

9.3.2 Nonlinear Diffusion Filtering

Diffusion Interpretation of the Kačanov Method

It is instructive to study also a finite difference discretization of the Kačanov method for variational image restoration: it follows that such a discretization can be regarded as a fixed point iteration for solving the implicitly time-discretized diffusion equation (9.10). This will shown next.

A discrete m-dimensional image can be regarded as a vector $\mathbf{g} \in \mathbb{R}^N$, whose components g_i, $i \in \{1, ..., N\}$, denote the grey-values at the pixels. Pixel i represents the location x_i. By u_i and ρ_i we denote approximations to $u(x_i, t)$ and $\rho(|\nabla u(x_i, t)|)$, respectively, where the gradient is replaced by central differences. For more details on the finite difference method for PDEs we refer to [256].

A spatial finite difference discretization of (9.10) is given by

$$\frac{u_i - g_i}{\alpha} = \sum_{l=1}^{m} \sum_{j \in \mathcal{N}_l(i)} \frac{\rho_j + \rho_i}{2} (u_j - u_i), \tag{9.28}$$

where $\mathcal{N}_l(i)$ consists of the two neighbours of pixel i along the l direction (boundary pixels may have only one neighbour). In vector–matrix notation this becomes

$$\frac{\mathbf{u} - \mathbf{g}}{\alpha} = \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u})\, \mathbf{u}, \tag{9.29}$$

where $\mathbf{A}_l$ describes the diffusive interaction in direction l.

This equation can be rewritten in fixed point structure as

$$\mathbf{u} \;=\; \Big(\mathbf{I} - \alpha \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u})\Big)^{-1} \mathbf{g}. \tag{9.30}$$

A corresponding fixed point iteration is given by

$$\mathbf{u}^{k+1} \;=\; \Big(\mathbf{I} - \alpha \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u}^k)\Big)^{-1} \mathbf{g}. \tag{9.31}$$

This is nothing else but the finite difference approximation

$$\frac{\mathbf{u}^{k+1} - \mathbf{g}}{\alpha} \;=\; \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u}^k)\, \mathbf{u}^{k+1} \tag{9.32}$$

to the Kačanov method (9.19), whose Euler equation is given by

$$\frac{u^{k+1} - g}{\alpha} \;=\; \mathrm{div}\,\Big(\rho(|\nabla u^k|)\, \nabla u^{k+1}\Big). \tag{9.33}$$

The fixed point iteration (9.31) requires to solve the sequence of linear systems

$$\Big(\mathbf{I} - \alpha \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u}^k)\Big)\, \mathbf{u}^{k+1} \;=\; \mathbf{g} \tag{9.34}$$

with $k = 1, ..., \infty$. Since this method is globally convergent, we may use an arbitrary initial vector $\mathbf{u}^0$. A natural choice is $\mathbf{u}^0 := \mathbf{g}$. In practice, one stops the iterations when $\mathbf{u}^k$ and $\mathbf{u}^{k+1}$ are sufficiently close, or when the residue obtained from plugging $\mathbf{u}^k$ into (9.29) is close to 0.
The system matrix for the finite difference Kačanov scheme (9.34) can be represented by a convolution stencil of the type

0	$-\alpha\,\frac{\rho(N)+\rho(C)}{2}$	0
$-\,\alpha\,\frac{\rho(W)+\rho(C)}{2}$	$1 + \alpha\,\frac{\rho(N)+\rho(C)}{2} + \alpha\,\frac{\rho(W)+\rho(C)}{2}$ $+\alpha\,\frac{\rho(E)+\rho(C)}{2} + \alpha\,\frac{\rho(S)+\rho(C)}{2}$	$-\alpha\,\frac{\rho(E)+\rho(C)}{2}$
0	$-\alpha\,\frac{\rho(S)+\rho(C)}{2}$	0

(9.35)

where the diffusivity $\rho(C)$ refers to some inner pixel, and $\rho(N)$, $\rho(W)$, $\rho(E)$, and $\rho(S)$ represent the northern, western, eastern, and southern neighbours, respectively. They are all functions of $\mathbf{u}^k$. We observe large structural similarities to the finite element Kačanov method. Both the finite difference and the finite element method boil down to the solution of a sequence of linear systems of equations with a sparse and symmetric positive definite system matrix.

A Semi-Implicit Scheme for Nonlinear Diffusion Filtering

If one is interested in solving the nonlinear diffusion equation (9.7) for some time α, one may proceed iteratively in n steps with step size $\tau = \alpha/n$. Let us denote by $\mathbf{u}^k$ the finite difference approximation at time $k\tau$. One possibility for a finite difference discretization to (9.7) is the scheme

$$\frac{\mathbf{u}^{k+1} - \mathbf{u}^k}{\tau} \;=\; \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u}^k)\, \mathbf{u}^{k+1} \tag{9.36}$$

with $\mathbf{u}^{(0)} := \mathbf{g}$. We observe that we have to solve n linear systems of type

$$\Big(\mathbf{I} - \tau \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u}^k)\Big)\, \mathbf{u}^{k+1} \;=\; \mathbf{u}^k. \tag{9.37}$$

Since this scheme does not give the solution $\mathbf{u}^{k+1}$ directly (explicitly), but requires to solve a linear system first, it is called a *linear-implicit (semi-implicit)* method. Its system matrix has the same structure as the one used in the finite difference Kačanov method.

In the 1-D case the system matrix is tridiagonal and diagonally dominant. For such a system a Gaussian algorithm for tridiagonal systems (also called *Thomas algorithm*) is stable and solves the problem with linear complexity, both with respect to memory and computational time [256, Section 2.9].

For dimensions $m \geq 2$, however, it is not possible to order the pixels in such a way that in the i-th row all nonvanishing elements of the system matrix can be found within the positions $[i, i-m]$ to $[i, i+m]$: Usually, the matrix possesses a much larger bandwidth. Applying direct algorithms such as Gaussian elimination would destroy the zeros within the band and would lead to an immense storage and computation effort. Classical iterative algorithms like the Jacobi or Gauß–Seidel method slow down for large τ, since this increases the condition number of the system matrix. Just like in the finite element case, a preconditioned conjugate gradient technique will be more efficient; see also [334].

This shows that there is also a large amount of structural similarities between discrete variational approaches and discrete nonlinear diffusion filters. In contrast to variational image restoration, however, the number of linear systems to be solved is *a-priori* fixed to n and no *a posteriori* stopping criterion is required. Clearly, choosing a larger n leads to smaller time step sizes and improves the approximation quality.

A possible speed advantage of semi-implicit nonlinear diffusion filtering as compared to the Kačanov method for variational image restoration may result from the fact that the time step size τ is usually smaller than the regularization parameter α. Therefore, nonlinear diffusion filtering has a better-conditioned system matrix

than variational image restoration. Thus, one may expect that iterative solvers for linear systems whose convergence depends on the condition number will converge faster for nonlinear diffusion filters. This also applies to preconditioned conjugate gradient methods, although to a much smaller degree than in classical iterative solvers.

AOS Schemes for Nonlinear Diffusion Filtering

Next we discuss a modification of the semi-implicit scheme

$$\mathbf{u}^{k+1} \;=\; \Big(\mathbf{I} - \tau \sum_{l=1}^{m} \mathbf{A}_l(\mathbf{u}^k)\Big)^{-1} \mathbf{u}^k \tag{9.38}$$

which leads to a further speed improvement. It takes advantage from two observations that we made in the last section:

1. It is computationally easier to solve linear equations stemming from 1-D diffusions than those from higher-dimensional diffusion processes.

2. For accuracy reasons, diffusion filtering often uses not too large time step sizes τ.

Now the idea is to replace (9.38) by the *additive operator splitting (AOS) scheme* [429]

$$\mathbf{u}^{k+1} \;=\; \frac{1}{m} \sum_{l=1}^{m} \Big(\mathbf{I} - m\tau \mathbf{A}_l(\mathbf{u}^k)\Big)^{-1} \mathbf{u}^k. \tag{9.39}$$

Such a scheme has several interesting properties:
First, the operators $\mathbf{I} - m\tau\mathbf{A}_l(\mathbf{u}^k)$ describe one-dimensional diffusion processes along the x_l axes. Under a consecutive pixel numbering along the direction l they lead to strictly diagonally dominant tridiagonal matrices. The corresponding linear systems of equations can be solved directly in an efficient and stable way by the Thomas algorithm. Its forward elimination and backward substitution step can be regarded as a causal and an anticausal recursive filter, respectively.
Moreover, (9.39) has the same first-order Taylor expansion in τ as the semi-implicit scheme: although both methods are algebraically different, they are $O(\tau + h_1^2 + \ldots + h_m^2)$ approximations to the continuous equation. Thus, the approximation quality improves with decreasing τ, and the solutions from the AOS scheme and the semi-implicit one become more and more similar.
Since AOS is an *additive* splitting, it is assured that all coordinate axes are treated in exactly the same manner. This is in contrast to the finite element discretization from Section 9.3.1 and to conventional splitting techniques from the literature, which are *multiplicative* [241]. They may produce different results if the image is rotated by 90 degrees.

Recently a general framework for discrete nonlinear diffusion scale-spaces has been established, which guarantees that the discretization reveals the same scale-space properties as its continuous counterpart [407, 412]. One can verify [429] that the AOS scheme creates such a discrete nonlinear diffusion scale-space for every step size τ. As a consequence, it preserves the average grey level μ, satisfies a causality property in terms of a maximum–minimum principle, and converges to a constant steady state. Moreover, the process is a simplifying, information-reducing transform with respect to many aspects: The p-norms

$$\|\mathbf{u}^k\|_p := \left(\sum_{i=1}^{N} |u_i^k|^p \right)^{1/p} \qquad (p \geq 2) \tag{9.40}$$

and all even central moments

$$M_{2n}[\mathbf{u}^k] := \frac{1}{N} \sum_{j=1}^{N} (u_j^k - \mu)^{2n} \tag{9.41}$$

are decreasing in k, and the entropy

$$S[\mathbf{u}^k] := -\sum_{j=1}^{N} u_j^k \ln u_j^k, \tag{9.42}$$

a measure of uncertainty and missing information, is increasing in k (if u_j is positive for all j).
For further algorithmic details and a performance analysis of AOS schemes we refer to [429]. There it is demonstrated that, under typical accuracy requirements, AOS schemes are one order of magnitude more efficient than the commonly used schemes for nonlinear diffusion filtering. In Section 9.4 we will see that a speed-up by another order of magnitude is possible by parallel implementations.

9.4 Parallel Implementations

9.4.1 Variational Approaches

To solve the linear system (9.25) successively as part of the two–step iteration (9.17)–(9.18), we implemented an inexact version of the well–known *conjugate gradient (CG)* method along with several preconditioners on a multi–processor SGI Power–Challenge machine at the Regional Computer Center (RRZ) of the University of Hamburg. We used the software package PETSc (Portable Extensible Toolkit for Scientific Computing) [32, 31], which is based on the message passing standard MPI [247].

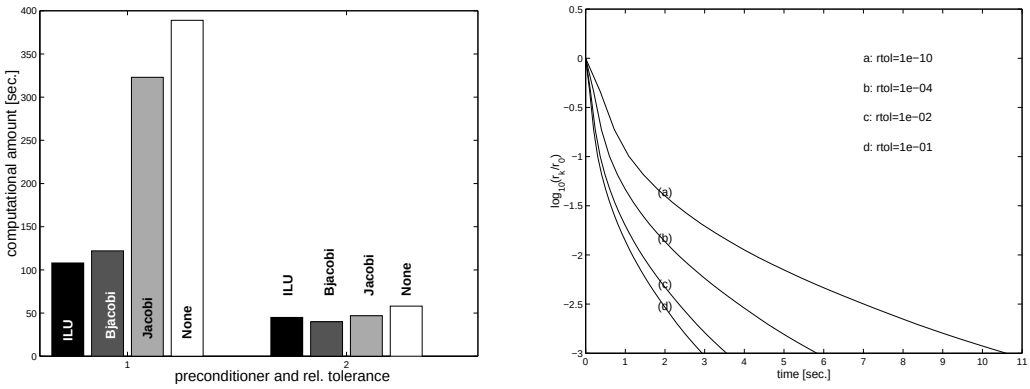


Figure 9.6: LEFT: Convergence of the variational method for various accuracies *rtol* of the linear solver for a 256×256 pixel sized image (computed with 16 processing units). RIGHT: Computational costs of different preconditioners for exact (left: $rtol = 10^{-6}$) and inexact linear solvers (right: $rtol = 10^{-1}$) (computed with 1 processing unit).

Among the class of iterative Krylov subspace solvers (cf., e.g., [203]), the CG–method is nearly optimal for the class of matrices $\mathbf{B}^k$ considered here (sparse, symmetric, positive definite). "Inexact" refers to the stopping criterion $rtol \cdot \|\mathbf{r}^k\|_2$, that is comparing the relative reduction of the initial *linear* residuals $\mathbf{r}^k = \mathbf{B}^k\mathbf{u}^k - \mathbf{b}$ to a threshold. It turned out in our experiments that the rather crude criterion $rtol = 0.1$ leads to a fast inner loop of the iteration (9.17)–(9.18) *without* loosing both convergence of the overall iteration and (consequently) quality of the computed result (Fig. 9.6, left). In all experiments the stopping criterion of the overall iteration (9.17)–(9.18) was $rtol = 0.001$ with respect to the *nonlinear* residuals of the system (9.24). We note that convergence of the Kačanov method using an "inexact" inner loop has not been proven yet. For hints on how this may be accomplished we refer to the work of Axelsson [27].
To improve the condition number and, in turn, the speed of convergence, preconditioners $\mathbf{L}, \mathbf{R}$ were applied to the linear system (9.25):

$$(\mathbf{L}^{-1}\mathbf{B}^k\mathbf{R}^{-1})\mathbf{R}\mathbf{u}^{k+1} = \mathbf{L}^{-1}\mathbf{b}$$

Classical preconditioners are obtained through either an additive matrix splitting (like Jacobi–, SOR–, or SSOR–preconditioning) or a multiplicative matrix splitting (e.g. ILU or ICC factorizations). The approach of *domain decomposition* is an alternative way of preconditioning (Block–Jacobi or Block–Gauss–Seidel, for example), which is more suited for parallel implementations [360]. Figure 9.6, right, shows the influence of various preconditioners on the convergence speed for both exact and inexact linear solvers. Interestingly, the picture in Figure 9.6 on the

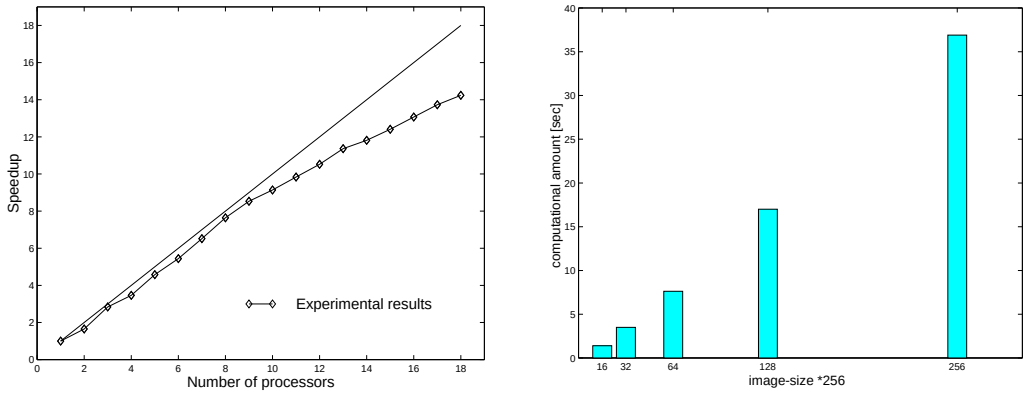


Figure 9.7: LEFT: Computational costs as a function of the image size (1 processing unit, $rtol = 0.1$, Block–Jacobi preconditioning). RIGHT: Speed–up factor as a function of the number of processing units used.

right shows that, by using an inexact linear solver, a simple preconditioner like Block–Jacobi performs much like a computationally more involved preconditioner (ILU factorization, for example). In fact, the less expensive Block–Jacobi preconditioner combined with the inexact solver turned out to be best and reduced the computational amount by one order of magnitude. Figure 9.7, left, depicts the dependency of the computation time on the image size. Fig. 9.7, right, finally shows a nearly optimal linear increase of the speed–up factor as a function of the number of processing units (18 units were available to us). This proves the efficiency of the parallel implementation.
These results show that the combination of appropriate problem–specific numerical concepts (Kačanov linearization, inexact linear CG solver, Block–Jacobi preconditioning) with an efficient parallel implementation yields a reduction of the overall computational costs of two orders of magnitude. Using more massive parallelism a further reduction of one order of magnitude should be feasible, enabling the processing of 2D images in quasi-real-time (2–3 frames per second). Furthermore, application of the AOS scheme to the variational approach, based on the relationship to nonlinear diffusion filters described in Sections 9.2.2 and 9.3.2, will be considered in future work.

9.4.2 Nonlinear Diffusion Filtering

For nonlinear diffusion filtering, the AOS schemes described in Section 9.3.2 work efficiently on sequential computer architectures. However, they also offer two intrinsic levels of parallelism: *coarse grain parallelism* and *mid grain parallelism*.
The coarse grain parallelism can be described as follows. The result $\mathbf{u}^{k+1}$ of an

m-dimensional AOS scheme can be regarded as the average of m filters of type

$$\mathbf{v_l}^{k+1} := \left(I - m\tau \mathbf{A_l}(\mathbf{u}^k)\right)^{-1} \mathbf{u}^k \qquad (l = 1, ..., m).$$

Since $\mathbf{v_1}^{k+1}$,...,$\mathbf{v_m}^{k+1}$ can be calculated independently from each other, it is possible to distribute their computation to different processors of a parallel machine.
Often parallel computers offer more processors than the dimensionality m of the problem. In this case it is interesting to exploit the mid grain parallelism which is described next. Let N_j denote the number of pixels in the j direction. Then $(\mathbf{I} - m\tau \mathbf{A}_l(u^k))^{-1}$ creates $\prod_{j \neq l} N_j$ one-dimensional diffusion processes along the l direction. Since these processes are completely independent from each other, they can be computed in parallel.
As a demonstrator for a mid grain parallelization we have implemented a three-dimensional AOS scheme on an SGI Power Challenge XL with eight 195 MHz R10000 processors and 1 MByte of joint secondary cache [430]. The test image is a 3-D ultrasound data set of size $138 \times 208 \times 138$ which depicts a 10-weeks old human fetus.
We used the AOS scheme for the regularized nonlinear diffusion filter from Section 9.2.2. The contrast parameter ν was set to the 40% quantile of the cumulative histogram for $\nabla(K_\sigma * f)$, and the standard deviation σ of the Gaussian was chosen as the unit length of the cubic voxel. For our test image 8 iterations with $\tau = 10$ were sufficient to yield satisfactory denoising, as is seen in Fig. 9.8. Note the significantly improved visibility of the skull and the hands.
The implementation of the parallel AOS scheme was done in two steps: First we have further optimized our existing sequential AOS code by using the C++-based dedicated TULIP library[1]. This feature has led to a speed-up by a factor of three. In a second step the optimized code has been split up into functions acting only within 2-D slices of the 3-D image. These slices were then distributed to the different processors.
Table 9.1 shows the measured execution times as a function of the number of processors. We observe that an implementation of AOS schemes on a few processors is already sufficient for filtering 3-D medical data sets within their typical acquisition time.
The total execution time T of an algorithm on a parallel system is given by

$$T = T_s + \frac{T_p}{p} + T_{sc}(p), \tag{9.43}$$

where T_s and T_p are the sequential and parallel components, respectively, and $T_{sc}(p)$ is the synchronization time for the parallel processes.

[1] TULIP has been developed by Karel Zuiderveld and Fred Appelman (Image Sciences Institute, Utrecht) using cache memory addressing in a sophisticated way in order to speed up the way of accessing neighbouring voxels.

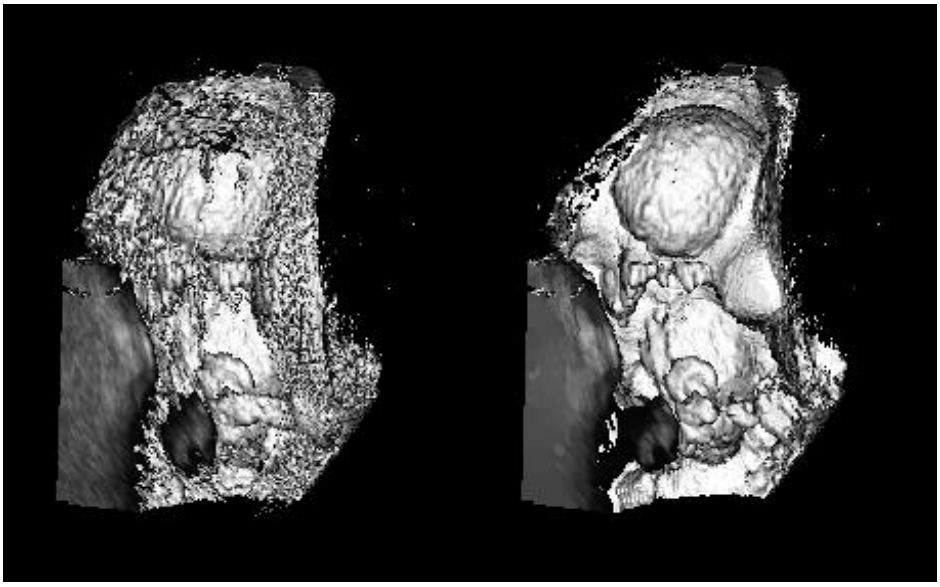


Figure 9.8: Rendering of a 3-D ultrasound image of a 10-week old fetus. LEFT: Original. RIGHT: Filtered.

Table 9.1: Execution time T for eight AOS iterations on an SGI Power Challenge for an image of size $138 \times 208 \times 138$ using a varying number of processors.

# proc.	exec. time T	speed-up	efficiency
1	158.22 s	1.00	100.0 %
2	93.11 s	1.70	85.0 %
3	71.48 s	2.21	73.7 %
4	59.75 s	2.64	66.0 %
5	51.96 s	3.05	61.0 %

If we neglect $T_{sc}(p)$, we find from Table 9.1 that $T_s \approx 25.5s$, that means about 84 % of the code is executed in parallel. The sequential part is mainly caused by I/O from and to the harddisk, and from the need to create copies of the volume data. For more iterations we may expect further improvement with respect to the scaling behaviour. The measured data indicate that due to parallel implementations, an efficiency increase by one order of magnitude is realistic.

9.5 Conclusion and Further Work

In the image processing literature, numerical methods for diffusion filtering and for variational restoration methods are treated separately, and often algorithms are applied that seem to be unrelated at first glance. It is also common for PDE methods in image processing that either finite element or finite difference methods are used.
In the present paper we have presented a unifying numerical viewpoint, both for variational approaches and diffusion filtering on the one hand, but also for finite element and finite difference methods on the other hand.
This allowed us to demonstrate a large degree of structural similarities between the different efficient algorithms. For two of these methods we have presented parallel implementations on general-purpose hardware and assessed their performance for 2D and 3D image restoration problems.
Our work has been triggered by recent results on the close relations between variational image restoration methods and nonlinear diffusion scale-spaces. Since these results have been established in a continuous context, it became interesting to study similar relations also at an algorithmic level. It is our hope that our work provides a unifying basis for efficient parallel implementations of these two approaches, and that it makes them attractive for many challenging real-life applications.
Our further work will include the investigation of numerical multigrid schemes and their connections to the algorithms described in the present paper, as well as corresponding implementations on more massively parallel architectures.

Acknowledgements. This research has been partly funded by the German National Science Foundation (DFG), as well as by the Real–World Computing Partnership (RWCP), the EU–TMR project VIRGO, the Fonds zur Förderung der Wissenschaftlichen Forschung (Austria), SFB F1310 (Austria), and the Fonds zur Förderung der Gewerblichen Forschung (FFF), Project 200354 (Austria).

Chapter 10

Efficient Image Segmentation Using Partial Differential Equations and Morphology

Joachim Weickert.
Pattern Recognition, in press.

Abstract

The goal of this paper is to investigate segmentation methods that combine fast preprocessing algorithms using partial differential equations (PDEs) with a watershed transformation with region merging. We consider two well-founded PDE methods: a nonlinear isotropic diffusion filter that permits edge enhancement, and a convex nonquadratic variational image restoration method which gives good denoising. For the diffusion filter, an efficient algorithm is applied using an additive operator splitting (AOS) that leads to recursive and separable filters. For the variational restoration method, a novel algorithm is developed that uses AOS schemes within a Gaussian pyramid decomposition. Examples demonstrate that preprocessing by these PDE techniques significantly improves the watershed segmentation, and that the resulting segmentation method gives better results than some traditional techniques. The algorithm has linear complexity and it can be used for arbitrary dimensional data sets. The typical CPU time for segmenting a 256^2 *image on a modern PC is far below one second.*

10.1 Introduction

Segmentation is one of the bottlenecks of many image analysis and computer vision tasks ranging from medical image processing to robot navigation. Ideally it should be efficient to compute and correspond well with the physical objects depicted in the image. This also requires that segmentation gives a complete partitioning of the image such that object contours are closed and no dangling edges exist.
In the last decade much research on PDE-based regularization methods has been carried out; see e.g. [70, 381, 274, 412] for recent overviews. Although the promising results suggest that they might be attractive as a preprocessing step for many subsequent image analysis methods, little research has actually been carried out which combines PDE-based preprocessing methods with other techniques. One of the problems was that PDE-based methods have been considered as being too slow in order to become an adequate partner for efficient other methods. This shows the need to further develop efficient algorithms for PDE-based techniques.
The goal of the present paper is to address these topics in the following way:

- Image segmentation is achieved by means of a watershed algorithm. This popular morphological method is more more than an edge detector: it gives a true image partitioning without dangling edges. The watershed segmentation is sufficiently fast for most applications, but it suffers from the limitation that many irrelevant minima cause an oversegmentation.

- In order to reduce the oversegmentation problem we study the use of two PDE-based techniques for preprocessing the image before segmentation: the nonlinear diffusion technique by Catté *et al.* [74] which allows edge enhancement, and a nonquadratic variational restoration technique of Schnörr [345] and Charbonnier *et al.* [80] which is well-suited for edge-preserving image denoising. These methods have been chosen as simple prototypes of PDEs that are mathematically well-founded: they are well-posed in the sense of Hadamard in that they have a unique solution which is stable with respect to perturbations of the original image.

- As an efficient algorithm for the nonlinear diffusion filter we apply a recently developed method based on an additive operator splitting (AOS) [429]. It leads to separable and recursive filters. For the nonquadratic variational image restoration method, we develop a novel algorithm: minimization of the energy functional is achieved by considering a steepest descent method that leads to a diffusion–reaction PDE. This PDE is then solved by a modified AOS algorithm that is embedded in a Gaussian pyramid decomposition.

The resulting segmentation algorithms can be generalized in a straightforward way to arbitrary dimensional data sets. Their complexity is linear in the pixel number,

and they produce identical results when the image is rotated by 90 degrees. An overall CPU time of less than one second for segmenting a 256^2 image on a typical PC or workstation makes them attractive for many time-critical applications.

The paper is organized as follows. Section 10.2 sketches the basic structure of the contrast-enhancing nonlinear diffusion filter and the convex nonquadratic restoration method. In Section 10.3 we discuss efficient and reliable numerical techniques for these methods. They are based on an additive operator splitting. For approximating the variational restoration method, these AOS techniques are extended to novel pyramid AOS schemes. In Section 10.4 we describe the watershed algorithm with region merging, and in Section 10.5 we illustrate the usefulness of the combined segmentation process by applying it to several examples and comparing it with classic approaches. The paper is concluded with a summary in Section 10.6.

Related work. The work presented here has been influenced by several related approaches. Closest in terms of efficient PDE-based regularization methods is the work of Acton [3] on multigrid versions for nonlinear diffusion filtering. They are, however not based on AOS schemes and they do not use methods with a reaction term. It is common to supplement watershed segmentations with tools for reducing the oversegmentation problem. An algorithm by Orphanoudakis *et al.* [288] also uses region merging for this purpose, but it applies statistical instead of PDE-based smoothing strategies. Promising results of combining watershed algorithms with nonlinear diffusion have been described recently by De Vleeschauer *et al.* [100] and Sijbers *et al.* [358]. Investigations of watershed algorithms within scale-space hierarchies have been carried out by Griffin *et al.* [155], Olsen [283], Sramek and Wrbka [374], and Olsen and Sporring [284] for the linear diffusion scale-space, and by Jackway [194] for the dilation–erosion scale-space. A nonmorphological segmentation algorithm based on nonlinear diffusion scale-spaces has been studied by Niessen *et al.* [277, 276]. This discussion shows that the novelty of our approach consist of developing pyramid AOS algorithms for efficient PDE-based regularization, and combining fast AOS-based algorithms with an important morphological segmentation tool, the watershed algorithm. This results in a fast segmentation method. A preliminary version of the present manuscript has been presented at a conference [413].

10.2 PDE-Based Regularization

Below two prototypes for well-posed PDE-based regularization techniques are presented. The first one allows contrast enhancement, while the second one can be expressed as an energy minimization method. These two methods are only representatives of a much larger class of diffusion-based smoothing methods. For a more detailed treatment of this topic the reader is referred to [412].

10.2.1 The Nonlinear Diffusion Filter of Catté *et al.*

In the m-dimensional case the filter of Catté, Lions, Morel and Coll [74] has the following structure:
Let $\Omega := (0, a_1) \times .... \times (0, a_m)$ be our image domain and consider a (scalar) image $f(x)$ as a bounded mapping from Ω into the real numbers $\mathbb{R}$. Then a filtered image $u(x,t)$ of $f(x)$ is calculated by solving the diffusion equation with the original image as initial state, and reflecting boundary conditions:

$$\partial_t u = \operatorname{div}\left(g(|\nabla u_\sigma|^2)\,\nabla u\right) \tag{10.1}$$

$$u(x,0) = f(x), \tag{10.2}$$

$$\partial_n u|_{\partial\Omega} = 0, \tag{10.3}$$

where n denotes the normal to the image boundary $\partial\Omega$.
The "time" t is a scale parameter: larger values lead to simpler image representations. In order to reduce smoothing at edges, the diffusivity g is chosen as a decreasing function of the edge detector $|\nabla u_\sigma|^2$, where ∇u_σ is the gradient of a Gaussian-smoothed version of u:

$$\nabla u_\sigma := \nabla(K_\sigma * u), \tag{10.4}$$

$$K_\sigma := \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \tag{10.5}$$

We use the diffusivity

$$g(s^2) := \begin{cases} 1 & (s^2 = 0) \\ 1 - \exp\left(\frac{-3.315}{(s/\lambda)^8}\right) & (s^2 > 0). \end{cases} \tag{10.6}$$

For such rapidly decreasing diffusivities, smoothing on both sides of an edge is much stronger than smoothing across it. This selective smoothing process prefers intraregional smoothing to interregional blurring. The factor 3.315 ensures that the flux $\Phi(s) := sg(s^2)$ is increasing for $|s| \leq \lambda$ and decreasing for $|s| > \lambda$. Thus, λ is a contrast parameter separating low-contrast regions with (smoothing) forward diffusion from high-contrast locations where backward diffusion may enhance edges [303]. After some time this filter creates segmentation-like results which are piecewise almost constant. For $t \to \infty$, however, the image becomes completely flat [412]. Well-posedness results for this filter can be found in [74, 412] and a scale-space interpretation in terms of an extremum principle as well as decreasing variance, decreasing energy, and increasing entropy is given in [412].
The effect of this diffusion filter is illustrated in Figure 10.1 (c),(d). We observe that it creates piecewise almost constant regions that are separated by sharp edges. If the images are very noisy, however, the filter performance deteriorates near edges

where it tends to preserve these noisy structures by decreasing the diffusivity. In the next scetion we are concerned with a related method that is better suited for noise elimination than isotropic nonlinear diffusion filtering.

10.2.2 Variational Image Restoration

Many variational methods for image restoration (such as [345, 80, 104]) obtain a filtered version of some degraded image f as the minimizer of an energy functional of type

$$E_f(u) := \int_\Omega \Big((u-f)^2 \, + \, \alpha \, \Psi(|\nabla u|^2) \Big) \, dx, \tag{10.7}$$

where the regularizer Ψ is an increasing function. The first summand encourages similarity between the restored image and the original one, while the second summand rewards smoothness. The smoothness weight $\alpha > 0$ is called *regularization parameter*.
From variational calculus it follows that the minimizer of $E_f(u)$ satisfies the so-called Euler–Lagrange equation

$$0 \; = \; \mathrm{div} \Big(\Psi'(|\nabla u|^2) \nabla u \Big) + \tfrac{1}{\alpha}(f-u) \tag{10.8}$$

This can be regarded as the steady-state ($t \to \infty$) of the diffusion–reaction process

$$\partial_t u \; = \; \mathrm{div} \Big(\Psi'(|\nabla u|^2) \nabla u \Big) + \tfrac{1}{\alpha}(f-u) \tag{10.9}$$

This shows the close connection between variational image restoration and diffusion filtering. Indeed, much more relations have been discovered recently; see [341] for more details. In our case the convex potential [268]

$$\Psi(|\nabla u|^2) = \lambda \sqrt{1 + |\nabla u|^2/\lambda^2} + \varepsilon |\nabla u|^2 \qquad (\lambda, \varepsilon > 0) \tag{10.10}$$

is used. The corresponding diffusivity in (10.9) is given by its derivative

$$\Psi'(|\nabla u|^2) = \frac{1}{\sqrt{1 + |\nabla u|^2/\lambda^2}} + \varepsilon. \tag{10.11}$$

Choosing a potential function $\Psi(s^2)$ that is convex in s allows to guarantee well-posedness and stable algorithms [345]. For nonconvex potentials as in [281, 140], several well-posedness questions are open. Moreover, the diffusion–reaction equation (10.2.2) converges globally (i.e. for all initial values) to the solution of the Euler–Lagrange equation (10.8). It should be noted that the convex potential implies that the corresponding diffusive flux $\Phi(s) = s\Psi'(s^2)$ is increasing in s. Thus,

Figure 10.1: (a) Top Left: test image. (b) Top Right: Gaussian noise with zero mean added. (c) Middle Left: nonlinear diffusion filtering of (a). (d) Middle Right: nonlinear diffusion filtering of (b). (e) Bottom Left: variational restoration of (a). (f) Bottom Right: variational restoration of (b).

backward diffusion does not appear and edge enhancement is not possible. Nevertheless, since the diffusivity $\Psi'(|\nabla u|^2)$ is decreasing in $|\nabla u|^2$, smoothing at edges is reduced and discontinuities are better preserved than in linear smoothing methods. Figure 10.1(e),(f) depicts the filter performance of this method. It has a remarkable robustness under noise, but, in contrast to the nonlinear diffusion filter of Catté *et al.*, it cannot enhance edges. This situation can be handled by more sophisticated diffusion filters such as edge-enhancing anisotropic diffusion [408]. They, however, require more complicated numerical algorithms that are beyond the scope of the present paper.

10.3 Efficient Algorithms for PDE-Based Regularization

10.3.1 Limitations of Conventional Schemes

Let us first consider finite difference approximations to the m-dimensional diffusion filter of Catté *et al.*.

A discrete m-dimensional image can be regarded as a vector $f \in \mathbb{R}^N$, whose components f_i, $i \in \{1, ..., N\}$ display the grey values at the pixels. Pixel i represents the location x_i. Let h_l denote the grid size in the l direction. We consider discrete times $t_k := k\tau$, where $k \in \mathbb{N}_0$ and τ is the time step size. By u_i^k and g_i^k we denote approximations to $u(x_i, t_k)$ and $g(|\nabla u_\sigma(x_i, t_k)|^2)$, respectively, where the gradient is replaced by central differences.

The simplest discretization of the diffusion equation with reflecting boundary conditions is given by

$$\frac{u_i^{k+1} - u_i^k}{\tau} = \sum_{l=1}^{m} \sum_{j \in \mathcal{N}_l(i)} \frac{g_j^k + g_i^k}{2h_l^2} (u_j^k - u_i^k). \tag{10.12}$$

where $\mathcal{N}_l(i)$ consists of the two neighbours of pixel i along the l direction (boundary pixels may have only one neighbour). In vector–matrix notation this becomes

$$\frac{u^{k+1} - u^k}{\tau} = \sum_{l=1}^{m} A_l(u^k)\, u^k. \tag{10.13}$$

A_l describes the diffusive interaction in l direction. One can calculate u^{k+1} directly (explicitly) from u^k without any matrix inversion:

$$u^{k+1} = \Big(I + \tau \sum_{l=1}^{m} A_l(u^k)\Big)\, u^k. \tag{10.14}$$

For this reason it is called *explicit scheme*. Each explicit iteration step can be performed very fast, but the step size has to be very small: one can show [429] that in order to guarantee stability, the step size must satisfy

$$\tau \leq \frac{1}{\sum_{l=1}^{m} \frac{2}{h_l^2}}. \tag{10.15}$$

For most practical applications, this restriction requires to use a very high number of iterations, such that the explicit scheme is rather slow.
Thus, we consider a slightly more complicated discretization next, namely

$$\frac{u^{k+1} - u^k}{\tau} = \sum_{l=1}^{m} A_l(u^k)\, u^{k+1}. \tag{10.16}$$

This scheme does not give the solution u^{k+1} directly (explicitly): It requires to solve a linear system first. It is called a *linear-implicit (semi-implicit)* scheme. The solution u^{k+1} is given by

$$u^{k+1} = \Big(I - \tau \sum_{l=1}^{m} A_l(u^k)\Big)^{-1} u^k. \tag{10.17}$$

This scheme is absolutely stable [412].
In the 1-D case the system matrix is tridiagonal and diagonally dominant. For such a system a Gaussian algorithm for tridiagonal systems (also called *Thomas algorithm*) solves the problem in linear complexity [429].
For dimensions $m \geq 2$, however, it is not possible to order the pixels in such a way that in the i-th row all nonvanishing elements of the system matrix can be found within the positions $[i, i-m]$ to $[i, i+m]$: Usually, the matrix reveals a much larger bandwidth. Applying direct algorithms such as Gaussian elimination would destroy the zeros within the band and would lead to an immense storage and computation effort. Typical iterative algorithms become slow for large τ, since this increases the condition number of the system matrix. Thus, in spite of its absolute stability, the semi-implicit scheme is often not much faster than the explicit one.

10.3.2 AOS Schemes

In order to address the preceding problem we consider a modification of (10.17), namely the *additive operator splitting (AOS) scheme* [429]

$$u^{k+1} = \frac{1}{m} \sum_{l=1}^{m} \Big(I - m\tau A_l(u^k)\Big)^{-1} u^k. \tag{10.18}$$

The operators $B_l(u^k) := I - m\tau A_l(u^k)$ describe one-dimensional diffusion processes along the x_l axes. Under a consecutive pixel numbering along the direction

l they come down to strictly diagonally dominant tridiagonal matrices which can be efficiently inverted by the Thomas algorithm.

Moreover, (10.18) has the same first-order Taylor expansion in τ as the explicit and semi-implicit scheme: all methods are $O(\tau + h_1^2 + ... + h_m^2)$ approximations to the continuous equation.

Since it is an *additive* splitting, all coordinate axes are treated in exactly the same manner. This is in contrast to conventional splitting techniques from the literature, which are *multiplicative* [241]. They may produce different results if the image is rotated by 90 degrees.

Recently a general framework for discrete nonlinear diffusion scale-spaces has been discovered, which guarantees that the discretization reveals the same scale-space properties as its continuous counterpart [412]. One can verify [429] that the AOS scheme creates such a discrete nonlinear diffusion scale-space for every (!) step size τ. As a consequence, it preserves the average grey level μ, satisfies a causality property in terms of a maximum–minimum principle, and converges to a constant steady state. Moreover, the process is a simplifying, information-reducing transform with respect to many aspects: The p-norms

$$\|u^k\|_p := (\sum_{i=1}^{N} |u_i^k|^p)^{1/p} \tag{10.19}$$

and all even central moments

$$M_{2n}[u^k] := \frac{1}{N} \sum_{j=1}^{N} (u_j^k - \mu)^{2n} \tag{10.20}$$

are decreasing in k, and the entropy

$$S[u^k] := -\sum_{j=1}^{N} u_j^k \ln u_j^k, \tag{10.21}$$

a measure of uncertainty and missing information, is increasing in k (if f_j is positive for all j).

Table 10.1 summarizes the features of the discussed schemes. Full algorithmic details of AOS schemes can be found in [429], and a parallel implementation for processing 3-D images is described in [419].

Many nonlinear diffusion problems require only the elimination of noise and some small-scale details. Often this can be accomplished with no more than 5 iterations in sufficient precision. We shall see that this takes about half a second for a 256^2 image on current PCs or workstations.

Table 10.1: Schemes that create discrete nonlinear diffusion scale-spaces.

scheme	formula	stability	costs/iter.	effic.
explicit	$u^{k+1} = \left(I + \tau \sum_{l=1}^{m} A_l(u^k)\right) u^k$	$\tau \leq \frac{1}{\sum_{l=1}^{m} \frac{2}{h_l^2}}$	very low	low
semi-impl.	$u^{k+1} = \left(I - \tau \sum_{l=1}^{m} A_l(u^k)\right)^{-1} u^k$	$\tau < \infty$	high	fair
AOS	$u^{k+1} = \frac{1}{m} \sum_{l=1}^{m} \left(I - m\tau A_l(u^k)\right)^{-1} u^k$	$\tau < \infty$	low	high

10.3.3 Pyramid AOS

Let us now investigate a novel extension of the AOS framework to the variational image restoration method. In matrix-vector notation a semi-implicit discretization of (10.9) is given by

$$\frac{u^{k+1} - u^k}{\tau} = \sum_{l=1}^{m} A_l(u^k)\, u^{k+1} + \alpha\,(f - u^{k+1}). \tag{10.22}$$

Solving for u^{k+1} yields

$$u^{k+1} = \left(I - \frac{\tau}{1 + \frac{\tau}{\alpha}} \sum_{l=1}^{m} A_l(u^k)\right)^{-1} \frac{u^k + \frac{\tau}{\alpha} f}{1 + \frac{\tau}{\alpha}}. \tag{10.23}$$

In analogy to the previous section we may approximate this scheme by its AOS variant

$$u^{k+1} = \frac{1}{m} \sum_{l=1}^{m} \left(I - \frac{m\tau}{1 + \frac{\tau}{\alpha}} A_l(u^k)\right)^{-1} \frac{u^k + \frac{\tau}{\alpha} f}{1 + \frac{\tau}{\alpha}} \tag{10.24}$$

which again comes down to recursive filtering.
In contrast to the pure diffusion filter, however, we are now interested in approximating the steady-state solution for $t \to \infty$. Even with large time step sizes, the diffusion process will mainly act within a fairly small vicinity around each pixel. Thus, many iterations are required if the image size is large. In order to speed up the process, we may embed the AOS scheme into a pyramid framework. The idea is as follows:

- create a Gaussian pyramid decomposition [62] with the smoothing mask $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$
- adapt the filter parameters to the downsampled image. The scaling behaviour of diffusion–reaction processes requires that α must be divided by 4 if one

reduces the image size by a factor 2. Since the smoothing mask $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ reduces the contrast of an ideal step edge by 25 %, it follows that the contrast parameter λ has to be multiplied by 0.75.

- start with the coarsest level (a 2×2 image), and apply a specified number of AOS iterations.
- expand this solution to the next finer level by linear interpolation, and use it as initial value for AOS iterations at this level.
- proceed in the same way until convergence at the finest level is reached.

Figure 10.2 illustrates the effect of pyramid AOS. Typically, five iterations are sufficient in order to obtain good convergence at each level. Since the Gaussian pyramid decomposition can be performed with linear complexity, the overall complexity remains linear as well. We shall see that regularizing a 256^2 image on a current PC or workstation with this pyramid AOS scheme requires only around 0.5 CPU seconds.
It should be noted that the pyramid embedding converges to the same regularized image than pure AOS iterations would do, since the convex variational approach is globally convergent. However, pyramid AOS converges faster because of its better initial data that are provided by the previous pyramid level.

10.4 Watershed Segmentation with Region Merging

The preceding PDE-based regularization techniques lead to simplified images where noise and unimportant fine-scale details have been removed.
In order to create a true segmentation, we have to postprocess the regularized image by a technique which gives an edge map without dangling edges. This edge map should lead to a partitioning of the entire image into a finite number of regions, it should handle the semantically important corners and junctions gracefully, and – last but not least – it should be fast. Classical gradient-based edge detectors such as a Sobel operator or Kirsch masks are not sufficient for this task, as they do not give closed contours. This also holds for more sophisticated variants such as the Canny edge detector [65].
We found a watershed technique [41, 248] based on the squared gradient magnitude very useful for these purposes. Such a technique regards an image as a landscape where the intensity values correspond to the elevation. Areas where a rain drop would drain to the same minimum are denoted as catchment basins, and the lines separating adjacent catchment basins are called watersheds. Watersheds

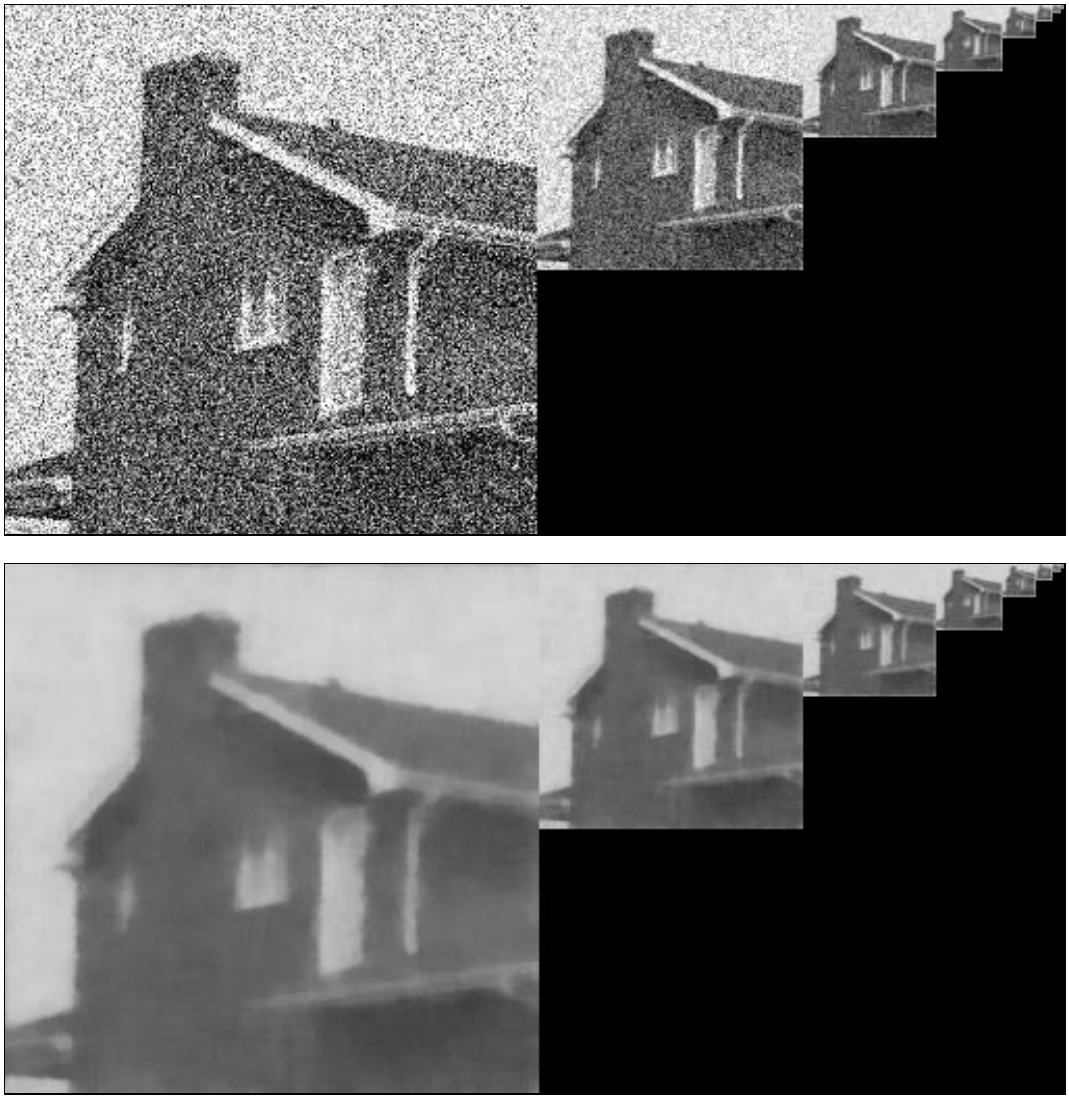


Figure 10.2: (a) Top: Gaussian pyramid of a noisy test image. (b) Bottom: regularized by pyramid AOS.

are a morphological technique, since they are invariant under monotone grey scale-transformations. They lead to an image segmentation into regions, and they can describe edge junctions [266]. This is in contrast to edge detectors based on zero-crossings of differential operators such as the Laplacian-of-Gaussian [242]: they do not allow to detect T-junctions [388].

We use Fairfield's watershed algorithm [118]. Our code is based on an implementation of Oltmans [285], where the Pascal code has been transferred into C and minor modifications have been included in order to optimize its performance.

The basic idea of this algorithm is sliding downhill on the gradient squared surface until one arrives at a local minimum. Then one replaces all pixels along this path by the image intensity at its corresponding extremum. This algorithm has linear complexity. The squared image gradient is calculated by Sobel operators.

Watershed algorithms often create too many segments. Although this oversegmentation is less dominant in the PDE-regularized image than in the original one, it may still lead to problems. Numerous ways have been proposed in order to deal with the oversegmentation problem, for instance by using markers [248, 267], region merging [397, 237, 288], or scale-space techniques [155, 194, 283]. In our case we shall see that a simple region merging strategy is adequate.

In such a step, adjacent regions are merged if their contrast difference is below a specified threshold. This contrast parameter can be related to the contrast parameter λ of the PDE-based regularization, thus it does not constitute an additional parameter. Finding a connected region where neighbouring pixels do not differ by more than a specified contrast value can be performed in linear complexity and the result is independent of the order in which the algorithm runs through the pixels. Thus, the entire segmentation algorithm is invariant under image rotations by 90 degrees and it reveals a linear total complexity.

A watershed segmentation of a 256^2 image with subsequent region merging takes about 0.2 CPU seconds on a PC or workstation. Thus, the overall segmentation time including the PDE-based regularization is far less than 1 second.

10.5 Experiments

Figure 10.3 illustrates how preprocessing by nonlinear diffusion filtering greatly reduces the number of segments in a medical MR image. We also observe that under nonlinear diffusion the segment boundaries remain well located and need not be traced back in order to improve their localization. As can be seen for instance at the cerebellum, the segments correspond well with the depicted physical objects. Moreover, segmentation of elongated objects does not create any problem. For comparison purposes with a classic approach, Figure 10.3(e) shows the result of an edge detector based on Sobel operators. Here the gradient magnitude has

to be postprocessed in order to give useful information. But even with sophisticated postprocessing strategies such as hysteresis thresholding and nonmaximum suppression, there remains one significant difference to a watershed segmentation: the contours are not closed. Hence such an operator does not give a partitioning of the image domain into segments. The latter one can be achieved by considering level sets of a differential operator such as the zero crossings of the Laplacian in Figure 10.3(f). In this case fairly large Gaussians are required in order to prevent oversegmentation. As a result, image structures become much more dislocated that in the watershed approach with nonlinear diffusion as preprocessing. This is clearly visible when comparing the contours of the cerebellum in Figs. 10.3(d) and (f). Another difference between these two approaches consists of the behaviour at junctions: as already mentioned and as is visible in Figure 10.3(f), zero crossings cannot meet at T-junctions, whereas watersheds do.

In Figure 10.4 it is shown that the merging step can be essential for avoiding the oversegmentation problems in the watershed algorithm. Nonlinear diffusion may create almost piecewise constant areas, but small fluctuations within such an area introduce many semantically irrelevant catchment basins. Such fluctuations can also be caused by quantization errors, e.g. by storing grey values in a bytewise manner. Merging adjacent regions with similar grey values constitutes a simple remedy for these problems.

Finally, Figure 10.5 gives a comparison between the two PDE-based regularization techniques. The results are in complete accordance with those from Figure 10.1. The contrast-enhancing nonlinear diffusion method gives more realistic results for images with less noise, as can be seen from the segmentation of the arms and legs in Figure 10.5(c). However, in a more noisy situation, the quality of this method degrades significantly. The variational method that does not allow contrast enhancement, on the other hand, does not reach the qualities of nonlinear diffusion preprocessing (Fig. 10.5(e)), but is very stable under noise (Fig. 10.5(f)). It is thus the better preprocessing method for noisy images. Again it should be emphasized that there exist more sophisticated nonlinear diffusion methods that combine the advantages of both approaches studied here [408]. Their efficient algorithmic realization, however, is more complicated and requires further research.

Table 10.2 presents precise CPU times for our segmentation algorithm both for a workstation (Sun Ultra 60) and a PC (Pentium II MMX, 440 Mhz), when 256^2 images are processed and a GNU C compiler is used. On both architectures, preprocessing by means of nonlinear diffusion or variational restoration can be achieved in about 0.5 seconds, while the watershed transformation with region merging takes 0.2 seconds. This shows that the complete algorithm allows to segment 256^2 images in much less than a second. With a PC with 700 MHz it is even possible to segment two such images in less than one second. Moreover, it should be taken into account that the AOS algorithm, which is the most time consuming subroutine

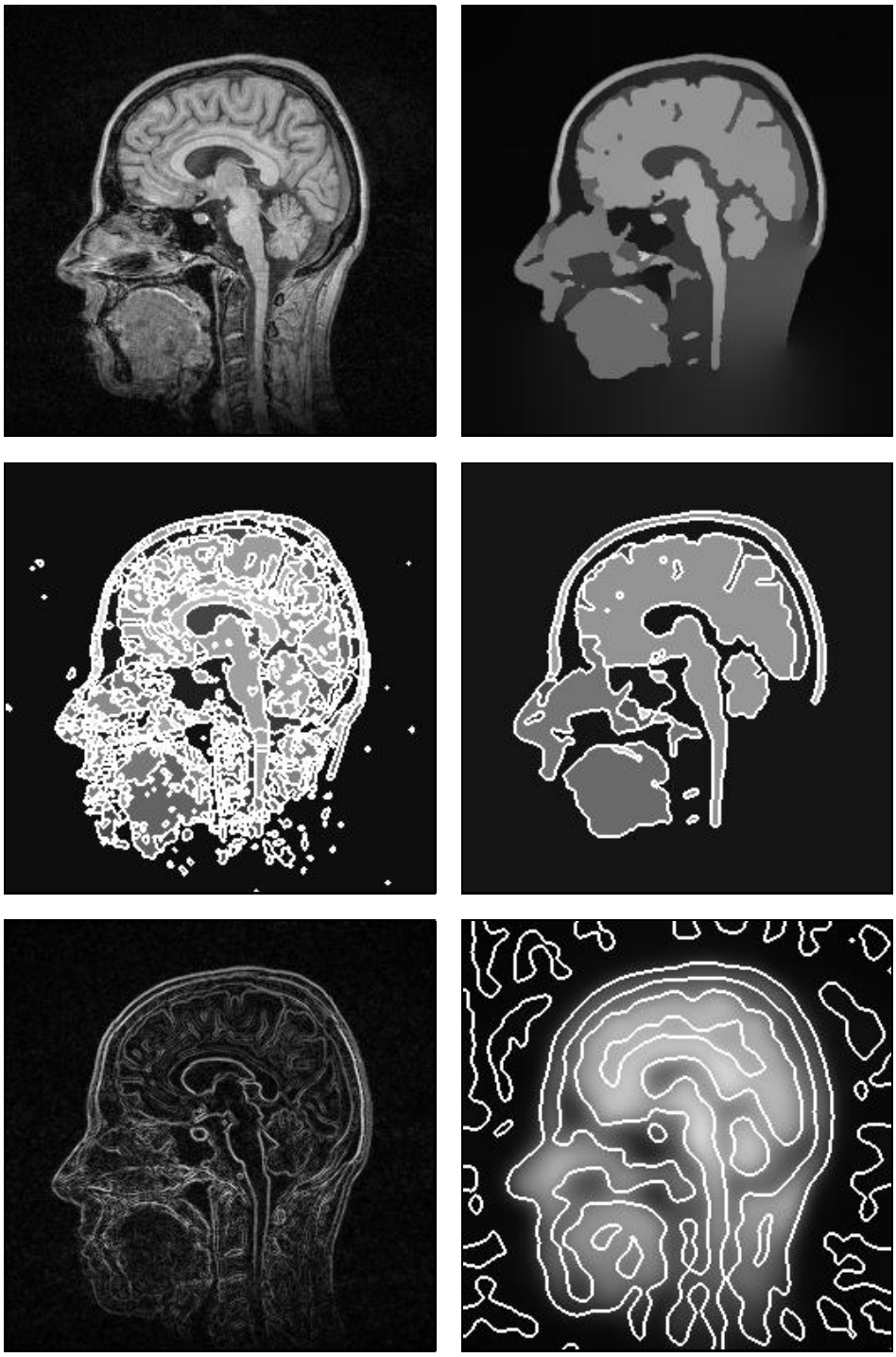


Figure 10.3: (a) Top Left: MR image. (b) Top Right: nonlinear diffusion filtering of (a). (c) Middle Left: segmentation of (a). (d) Middle Right: segmentation of (b). (e) Bottom Left: Sobel operator applied to (a). (f) Bottom Right: Zero crossings of the Laplacian-of-Gaussian of (a).

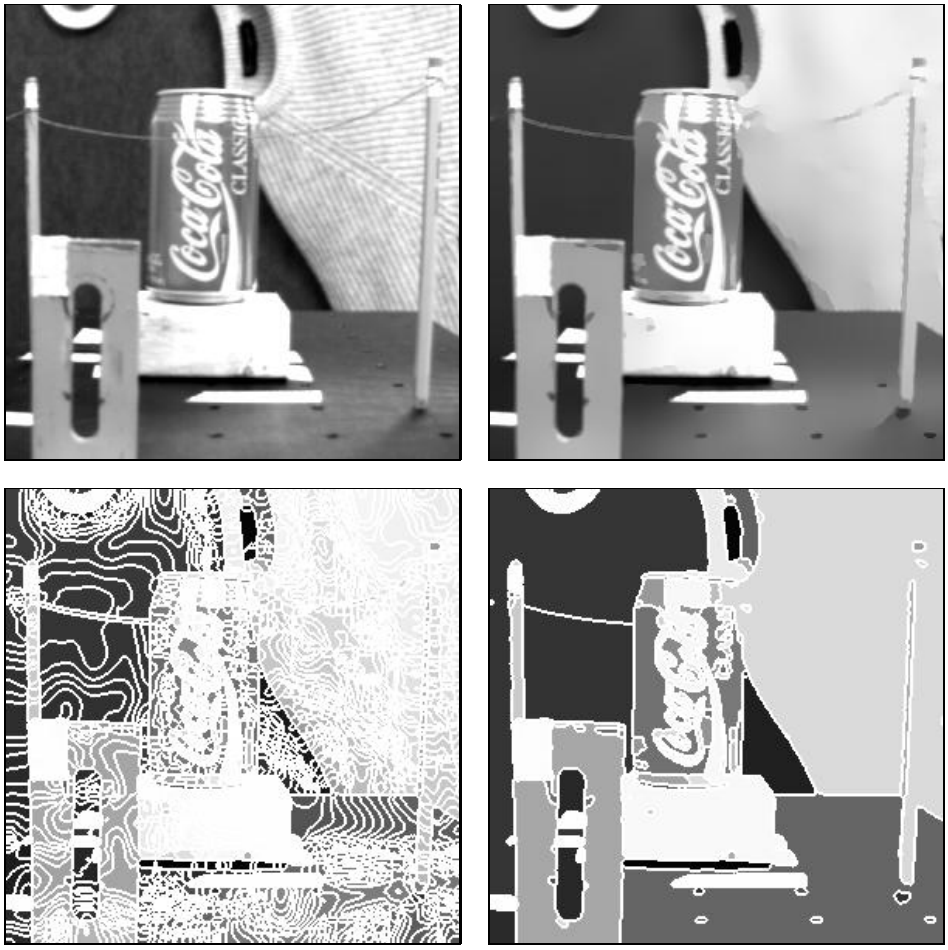


Figure 10.4: (a) Top Left: test image. (b) Top Right: nonlinear diffusion filtering of (a). (c) Bottom Left: segmentation of (b) without merging. (d) Bottom Right: segmentation of (b) with merging.

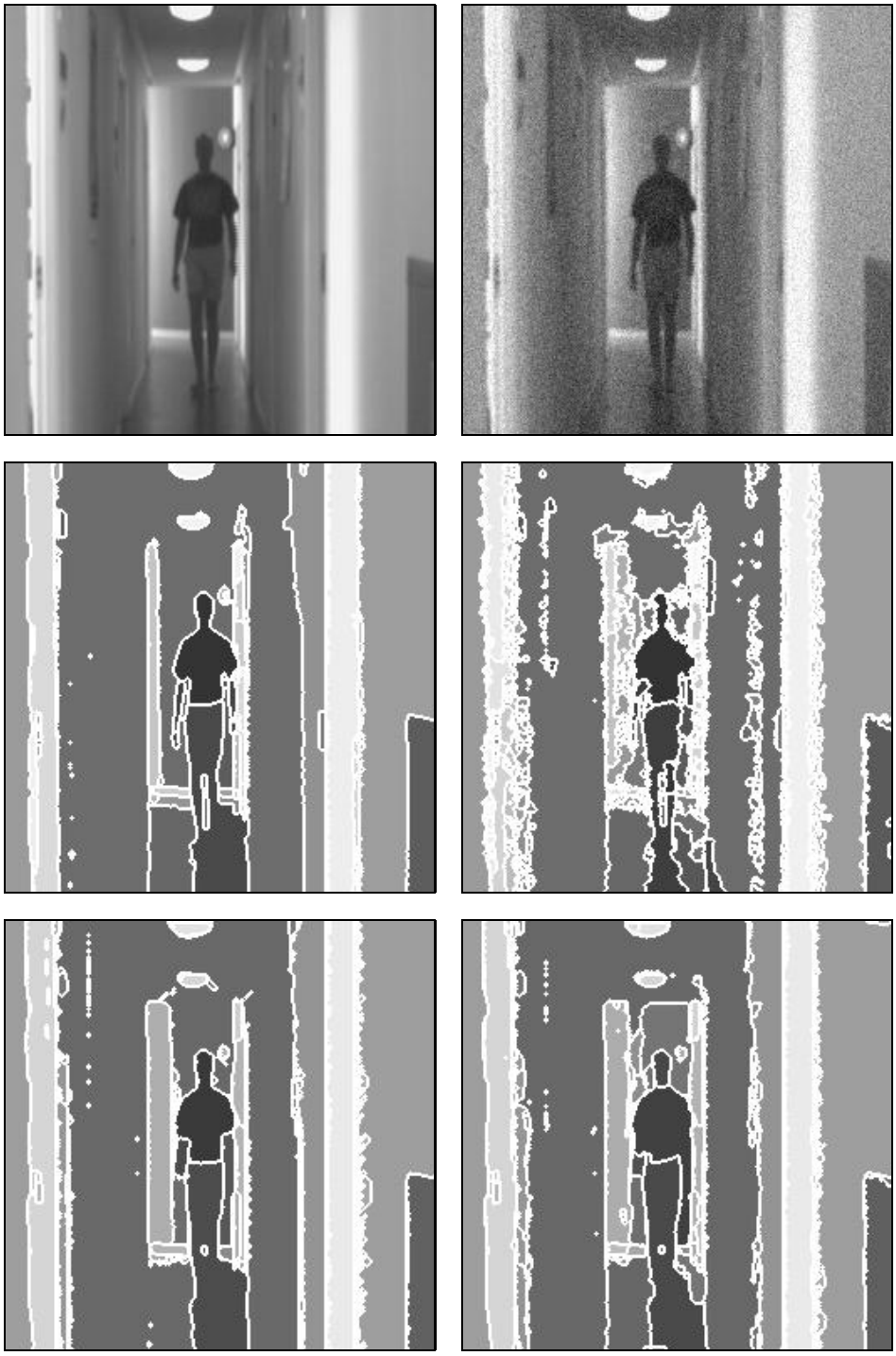


Figure 10.5: (a) Top Left: hallway scene. (b) Top Right: Gaussian noise added. (c) Middle Left: segmentation of (a) with nonlinear diffusion as preprocessing. (d) Middle Right: segmentation of (b) with nonlinear diffusion as preprocessing. (e) Bottom Left: segmentation of (a) with variational restoration as preprocessing. (f) Bottom Right: segmentation of (b) with variational restoration as preprocessing.

Table 10.2: CPU times for the different steps of the segmentation algorithm when 256^2 images are processed.

method	Sun Ultra 60	PC (Pentium II MMX, 440 Mhz)
nonlinear diffusion (5 iterations)	0.494 s	0.591 s
variational restoration (5 iterations per level)	0.407 s	0.516 s
watershed transformation with region merging	0.162 s	0.199 s

of the entire method, is very well-suited for parallel computing [419]. Therefore, further speed up can be achieved in a straightforward way.

10.6 Summary

We have presented efficient algorithms for two prototypes of PDE-based regularization techniques. These regularizations simplify subsequent segmentation tasks significantly, such that already a simple watershed algorithm with region merging gives good results. These segmentation techniques are very fast thanks to the use of AOS schemes and a novel pyramid AOS algorithm. This makes them attractive for many time-critical applications. All axes are treated equally, since the result is independent of the pixel ordering. The entire algorithm can be extended in a straightforward way to m-dimensional data, and the linear complexity in the pixel number remains valid in any dimension.

Acknowledgements. This work has been carried out within the EU-TMR network VIRGO. The author thanks Ole Fogh Olsen and Mads Nielsen for the hallway image, and Hans Oltmans for providing his implementation of Fairfield's watershed algorithm.

Part IV

Variational Optic Flow Computation

Chapter 11

Reliable Estimation of Dense Optical Flow Fields with Large Displacements

Luis Alvarez, Joachim Weickert, and Javier Sánchez.
International Journal of Computer Vision,
Vol. 39, No. 1, 41–56, August 2000.

Abstract

In this paper we show that a classic optical flow technique by Nagel and Enkelmann (1986) can be regarded as an early anisotropic diffusion method with a diffusion tensor. We introduce three improvements into the model formulation that (i) avoid inconsistencies caused by centering the brightness term and the smoothness term in different images, (ii) use a linear scale-space focusing strategy from coarse to fine scales for avoiding convergence to physically irrelevant local minima, and (iii) create an energy functional that is invariant under linear brightness changes. Applying a gradient descent method to the resulting energy functional leads to a system of diffusion–reaction equations. We prove that this system has a unique solution under realistic assumptions on the initial data, and we present an efficient linear implicit numerical scheme in detail. Our method creates flow fields with 100 % density over the entire image domain, it is robust under a large range of parameter variations, and it can recover displacement fields that are far beyond the typical one-pixel limits which are characteristic for many differential methods for determining optical flow. We show that it performs better than the optical flow methods with 100 % density that are evaluated by Barron et al. *(1994). Our software is available from the Internet.*

11.1 Introduction

The goal of optical flow computations is to estimate a so-called optical flow field which represents the apparent shift of greyvalue structures in the image plane. Recovering this displacement field is a key problem in computer vision and much research has been devoted to this field during the last two decades. For a survey of these activities we refer to Mitiche and Bouthemy [251], and performance evaluations of some of the most popular algorithms include papers of Barron *et al.* [35], Jähne and Haussecker [198], and Galvin *et al.* [139].
One important class of optical flow methods consists of so-called differential methods. Often they are considered as useful only in the case of small displacement fields. The goal of the present paper is to show that a combination of linear and nonlinear scale-space ideas may lead to a well-posed differential method that allows to recover the optical flow between two images with high accuracy, even in the case of large displacement fields.
We consider two images $I_1(x, y)$ and $I_2(x, y)$ (defined on $\mathbb{R}^2$ to simplify the discussion) which represent two consecutive views in a sequence of images. Under the assumption that corresponding pixels have equal grey values, the determination of the optical flow from I_1 to I_2 comes down to finding a function $\overline{h}(x, y) = (u(x, y), v(x, y))$ such that

$$I_1(x, y) \;=\; I_2(x + u(x, y), y + v(x, y)), \qquad \forall (x, y) \in \mathbb{R}^2. \tag{11.1}$$

To compute $\overline{h}(x, y)$ the preceding equality is usually linearized yielding the so-called *linearized optical flow constraint*

$$I_1(\overline{x}) - I_2(\overline{x}) \approx \left\langle \nabla I_2(\overline{x}), \overline{h}(\overline{x}) \right\rangle \qquad \forall \overline{x} \tag{11.2}$$

where $\overline{x} := (x, y)$. The linearized optical flow constraint is based on the assumption that the object displacements $\overline{h}(\overline{x})$ are small or that the image is slowly varying in space. In other cases, this linearization is no longer valid.
Frequently, instead of equation (11.1), the alternative equality

$$I_1(x - u(x, y), y - v(x, y)) \;=\; I_2(x, y), \qquad \forall (x, y) \in \mathbb{R}^2 \tag{11.3}$$

is used. In this case the displacement $\overline{h}(x, y)$ is centred in the image $I_2(x, y)$.
The determination of optical flow is a classic ill-posed problem in computer vision [40], and it requires to be supplemented with additional regularizing assumptions. The regularization by Horn and Schunck [171] reflects the assumption that the optical flow field varies smoothly in space. However, since many natural image sequences are better described in terms of piecewise smooth flow fields separated by discontinuities, much research has been done to modify the Horn and Schunck approach in order to permit such discontinuous flow fields [26, 44, 45, 49, 85, 105,

159, 166, 221, 246, 265, 270, 309, 344, 414]. A survey of these approaches will be presented at the end of this section.
An important improvement of the Horn and Schunck method has been achieved by Nagel and Enkelmann [265] in 1986 (see also [260]). They consider the following minimization problem:

$$\begin{aligned} E_{NE}(\overline{h}) &= \int_{\mathbb{R}^2} \left(I_1(x-u(x,y),y-v(x,y)) - I_2(x,y)\right)^2 \, dx\, dy \\ &\quad + c \int_{\mathbb{R}^2} \left(\nabla u^T D\left(\nabla I_1\right) \nabla u + \nabla v^T D\left(\nabla I_1\right) \nabla v\right) \, dx\, dy \end{aligned} \tag{11.4}$$

where c is a positive constant and $D\left(\nabla I_1\right)$ is a regularized projection matrix in the direction perpendicular of ∇I_1:

$$D\left(\nabla I_1\right) = \frac{1}{|\nabla I_1|^2 + 2\lambda^2} \left\{ \begin{pmatrix} \frac{\partial I_1}{\partial y} \\ -\frac{\partial I_1}{\partial x} \end{pmatrix} \begin{pmatrix} \frac{\partial I_1}{\partial y} \\ -\frac{\partial I_1}{\partial x} \end{pmatrix}^T + \lambda^2 Id \right\}. \tag{11.5}$$

In this formulation, Id denotes the identity matrix. The advantage of this method is that it inhibits blurring of the flow across boundaries of I_1 at locations where $|\nabla I_1| \gg \lambda$.
In spite of its merits, however, this method still leaves room for improvements:

(i) The Nagel–Enkelmann model uses an optical flow constraint which is centred in I_2, while the projection matrix D in the smoothness term depends on I_1. This inconsistency may create artifacts for large displacement fields.

(ii) Refraining from a linearization of the optical flow constraint has the consequence that the energy functional (11.6) may be nonconvex. In this case popular algorithms such as gradient descent methods may get trapped in physically irrelevant local minima.

(iii) Minimizers of the energy functional (11.4) are not invariant under linear brightness changes of the images I_1 and I_2.

In the present paper we will address these points by introducing three improvements into the Nagel–Enkelmann model:

(i) We design an energy functional that consistently centers both the optical flow constraint and the smoothness constraint in the same image.

(ii) We encourage convergence to the global energy minimum by embedding the method into a linear scale-space framework that allows to focus down from coarse to fine scales in small steps.

(iii) We introduce an adaptation of the parameters c and λ to the dynamic range of the images such that the resulting energy functional is invariant under linear brightness rescalings. This adaptation is particularly useful in the context of our scale-space focusing which alters the dynamic range of the images.

Applying the gradient descent method to our model leads to a coupled system of two diffusion–reaction equations, for which we establish the existence of a unique solution. Interestingly, these equations can be related to anisotropic diffusion filtering with a diffusion tensor. We present an efficient numerical scheme that is based on a linear implicit finite difference discretization. Afterwards, we discuss the role of the model parameters and demonstrate that our model allows very accurate recovery of optic flow fields for a large range of parameters. This is done by considering both synthetic image sequences, for which ground truth flow fields exist, as well as a real-world test sequence. Owing to the scale-space focusing, our model is particularly suited for recovering large displacement fields.
The paper is organized as follows: In Section 11.2 we describe our optical flow method that incorporates the three improvements, and we show that the Nagel–Enkelmann method and its modifications are closely related to anisotropic diffusion filtering. In Section 11.3 we present existence and uniqueness results for the nonlinear parabolic system that arises from using the gradient descent method for minimizing the energy functionals. Section 11.4 describes an efficient numerical discretization of this system based on a linear implicit finite difference scheme. Section 11.5 clarifies the role of the model parameters, and in Section 11.6 we present experimental results on synthetic and real-world image sequences. Finally, in Section 11.7 we conclude with a summary.

Related work. Proesmans *et al.* [309, 308] studied a related approach that also dispenses with a linearization of the optical flow constraint in order to allow for larger displacements. Their method, however, requires six coupled partial differential equations and its nonlinear diffusion process uses a scalar-valued diffusivity instead of a diffusion tensor. Their discontinuity-preserving smoothing is flow-driven while ours is image-driven. Another PDE technique that is similar in vein to the work of Proesmans *et al.* is a stereo method by Shah [354]. Other flow-driven regularizations with discontinuity-preserving properties include the work of Aubert *et al.* [26], Cohen [85], Deriche *et al.* [105], Hinterberger [170], Kumar *et al.* [221], Schnörr [344], Weickert [414], and Weickert and Schnörr [423]. Related stochastic regularization approaches have been studied by Black and Anandan [44, 45], Blanc–Féraud *et al.* [49], Heitz and Bouthemy [166], and Mémin and Pérez [246]. The image-driven anisotropic Nagel–Enkelmann approach has been subject to many subsequent studies. Examples include later work by Nagel [261, 264] as well as research by Schnörr [342, 343] and Snyder [361]. A multigrid realization of this method has been described by Enkelmann [115], and a related pyramid

framework is studied by Anandan [20]. An isotropic image-driven optic flow regularization is investigated by Alvarez *et al.* [11]. With respect to embeddings into a linear scale-space framework our method can be also be related to the optical flow approach of Florack *et al.* [129]. Their method differs from ours in that it is purely linear, applies scale selection mechanisms and does not use discontinuity-preserving nonlinear smoothness terms. Our focusing strategy for avoiding to end up in irrelevant local minima also resembles the *graduated non-convexity (GNC) algorithms* of Blake and Zisserman [48]. A preliminary version of our work has been presented at a conference [17], and a related optical flow method has been used by Hinterberger [170] to generate a movie between two images.

11.2 The Model

In this section we consider three modifications of the Nagel–Enkelmann model in order to improve its performance in the case of large displacement fields. We also discuss relations between this method and anisotropic diffusion filtering.

11.2.1 Consistent Centering

We have seen that the energy functional (11.4) uses an optical flow constraint and a smoothness term that are centred in different images. Our experiments showed that this inconsistency may lead to artifacts when the displacement field is large. As a remedy, we consider a modified energy functional where both the optical flow constraint and the smoothness constraint are related to I_1:

$$\begin{aligned} E(\overline{h}) &= \int_{\mathbb{R}^2} (I_1(x,y) - I_2(x+u(x,y), y+v(x,y)))^2 \, dx\, dy \\ &\quad + c \int_{\mathbb{R}^2} \left(\nabla u^T D\left(\nabla I_1\right) \nabla u + \nabla v^T D\left(\nabla I_1\right) \nabla v \right) dx\, dy. \end{aligned} \tag{11.6}$$

The associated Euler-Lagrange equations are given by the PDE system

$$c \operatorname{div}\left(D\left(\nabla I_1\right) \nabla u\right) + \left(I_1(\overline{x}) - I_2(\overline{x}+\overline{h}(\overline{x}))\right) \frac{\partial I_2}{\partial x}(\overline{x}+\overline{h}(\overline{x})) = 0, \tag{11.7}$$

$$c \operatorname{div}\left(D\left(\nabla I_1\right) \nabla v\right) + \left(I_1(\overline{x}) - I_2(\overline{x}+\overline{h}(\overline{x}))\right) \frac{\partial I_2}{\partial y}(\overline{x}+\overline{h}(\overline{x})) = 0. \tag{11.8}$$

In this paper, we are interested in solutions of the equations (11.7)-(11.8) in the case of *large* displacement fields and images that are not necessarily slowly varying in space. Therefore, we do not use the linearized optic flow constraint (11.2) in the above system.

11.2.2 Relations to Anisotropic Diffusion Filtering

We obtain the solutions of the Euler–Lagrange equations (11.7)-(11.8) by calculating the asymptotic state ($t \to \infty$) of the parabolic system

$$\frac{\partial u}{\partial t} = c\,\mathrm{div}\,(D\,(\nabla I_1)\;\nabla u) + \big(I_1(\overline{x}) - I_2(\overline{x}+\overline{h}(\overline{x}))\big)\,\frac{\partial I_2}{\partial x}(\overline{x}+\overline{h}(\overline{x})), \quad (11.9)$$

$$\frac{\partial v}{\partial t} = c\,\mathrm{div}\,(D\,(\nabla I_1)\;\nabla v) + \big(I_1(\overline{x}) - I_2(\overline{x}+\overline{h}(\overline{x}))\big)\,\frac{\partial I_2}{\partial y}(\overline{x}+\overline{h}(\overline{x})). \quad (11.10)$$

These equations do also arise when the steepest descent method is applied in order to minimize the energy (11.6).

Interestingly, this coupled system of diffusion–reaction equations reveals a diffusion tensor which resembles the one used for edge-enhancing anisotropic diffusion filtering. Indeed, $D(\nabla I_1)$ has the eigenvectors $v_1 := \nabla I_1$ and $v_2 := \nabla I_1^{\perp}$. The corresponding eigenvalues are given by

$$\lambda_1(|\nabla I_1|) = \frac{\lambda^2}{|\nabla I_1|^2 + 2\lambda^2}, \quad (11.11)$$

$$\lambda_2(|\nabla I_1|) = \frac{|\nabla I_1|^2 + \lambda^2}{|\nabla I_1|^2 + 2\lambda^2}. \quad (11.12)$$

We observe, that $\lambda_1 + \lambda_2 = 1$ holds independently of ∇I_1. In the interior of objects we have $|\nabla I_1| \to 0$, and therefore $\lambda_1 \to 1/2$ and $\lambda_2 \to 1/2$. At ideal edges where $|\nabla I_1| \to \infty$, we obtain $\lambda_1 \to 0$ and $\lambda_2 \to 1$. Thus, we have isotropic behaviour within regions, and at image boundaries the process smoothes anisotropically along the edge. This behaviour is very similar to edge-enhancing anisotropic diffusion filtering [408], and it is also close in spirit to the modified mean-curvature motion considered in [13]. In this sense, one may regard the Nagel–Enkelmann method as an early predecessor of modern PDE techniques for image restoration.

One structural difference, however, should be observed: the optical flow equations (11.9)–(11.10) use a temporally constant diffusion tensor, while the nonlinear diffusion tensor of anisotropic diffusion filtering is a function of the evolving image itself. Hence, the Nagel–Enkelmann model is anisotropic and space-variant, but it remains linear in its diffusion part. Related linear anisotropic diffusion filters have been pioneered by Iijima in the sixties and seventies in the context of optical character recognition; see [421] and the references therein. For a detailed treatment of anisotropic diffusion filtering we refer to [412], an axiomatic classification of mean-curvature motion and related morphological PDEs for image analysis is presented in [12], and recent collections of papers on PDE-based image smoothing methods include [38, 70, 381, 274].

11.2.3 Recovering Large Displacements by Scale-Space Focusing

The energy functional (11.6) may be nonconvex due to its data term without linearization. In this case we cannot expect the uniqueness of solutions of the elliptic system (11.7)-(11.8). As a consequence, the asymptotic state of the parabolic system (11.9)-(11.10), which we use for approximating the optical flow, depends on the initial data. Typically, we may expect that the algorithm converges to a local minimizer of the energy functional (11.6) that is located in the vicinity of the initial data. When we have small displacements in the scene, the natural choice is to take $u \equiv v \equiv 0$ as initialization of the flow. For large displacement fields, however, this may not work, and we need better initial data. To this end, we embed our method into a linear scale-space framework [175, 421]. Considering the problem at a coarse scale avoids that the algorithm gets trapped in physically irrelevant local minima. The coarse-scale solution serves then as initial data for solving the problem at a finer scale. Scale focusing has a long tradition in linear scale-space theory (see e.g. Bergholm [39] for an early approach), and in spite of the fact that some theoretical questions remain open, it has not lost its popularity. For more details on linear scale-space theory we refer to [127, 187, 188, 230, 369]. Using a scale-space approach enables us also to perform a finer and more reliable scale focusing as would be the case for related pyramid or multigrid approaches.
We proceed as follows. First, we introduce a linear scale factor in the parabolic PDE system in order to end up with

$$\begin{aligned} \frac{\partial u_\sigma}{\partial t} &= c \operatorname{div} \left(D \left(\nabla I_1^\sigma\right) \nabla u_\sigma\right) \\ &\quad + \left(I_1^\sigma(\overline{x}) - I_2^\sigma(\overline{x} + \overline{h}_\sigma(\overline{x}))\right) \frac{\partial I_2^\sigma}{\partial x}(\overline{x} + \overline{h}_\sigma(\overline{x})), \qquad (11.13) \\ \frac{\partial v_\sigma}{\partial t} &= c \operatorname{div} \left(D \left(\nabla I_1^\sigma\right) \nabla v_\sigma\right) \\ &\quad + \left(I_1^\sigma(\overline{x}) - I_2^\sigma(\overline{x} + \overline{h}_\sigma(\overline{x}))\right) \frac{\partial I_2^\sigma}{\partial y}(\overline{x} + \overline{h}_\sigma(\overline{x})) \qquad (11.14) \end{aligned}$$

where $I_1^\sigma := G_\sigma * I_1$, $I_2^\sigma := G_\sigma * I_2$, $\overline{h}_\sigma(\overline{x}) := (u_\sigma(\overline{x}), v_\sigma(\overline{x}))$, and $G_\sigma * I_j$ represents the convolution of I_j with a Gaussian of standard deviation σ.
The convolution with a Gaussian blends the information in the images and allows us to recover a connection between the objects in I_1 and I_2. In our application, this global support property that is characteristic for linear diffusion scale-spaces is very important. It makes them favourable over morphological scale-spaces in the sense of [12], since the latter ones cannot transport information between topologically disconnected objects.
We start with a large initial scale σ_0. Then we compute the optical flow $(u_{\sigma_0}, v_{\sigma_0})$ at scale σ_0 as the asymptotic state of the solution of the above PDE system using

as initial data $u \equiv v \equiv 0$. Next, we choose a number of scales $\sigma_n < \sigma_{n-1} < .... < \sigma_0$, and for each scale σ_i we compute the optical flow $(u_{\sigma_i}, v_{\sigma_i})$ as the asymptotic state of the above PDE system with initial data $(u_{\sigma_{i-1}}, v_{\sigma_{i-1}})$. The final computed flow corresponds to the smallest scale σ_n. In accordance with the logarithmic sampling strategy in linear scale-space theory [213], we choose $\sigma_i := \eta^i \sigma_0$ with some decay rate $\eta \in (0, 1)$.

11.2.4 Invariance Under Linear Greyvalue Transformations

A remaining shortcoming of the modified model is that the energy $E(\overline{h})$ is not invariant under grey level transformation of the form $(I_1, I_2) \to (kI_1, kI_2)$, where k is a constant. Therefore, the choice of the parameters depends strongly on the image contrast. This is especially problematic when the method is embedded in the scale-space focusing strategy, since the amount of smoothing influences the contrast range in the regularized images $G_\sigma * I_1$ and $G_\sigma * I_2$.
We address this problem by normalizing the parameters c and λ in such a way that the energy $E(\overline{h})$ becomes invariant under grey level transformation of the form $(I_1, I_2) \to (kI_1, kI_2)$. We compute c and λ by means of two parameters α and $s \in (0, 1)$ that are calculated via

$$\begin{aligned} c &= \frac{\alpha}{\max_{\overline{x}}(|(\nabla G_\sigma * I_1)(\overline{x})|^2)}, \\ s &= \int_0^\lambda \mathcal{H}_{|\nabla G_\sigma * I_1|}(z) dz \end{aligned}$$

where $\mathcal{H}_{|\nabla G_\sigma * I_1|}(z)$ represents the normalized histogram of $|\nabla G_\sigma * I_1|$. We name s the *isotropy fraction*. When $s \to 0$, the diffusion operator becomes anisotropic at all locations, and when $s \to 1$, it leads to isotropic diffusion everywhere. So now $c = c(\alpha, \nabla G_\sigma * I_1)$, and $\lambda = \lambda(s, \nabla G_\sigma * I_1)$. With this normalization of c and λ, the energy $E(\overline{h})$ is invariant under grey level transformation of the form $(I_1, I_2) \to (kI_1, kI_2)$. In practical applications of our method it is thus sufficient to specify the parameters α and s instead of c and λ. The parameters c and λ are then automatically adjusted to the dynamic image range in each step of the focusing procedure.

11.3 Existence and Uniqueness of the Parabolic System

In this section we show the existence and uniqueness of solutions of the parabolic system (11.13)-(11.14) where $D(\nabla I_1^\sigma)$ is given by (11.5). The parameters c and λ can be arbitrary positive real numbers. In particular, they may be determined as

described in the previous section. First we introduce an abstract framework where we study the above system. This framework is used to show the existence and uniqueness of the solutions afterwards.

11.3.1 Abstract Framework

For simplicity we assume that the images are defined on the entire space $\mathbb{R}^2$. We assume that the input images I_1 and I_2 belong to the functional space $L^2(\mathbb{R}^2)$. Let $H = L^2(\mathbb{R}^2) \times L^2(\mathbb{R}^2)$, and let us denote by $A : \mathcal{D}(A) \subset H \to H$ the differential operator defined by

$$A(\overline{h}) = -c \begin{pmatrix} \operatorname{div}\left(D\left(\nabla I_1^\sigma\right) \nabla u_\sigma\right) \\ \operatorname{div}\left(D\left(\nabla I_1^\sigma\right) \nabla v_\sigma\right) \end{pmatrix},$$

where $\mathcal{D}(A)$ is the domain of the mapping A. If $I_1 \in L^2(\mathbb{R}^2)$ then $I_1^\sigma \in W^{1,\infty}(\mathbb{R}^2)$, so ∇I_1^σ is bounded and the eigenvalues of the diffusion tensor $D\left(\nabla I_1^\sigma\right)$ are strictly positive. Therefore, as $c > 0$, the operator $A(\overline{h})$ is a maximal monotone operator. For more details about maximal monotone operators we refer to Brezis [51]. Next, let us introduce the function $F : H \to H$ defined by

$$F(\overline{h}) = \left(I_1^\sigma - I_2^\sigma(Id + \overline{h})\right) \nabla I_2^\sigma(Id + \overline{h}).$$

Then the abstract evolution problem can be written as

$$\begin{cases} \frac{d\overline{h}_\sigma}{dt} + A\overline{h}_\sigma = F(\overline{h}_\sigma) \text{ in } H, \ \forall t \in [0,T] \\ \overline{h}_\sigma(0) = \overline{h}^0 \text{ in } H. \end{cases} \tag{11.15}$$

Any classical solution $\overline{h}_\sigma \in C^1([0,T];H) \cap C([0,T];\mathcal{D}(A))$ of (11.15) is given by

$$\overline{h}_\sigma(t) = S(t)\overline{h}^0 + \int_0^t S(t-s)F(\overline{h}_\sigma(s))ds, \tag{11.16}$$

where $\{S(t)\}_{t>0}$ is the contraction semi-group associated to the homogeneous problem.

Definition. *We say that $h \in C([0,T];H)$ is a generalized solution of (11.15) if it satisfies (11.16).*

11.3.2 Existence and Uniqueness Result

In order to prove existence and uniqueness, we have to establish a lemma first.

Lemma 8 *Suppose that $I_1, I_2 \in L^2(\mathbb{R}^2)$, then F is Lipschitz-continuous, and the Lipschitz constant L depends on the functions I_1 and I_2 and on σ.*

Proof:
First we note that if $I_1, I_2 \in L^2(\mathbb{R}^2)$, then we have in particular that $I_2^\sigma \in W^{1,\infty}(\mathbb{R}^2)$ and $I_1^\sigma \in L^\infty(\mathbb{R}^2)$. Let $\overline{h}_1, \overline{h}_2 \in H$. For the i-th component of $F(\overline{h}_1) - F(\overline{h}_2)$, $i = 1, 2$, we have the following pointwise estimate.

$$\begin{aligned}
|F_i(\overline{h}_1) - F_i(\overline{h}_2)| &= |(I_1^\sigma - I_2^\sigma(Id + \overline{h}_1))\partial_i I_2^\sigma(Id + \overline{h}_1) \\
&\quad -(I_1^\sigma - I_2^\sigma(Id + \overline{h}_2))\partial_i I_2^\sigma(Id + \overline{h}_2)|, \\
&\leq |I_2^\sigma(Id + \overline{h}_1)\partial_i I_2^\sigma(Id + \bar{h}_1) - I_2^\sigma(Id + \overline{h}_2)\partial_i I_2^\sigma(Id + \bar{h}_2)| \\
&\quad +|I_1^\sigma| \cdot |\partial_i I_2^\sigma(Id + \overline{h}_1) - \partial_i I_2^\sigma(Id + \overline{h}_2)|, \\
&\leq \frac{1}{2}|\partial_i(|I_2^\sigma|^2)(Id + \overline{h}_1) - \partial_i(|I_2^\sigma|^2)(Id + \overline{h}_2)| \\
&\quad +\|I_1^\sigma\|_\infty \cdot |\partial_i I_2^\sigma(Id + \bar{h}_1) - \partial_i I_2^\sigma(Id + \overline{h}_2)|, \\
&\leq \frac{1}{2}C_{Lip}(\partial_i(|I_2^\sigma|^2)) \cdot |\overline{h}_1 - \overline{h}_2| + \|I_1^\sigma\|_\infty \cdot C_{Lip}(\partial_i I_2^\sigma) \cdot |\overline{h}_1 - \overline{h}_2| \\
&\leq \left(\frac{1}{2}C_{Lip}(\partial_i(|I_2^\sigma|^2)) + \|I_1^\sigma\|_\infty \cdot C_{Lip}(\partial_i I_2^\sigma)\right) \cdot |\overline{h}_1 - \overline{h}_2|,
\end{aligned}$$

where $C_{Lip}(f)$ denotes the Lipschitz constant of the function f.
We finally deduce that

$$\begin{aligned}
\|F(\overline{h}_1) - F(\overline{h}_2)\|_H &= \|F_1(\overline{h}_1) - F_1(\overline{h}_2)\|_{L^2} + \|F_2(\overline{h}_1) - F_2(\overline{h}_2)\|_{L^2} \\
&\leq \sum_{i=1}^{2}\left(\frac{1}{2}C_{Lip}(\partial_i(|I_2^\sigma|^2)) + \|I_1^\sigma\|_\infty \cdot C_{Lip}(\partial_i I_2^\sigma)\right) \cdot \|\overline{h}_1 - \overline{h}_2\|_H.
\end{aligned}$$

We conclude the proof of the lemma by setting

$$L := \sum_{i=1}^{2}\left(\frac{1}{2}C_{Lip}\left(\partial_i(|I_2^\sigma|^2)\right) + \|I_1^\sigma\|_\infty \cdot C_{Lip}(\partial_i I_2^\sigma)\right).$$

This shows the assertion. □

Now we can state the existence and uniqueness result for problem (11.13)-(11.14).

Theorem 15 *Suppose that $I_1, I_2 \in L^2(\mathbb{R}^2)$ then, for all $\overline{h}^0 \in H$, there exists a unique generalized solution $\overline{h}_\sigma(t) \in C([0,\infty); H)$ of (11.13)-(11.14).*

Proof:
The assumptions on I_1 and I_2 allow us to apply Lemma 8. Assume that $\overline{h}_1(t)$ and $\overline{h}_2(t)$ are solutions of (11.16) for initial conditions $\overline{h}_1(0)$ and $\overline{h}_2(0)$, then we have, using the fact that A is monotone (which yields $\|S(t)f\|_H \leq \|f\|_H$), and the Lipschitz continuity of F the following estimate.

$$\|\overline{h}_1(t) - \overline{h}_2(t)\|_H \leq \|\overline{h}_1(0) - \overline{h}_2(0)\|_H + L\int_0^t \|\overline{h}_1(s) - \overline{h}_2(s)\|_H \, ds.$$

Applying the Gronwall–Bellman lemma [51] gives

$$\|\overline{h}_1(t) - \overline{h}_2(t)\|_H \leq e^{Lt}.\|\overline{h}_1(0) - \overline{h}_2(0)\|_H,$$

which yields uniqueness of the solution if it exists. Now consider the Banach space defined by

$$E = \{\overline{h} \in C\left([0,\infty); H\right) \ : \ \sup_{t\geq 0} \|\overline{h}(t)\|_H e^{-Kt} < \infty\}$$

endowed with the norm $\|\overline{h}\|_E = \sup_{t\geq 0} \|\overline{h}(t)\|_H e^{-Kt}$.
Let $\phi : E \to C\left([0,\infty); H\right)$ be defined by

$$\phi(\overline{h})(t) = S(t)\overline{h}^0 + \int_0^t S(t-s)F(\overline{h}(s))ds.$$

If $K > L$, then $\phi(E) \subset E$, and ϕ is $\frac{L}{K}$-Lipschitz since

$$\begin{aligned}
\|\phi(\overline{h}_1) - \phi(\overline{h}_2)\|_E &= \sup_{t\geq 0} \|\phi(\bar{h}_1)(t) - \phi(\overline{h}_2)(t)\|_H e^{-Kt}, \\
&\leq \sup_{t\geq 0} \int_0^t L\|\overline{h}_1(s) - \overline{h}_2(s)\|_H \, ds \, e^{-Kt} \\
&\leq \sup_{t\geq 0} L\|\overline{h}_1 - \overline{h}_2\|_E \cdot e^{-Kt} \int_0^t e^{Ks} \, ds \\
&\leq \sup_{t\geq 0} \frac{L}{K}\|\overline{h}_1 - \overline{h}_2\|_E \cdot e^{-Kt}(e^{Kt} - 1) \\
&\leq \frac{L}{K}\|\overline{h}_1 - \overline{h}_2\|_E.
\end{aligned}$$

We deduce that ϕ is a contraction, and by Banach's fixed point theorem there exists a unique $\overline{h}_\sigma$ such that $\phi(\overline{h}_\sigma) = \overline{h}_\sigma$. This is the generalized solution of (11.15), and the proof is concluded. □

Remark. We notice that our existence and uniqueness proof is based on rather weak assumptions on the initial images I_1 and I_2. We only assumed square integrability. They do not have to be continuous and may even be corrupted by noise or quantization artifacts, as is common for real-world images. The behaviour of the solution when σ goes to 0 is a challenging mathematical problem. If I_1 and I_2 are not sufficiently smooth, we cannot expect a good asymptotic behaviour when σ tends to 0. This suggests that the original images I_1 and I_2 should always be preprocessed by some small amount of Gaussian smoothing. In our experiments we shall also observe that it can be advantageous to stop the focusing procedure when σ attains the order of the pixel size.

11.4 Numerical Scheme

Next we describe an efficient algorithm for our optical flow model. We discretize the parabolic system (11.13)–(11.14) by finite differences (see e.g. [256] for an introduction to this subject). All spatial derivatives are approximated by central differences, and for the discretization in t direction we use a linear implicit scheme. Let $D(\nabla G_\sigma * I_1) =: \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$. Then our linear implicit scheme has the structure

$$\begin{aligned}
\frac{u_{i,j}^{k+1} - u_{i,j}^{k}}{\tau} &= c\left(\frac{a_{i+1,j} + a_{i,j}}{2}\,\frac{u_{i+1,j}^{k+1} - u_{i,j}^{k+1}}{h_1^2} + \frac{a_{i-1,j} + a_{i,j}}{2}\,\frac{u_{i-1,j}^{k+1} - u_{i,j}^{k+1}}{h_1^2}\right.\\
&\quad + \frac{c_{i,j+1} + c_{i,j}}{2}\,\frac{u_{i,j+1}^{k+1} - u_{i,j}^{k+1}}{h_2^2} + \frac{c_{i,j-1} + c_{i,j}}{2}\,\frac{u_{i,j-1}^{k+1} - u_{i,j}^{k+1}}{h_2^2}\\
&\quad + \frac{b_{i+1,j+1} + b_{i,j}}{2}\,\frac{u_{i+1,j+1}^{k+1} - u_{i,j}^{k+1}}{2h_1h_2} + \frac{b_{i-1,j-1} + b_{i,j}}{2}\,\frac{u_{i-1,j-1}^{k+1} - u_{i,j}^{k+1}}{2h_1h_2}\\
&\quad \left. - \frac{b_{i+1,j-1} + b_{i,j}}{2}\,\frac{u_{i+1,j-1}^{k+1} - u_{i,j}^{k+1}}{2h_1h_2} - \frac{b_{i-1,j+1} + b_{i,j}}{2}\,\frac{u_{i-1,j+1}^{k+1} - u_{i,j}^{k+1}}{2h_1h_2}\right)\\
&\quad + I_{2,x}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\left(I_1(\overline{x}_{i,j}) - I_2(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\right.\\
&\quad \left. + u_{i,j}^{k} I_{2,x}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}) + v_{i,j}^{k} I_{2,y}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\right)\\
&\quad - u_{i,j}^{k+1} I_{2,x}^2(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}) - v_{i,j}^{k+1} I_{2,y}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\, I_{2,x}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}), \qquad (11.17)
\end{aligned}$$

$$\begin{aligned}
\frac{v_{i,j}^{k+1} - v_{i,j}^{k}}{\tau} &= c\left(\frac{a_{i+1,j} + a_{i,j}}{2}\,\frac{v_{i+1,j}^{k+1} - v_{i,j}^{k+1}}{h_1^2} + \frac{a_{i-1,j} + a_{i,j}}{2}\,\frac{v_{i-1,j}^{k+1} - v_{i,j}^{k+1}}{h_1^2}\right.\\
&\quad + \frac{c_{i,j+1} + c_{i,j}}{2}\,\frac{v_{i,j+1}^{k+1} - v_{i,j}^{k+1}}{h_2^2} + \frac{c_{i,j-1} + c_{i,j}}{2}\,\frac{v_{i,j-1}^{k+1} - v_{i,j}^{k+1}}{h_2^2}\\
&\quad + \frac{b_{i+1,j+1} + b_{i,j}}{2}\,\frac{v_{i+1,j+1}^{k+1} - v_{i,j}^{k+1}}{2h_1h_2} + \frac{b_{i-1,j-1} + b_{i,j}}{2}\,\frac{v_{i-1,j-1}^{k+1} - v_{i,j}^{k+1}}{2h_1h_2}\\
&\quad \left. - \frac{b_{i+1,j-1} + b_{i,j}}{2}\,\frac{v_{i+1,j-1}^{k+1} - v_{i,j}^{k+1}}{2h_1h_2} - \frac{b_{i-1,j+1} + b_{i,j}}{2}\,\frac{v_{i-1,j+1}^{k+1} - v_{i,j}^{k+1}}{2h_1h_2}\right)\\
&\quad + I_{2,y}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\left(I_1(\overline{x}_{i,j}) - I_2(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\right.\\
&\quad \left. + u_{i,j}^{k} I_{2,x}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}) + v_{i,j}^{k} I_{2,y}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\right)\\
&\quad - u_{i,j}^{k+1} I_{2,x}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k})\, I_{2,y}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}) - v_{i,j}^{k+1} I_{2,y}^2(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}). \qquad (11.18)
\end{aligned}$$

Although this scheme might look fairly complicated at first glance, it is actually straightforward to implement. The notations are almost selfexplaining: for instance, τ is the time step size, h_1 and h_2 denote the pixel size in x and y direction, respectively, $u_{i,j}^k$ approximates u_σ in some grid point $\overline{x}_{i,j}$ at time $k\tau$, and $I_{1,x}$ is an approximation to $G_\sigma * \frac{\partial I_1}{\partial x}$. We calculate values of type $I_2(\overline{x}_{i,j} + \overline{h}_{i,j}^k)$ by linear

interpolation.
The idea behind linear implicit schemes is to use implicit discretizations in order to improve stability properties, as long that they lead to linear systems of equations. Implicit means that unknowns of the new time level appear on both sides of the equation. For nonlinear equations, fully implicit methods would require a computationally expensive solution of nonlinear systems of equations. The computationally less expensive linear implicit methods avoid this by using suitable Taylor expansions, such that the resulting system of equations become linear. In our case we achieved this by using the first-order Taylor expansion

$$\begin{aligned} I_2(\overline{x}_{i,j} + \overline{h}_{i,j}^{k+1}) &\approx I_2(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}) \\ &+ (u_{i,j}^{k+1} - u_{i,j}^{k})\, I_{2,x}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}) \\ &+ (v_{i,j}^{k+1} - v_{i,j}^{k})\, I_{2,y}(\overline{x}_{i,j} + \overline{h}_{i,j}^{k}) \end{aligned}$$

in a fully implicit discretization, and by discretizing $G_\sigma * \frac{\partial I_2}{\partial x}$ and $G_\sigma * \frac{\partial I_2}{\partial y}$ in an explicit way. A consistency analysis shows that the preceding scheme is of second order in space and of first order in time.
We solve the resulting linear system of equations iteratively by a symmetric Gauß–Seidel algorithm. In order to explain its structure let us suppose that we want to solve a linear system $A\overline{w} = \overline{b}$ where $A = D - L - U$ and D is a diagonal matrix, L a strictly lower triangular matrix, and U a strictly upper triangular matrix. Then the symmetric Gauß–Seidel iterations are given by

$$\begin{aligned} (D - L)\, \overline{w}^{(n+1/2)} &= \overline{b} + U\, \overline{w}^{(n)}, \\ (D - U)\, \overline{w}^{(n+1)} &= \overline{b} + L\, \overline{w}^{(n+1/2)} \end{aligned}$$

where the upper index denotes the iteration index. The systems are solved directly using forward and backward elimination, respectively.
In an earlier version of our work [17] we have studied an explicit scheme that did not require to solve linear systems of equations. The linear implicit approach that we employ in the meantime, however, has led to a speed-up of one to two orders of magnitude, since it allows significantly larger time step sizes without creating stability problems.

11.5 Parameters

Our algorithm for computing the optical flow depends on a number of parameters that have an intuitive meaning:

- The regularization parameter α specifies the balance between the smoothing term and the optical flow constraint. Larger values lead to smoother flow

fields by filling in information from image edges where flow measurements with higher reliability are available.

- The isotropy fraction s determines the contrast parameter λ via the cumulative histogram of the image gradient magnitude. Choosing e.g. $s := 0.7$ means that the smoothness term diffuses isotropically at 70 % of all image locations, while 30 % of all locations are assumed to belong to image edges, where smoothing is performed anisotropically along the edge.

- The scale σ_0 denotes the standard deviation of the largest Gaussian. In general, σ_0 is chosen according to the maximum displacement expected.

- The decay rate $\eta \in (0, 1)$ for the computation of the scales $\sigma_m := \eta^m \sigma_0$. We may expect a good focusing if η is close to 1.

- The smallest scale is given by σ_n. It should be close to the inner scale of the image in order to achieve optimal flow localization.

- The time step size τ and the stopping time T for solving the system (11.13)–(11.14) at each scale σ_m are pure numerical parameters. We experienced that fixing $\tau := 10$ and $T := 500$ creates results that are sufficiently close to the asymptotic state. Using smaller values of τ or larger values of T slowed down the algorithm without improving the quality of the flow fields.

In the next section we will see that the results of our method are hardly affected by fairly large parameter variations. As a consequence default values can be used for most of the parameters.

11.6 Experimental Results

Figure 11.1 shows our first experiment. We use a synthetic image composed of four black squares on a white background. Each square moves in a different direction and with a different displacement magnitude: under the assumption that the x axis is oriented from left to right and the y axis from top to bottom, the left square on the top moves with $(u, v) = (5, 10)$, the right square on the top is displaced with $(u, v) = (-10, 0)$, the left square on the bottom is shifted by $(u, v) = (0, -5)$, and the right square on the bottom undergoes a translation by $(-10, -10)$. In order to visualize the flow field (u, v) we use two grey level images (u_{gl}, v_{gl}) defined by $u_{gl} := 128 + 12u$ and $v_{gl} := 128 + 12v$. In Figure 11.1 we notice that the flow estimates improve significantly by focusing down from $\sigma_0 := 10$ to $\sigma_{50} := 0.8$: flow discontinuities evolve and the calculated flow fields approximate the true motion field more and more.

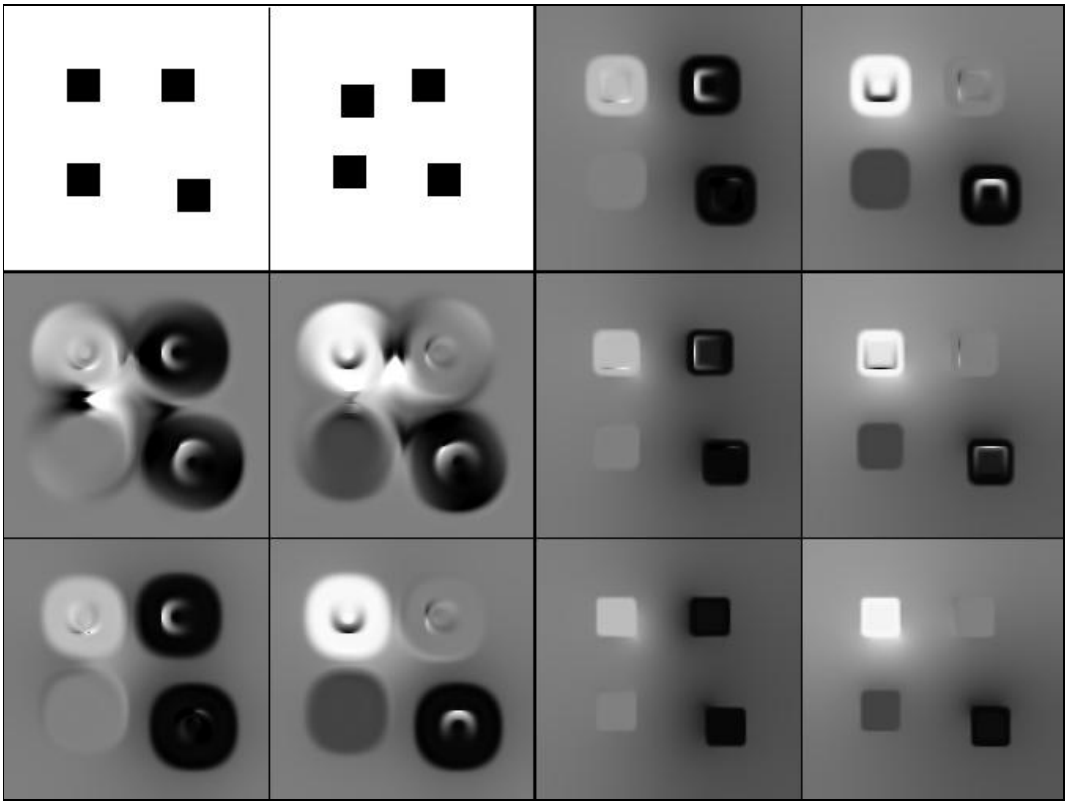


Figure 11.1: Computation of the optical flow for the square images with $\alpha = 0.6$, $s = 0.1$, and $\eta = 0.95$. (a) First Row, Left Pair: Original image pair. (b) Second Row, Left Pair: Optical flow components (u, v) for $\sigma_0 = 10$. (c) Third Row, Left Pair: Optical flow result after focusing down to $\sigma_{12} = 5.7$. (d) First Row, Right Pair: $\sigma_{25} = 2.9$. (e) Second Row, Right Pair: $\sigma_{37} = 1.4$, (f) Third Row, Right Pair: $\sigma_{50} = 0.8$.

This qualitative observation is confirmed in the quantitative evaluations carried out in Figure 11.2. The left plot shows the average angular errors in the four squares of the first frame. The angular error Ψ_e has been calculated in the same way as in Barron *et al.* [35] using

$$\Psi_e := \arccos\left(\frac{u_c u_e + v_c v_e + 1}{\sqrt{(u_c^2 + v_c^2 + 1)(u_e^2 + v_e^2 + 1)}}\right) \tag{11.19}$$

where (u_c, v_c) denotes the correct flow, and (u_e, v_e) is the estimated flow. The right plot depicts the Euclidean error $\sqrt{(u_e - u_c)^2 + (v_e - v_c)^2}$ averaged over all pixels within the four squares of the first frame.
In both cases we observe that the error is reduced drastically by focusing down in scale-space until it reaches a very small value when the Gaussian width σ ap-

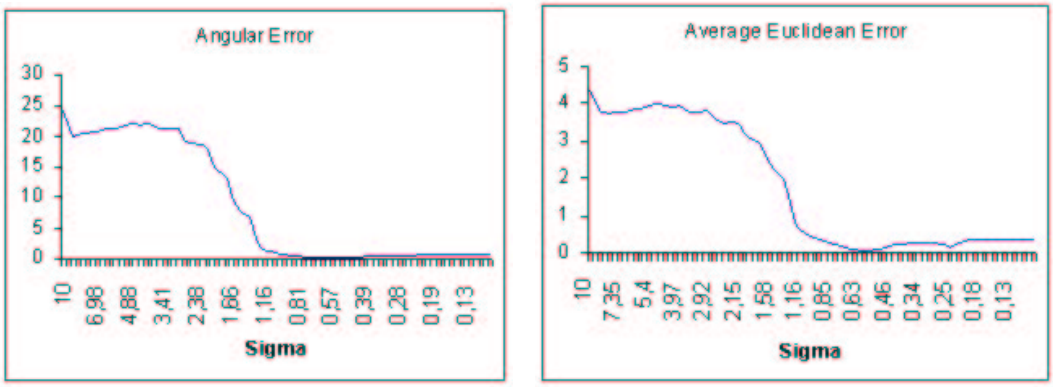


Figure 11.2: LEFT: Average angular error of the optic flow calculations for the squares in the first frame of Figure 11.1. RIGHT: Corresponding average Euclidean error.

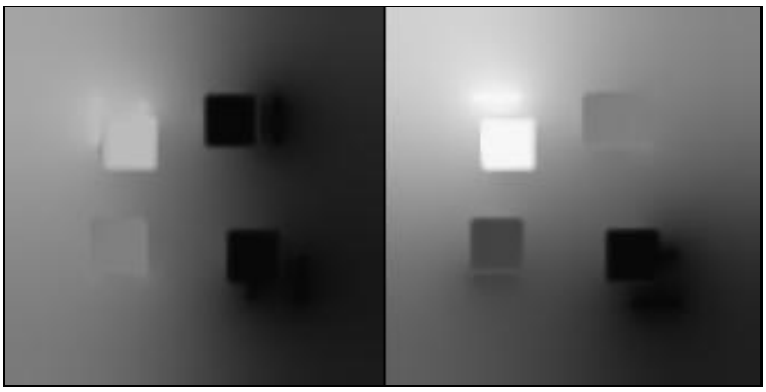


Figure 11.3: Result when the optical flow constraint and the smoothness constraint are centred in different images. Images and parameters are identical with those in Figure 11.1.

proaches the inner scale of the image. Further reduction of σ leads to slightly larger errors. It appears that this is caused by discretization and quantization effects. We evaluated the error only in the interior of the squares because of the constant background. The flow is not defined correctly in this area in the sense that any displacement of the background is compatible with the image sequence.

Figure 11.3 demonstrates the importance of consistent centering of the optical flow constraint and the smoothness constraint. In this experiment the smoothness term follows the gradient of I_2, while the optic flow term is centred in I_1, i.e. we are looking for displacements from I_1 to I_2. Since the boundary locations of I_1 and I_2 differ for large displacements, it is not surprising that the optical flow field that

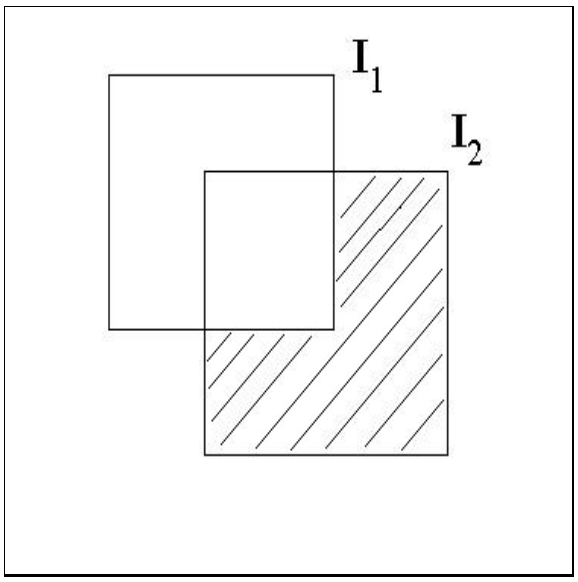


Figure 11.4: Illustration of the occlusion problem. A square is moving from I_1 to I_2. The shadowed region in the image I_1 has no correspondence in I_2.

respects the discontinuities of I_2 is "leaking" at the boundaries of I_1.

We notice that when an object moves across the image sequence, the background is partially occluded. This occlusion problem is illustrated in Figure 11.4. In the direction of the object motion a region of the background is occluded, so the points of this region (the shadowed area of Figure 11.4) have no correspondence in I_2, and the optical flow constraint is no longer valid. In this background region some slight inhomogeneities appear as can be seen in Figure 11.1. However, we observed that the smoothness term of the energy helps to reduce such effects.

For the following experiment we use the classical Hamburg taxi test sequence. These data are available from `ftp://csd.uwo.ca` in the directory `pub/vision`. Instead of taking two consecutive frames – as is usually done – we consider the frames 15 and 19. The dark car at the left creates a largest displacement magnitude of approximately 12 pixels. In Figures 11.5 and 11.6 we present the computed flow. The computed maximal flow magnitude is 11.68, which is a good approximation of the actual displacement of the dark car. It is interesting to note that, although the movement of the pedestrian in the upper left part of the scene is difficult to recognize in the greyscale plot in Figure 11.5, the vector plot in Figure 11.6 shows that this motion is not suppressed.

In Figs. 11.7 and 11.8 we show the results of our method for the marbled block sequence. This sequence is copyright by H.-H. Nagel (KOGS/IAKS, University of Karlsruhe, Germany). It has been recorded and first evaluated by Michael Otte [293] and it is available from `http://i21www.ira.uka.de/image_sequences`. The

Figure 11.5: Computation of the optical flow for the taxi sequence (frames 15 and 19) with $\alpha = 0.6$, $s = 0.1$, $\sigma_0 = 10$, $\sigma_n = 0.8$, and $\eta = 0.95$.

marbled block sequence described a real-world scene where the camera is moving. In our experiments we took the frame 20 and 25, and we used the same parameters as for the taxi scene.

Next we perform quantitative comparisons with classic optical flow techniques from the survey paper of Barron *et al.* [35]. This is done using their ground truth data as well as the evaluation utilities that are available from `ftp://csd.uwo.ca` in the directory `pub/vision`. It should be noted that the results in [35] have been achieved with test sequences where the displacements are small, while our method is designed for large displacement fields. Moreover, their methods also used a presmoothing in time which involves more than two frames, whereas we use only two frames. In spite of these limitations we are going to show that we can obtain competitive results with our method.

In the comparison we focus on those methods in [35] that create flow fields with 100% density. For many subsequent tasks such as the inference of egomotion and surface structure this is a very desirable property. Local methods that yield a lower density may have to be supplemented with additional strategies for filling in infor-

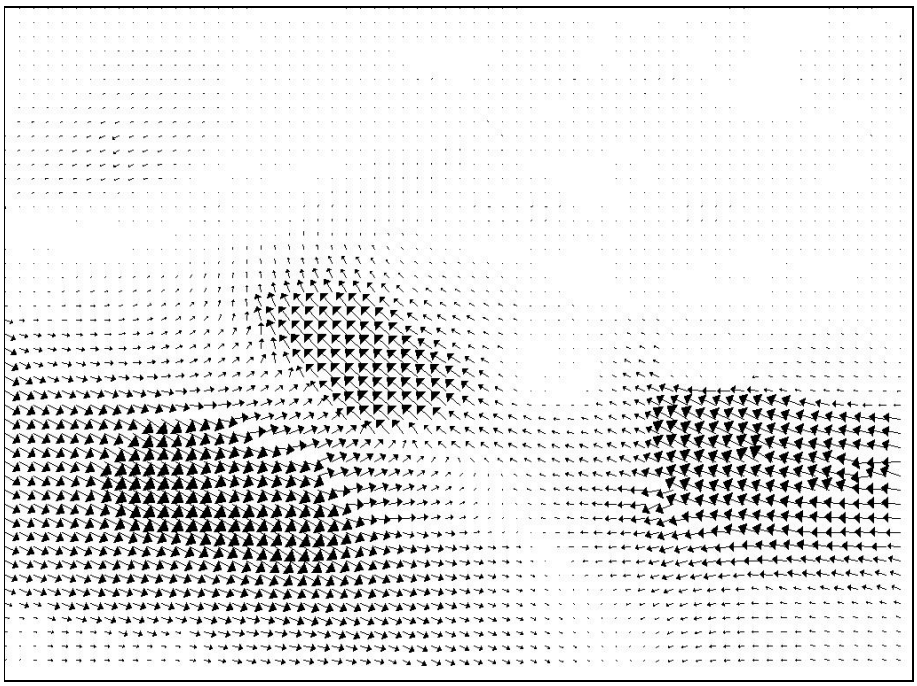


Figure 11.6: Vector plot of the optical flow from Figure 11.5.

mation at locations where no results are available. Their practical performance may thus depend heavily on this postprocessing. Variational approaches with smoothness terms do not require such a postprocessing step as they automatically yield flow fields with 100 % density.

In Figures 11.9 and 11.10 we show the computed optical flow for the *Square2* sequence that depicts a square moving with velocity $(4/3, -4/3)$. Table 11.1 gives a comparison with the results of Barron *et al.* for some classic optic flow techniques that create flow fields with 100 % density. It can be seen that our proposed technique reveals smaller errors than these methods. In particular, this also shows that our three modifications improve Nagel's method substantially. While the implementation of Nagel's method in [35] gives an angular error of 34.57°, our method reveals an error of 10.97°. In this example Barron *et al.* assume that the background moves in the same direction as the square. However, as the background is constant the displacement is not well defined in this area. If we focus our attention on the error of the computed flow within the interior of the square we obtain an average angular error of 0.85. This shows that the computed flow is very accurate in the interior of the square.

Next we draw our attention to the most complex synthetic test sequence from [35], the Yosemite sequence with cloudy sky. It contains displacements of up to five pixels. Our optical flow results are shown in Figures 11.11 and 11.12, and a

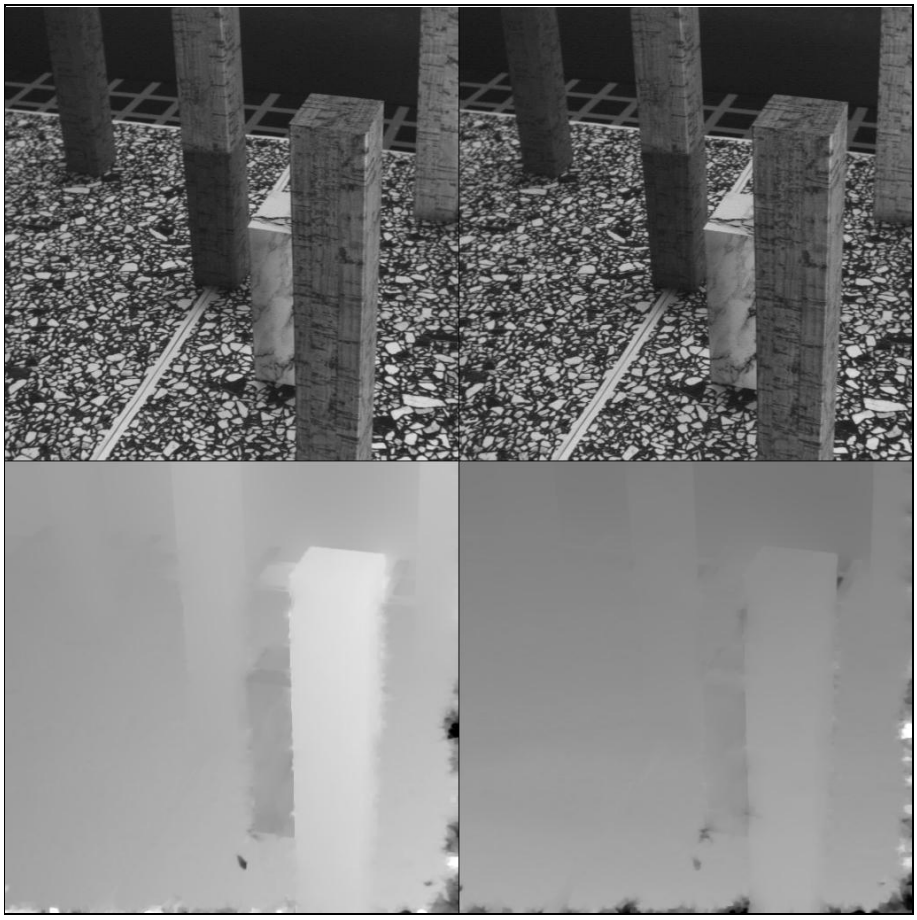


Figure 11.7: Optical flow results for the marbled block sequence (frames 20 and 25) with $\alpha = 0.6$, $s = 0.1$, $\sigma_0 = 10$, $\sigma_n = 0.8$, and $\eta = 0.95$.

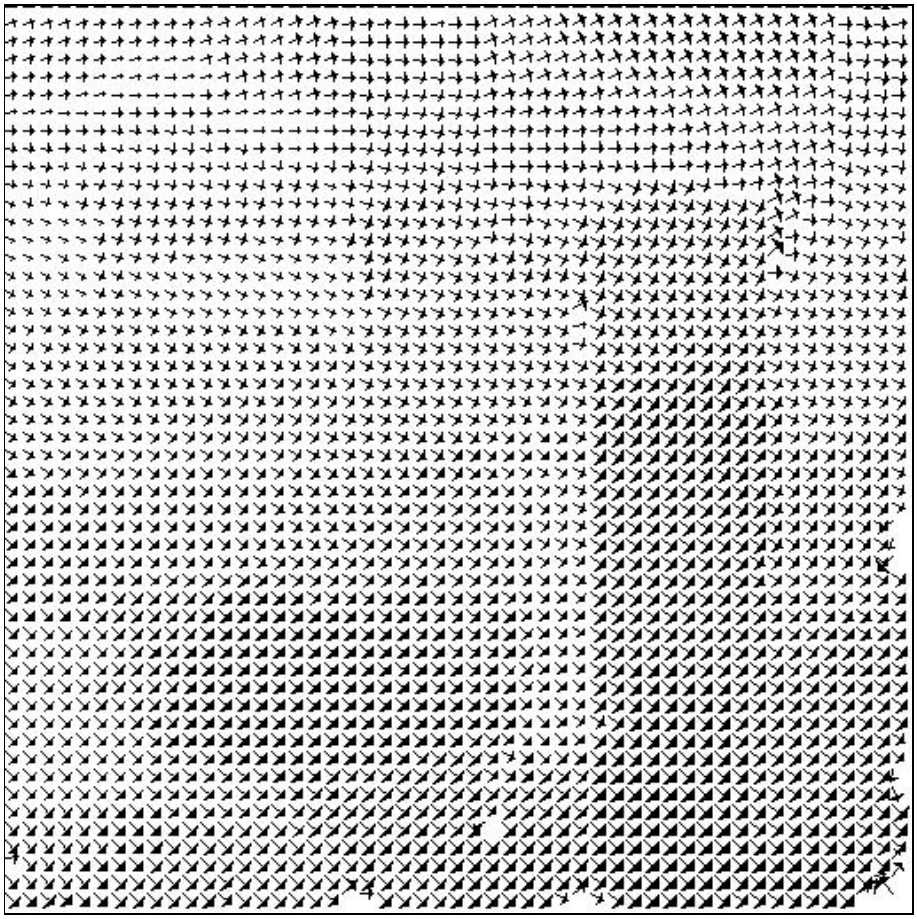


Figure 11.8: Vector plot of the optical flow from Figure 11.7.

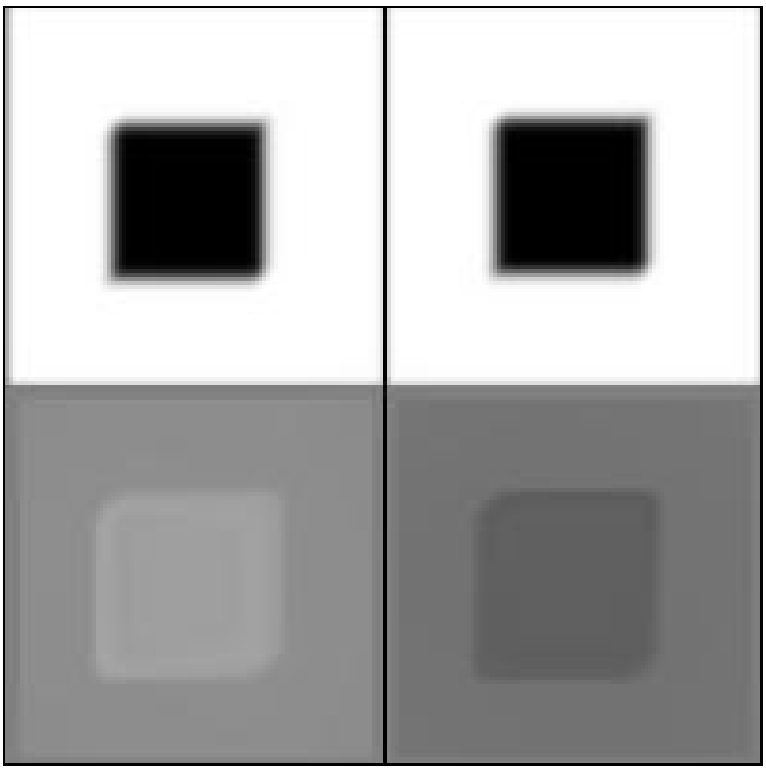


Figure 11.9: Computation of the optical flow for the *Square2* sequence with $\alpha = 0.6$, $s = 0.1$, $\sigma_0 = 10$, $\sigma_n = 1$, and $\eta = 0.95$.

juxtaposition with other methods can be found in Table 11.2. Again our technique outperforms all methods from [35] which yield flow fields with 100 % density. With an angular error of 5.53° it even reaches the estimation quality of typical methods with 30 % density, and the standard deviation of 7.40° is lower than the standard deviation of all methods that have been evaluated in [35]: the best method (Lucas and Kanade with $\lambda_2 \geq 5.0$) had an average angular error of 3.22° with a standard deviation of 8.92° and a density of only 8.7 %.

In order to evaluate the robustness of our algorithm with respect to the choice of parameters we present in Table 11.3 the errors for the Yosemite sequence taking different values of the parameters. To simplify the presentation, we fixed the finest scale to $\sigma_n := 1$, and as numerical parameters we used $\tau := 10$ and $T := 500$. These parameters are almost independent of the image and can therefore be set to default values. Hence, we vary only the parameters α, s, η and σ_0 in Table 11.3. First of all it can be seen that our method outperforms all methods in [35] with 100 % density not only in case of optimized parameters, but also for a rather large range of parameter settings. Let us now study the parameter influence in more detail.

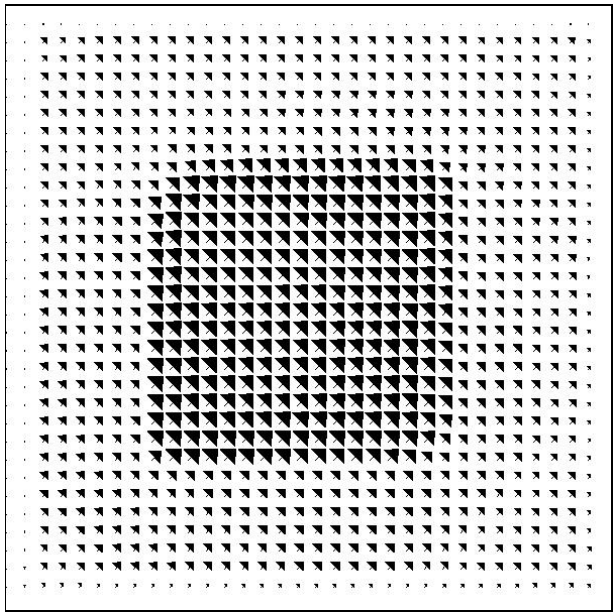


Figure 11.10: Vector plot of the optical flow from Figure 11.9.

Table 11.1: Comparison between the results from [35] with 100 % density and our method for the *Square2* sequence.

Technique	Aver. Error	Stand. Deviat.	Density
Horn and Schunck (original)	47.21°	14.60°	100%
Horn and Schunck (modified)	32.81°	13.67°	100%
Nagel	34.57°	14.38°	100%
Anandan (unthresholded)	31.46°	18.31°	100%
Singh (step 1)	49.03°	21.38°	100%
Singh (step 2)	46.12°	18.64°	100%
our method	10.97°	9.60°	100%

One important observation from Table 11.3 is that the decay parameter η has an important influence of the result: values around 0.5, as are implicitly used by typical pyramid-based focusing algorithms, are by far not optimal. A slow focusing with $\eta = 0.95$ gives significantly better results. Our experience with other images suggests that η may be fixed to this value for all applications.
Choosing too a small value for the isotropy fraction s does hardly worsen the results, while for larger values the smoothness term becomes isotropic almost everywhere and approximates the Horn and Schunck scheme [171]. In order to avoid the resulting deteriorations, we propose to fix $s := 0.1$, which means that the

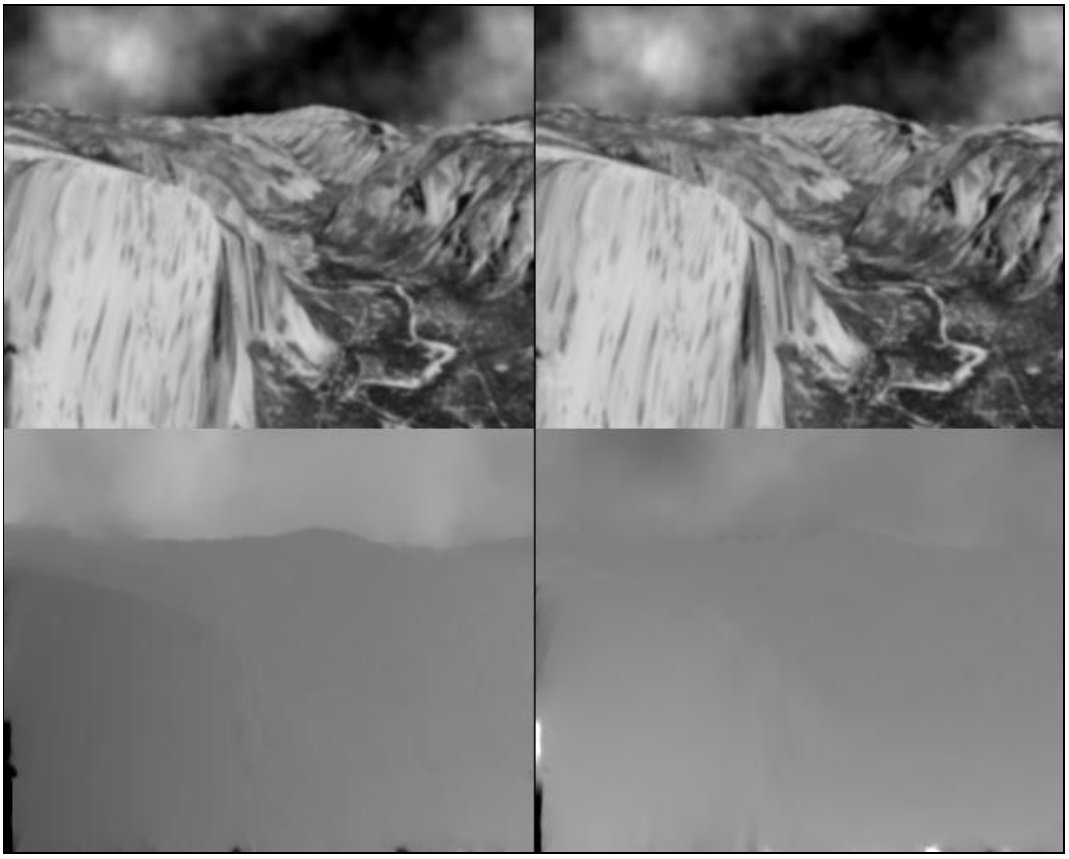


Figure 11.11: Computation of the optical flow for the Yosemite sequence with $\alpha = 0.6$, $s = 0.1$, $\sigma_0 = 5$, $\sigma_n = 1$, and $\eta = 0.95$.

method smoothes anisotropically at 90% of all locations.
Regarding the smoothness parameter α, our method appeared to be rather robust with respect to over- and underestimations. We have thus used a fixed value of 0.6 for all experiments in the present paper.
As already mentioned, the initial scale σ_0 should be chosen such that it covers the largest expected displacements. We found that overestimations are less critical than underestimations. This also confirms the use of the focusing strategy. Too small values increase the danger of ending up in a physically irrelevant local minimum. Actually, σ_0 was basically the only parameter that we had to adapt in order to analyse different image sequences. Since it has a clear physical interpretation, this adaptation was simple.

Remark. More detailed information about the experiments in this section can be found at `http://serdis.dis.ulpgc.es/~lalvarez/research/demos`. In particular, some movies to illustrate the focusing strategy are presented. At this site we also provide a window oriented image processing software named `XMegaWave` (see

Table 11.2: Comparison between the results from [35] with 100 % density and our method for the *Yosemite* sequence.

Technique	Aver. Error	Stand. Deviat.	Density
Horn and Schunck (original)	31.69°	31.18°	100%
Horn and Schunck (modified)	9.78°	16.19°	100%
Nagel	10.22°	16.51°	100%
Anandan (unthresholded)	13.36°	15.64°	100%
Uras et al. (unthresholded)	8.94°	15.61°	100%
Singh (step 2)	10.03°	13.13°	100%
our method	5.53°	7.40°	100%

Table 11.3: Errors for the Yosemite sequence, using different parameters of the algorithm

smoothness α	init. scale σ_0	isotr. fract. s	decay rate η	angul. error	stand. dev.
0.4	5	0.1	0.90	5.61°	7.46°
0.5	”	”	”	5.57°	7.41°
0.6	”	”	”	5.55°	7.37°
0.7	”	”	”	5.56°	7.33°
1.0	”	”	”	5.69°	7.24°
0.6	1	0.1	0.90	16.83°	15.23°
”	2.5	”	”	5.92°	7.31°
”	5	”	”	5.55°	7.37°
”	10	”	”	5.54°	7.37°
”	15	”	”	5.81°	8.45°
0.6	5	0.01	0.90	5.70°	7.92
”	”	0.1	”	5.55°	7.37°
”	”	0.2	”	5.70°	7.31°
”	”	0.5	”	6.38°	8.14°
”	”	0.8	”	7.31°	9.76°
”	”	0.9	”	7.64°	10.37°
”	”	0.99	”	8.04°	11.21°
0.6	5	0.1	0.50	7.25°	7.58°
”	”	”	0.70	6.14°	7.36°
”	”	”	0.80	5.75°	7.33°
”	”	”	0.95	5.53°	7.40°
”	”	”	0.99	5.56°	7.45°

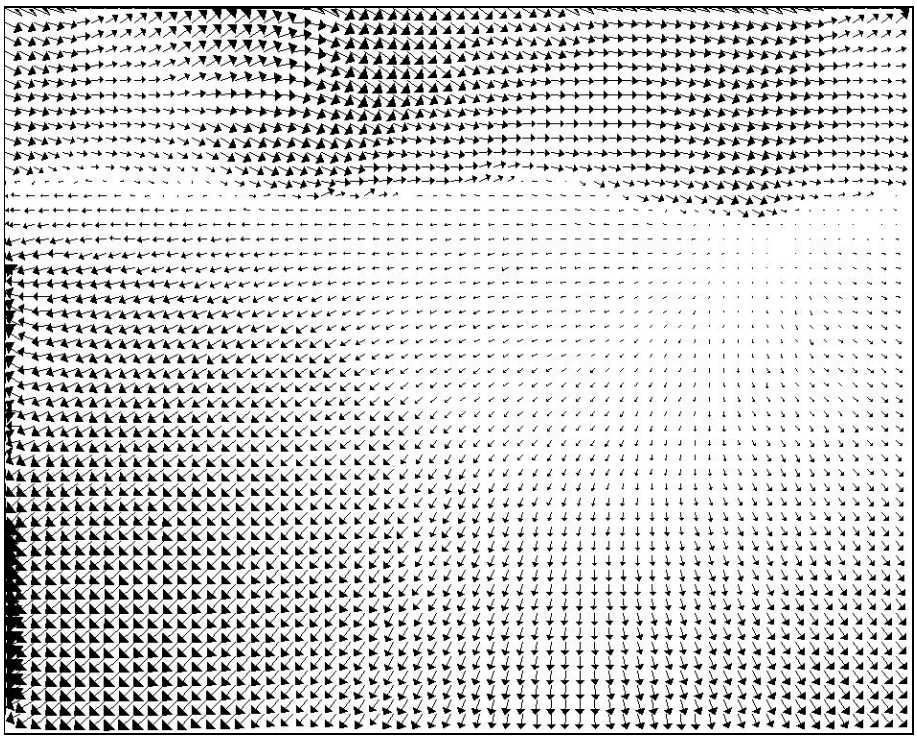


Figure 11.12: Vector plot of the optical flow from Figure 11.11.

[149]) which includes the algorithm that we have developed in this paper.

11.7 Conclusions

Usually, when computer vision researchers deal with variational methods for optical flow calculations, they linearize the optical flow constraint. Except for those cases where the images vary sufficiently slowly in space, linearization, however, does only work for small displacements. In this paper we introduced three improvements into a classical method by Nagel and Enkelmann where no linearization is used. We identified this method as two coupled linear anisotropic diffusion filters with a nonlinear reaction term. We showed that this parabolic system is well-posed from a mathematical viewpoint, and we presented a linear implicit finite difference scheme for its efficient numerical solution. In order to avoid that the algorithms converges to physically irrelevant local minima, we embedded it into a linear scale-space approach for focusing the solution from a coarse to a fine scale. A detailed quantitative analysis using test sequences with ground truth data showed the following results.

- The method can recover displacements of more than 10 pixels with good

accuracy.

- It performs significantly better than Nagel's original method and all other methods with 100 % density that are evaluated by Barron *et al.* [35].
- The performance hardly deteriorated for quite a large range of parameters. This allows to use default parameter settings for many applications.

In spite of these favourable properties, there are still possibilities for further improvements and extensions of this algorithm. For instance, we expect that our method can benefit from results obtained from preceding frames. One can interpolate the flow in the previous frames in order to have a first estimation of the flow in the current frame [44]. If this first estimation is a reasonable initial guess for our algorithm, there is no need to introduce large scale corrections by using large values for σ_0. A smaller initial scale σ_0 speeds up the focusing procedure and makes the algorithm faster. As an example for another extension possibility, we are currently investigating the use of our method for related matching problems such as 3D reconstructions from greyscale and colour stereo pairs [10, 16]. It is our hope that our method that combines anisotropic diffusion–reaction equations with linear scale-space techniques may serve as a motivation to study other combinations of linear and nonlinear scale-space approaches for solving computer vision problems.

Acknowledgement. This work has been supported by the European TMR network *Viscosity Solutions and Their Applications*.

Chapter 12

Variational Optic Flow Computation with a Spatio-Temporal Smoothness Constaint

Joachim Weickert and Christoph Schnörr.
Journal of Mathematical Imaging and Vision, in press.

Abstract

Nonquadratic variational regularization is a well-known and powerful approach for the discontinuity-preserving computation of optic flow. In the present paper, we consider an extension of flow-driven spatial smoothness terms to spatio-temporal regularizers. Our method leads to a rotationally invariant and time symmetric convex optimization problem. It has a unique minimum that can be found in a stable way by standard algorithms such as gradient descent. Since the convexity guarantees global convergence, the result does not depend on the flow initialization. Two iterative algorithms are presented that are not difficult to implement. Qualitative and quantitative results for synthetic and real-world scenes show that our spatio-temporal approach (i) improves optic flow fields significantly, (ii) smoothes out background noise efficiently, and (iii) preserves true motion boundaries. The computational costs are only 50 % higher than for a pure spatial approach applied to all subsequent image pairs of the sequence.

12.1 Introduction

Variational methods in image processing and computer vision have attracted a lot of interest in recent years. They offer the advantage of providing a clear mathematical formalism for all model assumptions. Minimizing the resulting energy functionals gives solutions that are optimal with respect to the specified assumptions.

One of the earliest application areas of variational methods within computer vision is the estimation of optic flow [171]. The optic flow field of an image sequence describes the displacement of brightness patterns over time. Applications of optic flow range from vision-based robot navigation to second-generation video compression. Numerous methods for calculating optic flow have been proposed in the last two decades; see e.g. the survey papers of Mitiche and Bouthemy [251], and Stiller and Konrad [377]. Performance evaluations of some of the most popular algorithms have been carried out by Barron *et al.* [35] and Galvin *et al.* [139]. These papers also showed that variational optic flow methods belong to the techniques that perform well. In contrast to many other optic flow methods, they offer the advantage of creating flow fields with 100 % density, such that no postprocessing by interpolation becomes necessary.

Variational optic flow methods have been pioneered by Horn and Schunck [171] and improved by Nagel [260] and many others. Approaches of this type calculate optic flow as the minimizer of an energy functional, which consists of a data term and a smoothness term. The data term involves optic flow constraints such as the assumption that corresponding pixels in different frames should reveal the same grey value. The smoothness term usually requires that the optic flow field should vary smoothly in space. Such a term may be modified in an *image-driven* way in order to suppress smoothing at image boundaries; see e.g. [11, 18, 260, 342, 361]. Recently also *flow-driven* modifications have been proposed which reduce smoothing at flow discontinuities [26, 85, 105, 221, 309, 344, 414]. These nonlinear methods have already led to rather good results in spite of the fact that the smoothness term imposed only *spatial* smoothness of the flow field. They work locally in time and do not make use of the temporal coherence within the sequence.

The goal of this paper is to investigate an extension of spatial flow-driven smoothness terms to *spatio-temporal flow-driven* regularizations. Such an extension makes consequent use of the available data, and it leads to equations which are hardly more complicated than in the pure spatial case. Our experiments on synthetic and real-world sequences, however, show that this approach leads to significantly more robust results.

Our paper is organized as follows. In Section 12.2 we review optic flow approaches with spatial smoothness terms, and Section 12.3 describes our novel method using a spatio-temporal smoothness constraint. A simple numerical algorithm is derived

in Section 12.4, and Section 12.5 analyses the performance of our approach by applying it to synthetic and real-world image sequences. The paper is concluded with a summary in Section 12.6. We have presented a shorter preliminary version of this work at a national symposium [423].

Related work. While spatial smoothness assumptions are common in the optic flow literature, spatio-temporal approaches are significantly less frequent.
An interesting extension of a smoothness constraint into the temporal domain has been proposed by Nagel [264]. He derived the model for a spatio-temporal oriented smoothness constraint, but did not present any experiments. Nagel's constraint was *image-driven*, since it reduces smoothing across *image* discontinuities. Our approach is *flow-driven* due to the more direct constraint that smoothing at *flow* discontinuities should be reduced. For a more detailed account on image- and flow-driven regularizers and a well-posedness framework for both types we refer to [425].
Other temporal smoothness assumptions that have been studied in the literature include the work of Murray and Buxton [259], Black and Anandan [44], and Black [46]. Their assumptions lead to *nonconvex* optimization problems which may have many local minima and for which it is difficult to find algorithms that are both efficient and converge to a global minimum. Algorithms that converge to a global minimum (such as simulated annealing [394]) are computationally very expensive, while methods that are more efficient (such as graduated non-convexity algorithms [48]) may get trapped in local minima. Our method leads to a nonquadratic *convex* optimization problem. It has a unique global minimum that can be found in a reliable way by using standard techniques from convex optimization, for instance gradient descent methods. Since their convergence is global, every arbitrary flow initialization leads to the same solution: the global minimum of the functional. This property is an important quality of a robust algorithm.
Another difference between the methods considered in [46, 44, 259] and our approach is that our method uses a genuinely *continuous* formulation and derives a discrete algorithm afterwards by discretizing the corresponding partial differential equations. The continuous formulation has the advantage of being rotationally invariant. Applying the well-established theory of discretization methods allows us to derive a numerically consistent scheme. It guarantees that rotational invariance is fulfilled up to an error of order $O(\frac{1}{N^2})$ where N denotes the number of pixels in x or y direction. Results of this type cannot be established in genuinely discrete formulations. The discrete models in [46, 44, 259], for instance, approximate continuous processes that are *not* rotationally invariant.
The approaches of Black and Anandan [44] and Black [46] use a model which applies incremental minimization over time. Such a technique is highly useful for tasks such as robotics where images have to be processed online and only information from the past is accessible. Our approach, however, is designed for batch

mode since it is symmetric with respect to past and future. Methods of this type are useful if a video sequence is to be processed offline. In this case there is information available both from the past and from the future and there is no reason for using only a part of it. Temporal symmetry also guarantees that the first frame is processed in the same way as the last one.
The nonlinear regularization approach that we use is similar to nonquadratic convex regularization methods for image restoration [79, 81, 345, 376]. Stochastic counterparts to these techniques are given by Bayesian methods using Markov random field approaches [101, 141].

12.2 Spatial Smoothness Terms

Let us denote an image sequence by some real-valued function $f(x, y, z)$ where (x, y) denotes the location within some rectangular image domain Ω and $z \in [0, T]$ is the time. Many variational optic flow calculations determine the optic flow vector $(u, v)^T$ based on two assumptions:

1. Corresponding features are supposed to maintain their intensity over time. A differential formulation of this brightness constancy assumption leads to the *optic flow constraint (OFC) equation*

$$f_x u + f_y v + f_z = 0, \tag{12.1}$$

   where the subscripts denote partial derivatives. Numerous generalizations exist where multiple constraint equations are used, or different "conserved quantities" (replacing intensity) are considered; see e.g. [24, 126, 389, 402]. One may also relax the OFC equation taking into account illumination changes [88, 252, 269]. In case of large magnitude displacements a nondifferential form of the optic flow constraint without any linearization should be used; see e.g. [18].
   Evidently, the single equation (12.1) is not sufficient to determine the two unknown functions u and v uniquely *(aperture problem)*. In order to obtain a unique flow field, a second constraint is needed.

2. Such a second constraint may impose that the flow field should vary (piecewise) smoothly in space. This can be achieved if

$$\int_\Omega \Psi\left(|\nabla u|^2 + |\nabla v|^2\right) dx\, dy \tag{12.2}$$

   is small, where $\Psi : \mathbb{R} \to \mathbb{R}$ is an increasing differentiable function and $\nabla := (\partial_x, \partial_y)^T$ denotes the 2D nabla operator. This assumption is called

smoothness constraint. In the sequel we shall assume that $\Psi(s^2)$ is convex in s, and that there exist constants $c_1, c_2 > 0$ with $c_1 s^2 \leq \Psi(s^2) \leq c_2 s^2$ for all s. In this case the optic flow problem becomes well-posed. Examples for Ψ will be presented at the end of this section.

In order to satisfy both the optic flow and the smoothness constraint as good as possible, they are assembled into a single energy functional to be minimized:

$$E(u,v) := \int_\Omega \Big((f_x u + f_y v + f_z)^2 + \alpha \, \Psi(|\nabla u|^2 + |\nabla v|^2) \Big) \, dx \, dy \tag{12.3}$$

where the *regularization parameter* $\alpha > 0$ specifies the weight of the second summand *(smoothness term, regularizer)* relative to the first one *(data term)*. Larger values for α lead to smoother flow fields. Some authors have proposed optimization tools to compute this parameter; see e.g. [386].
Using steepest descent for the minimization of (12.3) gives the diffusion–reaction system

$$u_t = \text{div} \left(\Psi' \left(|\nabla u|^2 + |\nabla v|^2 \right) \nabla u \right) - \tfrac{1}{\alpha} f_x (f_x u + f_y v + f_z), \tag{12.4}$$

$$v_t = \text{div} \left(\Psi' \left(|\nabla u|^2 + |\nabla v|^2 \right) \nabla v \right) - \tfrac{1}{\alpha} f_y (f_x u + f_y v + f_z), \tag{12.5}$$

where Ψ' is the derivative of Ψ with respect to its argument, and div denotes the 2D divergence operator, i.e. $\text{div} \binom{a}{b} := \partial_x a + \partial_y b$. The diffusion time t is an artificial evolution parameter which should not be mixed up with the time z of the image sequence $f(x,y,z)$. For $t \to \infty$, the solution (u,v) gives the minimum of $E(u,v)$. It is unique since $\Psi(s^2)$ is convex in s.
The diffusivity in both equations is given by $\Psi' \left(|\nabla u|^2 + |\nabla v|^2 \right)$. It steers the activity of the smoothing process: diffusion is strong at locations where the diffusivity is large, and smoothing is reduced at places where the diffusivity is small. We shall now consider some examples which demonstrate how the choice of Ψ influences the smoothing process.

1. Horn and Schunck [171] considered the linear case $\Psi(s^2) = s^2$. This corresponds to the constant diffusivity $\Psi'(s^2) = 1$. Therefore, the smoothing activity of the Horn and Schunck method does not depend on the flow variation $s^2 = |\nabla u|^2 + |\nabla v|^2$. As a consequence, the flow is also smoothed across motion boundaries. This explains a well-known drawback of this method: a blurry flow field which is ignorant of the true motion boundaries.

2. Many modifications have been proposed to alleviate this problem. Nagel [260] for instance reduced diffusion across image boundaries with large $|\nabla f|$. Thus, this method considers an *image-driven* smoothness term for the flow field. In many cases this modification outperforms the Horn and Schunck approach.

In specific situations, however, image discontinuities may not coincide with flow discontinuities: strongly textured rigid objects, for example, have numerous texture edges which are not motion boundaries. Then an image-driven smoothness term would lead to an oversegmentation and it would be desirable to replace it by one which respects flow discontinuities instead of image discontinuities.

3. A *flow-driven* smoothness term can be constructed by using a nonlinear convex regularizer $\Psi(s^2)$ which creates a decreasing diffusivity $\Psi'(s^2)$. This ensures that the smoothing is reduced at locations where the flow magnitude is small. In the context of optic flow, such methods have been considered by Schnörr [344] and Weickert [414]. One may for instance consider the regularizer

$$\Psi(s^2) := \varepsilon s^2 + (1-\varepsilon)\lambda^2 \sqrt{1 + s^2/\lambda^2} \tag{12.6}$$

with $0 < \varepsilon \ll 1$ and $\lambda > 0$. It leads to the diffusivity

$$\Psi'(s^2) = \varepsilon + \frac{1-\varepsilon}{\sqrt{1 + s^2/\lambda^2}}. \tag{12.7}$$

We observe that λ can be regarded as a *contrast parameter*: If the flow variation $s^2 = |\nabla u|^2 + |\nabla v|^2$ is large compared to λ^2, then the diffusivity is close to 0, and for $s^2 \ll \lambda^2$ the diffusivity tends to 1. Choosing a very small value for λ relates this method to total variation regularization, a powerful denoising technique permitting discontinuous solutions [326]. The parameter ε is only required for proving well-posedness. In practical applications it can be fixed to some small value, e.g. $\varepsilon := 10^{-6}$.

4. Other flow-driven smoothness terms from the literature [26, 85, 105, 221] replace the regularizer $\Psi(|\nabla u|^2 + |\nabla v|^2)$ by $\Psi(|\nabla u|^2) + \Psi(|\nabla v|^2)$. This leads to two diffusion–reaction equations where the joint diffusivity $\Psi'(|\nabla u|^2 + |\nabla v|^2)$ is replaced by $\Psi'(|\nabla u|^2)$ and $\Psi'(|\nabla v|^2)$, respectively. Hence, the coupling between the two equations becomes weaker and flow discontinuities may be formed at different locations for u and v. It should also be mentioned that in general such models are not rotationally invariant.

12.3 Spatio-Temporal Smoothness Terms

The methods that we have discussed so far work locally in time: two frames are sufficient to calculate the optic flow field. In general, however, we have much more data at our disposal, namely the entire image sequence. It would thus be consequent if we use this full information for computing the optic flow field. In this way we may expect more robust results.

Using the knowledge from the previous section it is not difficult to extend the smoothness constraint into the temporal domain. Instead of calculating the optic flow (u, v) as the minimizer of the two-dimensional integral (12.3) for each time frame z, we now minimize a single three-dimensional integral whose solution is the optic flow for *all* frames $z \in [0, T]$:

$$E(u,v) := \int_{\Omega\times[0,T]} \Big((f_x u + f_y v + f_z)^2 + \alpha\, \Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2) \Big)\, dx\, dy\, dz \qquad (12.8)$$

where $\nabla_3 := (\partial_x, \partial_y, \partial_z)^T$ denotes the spatio-temporal nabla operator. The corresponding steepest descent equations are given by

$$u_t = \nabla_3 \cdot \left(\Psi' \left(|\nabla_3 u|^2 + |\nabla_3 v|^2 \right) \nabla_3 u \right) - \tfrac{1}{\alpha} f_x (f_x u + f_y v + f_z), \qquad (12.9)$$

$$v_t = \nabla_3 \cdot \left(\Psi' \left(|\nabla_3 u|^2 + |\nabla_3 v|^2 \right) \nabla_3 v \right) - \tfrac{1}{\alpha} f_y (f_x u + f_y v + f_z). \qquad (12.10)$$

In contrast to the two-dimensional diffusion–reaction system (12.4)–(12.5) we now have a three-dimensional problem. In the present paper we study this process for the nonlinear regularizer given in (12.6).
The diffusion part in (12.9)–(12.10) has the same structure as nonlinear diffusion filters for regularizing three-dimensional vector-valued images. Such methods have first been applied by Gerig *et al.* [144] in the context of medical imaging. The latter approach, however, uses diffusivities from [303] which may create ill-posed processes. This cannot happen in our case, where convex smoothness terms in the energy functional create well-posed diffusion–reaction processes. The well-posedness proof is a straightforward extension of the results in [344] to the spatio-temporal case. For a more detailed discussion of nonlinear diffusion filtering we refer to [382, 412].

12.4 Numerical Aspects

We approximate the 2-D diffusion–reaction system (12.4)–(12.5) and its 3-D counterpart (12.9)–(12.10) by finite differences. Derivatives in x, y and z are approximated by central differences, and for the discretization in t direction we use a slightly modified explicit (Euler forward) scheme.
Each iteration step proceeds as follows. Let τ be the step size in t direction and let f_{xi}, f_{yi} and f_{zi} denote central difference approximations of f_x, f_y and f_z in some pixel i, respectively. Let the flow components for the first iteration be initialized by 0. The $(k+1)$-th iteration calculates the unknown flow components u_i^{k+1} and v_i^{k+1} using known values from level k:

$$\frac{u_i^{k+1} - u_i^k}{\tau} = \sum_j a_{ij}^k\, u_j^k - \tfrac{\tau}{\alpha} f_{xi}\, (f_{xi} u_i^{k+1} + f_{yi} v_i^k + f_{zi}), \qquad (12.11)$$

$$\frac{v_i^{k+1} - v_i^k}{\tau} = \sum_j a_{ij}^k v_j^k - \tfrac{\tau}{\alpha} f_{yi} \left(f_{xi} u_i^k + f_{yi} v_i^{k+1} + f_{zi}\right), \tag{12.12}$$

The matrix entries a_{ij}^k result from a standard discretization of the divergence expressions:

$$\sum_j a_{ij}^k u_j^k := \sum_{j \in \mathcal{N}(i)} \frac{\Psi'^k_j + \Psi'^k_i}{2} (u_j^k - u_i^k), \tag{12.13}$$

where $\mathcal{N}(i)$ denotes the set of (4 in 2-D, 6 in 3-D) neighbours of pixel i, and Ψ'^k_i approximates the diffusivity $\Psi'(|\nabla u|^2 + |\nabla v|^2)$ in pixel i at time level k. In (12.11) and (12.12), one expression in the reaction term is approximated at time level $k+1$ in order to improve stability. Note that this scheme can still be solved explicitly for u_i^{k+1} and v_i^{k+1}:

$$u_i^{k+1} = \frac{u_i^k + \tau \sum_j a_{ij}^k u_j^k - \frac{\tau}{\alpha} f_{xi} \left(f_{yi} v_i^k + f_{zi}\right)}{1 + \frac{\tau}{\alpha} f_{xi}^2}, \tag{12.14}$$

$$v_i^{k+1} = \frac{v_i^k + \tau \sum_j a_{ij}^k v_j^k - \frac{\tau}{\alpha} f_{yi} \left(f_{xi} u_i^k + f_{zi}\right)}{1 + \frac{\tau}{\alpha} f_{yi}^2}. \tag{12.15}$$

We used the time step size $\tau = 1/4$ in the 2-D case and $\tau = 1/6$ in the 3-D case. For the linear case with $\Psi(s^2) = s^2$, one can verify that this choice of step sizes comes down to applying a Jacobi algorithm to the discretized Euler–Lagrange equations. If the discrete image gradient does not vanish at one point, the system matrix of these equations is irreducibly diagonally dominant. This guarantees the existence of a unique solution of the linear system and global convergence of the Jacobi iterations [256]. The iterations were stopped when the Euclidean norm of the relative residue dropped below 0.001. In the 2-D case with an image pair with 200×200 pixels, one can perform 20 iterations per CPU second on a 700 MHz PC. The itypical number of iterations to reach the desired accuracy is in the range between 100 and 1000. By relaxing the accuracy requirements, one can also stop the process much earlier.

The explicit scheme (12.11),(12.12) has been chosen for simplicity reasons. In order to gain more efficiency, it is also possible to replace it by a slightly more complicated semi-implicit approximation, for instance the additive operator splitting (AOS) scheme considered in [414]. In matrix–vector notation, such an AOS scheme is given by

$$u^{k+1} = \frac{1}{m} \sum_{l=1}^{m} \left(I + \tfrac{m\tau}{\alpha} f_x^2 I - m\tau A_l(u^k, v^k)\right)^{-1} \left(u^k - \tfrac{\tau}{\alpha} f_x \left(f_y v^k + f_z\right)\right), \tag{12.16}$$

$$v^{k+1} = \frac{1}{m}\sum_{l=1}^{m}\left(I + \tfrac{m\tau}{\alpha}f_y^2 I - m\tau A_l(u^k, v^k)\right)^{-1} \left(v^k - \tfrac{\tau}{\alpha}f_y\left(f_x u^k + f_z\right)\right), \quad (12.17)$$

where m denotes the dimension ($m = 2$ for the spatial case, $m = 3$ for the spatio–temporal one), I is the unit matrix, and the matrix A_l describes the diffusive interaction in l direction. Although one AOS step is about twice as expensive as an explicit step, the AOS scheme is one order of magnitude more efficient than the explicit scheme since one may use 20 times larger time step sizes. For more details on AOS schemes we refer to [429]. The computational complexity and the memory requirement of both the explicit and the AOS scheme are linear in the number of pixels respectively voxels.

The explicit scheme (12.11),(12.12) has been chosen for simplicity reasons. In order to gain absolute stability, it is also possible to replace it by a slightly more complicated semi-implicit approximation, for instance the additive operator splitting (AOS) scheme considered in [414].

It should be noted that the 3-D scheme requires only about 50 % more computing time than a corresponding 2-D scheme that is applied to all subsequent frame pairs of an image sequence: 2-D diffusion within a 4-neighbourhood is replaced by 3-D diffusion within a 6-neighbourhood. The main difference is an increased memory requirement, since, in the 3-D case, the whole sequence is processed simultaneously. For the typical test sequences in computer vision, this does not lead to problems when modern PCs or workstations are used: On a computer with 512 MB memory one can process sequences with sizes up to $256 \times 256 \times 128$.

12.5 Experiments

In this section we illustrate the behaviour of our method by applying it to three test image sequences. We compare pure 2D processing (eqns. (12.4)–(12.5)) with 3D processing (eqns. (12.9)–(12.10)).

Figure 1 depicts one frame from a hallway scene where a person is moving towards the camera. The calculated optic flow results are shown in Figure 2. For pure spatial regularization we observe that outliers dominate, and that it is difficult to achieve good motion segmentation by thresholding the optic flow vectors. Spatio-temporal regularization, on the other hand, creates a more homogeneous motion field within the contour of the person, and motion segmentation is much more realistic.

Figure 3 shows the results of our comparison for the famous Hamburg taxi sequence. It is available via anonymous ftp from the site `ftp://csd.uwo.ca` under the directory `pub/vision`. Also for this sequence one observes that spatio-temporal

processing leads to more realistic motion segmentation. Moreover, it is less sensitive to noise than a pure spatial processing. It is worth emphasizing that in this and in the previous example, the same parameters have been used for both methods, and no presmoothing of any kind has been applied. All smoothing effects are thus caused by the regularizers. Figure 4 shows that the better smoothing behaviour of spatio-temporal regularizers may also reduce temporal aliasing problems: for the pedestrian in the taxi scene, pure spatial processing gives an optic flow field that points in the wrong direction, while spatio-temporal processing creates coherent optic flow vectors pointing in the correct direction.

After these qualitative comparisons, let us now turn our attention to a quantitative validation. To this end we consider a synthetic street sequence for which ground truth flow data are available. We obtained it from

`http://www.cs.otago.ac.nz/research/vision/Downloads/`

It has been created by Galvin *et al.* [139] for evaluating eight optic flow algorithms, and it is one of a few nontrivial test sequences with ground truth data, where motion boundaries are important. We used the full sequence from the web, consisting of 20 frames of size 200×200 pixels. An interesting detail is depicted in Figure 5(a), and the corresponding ground truth flow field and the calculated ones are given in Figures 5(b),(c),(d), respectively. For assessing the performance of our method, we calculated the angular error

$$\Psi_e := \arccos\left(\frac{u_c u_e + v_c v_e + 1}{\sqrt{(u_c^2 + v_c^2 + 1)(u_e^2 + v_e^2 + 1)}}\right) \tag{12.18}$$

where (u_c, v_c) denotes the correct flow, and (u_e, v_e) is the estimated flow (cf. also [35]). In order to make our method comparable with the other approaches, we applied some presmoothing by convolving the images with a Gaussian with standard deviation σ. Preprocessing steps of this type are common for evaluating optic flow algorithms [35].

With optimized parameters for σ, α and λ we obtained an average angular error of $6.62°$ for the spatial approach, and $4.85°$ for the spatio-temporal approach. Table 12.1 shows that best method that has been reported by Galvin *et al.* [139] was a thresholded version of the Lucas–Kanade algorithm [234]. It achieved an average angular error of $4.98°$, but the density of its flow field was only 32.1 %, while our method creates flow fields with 100 % density. The best full density method in [139] was an algorithm by Proesmans *et al.* [309] with an average angular error of $7.41°$. This shows that our spatio-temporal method has very good performance. Also for less optimal parameter settings, the results remained competitive.

Table 12.1: Comparison between the results from Galvin *et al.* [139] and our results for the synthetic street sequence.

Technique	Angular Error	Density
Lucas & Kanade T	4.98°	32.1 %
Singh (step 2)	6.18°	77.6 %
Horn & Schunck MT	6.62°	46.3 %
Uras et al. T	6.93°	54.2 %
Proesmans et al.	7.41°	100.0 %
Singh (step 1)	7.87°	81.2 %
Lucas & Kanande	8.93°	74.1 %
Horn & Schunck M	9.24°	74.1 %
Uras et al.	10.19°	77.6 %
Anandan T	10.58°	53.7 %
Lucas & Kanade MT	11.17°	27.7 %
Anandan	11.63°	77.6 %
Camus (3 images)	13.69°	100.0 %
Horn & Schunck T	14.67°	76.7 %
Camus (15 images)	14.73°	100.0 %
Horn & Schunck	15.02°	92.3 %
Nagel T	19.23°	32.1 %
Lucas & Kanade M	23.35°	74.1 %
Nagel	24.43°	65.8 %
Weickert & Schnörr 2D	6.62°	100.0 %
Weickert & Schnörr 3D	4.85°	100.0 %

Figure 12.1: (a) LEFT: Frame 8 of a hallway sequence of size $256 \times 256 \times 16$ pixels. A person is moving towards the camera. (b) RIGHT: Detail.

12.6 Conclusions and Further Work

We have presented a nonlinear spatio–temporal regularization approach for the computation of piecewise smooth optic flow. It leads to a convex nonquadratic optimization problem which has a unique minimum that can be recovered by a globally convergent gradient descent algorithm. The model has a rotationally invariant continuous formulation, it is symmetric in time and it avoids smoothing over spatial and temporal flow discontinuities. Qualitative and quantitative comparisons showed a significant improvement over pure 2D processing at low additional computational costs.

It appears that the limited computer memory was the main reason why spatio-temporal optic flow regularizers have been used so rarely in the past. Since this is no longer a problem, it is likely that these methods will gain more importance in the future. It should also be noted that the spatio-temporal extension that we studied here is of course not limited to the specific nonlinear flow-driven regularizer that we used in this paper. It is a general strategy for exploiting the entire image sequence data for reliable optic flow estimation within a variational framework.

Based on these encouraging results we are currently investigating the design of highly efficient optic flow algorithms for sequential and parallel computer architectures. Some of these techniques will be based on our recent research on efficient algorithms for variational image restoration and nonlinear diffusion filtering [419].

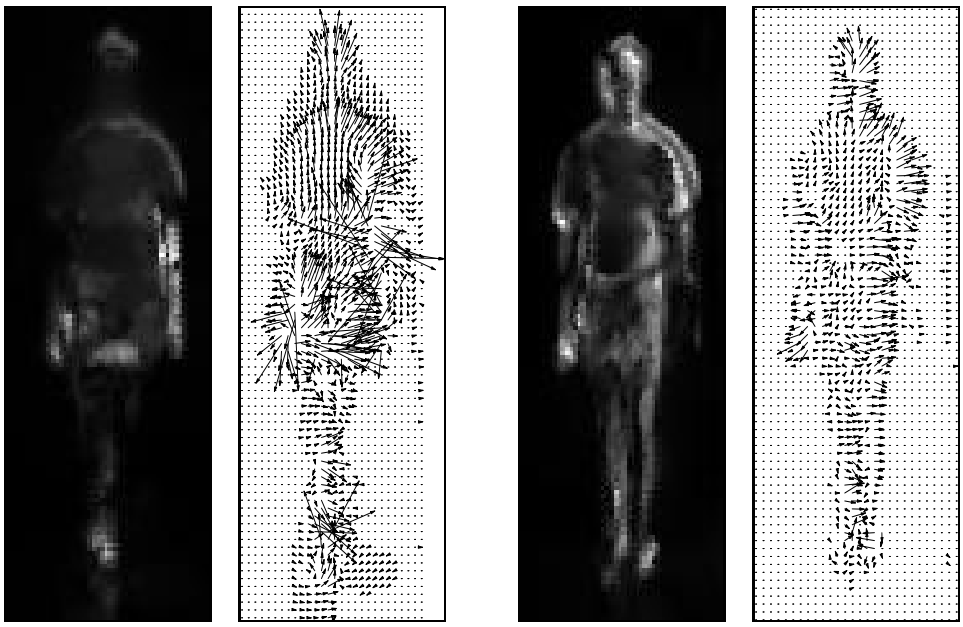


Figure 12.2: Computed flow fields using the hallway sequence. FROM LEFT TO RIGHT: (a) Grey value plot of the optic flow magnitude for 2-D processing (eqns. (12.4)–(12.5)). Note how outliers dominate the image such that other regions get scaled down. (b) Vector plot of the optic flow field for 2-D processing, subsampled by a factor 2. For better visibility, vectors w with $|w| < 0.2$ pixels have not been drawn. (c) Optic flow magnitude for 3-D processing (eqns. (12.9)–(12.10)). (d) Vector plot for 3-D processing, subsampled by a factor 2, and thresholded at 0.2 pixels. The proposed extension of adaptive smoothing to the temporal axis gives a much more coherent and complete result.

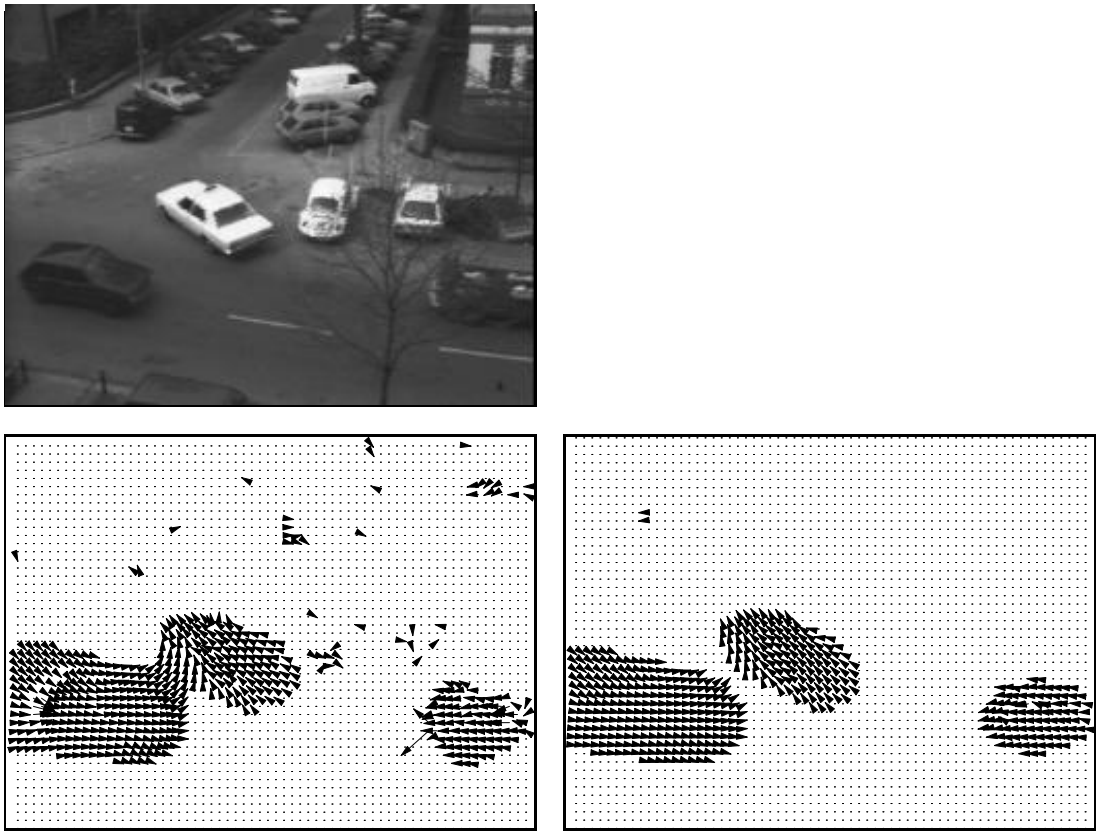


Figure 12.3: (a) TOP: Frame 10 of the well–known Hamburg taxi scene (20 frames of size 256×190). (b) BOTTOM LEFT: Optic flow field for 2-D processing, subsampled by factor 4 and thresholded at 0.2 pixels for better visibility. (c) BOTTOM RIGHT: Result for 3-D processing. Spatio-temporal regularization improves the vector fields significantly, smoothes out background noise, and preserves true motion boundaries.

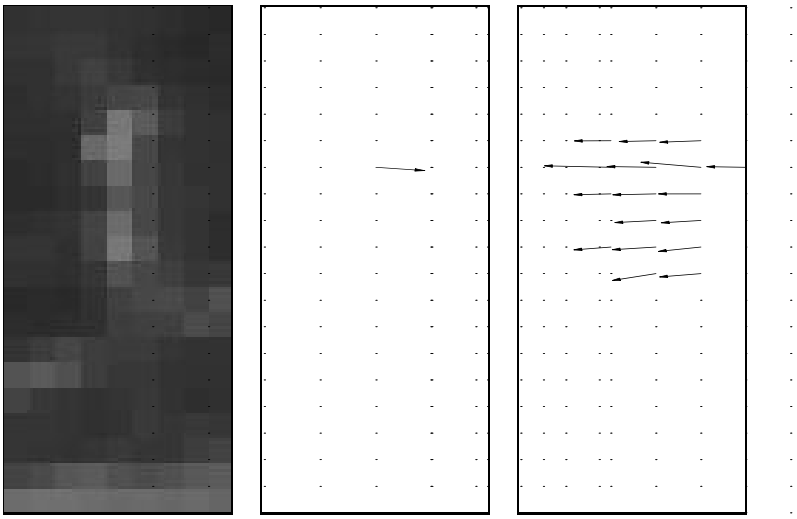


Figure 12.4: (a) LEFT: Magnification showing the pedestrian in the upper left part of the taxi scene (9×20 pixels). The pedestrian moves to the left. (b) MIDDLE: Optic flow field for 2-D processing, subsampled by a factor 2, and thresholded at 0.2 pixels. The 2-D processing smoothes out the noisy local motion data (normal flow). Temporal aliasing, however, creates an erroneous motion to the right. (c) RIGHT: Optic flow field for 3-D processing, subsampled by a factor 2. The 3D processing computes a coherent flow field for the "rigid part" of the pedestrian. The motion is in the correct direction. Regions with moving limbs are interpreted as noise at such a small spatial scale.

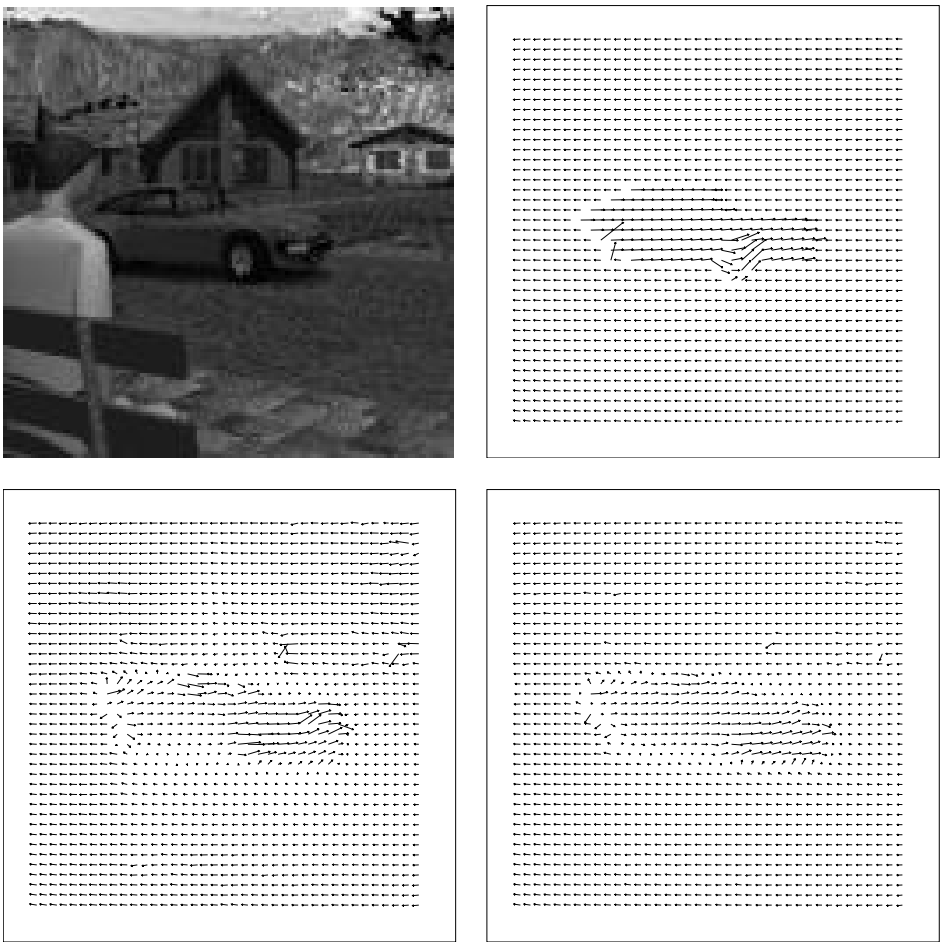


Figure 12.5: (a) TOP LEFT: Detail from a synthetic street scene (128 × 128 pixels). (b) TOP RIGHT: Exact optic flow field, subsampled by a factor 3 and scaled by 2 for improving visibility. (c) BOTTOM LEFT: Optic flow field for 2-D processing ($\sigma = 0.8$, $\alpha = 800$, $\lambda = 0.04$). (d) BOTTOM RIGHT: Result for 3-D processing ($\sigma = 0.6$, $\alpha = 250$, $\lambda = 0.05$).

Acknowledgements. We thank Ole Fogh Olsen and Mads Nielsen (IT University of Copenhagen) for providing the hallway sequence, as well as Brendan McCane and Simon Brady (University of Otago) for their kindness of making the precise numbers of their optic flow evaluations available to us. J.W. also thanks Jens Arnspang (Department of Computer Science, University of Copenhagen) for interesting discussions on optic flow. Part of this work has been supported by the EU-TMR Project VIRGO.

Chapter 13

A Theoretical Framework for Convex Regularizers in PDE-Based Computation of Image Motion

Joachim Weickert and Christoph Schnörr.
Techn. Rep. 13/2000, Computer Science Series, Dept. of Mathematics and Computer Science, University of Mannheim, Germany, June 2000.

Abstract

Many differential methods for the recovery of the optic flow field from an image sequence can be expressed in terms of a variational problem where the optic flow minimizes some energy. Typically, these energy functionals consist of two terms: a data term, which requires e.g. that a brightness constancy assumption holds, and a regularizer that encourages global or piecewise smoothness of the flow field. In this paper we present a systematic classification of rotation invariant convex regularizers by exploring their connection to diffusion filters for multichannel images. This taxonomy provides a unifying framework for data-driven and flow-driven, isotropic and anisotropic, as well as spatial and spatio-temporal regularizers. While some of these techniques are classic methods from the literature, others are derived here for the first time. We prove that all these methods are well-posed: they posses a unique solution that depends in a continuous way on the initial data. An interesting structural relation between isotropic and anisotropic flow-driven regularizers is identified, and a design criterion is proposed for constructing anisotropic flow-driven regularizers in a simple and direct way from isotropic ones. Its use is illustrated by several examples.

13.1 Introduction

Even after two decades of intensive research, robust motion estimation continues to be a key problem in computer vision. Motion is linked to the notion of optic flow, the displacement field of corresponding pixels in subsequent frames of an image sequence. Optic flow provides information that is important for many applications, ranging from the estimation of motion parameters for robot navigation to the design of second generation video coding algorithms. Surveys of the state-of-the-art in motion computation can be found in papers by Mitiche and Bouthemy [251], and Stiller and Konrad [377]. For a performance evaluation of some of the most popular algorithms we refer to Barron *et al.* [35] and Galvin *et al.* [139].

Bertero *et al.* [40] pointed out that, depending on its formulation, optic flow calculations may be ill-conditioned or even ill-posed. It is therefore natural to introduce additional smoothness constraints in order to stabilize or regularize the process. This way has been pioneered by Horn and Schunck [171] and improved by Nagel [260] and many others. Variationals approaches of this type calculate optic flow as the minimizer of an energy functional, which consists of a data term and a smoothness term. Formulations in terms of energy functionals allow a conceptually clear formalism without any hidden model assumptions, and several evaluations have shown that these methods perform well [35, 139].

The data term in the energy functional involves optic flow constraints such as the assumption that corresponding pixels in different frames should reveal the same grey value. The smoothness term usually requires that the optic flow field should vary smoothly in space [171]. Such a term may be modified in an *image-driven* way in order to suppress smoothing at or across image boundaries [11, 260]. As an alternative, *flow-driven* modifications have been proposed which reduce smoothing across flow discontinuities [45, 85, 105, 221, 309, 344, 414]. Most smoothness terms require only *spatial* smoothness. *Spatio-temporal* smoothness terms have been considered to a much smaller extent [44, 259, 264, 426]. Since smoothness terms fill in information from regions where reliable flow estimates exist to regions where no estimates are possible, they create dense flow fields. In many applications, this is a desirable quality which distinguishes regularization methods from other optic flow algorithms. The latter ones create non-dense flow fields, that have to be postprocessed by interpolation, if 100 % density is required.

Modeling the optic flow recovery problem in terms of *continuous* energy functionals offers the advantage of having a formulation that is as independent of the pixel grid as possible. A correct continuous model can be rotation invariant, and the use of well-established numerical methods shows how this rotation invariance can be approximated in a mathematically consistent way.

From both a theoretical and practical point of view, it can be attractive to use energy functionals that are *convex*. They have a unique minimum, and this global

minimum can be found in a stable way by using standard techniques from convex optimization, for instance gradient descent methods. Having a unique minimum allows to use globally convergent algorithms, where every arbitrary flow initialization leads to the same solution: the global minimum of the functional. This property is an important quality of a robust algorithm. *Nonconvex* energy functionals, on the other hand, may have many local minima, and it is difficult to find algorithms that are both efficient and converge to a global minimum. Typical algorithms which converge to a global minimum (such as simulated annealing [394]) are computationally very expensive, while methods which are more efficient (such as graduated non-convexity algorithms [48]) may get trapped in a local minimum. Minimizing continuous energy functionals leads in a natural way to partial differential equations (PDEs): applying gradient descent, for instance, yields a system of coupled diffusion–reactions equations for the two flow components. The fastly emerging use of PDE-based image restoration methods [381, 274], such as nonlinear diffusion filtering and total variation denoising, has motivated many researchers to apply similar ideas to estimate optic flow [11, 26, 85, 105, 170, 221, 270, 309, 344, 414]. A systematic framework that links the diffusion and optic flow paradigms, however, has not been studied so far. Furthermore, from the framework of diffusion filtering it is also well-known that anisotropic filters with a diffusion tensor have more degrees of freedom than isotropic ones with scalar-valued diffusivities. These additional degrees of freedom can be used to obtain better results in specific situations [412]. However, similar nonlinear anisotropic regularizers have not been considered in the optic flow literature so far.

The goal of the present paper is to address these issues. We present a theoretical framework for a broad class of regularization methods for optic flow estimation. For the reasons explained above, we focus on models that allow a formulation in terms of convex and rotation invariant continuous energy functionals. We consider image-driven and flow-driven models, isotropic and anisotropic ones, as well as models with spatial and spatio-temporal smoothing terms. We prove that all these approaches are well-posed in the sense of Hadamard: they have a unique solution that depends in a continuous (and therefore predictable) way on the input data. We shall see that our taxonomy includes not only many existing models, but also interesting novel ones. In particular, we will derive novel regularization functionals for optic flow estimation that are flow-driven and anisotropic. They are the optic flow analogues of anisotropic diffusion filters with a diffusion tensor. Many of the spatio-temporal methods have not been proposed before as well. With the increased computational possibilities of modern computers it is likely that they will become more important in the future. In the present paper we also focus on interesting relations between isotropic and anisotropic flow-driven methods. They allow us to formulate a general design principle which explains how one can create anisotropic optic flow regularizers from isotropic ones.

Our paper is organized as follows. In Section 13.2 we first review and classify existing image-driven and isotropic flow-driven models, before we derive a novel energy functional leading to anisotropic flow driven models. Then we show how one has to modify all models with a spatial smoothness term in order to obtain methods with spatio-temporal regularization. A unifying energy functional is derived that incorporates the previous models as well as novel ones. Its well-posedness is established in Section 13.3. In Section 13.4 we take advantage of structural similarities between isotropic and anisotropic approaches in order to formulate a design principle for anisotropic optic flow regularizers. Section 13.5 illustrates the behaviour of the different optic flow regularizers by applying them to two real-world examples. We conclude with a summary in Section 13.6. Some preliminary results of our paper have been published in a conference volume [351].

13.2 A Framework for Convex Regularizers

13.2.1 Spatial Regularizers

Basic Structure

In order to formalize the optic flow estimation problem, let us consider a real-valued image sequence $f(x, y, \theta)$, where (x, y) denotes the location within the image domain $\Omega \in \mathbb{R}^2$, and the time parameter $\theta \in [0, T]$ specifies the frame. The optic flow field $(u_1(x, y, \theta), u_2(x, y, \theta))$ describes the displacement between two subsequent frames θ and $\theta + 1$, i.e. $f(x, y, \theta)$ and $f(x + u_1(x, y, \theta), y + u_2(x, y, \theta), \theta + 1)$ should depict the same image detail. Frequently it is assumed that image objects keep their grey value over time:

$$f(x, y, \theta) - f(x + u_1(x, y, \theta), y + u_2(x, y, \theta), \theta + 1) = 0. \tag{13.1}$$

Such a model assumes that illumination changes do not appear, and that occlusions or disocclusions do not happen. Numerous generalizations to multiple constraint equations and/or different "conserved quantities" (replacing intensity) exist; see e.g. [126, 402]. However, since the goal of the present paper is to study different regularizers, we restrict ourselves to (13.1). If the spatial and temporal sampling is sufficiently fine, we may replace (13.1) by its first order Taylor approximation

$$f_x u_1 + f_y u_2 + f_\theta = 0, \tag{13.2}$$

where the subscripts x, y and θ denote partial derivatives. This so-called *optic flow constraint (OFC)* forms the basis of many differential methods for estimating the optic flow. Evidently such a single equation is not sufficient to determine the two unknown functions u_1 and u_2 uniquely. In order to recover a unique flow field, we

need an additional assumption. Regularization-based optic flow methods use as additional assumption the requirement that the optic flow field should be smooth (or at least piecewise smooth). The basic idea is to recover the optic flow as a minimizer of some energy functional of type

$$E(u_1, u_2) := \int_\Omega \Big((f_x u_1 + f_y u_2 + f_\theta)^2 + \alpha V(\nabla f, \nabla u_1, \nabla u_2) \Big) \, dx \, dy \tag{13.3}$$

where $\nabla := (\partial_x, \partial_y)^T$ denotes the spatial nabla operator, and $u := (u_1, u_2)^T$. The first term in the energy functional is a data term requiring that the OFC be fulfilled, while the second term penalizes deviations from (piecewise) smoothness. The smoothness term $V(\nabla f, \nabla u_1, \nabla u_2)$ is called *regularizer*, and the positive smoothness weight α is the *regularization parameter*. One would expect that the specific choice of the regularizer has a strong influence on the result. Therefore, let us discuss different classes of convex regularizers next.

Homogeneous Regularization

In 1981 Horn and Schunck [171] pioneered the field of regularization methods for optic flow computations. They used the regularizer

$$V_H(\nabla f, \nabla u_1, \nabla u_2) := |\nabla u_1|^2 + |\nabla u_2|^2. \tag{13.4}$$

It is a classic result from the calculus of variations [92, 113] that – under mild regularity conditions – a minimizer (u_1, u_2) of some energy functional

$$E(u_1, u_2) := \int_\Omega G(x, y, u_1, u_2, \nabla u_1, \nabla u_2) \, dx \, dy \tag{13.5}$$

satisfies necessarily the so-called *Euler–Lagrange equations*

$$\partial_x G_{u_{1x}} + \partial_y G_{u_{1y}} - G_{u_1} = 0, \tag{13.6}$$
$$\partial_x G_{u_{2x}} + \partial_y G_{u_{2y}} - G_{u_2} = 0 \tag{13.7}$$

with homogeneous Neumann boundary conditions:

$$\partial_n u_1 = 0 \quad \text{on } \partial\Omega, \tag{13.8}$$
$$\partial_n u_2 = 0 \quad \text{on } \partial\Omega. \tag{13.9}$$

Hereby, n is a vector normal to the image boundary $\partial\Omega$.
Applying this framework to the minimization of the Horn and Schunck functional leads to the PDEs

$$\Delta u_1 - \tfrac{1}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta) = 0, \tag{13.10}$$
$$\Delta u_2 - \tfrac{1}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta) = 0, \tag{13.11}$$

where $\Delta := \partial_{xx} + \partial_{yy}$ denotes the Laplace operator. These equations can be regarded as the steady state ($t \to \infty$) of the diffusion–reaction system

$$\partial_t u_1 = \Delta u_1 - \tfrac{1}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta), \tag{13.12}$$

$$\partial_t u_2 = \Delta u_2 - \tfrac{1}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta), \tag{13.13}$$

where t denotes an artificial evolution parameter that should not be mixed up with the time θ of the image sequence. These equations also arise when minimizing the Horn and Schunck functional using steepest descent. Schnörr [342] has established well-posedness by showing that this functional has a unique minimizer in $\mathrm{H}^1(\Omega) \times \mathrm{H}^1(\Omega)$ that depends continuously on the input data f. Recently, Hinterberger [170] proved similar well-posedness results for a related model where the $\mathrm{L}^2(\Omega)$ norm in the data term is replaced by the $\mathrm{L}^1(\Omega)$ norm.
We observe that the underlying diffusion process in the Horn and Schunck approach is the linear diffusion equation

$$\partial_t u_i = \Delta u_i = \operatorname{div}(g \nabla u_i) \tag{13.14}$$

with $g := 1$ and $i = 1, 2$. This equation is well-known for its regularizing properties and has been extensively used in the context of Gaussian scale-space; see [369] and the references therein. It smoothes, however, in a completely homogeneous way, since its diffusivity g equals 1 everywhere. As a consequence, it also blurs across semantically important flow discontinuities. This is the reason why the Horn and Schunck approach creates rather blurry optic flow fields. The regularizers described in the sequel are attempts to overcome this limitation.

Isotropic Image-Driven Regularization

It seems plausible that motion boundaries are a subset of the image boundaries. Thus, a simple way to prevent smoothing at motion boundaries consists of introducing a weight function into the Horn and Schunck regularizers that becomes small at image edges. This modification yields the regularizer

$$V_{II}(\nabla f, \nabla u_1, \nabla u_2) := g(|\nabla f|^2) \left(|\nabla u_1|^2 + |\nabla u_2|^2 \right), \tag{13.15}$$

where g is a decreasing, strictly positive function. This regularizer has been proposed and theoretically analysed by Alvarez *et al.* [11]. They used a data term based on (13.1) instead of (13.2). Although this leads to a nonconvex optimization problem, they could establish the existence of a unique solution $(u_1, u_2) \in \mathrm{C}([0, \mathrm{T}]; \mathrm{H}^1(\Omega) \times \mathrm{H}^1(\Omega))$ of the corresponding diffusion–reaction system. Here we use the linearized data term (13.2), such that the corresponding diffusion–reaction equations are given by

$$\partial_t u_1 = \operatorname{div}\left(g(|\nabla f|^2) \nabla u_1\right) - \tfrac{1}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta), \tag{13.16}$$

$$\partial_t u_2 = \operatorname{div}\left(g(|\nabla f|^2) \nabla u_2\right) - \tfrac{1}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta). \tag{13.17}$$

The underlying diffusion process is

$$\partial_t u_i = \operatorname{div}\left(g(|\nabla f|^2)\,\nabla u_i\right) \qquad (i = 1, 2). \tag{13.18}$$

It uses a scalar-valued diffusivity g that depends on the image gradient. Such a method can therefore be classified as inhomogeneous, isotropic and image-driven. Isotropic refers to the fact that a scalar-valued diffusivity guarantees a direction-independent smoothing behaviour, while inhomogeneous means that this behaviour may be space-dependent. Since the diffusivity does not depend on the flow itself, the diffusion process is linear. For more details on this terminology and diffusion filtering in image processing, we refer to [412]. Homogeneous regularization arises as a special case of (13.15) when $g(|\nabla f|^2) := 1$ is considered.

Anisotropic Image-Driven Regularization

An early anisotropic modification of the Horn and Schunck functional is due to Nagel [260]; see also [18, 115, 261, 265, 342, 343, 361]. The basic idea is to reduce smoothing across image boundaries, while encouraging smoothing along image boundaries. This is achieved by considering the regularizer

$$V_{AI}(\nabla f, \nabla u_1, \nabla u_2) \;:=\; \nabla u_1^T D(\nabla f) \nabla u_1 + \nabla u_2^T D(\nabla f) \nabla u_2. \tag{13.19}$$

$D(\nabla f)$ is a regularized projection matrix perpendicular to ∇f:

$$D(\nabla f) := \frac{1}{|\nabla f|^2 + 2\lambda^2} \left(\nabla f^{\perp} \nabla f^{\perp T} + \lambda^2 I\right), \tag{13.20}$$

where I denotes the unit matrix. This methods leads to the diffusion–reaction equations

$$\partial_t u_1 \;=\; \operatorname{div}\left(D(\nabla f)\,\nabla u_1\right) - \tfrac{1}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta), \tag{13.21}$$

$$\partial_t u_2 \;=\; \operatorname{div}\left(D(\nabla f)\,\nabla u_2\right) - \tfrac{1}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta). \tag{13.22}$$

The usage of a diffusion tensor $D(\nabla f)$ instead of a scalar-valued diffusivity allows a direction-dependent smoothing behaviour. This method can therefore be classified as anisotropic. Since the diffusion tensor depends on the image f but not on the unknown flow, it is a purely image-driven process that is linear in its diffusion part. Well-posedness in the function space $\mathrm{H}^1(\Omega) \times \mathrm{H}^1(\Omega)$ has been established by Schnörr [342].

The eigenvectors of D are $v_1 := \nabla f$, $v_2 := \nabla f^{\perp}$, and the corresponding eigenvalues are given by

$$\lambda_1(|\nabla f|) \;=\; \frac{\lambda^2}{|\nabla f|^2 + 2\lambda^2}, \tag{13.23}$$

$$\lambda_2(|\nabla f|) \;=\; \frac{|\nabla f|^2 + \lambda^2}{|\nabla f|^2 + 2\lambda^2}. \tag{13.24}$$

In the interior of objects we have $|\nabla f| \to 0$, and therefore $\lambda_1 \to 1/2$ and $\lambda_2 \to 1/2$. At ideal edges where $|\nabla f| \to \infty$, we obtain $\lambda_1 \to 0$ and $\lambda_2 \to 1$. Thus, we have isotropic behaviour within regions, and at image boundaries the process smoothes anisotropically along the edge. This behaviour is very similar to edge-enhancing anisotropic diffusion filtering [412]. In contrast to edge-enhancing anisotropic diffusion, however, Nagel's optic flow technique is linear. It is interesting to note that only recently it has been pointed out that the Nagel method may be regarded as an early predecessor of anisotropic diffusion filtering [18].
Homogeneous and isotropic image-driven regularizers are special cases of (13.19), where $D(\nabla f) := I$ and $D(\nabla f) := g(|\nabla f|^2)I$ are chosen.

Isotropic Flow-Driven Regularization

Image-driven regularization methods may create oversegmentations for strongly textured objects: in this case we have much more image boundaries than motion boundaries. In order to reduce smoothing only at motion boundaries, one may consider using a purely flow-driven regularizer. This, however, is at the expense of refraining from quadratic optimization problems. In earlier work [344, 414], the authors considered regularizers of type

$$V_{IF}(\nabla f, \nabla u_1, \nabla u_2) \;:=\; \Psi\left(|\nabla u_1|^2 + |\nabla u_2|^2\right), \tag{13.25}$$

where $\Psi(s^2)$ is a differentiable and increasing function that is convex in s, for instance

$$\Psi(s^2) \;:=\; \varepsilon s^2 + (1-\varepsilon)\lambda^2\sqrt{1 + s^2/\lambda^2} \qquad (0 < \varepsilon \ll 1,\ \lambda > 0). \tag{13.26}$$

Regularizer of type (13.25) lead to the diffusion–reaction system

$$\partial_t u_1 \;=\; \mathrm{div}\left(\Psi'(|\nabla u_1|^2 + |\nabla u_2|^2)\,\nabla u_1\right) - \tfrac{1}{\alpha} f_x(f_x u_1 + f_y u_2 + f_\theta), \tag{13.27}$$

$$\partial_t u_2 \;=\; \mathrm{div}\left(\Psi'(|\nabla u_1|^2 + |\nabla u_2|^2)\,\nabla u_2\right) - \tfrac{1}{\alpha} f_y(f_x u_1 + f_y u_2 + f_\theta), \tag{13.28}$$

where Ψ' denotes the derivative of Ψ with respect to its argument. The scalar-valued diffusivity $\Psi'(|\nabla u_1|^2 + |\nabla u_2|^2)$ shows that this model is isotropic and flow-driven. In general, the diffusion process is nonlinear now. For the specific regularizer (13.26), for instance, the diffusivity is given by

$$\Psi'(s^2) \;=\; \varepsilon + \frac{1-\varepsilon}{\sqrt{1 + s^2/\lambda^2}} \tag{13.29}$$

Since this nonlinear diffusivity is decreasing in its argument, smoothing at flow discontinuities is inhibited. For the specific choice $\Psi(s^2) := s^2$, however, homogeneous regularization with diffusivity $\Psi'(s^2) = 1$ is recovered again.

The preceding diffusion–reaction system uses a common diffusivity for both channels. This avoids that edges are formed at different locations in each channel. The same coupling also appears in isotropic nonlinear diffusion filters for vector-valued images as considered by Gerig *et al.* [144], and Whitaker and Gerig [432]. Nonlinear flow-driven regularizers with different diffusivities for each channel are discussed in Section 13.4.

Anisotropic Flow-Driven Regularization

We have seen that there exist isotropic and anisotropic image-driven regularizers as well as isotropic flow-driven ones. Thus, our taxonomy would be incomplete without having discussed anisotropic flow-driven regularizers. In the context of nonlinear diffusion filtering, anisotropic models with a diffusion tensor instead of a scalar-valued diffusivity offer advantages for images with noisy edges or interrupted structures [416].

How can one construct related optic flow methods? Let us first have a look at diffusion filtering of multichannel images. In the nonlinear anisotropic case, Weickert [405, 416] and Kimmel *et al.* [209] proposed to filter a multichannel image by using a joint diffusion tensor that depends on the gradients of all image channels. Our goal is thus to find an optic flow regularizer that leads to a coupled diffusion–reaction system where the same flux-dependent diffusion tensor $D(\nabla u_1, \nabla u_2)$ is used in each equation.

In order to derive this novel class of regularizers, we have to introduce some definitions first. As in the previous section, we consider an increasing smooth function $\Psi(s^2)$ that is convex in s. Let us assume that A is some symmetric $n \times n$ matrix with orthonormal eigenvectors $w_1,...,w_n$ and corresponding eigenvalues $\sigma_1,...,\sigma_n$. Then we may formally extend the scalar-valued function $\Psi(z)$ to a matrix-valued function $\Psi(A)$ by defining $\Psi(A)$ as the matrix with eigenvectors $w_1,...,w_n$ and eigenvalues $\Psi(\sigma_1),...,\Psi(\sigma_n)$:

$$\Psi(A) \;:=\; \sum_i \Psi(\sigma_i)\, w_i w_i^T. \tag{13.30}$$

This definition can be motivated from the case where $\Psi(z)$ is represented by a power series $\sum_{k=0}^{\infty} c_k\, z^k$. It is easy to show that the corresponding matrix-valued power series $\sum_{k=0}^{\infty} c_k A^k$ has the eigenvectors $w_1,...,w_n$ with the eigenvalues $\Psi(\sigma_1),...,\Psi(\sigma_n)$. Considering power series a common way to introduce matrix-valued functions; see [125] for a more details. The interpretation in terms of power series allows us to identify the scalar-valued function Ψ with its matrix-valued counterpart. In this way one may also define the derivative of $\Psi(A)$ by differentiating its power series. This leads to the simple representation

$$\Psi'(A) \;=\; \sum_i \Psi'(\sigma_i)\, w_i w_i^T. \tag{13.31}$$

Another definition that is useful for our considerations below is the trace of a quadratic matrix $A = (a_{ij})$. It is the sum of its diagonal elements, or – equivalently – the sum of its eigenvalues:

$$\mathrm{tr}(A) := \sum_i a_{ii} = \sum_i \sigma_i. \tag{13.32}$$

With these notations we consider the regularizer

$$V_{AF}(\nabla f, \nabla u_1, \nabla u_2) \;:=\; \mathrm{tr}\Psi\left(\nabla u_1 \nabla u_1^T + \nabla u_2 \nabla u_2^T\right). \tag{13.33}$$

Its argument

$$J := \nabla u_1 \nabla u_1^T + \nabla u_2 \nabla u_2^T \tag{13.34}$$

is a symmetric and positive semidefinite 2×2 matrix. Hence, there exist two orthonormal eigenvectors v_1, v_2 with corresponding nonnegative eigenvalues μ_1, μ_2. These eigenvalues specify the contrast of the vector-valued image (u_1, u_2) in the directions v_1 and v_2, respectively. This concept has been introduced by Di Zenzo for edge analysis of multichannel images [107]. It can be regarded as a generalization of the structure tensor [132], and it is related to the first fundamental form in differential geometry [216].
Our result below states that the regularizer (13.33) leads to the desired nonlinear anisotropic diffusion-reaction system.

Proposition 4 (Anisotropic Flow-Driven Regularization)
For the energy functional (13.3) with the regularizer (13.33), the corresponding steepest descent diffusion–reaction system is given by

$$\begin{aligned} \partial_t u_1 &= \mathrm{div}\left(\Psi'\left(\nabla u_1 \nabla u_1^T + \nabla u_2 \nabla u_2^T\right) \nabla u_1\right) \\ &- \tfrac{1}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta), \end{aligned} \tag{13.35}$$

$$\begin{aligned} \partial_t u_2 &= \mathrm{div}\left(\Psi'\left(\nabla u_1 \nabla u_1^T + \nabla u_2 \nabla u_2^T\right) \nabla u_2\right) \\ &- \tfrac{1}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta). \end{aligned} \tag{13.36}$$

Proof. Calculating the Euler–Lagrange equations to

$$E(u_1, u_2) \;:=\; \int_\Omega \left((f_x u_1 + f_y u_2 + f_\theta)^2 \;+\; \alpha\, \mathrm{tr}\Psi\left(\nabla u_1 \nabla u_1^T + \nabla u_2 \nabla u_2^T\right) \right) dx\, dy \tag{13.37}$$

yields

$$\partial_x \partial_{u_{1_x}}(\mathrm{tr}\Psi(J)) + \partial_y \partial_{u_{1_y}}(\mathrm{tr}\Psi(J)) - \tfrac{2}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta) \;=\; 0, \tag{13.38}$$

$$\partial_x \partial_{u_{2_x}}(\mathrm{tr}\Psi(J)) + \partial_y \partial_{u_{2_y}}(\mathrm{tr}\Psi(J)) - \tfrac{2}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta) \;=\; 0 \tag{13.39}$$

with J as defined in (13.34).
In order to simplify the evaluation of the first and second summand in both equations, we replace (x, y) by (x_1, x_2), and denote by e_i the unit vector in x_i direction. Together with the identities

$$\begin{aligned} \Psi'(J) &= \Psi'(\mu_1)\, v_1 v_1^T + \Psi'(\mu_2)\, v_2 v_2^T, && (13.40)\\ \mathrm{tr}(ab^T) &= a^T b, && (13.41)\\ \mathrm{div}\,(a) &= \sum_i \partial_{x_i}(e_i^T a) && (13.42) \end{aligned}$$

it follows that

$$\begin{aligned} \sum_i \partial_{x_i}\partial_{u_{kx_i}}(\mathrm{tr}\Psi(J)) &= \sum_i \partial_{x_i}\mathrm{tr}\left(\Psi'(J)\,\partial_{u_{kx_i}} J\right)\\ &\overset{(13.34)}{=} \sum_i \partial_{x_i}\mathrm{tr}\left(\Psi'(J)\,(e_i\nabla u_k^T + \nabla u_k e_i^T)\right)\\ &\overset{(13.40)}{=} \sum_i \partial_{x_i}\mathrm{tr}\sum_j \Psi'(\mu_j)\, v_j v_j^T (e_i\nabla u_k^T + \nabla u_k e_i^T)\\ &= \sum_i \partial_{x_i}\mathrm{tr}\sum_j \Psi'(\mu_j)\left((v_j^T e_i)(v_j\nabla u_k^T) + (v_j^T\nabla u_k)(v_j e_i^T)\right)\\ &\overset{(13.41)}{=} 2\sum_i \partial_{x_i}\sum_j \Psi'(\mu_j)(e_i^T v_j)(v_j^T\nabla u_k)\\ &\overset{(13.40)}{=} 2\sum_i \partial_{x_i}\left(e_i^T\Psi'(J)\nabla u_k\right)\\ &\overset{(13.42)}{=} 2\ \mathrm{div}\ (\Psi'(J)\,\nabla u_k) \qquad (k=1,2). \qquad (13.43) \end{aligned}$$

Plugging this result into the Euler–Lagrange equations concludes the proof. □

It should be noted that, in general, the eigenvalues $\Psi'(\mu_1)$ and $\Psi'(\mu_2)$ of the diffusion tensor are not equal. Therefore, we have a real anisotropic diffusion process with different behaviour in different directions. Homogeneous regularization is a special case of the regularizer (13.33), if $\Psi(s^2) := s^2$.
An interesting similarity between the isotropic regularizer (13.25) and its anisotropic counterpart (13.33) becomes explicit when writing (13.25) as

$$V_{IF}(\nabla f, \nabla u_1, \nabla u_2) \;=\; \Psi\left(\mathrm{tr}(\nabla u_1\nabla u_1^T + \nabla u_2\nabla u_2^T)\right). \qquad (13.44)$$

This shows that it is sufficient to exchange the role of the trace operator and the penalty function Ψ to switch between both regularization techniques. Another structural similarity will be discussed in Section 13.4.

13.2.2 A Unifying Framework

Let us now make a synthesis of all previously discussed models. Table 13.1 gives an overview of the smoothness terms that we have investigated so far.

Table 13.1: Classification of regularizers for optic flow models.

	isotropic	anisotropic
image-driven	$g(\|\nabla f\|^2) \sum_{i=1}^{2} \|\nabla u_i\|^2$	$\sum_{i=1}^{2} \nabla u_i^T D(\nabla f) \nabla u_i$
flow-driven	$\Psi\left(\sum_{i=1}^{2} \|\nabla u_i\|^2\right)$	$\mathrm{tr}\,\Psi\left(\sum_{i=1}^{2} \nabla u_i \nabla u_i^T\right)$

One may regard these regularizers as special cases of two more general models. Using the compact notation $\nabla u := (\nabla u_1, \nabla u_2)$, the first model has the structure

$$V_1(\nabla f, \nabla u) \;:=\; \Psi\left(\mathrm{tr} \nabla u^T D(\nabla f) \nabla u\right). \tag{13.45}$$

For $\Psi(s^2) := s^2$, this model comprises pure image-driven models, regardless whether they are isotropic ($D(\nabla f) := g(|\nabla f|^2) I$) or anisotropic. Isotropic flow-driven models arise for $D := I$. In the general case, the model may be both image-driven and flow-driven.
The second model can be written as

$$V_2(\nabla f, \nabla u) \;:=\; \mathrm{tr}\Psi\left(\nabla u\, D(\nabla f) \nabla u^T\right). \tag{13.46}$$

It comprises the anisotropic flow-driven case and its combinations with image-driven approaches. Note the large structural similarities between (13.45) and (13.46). Both models can be assembled to the regularizer

$$\begin{aligned} V(\nabla f, \nabla u) \;&:=\; (1-\beta)\, \Psi\left(\mathrm{tr} \nabla u^T D(\nabla f) \nabla u\right) \\ &+\; \beta\, \mathrm{tr}\Psi\left(\nabla u\, D(\nabla f) \nabla u^T\right) \end{aligned} \tag{13.47}$$

where the paramter $\beta \in [0, 1]$ determines the anisotropy. This regularizer is embedded into the general optic flow functional

$$E(u) \;=\; \int_\Omega \Big((f_x u_1 + f_y u_2 + f_\theta)^2 + \alpha\, V(\nabla f, \nabla u) \Big)\, dx\, dy. \tag{13.48}$$

From the perspective of regularization, Table 1 reveals another useful classification in this context: while image–driven models correspond to the class of *quadratic* regularizers [40], flow–driven models belong to the more general class of *non–quadratic convex* regularizers. This latter class has been suggested in [80, 349, 376] for generalizing the well-known quadratic regularization approaches used for early computational vision.

13.2.3 Spatio-Temporal Regularizers

All regularizers that we have discussed so far use only *spatial* smoothness constraints. Thus, it would be natural to impose some amount of (piecewise) temporal smoothness as well. Using our results from the previous section it is straightforward to extend the smoothness constraint into the temporal domain. Instead of calculating the optic flow (u_1, u_2) as the minimizer of the two-dimensional integral (13.48) for each time frame θ, we now minimize a single three-dimensional integral whose solution is the optic flow for *all* frames $\theta \in [0, T]$:

$$E(u) := \int_{\Omega \times [0,T]} \Big((f_x u_1 + f_y u_2 + f_\theta)^2 + \alpha\, V(\nabla_3 f, \nabla_3 u) \Big)\, dx\, dy\, d\theta \tag{13.49}$$

where $\nabla_3 := (\partial_x, \partial_y, \partial_\theta)^T$ denotes the spatio-temporal nabla operator.
The corresponding diffusion–reaction systems of spatio–temporal energy functionals have the same structure as the pure spatial ones that we investigated so far. The only difference is that the spatial nabla operator ∇ has to be replaced by its spatio-temporal analogue ∇_3. Thus, one has to solve 3D diffusion–reaction systems instead of 2D ones.
Not many spatio-temporal regularizers have been studied in the literature so far. To the best of our knowledge, there have been no attempts to investigate rotation invariant spatio-temporal models that use homogeneous, isotropic image-driven, or anisotropic flow-driven regularizers.
Nagel [264] suggested an extension of his anisotropic image-driven smoothness constraint, where the diffusion tensor (13.20) is replaced by

$$D(\nabla_3 f) := \frac{1}{2|\nabla_3 f|^2 + 3\lambda^2} \left(\nabla_3 f^\perp \nabla_3 f^{\perp T} + \lambda^2 I \right), \tag{13.50}$$

Its eigenvalues are given by

$$\lambda_1(|\nabla_3 f|) = \frac{\lambda^2}{2|\nabla_3 f|^2 + 3\lambda^2}, \tag{13.51}$$

$$\lambda_2(|\nabla_3 f|) = \frac{|\nabla_3 f|^2 + \lambda^2}{2|\nabla_3 f|^2 + 3\lambda^2} = \lambda_3(|\nabla_3 f|). \tag{13.52}$$

Isotropic flow-driven spatio-temporal regularizers have been studied by the authors in [426]. They showed that it outperforms a corresponding spatial regularizer at low additional computing time, if an entire image stack is to be processed.
It appears that the limited memory of previous computer architectures prevented many researchers from studying approaches with spatio-temporal regularizers, since they require to keep the entire image stack in the computer memory. On contemporary PCs or workstations, however, this is no longer a problem, if typical stack sizes are used (e.g. 32 frames with 256×256 pixels). It is thus likely that spatio-temporal regularizers will become more important in the future.

13.3 Well-Posedness Properties

In this section we shall prove that the energy functionals (13.48) and (13.49), respectively, admit a unique solution that continuously depends on the initial data. These favourable properties are the consequence of embedding the optic flow constraint (13.2) into a convex regularization approach.

13.3.1 Prerequisites and Main Result

We assume that the regularizer $\Psi : \mathbb{R} \to \mathbb{R}$ satisfies the following properties:

- $\Psi(.)$ is differentiable and increasing.
- $\Psi(s^2)$ is strictly convex in s.
- There exist constants $c_1, c_2 > 0$ such that

$$c_1 s^2 \leq \Psi(s^2) \leq c_2 s^2 \;, \quad \forall s \;. \tag{13.53}$$

Furthermore, we consider only matrices $D(\nabla f)$ that are symmetric and positive definite.
As function space of admissible optic flow fields we use

$$\mathcal{H} \;:=\; \left\{ u = (u_1, u_2)^T \mid u_i,\, \partial_{x_j} u_i \in L^2(\Omega),\; \forall i, j \right\}, \tag{13.54}$$

endowed with the scalar product

$$(u, v)_{\mathcal{H}} := \int_\Omega \left(u^T v + \mathrm{tr} \nabla u \nabla v^T \right) \, dx_1 \cdots dx_n \tag{13.55}$$

and its induced norm

$$\|u\|_{\mathcal{H}} := (u, u)_{\mathcal{H}}^{1/2}. \tag{13.56}$$

In what follows, $\langle b, u \rangle$ denotes the action of some linear continuous functional $b \in \mathcal{H}^*$, i.e. some element of the dual space $\mathcal{H}^*$, on some vector field $u \in \mathcal{H}$.
We consider image sequences $f \in \mathcal{H}$, whose spatial derivatives f_{x_1} and f_{x_2} are linearly independent in $L^2(\Omega)$ and have finite $L^\infty(\Omega)$ norm.

These prerequisites enable us to state the following well-posedness result.

Theorem 16 (Well-Posedness)
Under the preceding assumptions, the optic flow methods (13.48) and (13.49) admit a unique solution in $\mathcal{H}$. This solution depends continuously on the image sequence f.

Before turning to the proof, we would like to comment on the choice of the regular function space $\mathcal{H}$. Camera sensors always create blurred image data, and its is quite common to smooth the image sequence even further, e.g. by convolving with a Gaussian before applying an optic flow algorithm [35]. With such input data one cannot expect to obtain discontinuous displacement fields in a strict mathematical sense. It should, however, be mentioned that one can still approximate the discontinuous situation with functions in $\mathcal{H}$ having large, but finite gradients. If one is interested in admitting real discontinuous flow fields, one can try to establish well-posedness results for functions of bounded variation, as has been done by Aubert, Deriche, and Kornprobst [26] for a specific optic flow model. In this case the mathematical analysis becomes more complicated.

In order to verify our theorem, we need to consider matrix-valued convexity estimates, and we have to pay special attention to the case when the data term degenerates. This shall be done next.

13.3.2 Convexity

We wish to show that the functional $E(u)$ is strictly convex over $\mathcal{H}$. To this end, we may disregard linear and constant terms in $E(u)$ and consider the functional $F(u)$ defined by

$$\begin{aligned} F(u) &:= \int_\Omega \left((\nabla f^T u)^2 + \alpha V(\nabla f, \nabla u) \right) dx_1 \cdots dx_n \\ &= E(u) + \langle b, u \rangle + c \end{aligned} \tag{13.57}$$

where

$$\langle b, u \rangle := -2 \int_\Omega f_\theta (\nabla f^T u)\, dx_1 \cdots dx_n, \tag{13.58}$$

$$c := -\int_\Omega f_\theta^2 \, dx_1 \cdots dx_n. \tag{13.59}$$

Strict convexity is a crucial property for the existence of a unique global minimizing optical flow field u of $E(u)$ determined as the root of the equation

$$F'(u) = b \tag{13.60}$$

for any linear functional $b \in \mathcal{H}^*$. We proceed in several steps. First, we consider the smoothness terms $V_1(\nabla f, \nabla u)$ and $V_2(\nabla f, \nabla u)$ separately. This can be done because the sum of convex functions is again convex. Then we consider all terms together, that is the functional $F(u)$.
The term

$$V_1(\nabla f, \nabla u) := \Psi \left(\mathrm{tr} \nabla u^T D(\nabla f) \nabla u \right) \tag{13.61}$$

belongs to the class of smoothness terms which were considered in earlier work on isotropic nonlinear diffusion of multichannel images (e.g. [346]). To see this, let

$$\mathrm{vec}\,(\nabla u) := \begin{pmatrix} \nabla u_1 \\ \nabla u_2 \end{pmatrix} \tag{13.62}$$

denote the vector obtained by stacking the columns of ∇u one upon the other, and let $|\cdot|_D$ denote the norm induced by the scalar product

$$(\mathrm{vec}\,(\nabla u), \mathrm{vec}\,(\nabla v))_D \;:=\; \mathrm{vec}\,(\nabla u)^T \begin{pmatrix} D(\nabla f) & 0 \\ 0 & D(\nabla f) \end{pmatrix} \mathrm{vec}\,(\nabla v) \tag{13.63}$$

Then V_1 can be rewritten as

$$\Psi\left(\mathrm{tr}\nabla u^T D(\nabla f)\nabla u\right) = \Psi(|\mathrm{vec}\,(\nabla u)|_D^2), \tag{13.64}$$

and the framework in [346] is applicable.

The second anisotropic and flow–driven smoothness term

$$V_2(\nabla f, \nabla u) \;:=\; \mathrm{tr}\Psi\left(\nabla u\, D(\nabla f)\nabla u^T\right) \tag{13.65}$$

is new in the context of optical flow computation. Note that by contrast to term V_1, the function Ψ is matrix-valued. The strict convexity of V_2 is stated in

Proposition 5 (Matrix-Valued Convexity)
Let $\Psi : \mathbb{R} \to \mathbb{R}$ be strictly convex, A and B two positive semidefinite symmetric matrices with $A \neq B$, and $\tau \in (0,1)$. Then

$$\mathrm{tr}\Psi((1-\tau)A + \tau B) \;<\; (1-\tau)\,\mathrm{tr}\Psi(A) + \tau\,\mathrm{tr}\Psi(B). \tag{13.66}$$

Proof. Put $C := (1-\tau)A+\tau B$. Since A, B, C are symmetric, there are orthonormal systems of eigenvectors $\{u_i\}, \{v_i\}, \{w_i\}$ and real-valued eigenvalues $\{\alpha_i\}, \{\beta_i\}, \{\gamma_i\}$ such that

$$A = \sum_i \alpha_i u_i u_i^T\,, \quad B = \sum_i \beta_i v_i v_i^T\,, \quad C = \sum_i \gamma_i w_i w_i^T\,. \tag{13.67}$$

Expanding the vectors u_i, v_i with respect to the system $\{w_i\}$ gives

$$u_i = \sum_j (u_i^T w_j) w_j\,, \quad v_i = \sum_j (v_i^T w_j) w_j. \tag{13.68}$$

With this we obtain

$$\begin{aligned}
\sum_j \gamma_j w_j w_j^T &= C = (1-\tau)A + \tau B \\
&= (1-\tau)\sum_i \alpha_i u_i u_i^T + \tau \sum_i \beta_i v_i v_i^T \\
&= (1-\tau)\sum_i \alpha_i \sum_j (u_i^T w_j)^2 w_j w_j^T + \tau \sum_i \beta_i \sum_j (v_i^T w_j)^2 w_j w_j^T \\
&= \sum_j \Big((1-\tau)\sum_i (u_i^T w_j)^2 \alpha_i + \tau \sum_i (v_i^T w_j)^2 \beta_i \Big) w_j w_j^T . \qquad (13.69)
\end{aligned}$$

Comparing the coefficients shows that γ_j is a convex combination of $\{\alpha_i\}$ and $\{\beta_i\}$:

$$\gamma_j = (1-\tau)\sum_i (u_i^T w_j)^2 \alpha_i + \tau \sum_i (v_i^T w_j)^2 \beta_i , \quad \forall j . \qquad (13.70)$$

Since $\Psi : \mathbb{R} \to \mathbb{R}$ is strictly convex and $\sum_j (u_i^T w_j)^2 = \sum_j (v_i^T w_j)^2 = 1$ for all i, we obtain

$$\begin{aligned}
\mathrm{tr}\Psi(C) &= \sum_j \Psi(\gamma_j) \qquad (13.71) \\
&= \sum_j \Psi\Big((1-\tau)\sum_i (u_i^T w_j)^2 \alpha_i + \tau \sum_i (v_i^T w_j)^2 \beta_i \Big) \\
&< \sum_j \Big((1-\tau)\sum_i (u_i^T w_j)^2 \Psi(\alpha_i) + \tau \sum_i (v_i^T w_j)^2 \Psi(\beta_i) \Big) \\
&= (1-\tau)\sum_i \sum_j (u_i^T w_j)^2 \Psi(\alpha_i) + \tau \sum_i \sum_j (v_i^T w_j)^2 \Psi(\beta_i) \\
&= (1-\tau)\,\mathrm{tr}\Psi(A) + \tau\,\mathrm{tr}\Psi(B) \qquad (13.72)
\end{aligned}$$

This concludes the convexity proof. □

13.3.3 Degeneracy of the Data Term

So far we have shown the strict convexity of the smoothness term $V(\nabla f, \nabla u)$ in (13.57). To show that also $F(u)$ is strictly convex, we may use the equivalent condition that $F'(u)$ is strongly monotone [443]:

$$\exists\, c_m > 0 \; : \quad \langle F'(u) - F'(v), u - v \rangle \;\geq\; c_m \|u - v\|_{\mathcal{H}}^2 , \quad \forall\, u, v \in \mathcal{H}. \qquad (13.73)$$

Note that the smoothness term fulfills this condition because it is convex, as we have just shown. Exploiting relation (13.73) for the smoothness term itself, we

obtain

$$\begin{aligned}
&\langle F'(u) - F'(v),\, u - v\rangle \\
&\geq \int_\Omega \Big(\nabla f^T u \nabla f^T (u-v) - \nabla f^T v \nabla f^T (u-v) + \alpha' \mathrm{tr}(\nabla(u-v)\nabla(u-v)^T)\Big)\, dx \\
&= \int_\Omega \Big((\nabla f^T w)^2 + \alpha' |\nabla w|^2\Big)\, dx\ , \qquad \forall w := u - v \in \mathcal{H}\ , \qquad (13.74)
\end{aligned}$$

with some constant α' and $dx := dx_1 dx_2$ (without loss of generality we consider the case $\Omega \subset \mathbb{R}^2$). Thus, it remains to show that

$$\int_\Omega \Big((\nabla f^T w)^2 + \alpha' |\nabla w|^2\Big) dx \ \geq\ c_m \|w\|^2_{\mathcal{H}}\ , \quad \forall w \in \mathcal{H}\ . \qquad (13.75)$$

This is not straightforward, because ∇f may vanish in some image regions. Therefore, the first term on the left side in (13.75) cannot simply be bounded away from zero with the smallest eigenvalue of $\nabla f \nabla f^T$. This difficulty has been dealt with in [342]. For the sake of completeness, we sketch the corresponding proof. We use the following abbreviations:

$$\forall w_1, w_2 \in L^2(\Omega), \qquad (w_1, w_2)_0 := \int_\Omega w_1 w_2\, dx\ , \qquad |w_1|_0 := (w_1, w_1)_0^{1/2} \qquad (13.76)$$

and $|w|_\infty := \mathrm{ess\,sup}_{x\in\Omega} |w(x)|$.

Now assume that (13.75) were not true. Then there exists a sequence $\{w_n\} \subset \mathcal{H}$ with $\|w_n\|_{\mathcal{H}} = 1$, such that

$$\int_\Omega \Big((\nabla f^T w_n)^2 + \alpha' |\nabla w_n|^2\Big)\, dx \ \to\ 0 \quad \text{for} \quad n \to \infty\ . \qquad (13.77)$$

Using the Poincaré inequality [448]

$$\int_\Omega (v - \bar{v})^2\, dx \ \leq\ C(\Omega) \int_\Omega |\nabla v|^2\, dx \qquad \forall v \in H^1(\Omega)\ , \qquad (13.78)$$

where $\bar{v} := \frac{1}{|\Omega|}\int_\Omega v\, dx$, it follows from (13.77) and (13.78) that

$$|w_n - \bar{w}_n|_0^2 \ \to\ 0, \qquad (13.79)$$

with $\bar{w}_n := \frac{1}{|\Omega|}\int_\Omega w_n\, dx$. By means of

$$|\nabla f^T w|_0^2 \ \leq\ 2\,|f_{x_1}|_\infty^2\, |w_1|_0^2 \ +\ 2\,|f_{x_2}|_\infty^2\, |w_2|_0^2 \qquad (13.80)$$

and (13.79), we obtain

$$|\nabla f^T (w_n - \bar{w}_n)|_0^2 \ \to\ 0\ . \qquad (13.81)$$

Hence, by virtue of (13.77) and (13.81) it follows that

$$\begin{aligned} |\nabla f^T \bar{w}_n|_0 &= |\nabla f^T w_n + \nabla f^T(\bar{w}_n - w_n)|_0 \\ &\leq |\nabla f^T w_n|_0 + |\nabla f^T(\bar{w}_n - w_n)|_0 \\ &\leq |\nabla f^T w_n|_0 + \alpha' \int_\Omega |\nabla w_n|^2 \, dx + |\nabla f^T(\bar{w}_n - w_n)|_0 \\ &\to 0 \,. \end{aligned} \tag{13.82}$$

Using temporarily the notation

$$p := f_{x_1} \bar{w}_{1n}, \qquad q := f_{x_2} \bar{w}_{2n} \tag{13.83}$$

we estimate:

$$\begin{aligned} |p+q|_0^2 &= |p|_0^2 + |q|_0^2 + 2\,(p,q)_0 \\ &\geq |p|_0^2 + |q|_0^2 - 2\,|p|_0\,|q|_0\,\frac{|(p,q)_0|}{|p|_0\,|q|_0} \\ &\geq |p|_0^2 + |q|_0^2 - (|p|_0^2 + |q|_0^2)\,\frac{|(p,q)_0|}{|p|_0\,|q|_0} \\ &= (|p|_0^2 + |q|_0^2)\,\left(1 - \frac{|(p,q)_0|}{|p|_0\,|q|_0}\right) \,. \end{aligned} \tag{13.84}$$

Resubstituting (13.83) gives

$$|\nabla f^T \bar{w}_n|_0^2 \;\geq\; (\bar{w}_{1n}^2 |f_{x_1}|_0^2 + \bar{w}_{2n}^2 |f_{x_2}|_0^2)\,\left(1 - \frac{|(f_{x_1}, f_{x_2})_0|}{|f_{x_1}|_0\,|f_{x_2}|_0}\right) \,. \tag{13.85}$$

Since f_{x_1} and f_{x_2} are linearly independent, it follows that

$$1 - \frac{|(f_{x_1}, f_{x_2})_0|}{|f_{x_1}|_0\,|f_{x_2}|_0} \;>\; 0. \tag{13.86}$$

With this and (13.82), we conclude from (13.85) that

$$\bar{w}_n \;\to\; 0, \tag{13.87}$$

and, by virtue of (13.77) and (13.79), we obtain

$$\|w_n\|_{\mathcal{H}} \;\leq\; \|w_n - \bar{w}_n\|_{\mathcal{H}} + \|\bar{w}_n\|_{\mathcal{H}} \;\to\; 0. \tag{13.88}$$

This contradicts our assumptions that $\|w_n\|_{\mathcal{H}} = 1$. Consequently, (13.75) must hold true.

13.3.4 Existence, Uniqueness, and Continuous Dependence on the Data

It is a well-established result (see, e.g., [443]) that property (13.73) together with the Lipschitz continuity of the operator F' (which holds true under mild conditions with respect to the data ∇f, f_θ, cf. [342, 350]) ensure the existence of a unique and globally minimizing optical vector field u that continuously depends on the data. To understand the latter property, suppose we are given two image sequences and corresponding functionals b_1, b_2 (cf. (13.57)) and minimizers u_1, u_2:

$$F'(u_1) = b_1, \tag{13.89}$$
$$F'(u_2) = b_2. \tag{13.90}$$

By virtue of (13.73) we have

$$\begin{aligned} c_m \|u_1 - u_2\|_{\mathcal{H}}^2 &\leq \langle F'(u_1) - F'(u_2),\, u_1 - u_2 \rangle \\ &\leq \|F'(u_1) - F'(u_2)\|_{\mathcal{H}^*}\, \|u_1 - u_2\|_{\mathcal{H}}. \end{aligned} \tag{13.91}$$

Thus,

$$\|u_1 - u_2\|_{\mathcal{H}} \leq \frac{1}{c_m} \|b_1 - b_2\|_{\mathcal{H}^*}\ , \quad \forall\, b_1, b_2\ . \tag{13.92}$$

This equation states that, for a slight change of the image sequence data, the corresponding optical flow field cannot arbitrarily jump but gradually changes, too. It is therefore an important robustness property.

13.4 Extensions

All regularizers that we have discussed so far can be motivated from existing nonlinear diffusion methods for multichannel images, where a joint diffusivity or diffusion tensor for all channels is used. As one might expect, this is not the only way to construct useful optic flow regularizers. In particular, there exists a more general design principle for anisotropic flow-driven regularizers which we will discuss next.

Our key observation for deriving this principle is an interesting relation between anisotropic flow-driven regularizers and isotropic flow-driven ones: the anisotropic regularizer $\mathrm{tr}\Psi(J)$ can be expressed by means of the eigenvalues μ_1, μ_2 of J as

$$V_{AF}(\nabla f, \nabla u_1, \nabla u_2) = \Psi(\mu_1) + \Psi(\mu_2), \tag{13.93}$$

while its isotropic counterpart $\Psi(\mathrm{tr}J)$ can be written as

$$V_{IF}(\nabla f, \nabla u_1, \nabla u_2) = \Psi(\mu_1 + \mu_2). \tag{13.94}$$

This observation motivates us to formulate the following design principle for rotationally invariant anisotropic flow-driven regularizers:

Design Principle (Rotationally Invariant Anisotropic Regularizers) *Assume that we are given some isotropic regularizer $\Psi(\sum_i |\nabla u_i|^2)$ with a non-quadratic function Ψ, and a decomposition of its argument*

$$\sum_i |\nabla u_i|^2 = \sum_j \rho_j, \tag{13.95}$$

where the ρ_j are rotationally invariant expressions. Then the regularizer $\sum_j \Psi(\rho_j)$ is rotationally invariant and anisotropic.

Examples

1. The decomposion that has been used in (13.93) and (13.94) to transit from an isotropic to an anisotropic model was the trace identity

$$|\nabla u_1|^2 + |\nabla u_2|^2 \;=\; \mu_1 + \mu_2, \tag{13.96}$$

where μ_1 and μ_2 are the eigenvalues of $J = \nabla u_1 \nabla u_1^T + \nabla u_2 \nabla u_2^T$.

2. Schnörr [344] proposed the regularizer

$$V_{AFS}(\nabla f, \nabla u_1, \nabla u_2) \;:=\; \Psi\left(\mathrm{div}^{\,2} u\right) + \Psi\left(\mathrm{rot}^{\,2} u\right) + \Psi\left(\mathrm{sh}^{\,2} u\right) \tag{13.97}$$

with $u := (u_1, u_2)^T$, the curl operator $\mathrm{rot}\, u := u_{2x} - u_{1y}$, and the shear operator $\mathrm{sh}\, u := \sqrt{(u_{2y} - u_{1x})^2 + (u_{1y} + u_{2x})^2}$. Applying the design principle, one can derive this expression from the identity [212]

$$|\nabla u_1|^2 + |\nabla u_2|^2 = \tfrac{1}{2}\left(\mathrm{div}^{\,2} u + \mathrm{rot}^{\,2} u + \mathrm{sh}^{\,2} u\right). \tag{13.98}$$

Using the regularizer (13.97) in the functional (13.3) leads to the highly anisotropic diffusion–reaction system

$$\begin{aligned}
\partial_t u_1 \;&=\; \partial_x \left(\left(\Psi'(\mathrm{div}^{\,2}) + \Psi'(\mathrm{sh}^{\,2} u)\right) u_{1x} + \left(\Psi'(\mathrm{div}^{\,2} u) - \Psi'(\mathrm{sh}^{\,2} u)\right) u_{2y}\right) \\
&+\; \partial_y \left(\left(\Psi'(\mathrm{sh}^{\,2} u) + \Psi'(\mathrm{rot}^{\,2} u)\right) u_{1y} + \left(\Psi'(\mathrm{sh}^{\,2} u) - \Psi'(\mathrm{rot}^{\,2} u)\right) u_{2x}\right) \\
&-\; \tfrac{1}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta),
\end{aligned} \tag{13.99}$$

$$\begin{aligned}
\partial_t u_2 \;&=\; \partial_x \left(\left(\Psi'(\mathrm{sh}^{\,2} u) - \Psi'(\mathrm{rot}^{\,2} u)\right) u_{1y} + \left(\Psi'(\mathrm{sh}^{\,2} u) + \Psi'(\mathrm{rot}^{\,2} u)\right) u_{2x}\right) \\
&+\; \partial_y \left(\left(\Psi'(\mathrm{div}^{\,2} u) - \Psi'(\mathrm{sh}^{\,2} u)\right) u_{1x} + \left(\Psi'(\mathrm{div}^{\,2} u) + \Psi'(\mathrm{sh}^{\,2} u)\right) u_{2y}\right) \\
&-\; \tfrac{1}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta).
\end{aligned} \tag{13.100}$$

Note that now the coupling between both equations is more complicated than in the previous cases, where a joint diffusivity or a joint diffusion tensor has been used. We are not aware of similar diffusion filters for multichannel images. Well-posedness properties in $\mathrm{H}^1(\Omega) \times \mathrm{H}^1(\Omega)$ and experimental results for this optic flow method are presented in [344, 350].

3. Requiring that the ρ_j in (13.95) be rotationally invariant ensures the rotation invariance of the anisotropic regularizer. If we dispense with rotation invariance, the design principle can still be used. As an example, let us study the flow-driven regularization methods that are considered in [26, 85, 105, 221]. They use a regularizer of type

$$V_C(\nabla f, \nabla u_1, \nabla u_2) \;:=\; \Psi(|\nabla u_1|^2) + \Psi(|\nabla u_2|^2). \tag{13.101}$$

According to our design principle, we may regard this regularizer as an anisotropic version of the isotropic regularizer (13.25). However, the decomposition of its argument into $|\nabla u_1|^2$ and $|\nabla u_2|^2$ is not rotationally invariant. The corresponding diffusion–reaction system is given by

$$\partial_t u_1 \;=\; \operatorname{div}\left(\Psi'(|\nabla u_1|^2)\,\nabla u_1\right) - \tfrac{1}{\alpha} f_x (f_x u_1 + f_y u_2 + f_\theta), \tag{13.102}$$
$$\partial_t u_2 \;=\; \operatorname{div}\left(\Psi'(|\nabla u_2|^2)\,\nabla u_2\right) - \tfrac{1}{\alpha} f_y (f_x u_1 + f_y u_2 + f_\theta), \tag{13.103}$$

which shows that both systems are completely decoupled in their diffusion terms. Thus, flow discontinuities may be created at different locations for each channel. The same decoupling appears also for some other PDE-based optic flow methods such as [309].

While each of the two diffusion processes is isotropic, the overall process reveals some anisotropy: in general, the two diffusivities $\Psi'(|\nabla u_1|^2)$ and $\Psi'(|\nabla u_2|^2)$ are not identical. For this approach, well-posedness results in $\mathrm{BV}(\Omega)$ have been established by Aubert *et al.* [26]. They used a regularized data term that cannot degenerate and where the $\mathrm{L}^2(\Omega)$ norm is replaced by the $\mathrm{L}^1(\Omega)$ norm.

Besides this model there is also a number of related stochastic methods that lead to discrete models which are not consistent approximations to rotation invariant processes [44, 45, 49, 166, 246, 259]. Nonconvex regularizers are typically used in these approaches. Discrete spatio-temporal versions of the regularizer (13.101) are investigated in [44, 259].

It is a challenging open question whether there exist more useful rotation invariant convex regularizers than the ones we have just discussed. This is one of our current research topics.

13.5 Experiments

Although the main goal of our paper is to provide a taxonomy for convex regularizers and to analyse their theoretical foundations, we will in this section present some qualitative experiments that show specific features of each method. We will

focus on spatial regularizers. We discretized the corresponding diffusion–reaction equations by means on an explicit finite difference scheme. For a more detailed discussion on numerical issues and a comparison between spatial and spatio-temporal regularization in the case of flow-driven isotropic smoothing, the reader is referred to [426].

In our first experiment we used the marbled block sequence of Otte and Nagel (KOGS/IAKS, University of Karlsruhe, Germany) [293]. These images can be downloaded from the web site `http://i21www.ira.uka.de/image-sequences`. The sequence consists of 31 frames of size 512×512. In our case we only used frame 16 and 17. The input images have been preprocessed by a convolution with a Gaussian of standard deviation $\sigma = 1$, and the regularization parameter α has been set to 10000. For image-driven models, we used as contrast parameter $\lambda := 1$, and for flow-driven methods $\lambda := 0.01$ was used. The iterations were stopped when the Euclidean norm of the residue dropped by a factor of 100. Figure 13.1 depicts the results for the optic flow magnitude, while Figure 13.2 shows the flow fields as vector plot. For better visibility, we also show a detail of the flow magnitude images in Figure 13.3.
As expected, one can observe that the homogeneous regularization of Horn and Schunck creates very smooth flow fields. It is, however, unsuited to respect any flow discontinuities.
Isotropic image-driven reduces smoothing at all image edges. This may create an oversegmentation of the flow fields, as can be seen from the flow artifacts resulting from the texture of the marbled floor. This oversegmentation influences in particular the flow magnitude, while the flow direction appears to be more stable.
Anisotropic image-driven regularization permits smoothing along image edges. This leads to a more homogeneous flow field than the one from isotropic image-driven smoothing. Larger structures of the marble texture, however, are still visible in this case as well.
Flow-driven models are performing better here. The marble texture, which corresponds to image discontinuities but not to flow discontinuities, does hardly perturb the flow field. Figure 13.3 shows that, similar to the image-driven case, anisotropic regularization is less affected by these texture artifacts that isotropic smoothing, although the differences are a bit smaller. This shows that anisotropic flow-driven regularization is an interesting technique for optic flow problems where flow discontinuities are important and highly textured image structures are present.

In a second example we sketch a possible application field for more advanced anisotropic flow-driven regularizers. Figure 13.4 shows the result of an optic flow technique with the div–rot–shear regularizion (13.97). This regularization is in particular useful for applications where the kinematic behaviour of moving objects plays an important role: The divergence term, for instance, may signal an

Figure 13.1: (a) TOP LEFT: Frame 16 of the marbled block sequence (512×512 pixels). (b) TOP RIGHT: Optic flow magnitude between Frame 16 and 17 for homogeneous regularization. (c) MIDDLE LEFT: Result for image-driven isotropic regularization (d) MIDDLE RIGHT: Image-driven anisotropic regularization. (e) BOTTOM LEFT: Flow-driven isotropic regularization (f) BOTTOM RIGHT: Flow-driven anisotropic regularization.

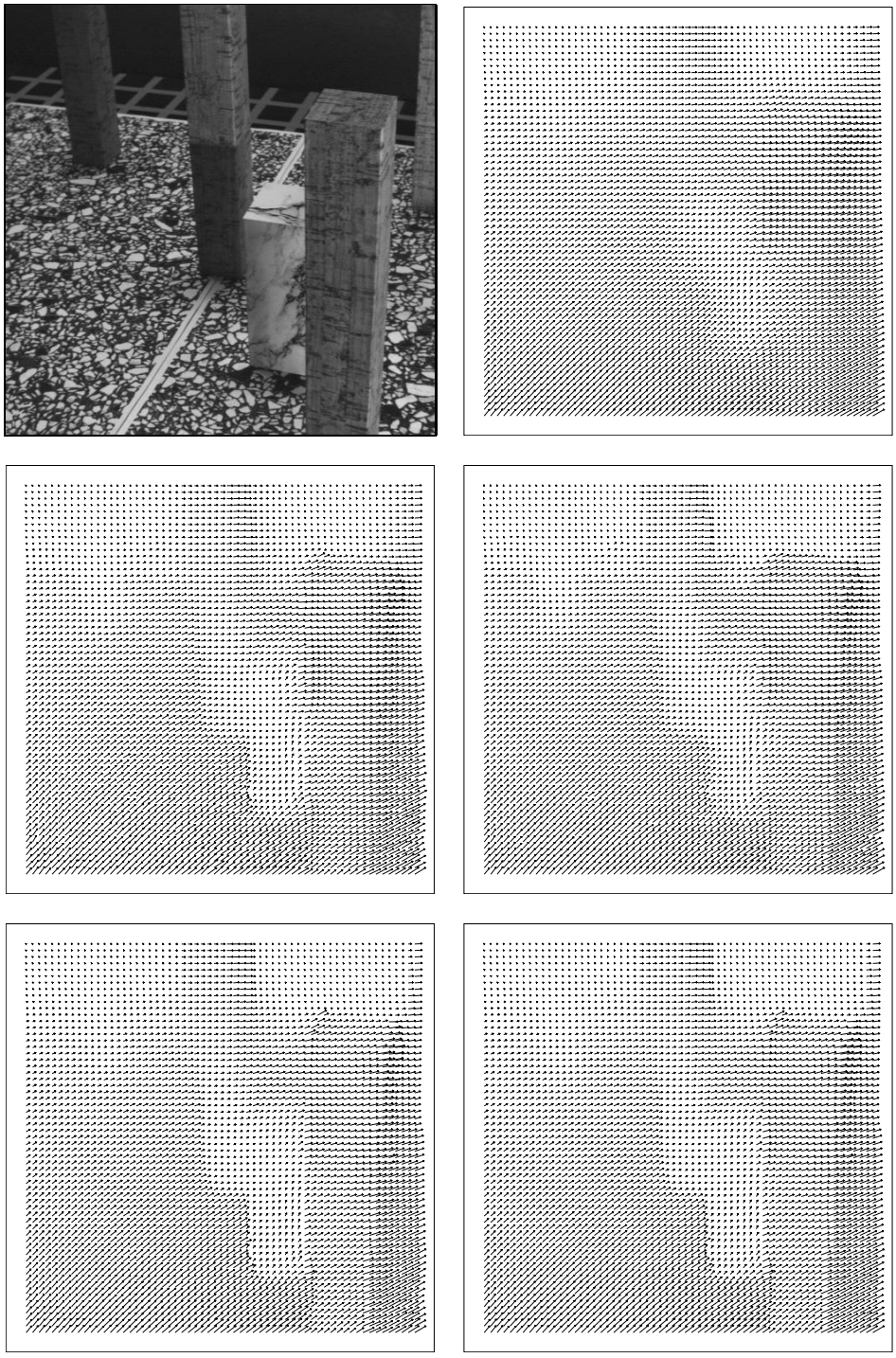


Figure 13.2: (a) TOP LEFT: Frame 16 of the marble block sequence. (b) TOP RIGHT: Optic flow magnitude for homogeneous regularization. (c) MIDDLE LEFT: Image-driven isotropic regularization (d) MIDDLE RIGHT: Image-driven anisotropic regularization. (e) BOTTOM LEFT: Flow-driven isotropic regularization (f) BOTTOM RIGHT: Flow-driven anisotropic regularization. In order to improved visibility, the flow fields have been subsampled by a factor 8, and their magnitude has been scaled by a factor 6.

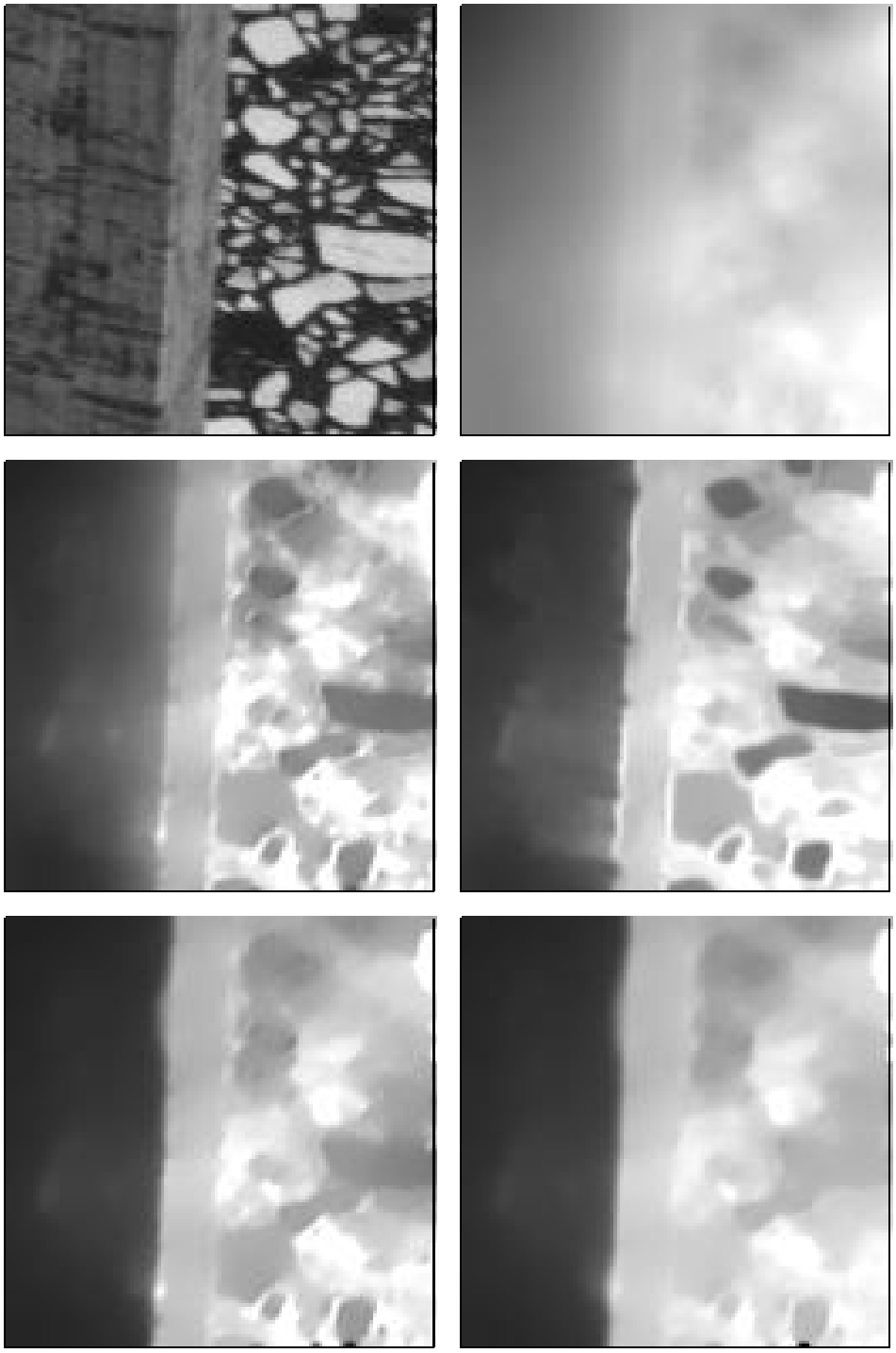


Figure 13.3: (a) TOP LEFT: Detail from the lower right part of Frame 16 (128×128 pixels). (b) TOP RIGHT: Optic flow magnitude for homogeneous regularization. (c) MIDDLE LEFT: Image-driven isotropic regularization (d) MIDDLE RIGHT: Image-driven anisotropic regularization. (e) BOTTOM LEFT: Flow-driven isotropic regularization (f) BOTTOM RIGHT: Flow-driven anisotropic regularization. For better visibility, the grey values of the optic flow results been transformed by a gamma correction with $\gamma = 0.4$.

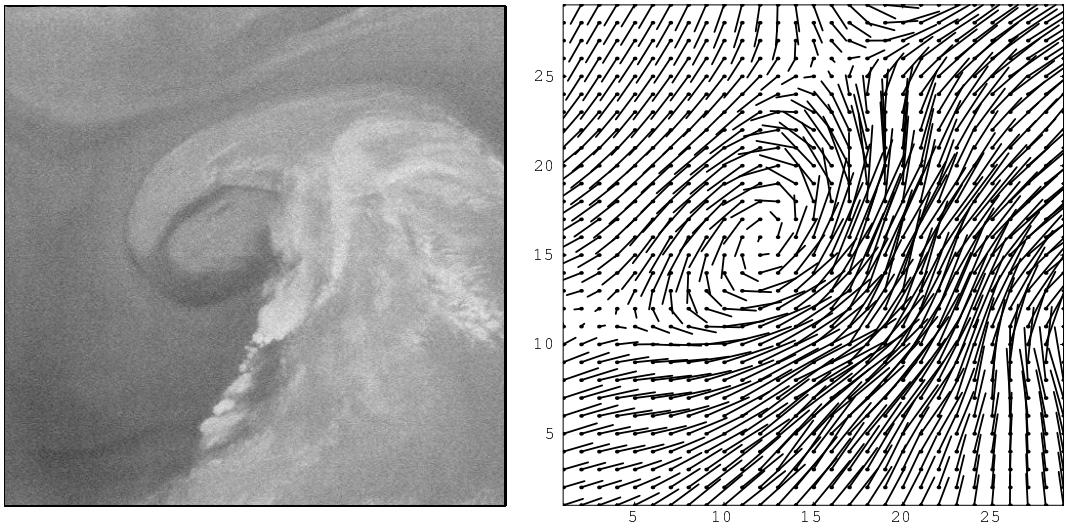


Figure 13.4: LEFT: Frame from a cyclone image sequence. RIGHT: Optic flow field with the regularizer (13.97).

approaching object, while the vorticity term analyses turbulent behaviour. It is thus not surprising that such an approach works well for the meteorological image in this figure.

13.6 Summary and Conclusions

The goal of this paper was to derive a diffusion theory for optic flow functionals. Minimizing optic flow functionals by steepest descent leads to a set of two coupled diffusion–reaction systems. Since similar equations appear for diffusion filtering of multi-channel images, the question arises whether there are optic flow analogues to the various kinds of diffusion filters.

We saw that image-driven optic flow regularizers correspond to linear diffusion filters, while flow-driven regularizers create nonlinear diffusion processes. Pure spatial regularizers can be expressed as 2D diffusion–reaction processes, and spatio-temporal regularizers may be regarded as generalizations to the 3D case. This taxonomy helped us not only to classify existing methods within a unifying framework, but also to identify gaps, where no models are available in the current literature. We filled these gaps by deriving suitable methods with the specified properties, and we proved well-posedness for the class of convex diffusion-based optic flow regularization methods.

One important novelty along these lines was the derivation of regularizers that can be related to anisotropic diffusion filters with a matrix-valued diffusion tensor.

This also enabled us to propose a design principle for anisotropic regularizers, and we discovered an interesting structural similarity between isotropic and anisotropic models: it is sufficient to exchange the role of the trace operator and the penalty function in order to switch between the two models.
We are convinced that these relations are only the starting point for many more fruitful interactions between the theories of diffusion filtering and variational optic flow methods. Diffusion filtering has progressed very much in recent years, and so it appears appealing to incorporate recent results from this area into optic flow methods. Conversely, it is clear that novel optic flow regularizers can also be regarded as energy functionals for suitable diffusion filters. It should also be mentioned that our taxonomy for regularizers is not restricted to optic flow problems. These regularizers can be used in all applications where vector-valued data are to be processed, be it colour images, multispectral data, or displacement fields in other applications (e.g. stereo reconstruction or registration of medical images).
We also hope that our taxonomy provides a unifying platform for algorithms for the entire class of convex variational optic flow methods. Our future plans are to use such a platform for a systematic comparison of different numerical methods. Another point on our agenda is an investigation of alternative rotation-invariant decompositions that can be applied to construct useful anisotropic regularizers.

Acknowledgement. C. S. completed his doctoral thesis under the supervision of Prof. Nagel in 1991. He is grateful to Prof. Nagel who introduced him to the field of computer vision.

Bibliography

[1] S. Abbasi, F. Mokhtarian, and J. Kittler. Reliable classification of chrysanthemum leaves through curvature scale space. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 284–295. Springer, Berlin, 1997.

[2] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill–posed problems. *Inverse Problems*, 10:1217–1229, 1994.

[3] S. T. Acton. Multigrid anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):280–291, Mar. 1998.

[4] S. T. Acton, A. C. Bovik, and M. M. Crawford. Anisotropic diffusion pyramids for image segmentation. In *Proc. 1994 IEEE International Conference on Image Processing*, volume 3, pages 478–482, Austin, TX, Nov. 1994. IEEE Computer Society Press.

[5] R. A. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[6] A. Almansa and T. Lindeberg. Enhancement of fingerprint images using shape-adapted scale-space operators. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*, pages 3–19. Kluwer, Dordrecht, 1997.

[7] H. W. Alt. *Lineare Funktionalanalysis.* Springer, Berlin, 1992.

[8] L. Alvarez. Images and PDE's. In M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, and J.-M. Morel, editors, *ICAOS '96: Images, Wavelets and PDEs*, volume 219 of *Lecture Notes in Control and Information Sciences*, pages 3–14. Springer, London, 1996.

[9] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. In *Proc. IAPR Workshop on Machine Vision*, pages 423–427, Tokyo, Japan, Nov. 2000.

[10] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation*, 2001. In press.

[11] L. Alvarez, J. Esclarín, M. Lefébure, and J. Sánchez. A PDE model for computing the optical flow. In *Proc. XVI Congreso de Ecuaciones Diferenciales y Aplicaciones*, pages 1349–1356, Las Palmas de Gran Canaria, Spain, Sept. 1999.

[12] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations in image processing. *Archive for Rational Mechanics and Analysis*, 123:199–257, 1993.

[13] L. Alvarez, P.-L. Lions, and J.-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. II. *SIAM Journal on Numerical Analysis*, 29:845–866, 1992.

[14] L. Alvarez and L. Mazorra. Signal and image restoration using shock filters and anisotropic diffusion. *SIAM Journal on Numerical Analysis*, 31:590–605, 1994.

[15] L. Alvarez and F. Morales. Affine morphological multiscale analysis of corners and multiple junctions. *International Journal of Computer Vision*, 25:95–107, 1994.

[16] L. Alvarez and J. Sánchez. 3-D geometry reconstruction using a color image stereo pair and partial differential equations. Technical Report 6, Instituto Universitario de Ciencias y Tecnologías Cibernéticas, University of Las Palmas de Gran Canaria, Spain, 2000.

[17] L. Alvarez, J. Weickert, and J. Sánchez. A scale-space approach to nonlocal optical flow calculations. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 235–246. Springer, Berlin, 1999.

[18] L. Alvarez, J. Weickert, and J. Sánchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39(1):41–56, Aug. 2000.

[19] L. Ambrosio and V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Bolletino della Unione Matematica Italiana*, 7:105–123, 1992.

[20] P. Ananden. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[21] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. Technical report, Dept. of Análisis Matemático, University of Valencia, Burjassot (Valencia), Spain, 1998.

[22] Y. Aoki and T. Iijima. Pluralizing method of simple similarity. *IEICE Transactions on Information and Systems*, E79-D(5):485–490, 1996. In English.

[23] A. B. Arehart, L. Vincent, and B. B. Kimia. Mathematical morphology: The Hamilton–Jacobi connection. In *Proc. Fourth International Conference on Computer Vision*, pages 215–219, Berlin, May 1993. IEEE Computer Society Press.

[24] J. Arnspang. Notes on local determination of smooth optic flow and the translational property of first order optic flow. Technical Report DIKU-88/1, Dept. of Computer Science, University of Copenhagen, Denmark, 1988.

[25] K. Aström and A. Heyden. Stochastic analysis of image acquisition and scale-space smoothing. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*, pages 129–136. Kluwer, Dordrecht, 1997.

[26] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal on Applied Mathematics*, 60(1):156–182, 1999.

[27] O. Axelsson. Globally convergent continuation methods for non-linear equations. In *Proc. First Workshop on Large-Scale Scientific Computations*, International House of Scientists "F. Curie", Bulgaria, 1997.

[28] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda. Uniqueness of the Gaussian kernel for scale space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:26–33, 1986.

[29] G. E. Bacon, P. J. Bacon, and R. K. Griffiths. The orientation of apatite crystals in bone. *Journal of Applied Crystallography*, 12:99–103, 1979.

[30] I. Bajla and I. Holländer. Nonlinear filtering of magnetic resonance tomograms by geometry-driven diffusion. *Machine Vision and Applications*, 10:243–255, 1998.

[31] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith. PETSc 2.0 Users Manual. Technical Report ANL-95/11 – Revision 2.0.22, Argonne National Laboratory, Argonne, IL, 1998.

[32] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith. PETSc Home Page, 1998. http://www.mcs.anl.gov/petsc.

[33] E. Bänsch and K. Mikula. A coarsening finite element strategy in image selective smoothing. *Computation and Visualization in Science*, 1:53–61, 1997.

[34] G. I. Barenblatt, M. Bertsch, R. D. Passo, and M. Ughi. A degenerate pseudoparabolic regularization of a nonlinear forward–backward heat equation arising in the theory of heat and mass exchange in stably stratified turbulent shear flow. *SIAM Journal on Mathematical Analysis*, 24:1414–1439, 1993.

[35] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77, 1994.

[36] E. Barth, C. Zetzsche, M. Ferraro, and I. Rentschler. Fractal properties from 2D-curvature on multiple scales. In B. Vemuri, editor, *Geometric Methods in Computer Vision*, volume 2031 of *Proceedings of SPIE*, pages 87–99. SPIE Press, Bellingham, 1993.

[37] B. Benhamouda. Parameter adaptation for nonlinear diffusion in image processing. Master's thesis, Dept. of Mathematics, University of Kaiserslautern, Germany, 1994.

[38] M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, and J.-M. Morel, editors. *ICAOS '96: Images, Wavelets and PDEs*, volume 219 of *Lecture Notes in Control and Information Sciences*. Springer, London, 1996.

[39] F. Bergholm. Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:726–741, 1987.

[40] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76:869–889, 1988.

[41] S. Beucher and C. Lantuéjoul. Use of watersheds in contour detection. In *Proc. International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation*, pages 2.1–2.12, Rennes, France, Sept. 1979. IRISA Report No. 131.

[42] J. Bigün and G. H. Granlund. Optimal orientation detection of linear symmetry. In *Proc. First International Conference on Computer Vision*, pages 433–438, London, England, June 1987. IEEE Computer Society Press.

[43] J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:775–790, 1991.

[44] M. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 292–302, Maui, HI, June 1991. IEEE Computer Society Press.

[45] M. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63:75–104, 1996.

[46] M. J. Black. Recursive non-linear estimation of discontinuous flow fields. In J.-O. Eklundh, editor, *Computer Vision – ECCV '94*, volume 800 of *Lecture Notes in Computer Science*, pages 138–145. Springer, Berlin, 1994.

[47] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 19(1):57–91, 1996.

[48] A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, Cambridge, MA, 1987.

[49] L. Blanc-Féraud, M. Barlaud, and T. Gaidon. Motion estimation involving discontinuities in a multiresolution scheme. *Optical Engineering*, 32(7):1475–1482, 1993.

[50] 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, Mar. 1998.

[51] H. Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* North Holland, Amsterdam, 1973.

[52] H. Brezis. *Analyse fonctionelle.* Masson, Paris, 1992.

[53] W. L. Briggs. *A Multigrid Tutorial.* SIAM, Philadelphia, 1987.

[54] A. D. Brink. Using spatial information as an aid to maximum entropy image threshold selection. *Pattern Recognition Letters*, 17(1):29–36, Jan. 1996.

[55] A. D. Brink and N. E. Pendock. Minimum cross-entropy threshold selection. *Pattern Recognition*, 29(1):179–188, Jan. 1996.

[56] R. W. Brockett and P. Maragos. Evolution equations for continuous-scale morphology. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 125–128, San Francisco, CA, Mar. 1992.

[57] R. W. Brockett and P. Maragos. Evolution equations for continuous-scale morphological filtering. *IEEE Transactions on Signal Processing*, 42:3377–3386, 1994.

[58] A. M. Bruckstein, G. Sapiro, and D. Shaked. Evolution of planar polygons. *International Journal of Pattern Recognition and Artificial Intelligence*, 9:991–1014, 1995.

[59] A. M. Bruckstein and D. Shaked. On projective invariant smoothing and evolution of planar curves and polygons. *Journal of Mathematical Imaging and Vision*, 7:225–240, 1997.

[60] R. Brügelmann and W. Förstner. Noise estimation for color edge extraction. In W. Förstner and S. Ruwiedel, editors, *Robust Computer Vision*, pages 90–107. Wichmann, Karlsruhe, 1992.

[61] B. Buck and V. Macaulay, editors. *Maximum Entropy in Action.* Clarendon Press, Oxford, 1991.

[62] P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.

[63] B. Cabral and L. Leedom. Imaging vector fields using line integral convolution. *Computer Graphics*, 27(4):263–272, 1993.

[64] L. D. Cai. Some notes on repeated averaging smoothing. In J. Kittler, editor, *Pattern Recognition*, volume 301 of *Lecture Notes in Computer Science*, pages 597–605. Springer, Berlin, 1988.

[65] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[66] R. Carmona and S. Zhong. Adaptive smoothing respecting feature directions. *IEEE Transactions on Image Processing*, 7(3):353–358, Mar. 1998.

[67] V. Caselles, F. Catté, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66:1–31, 1993.

[68] V. Caselles and T. Coll. Snakes in movement. *SIAM Journal on Numerical Analysis*, 33:2445–2456, 1996.

[69] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. Fifth International Conference on Computer Vision*, pages 694–699, Cambridge, MA, June 1995. IEEE Computer Society Press.

[70] V. Caselles, J.-M. Morel, G. Sapiro, and A. Tannenbaum, editors. *Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis*, volume 7(3) of *IEEE Transactions on Image Processing*. IEEE Signal Processing Society Press, Mar. 1998.

[71] V. Caselles, J.-M. Morel, and C. Sbert. An axiomatic approach to image interpolation. *IEEE Transactions on Image Processing*, 7(3):376–386, Mar. 1998.

[72] F. Catté. Convergence of iterated affine and morphological filters by nonlinear semi-group theory. In M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, and J.-M. Morel, editors, *ICAOS '96: Images, Wavelets and PDEs*, volume 219 of *Lecture Notes in Control and Information Sciences*, pages 125–133. Springer, London, 1996.

[73] F. Catté, F. Dibos, and G. Koepfler. A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets. *SIAM Journal on Numerical Analysis*, 32:1895–1909, 1995.

[74] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 32:1895–1909, 1992.

[75] A. Chambolle. Partial differential equations and image processing. In *Proc. 1994 IEEE International Conference on Image Processing*, volume 1, pages 16–20, Austin, TX, Nov. 1994. IEEE Computer Society Press.

[76] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[77] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal–dual method for total-variation based image restoration. In M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, and J.-M. Morel, editors, *ICAOS '96: Images, Wavelets and PDEs*, volume 219 of *Lecture Notes in Control and Information Sciences*, pages 241–252. Springer, London, 1996.

[78] T. F. Chan and C. K. Wong. Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7:370–375, 1998.

[79] P. Charbonnier. *Reconstruction d'image: Régularisation avec prise en compte des discontinuités*. PhD thesis, Laboratoire Informatique, Signaux et Systémes de Sophia Antipolis, Université Nice – Sophia Antipolis, France, 1994.

[80] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. 1994 IEEE International Conference on Image Processing*, volume 2, pages 168–172, Austin, TX, Nov. 1994. IEEE Computer Society Press.

[81] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.

[82] B. B. Chaudhuri and N. Sarkar. Texture segmentation using fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):72–77, Jan. 1995.

[83] S. Chinveeraphan, R. Takamatsu, and M. Sato. A hierarchical description of digital grayscale images based on image dipoles. In *Proc. 13th International Conference on Pattern Recognition*, volume B, pages 246–250, Vienna, Austria, Aug. 1996.

[84] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[85] I. Cohen. Nonlinear variational method for optical flow computation. In *Proc. Eighth Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, Tromsø, Norway, May 1993.

[86] T. Cohignac, F. Eve, F. Guichard, C. Lopez, and J.-M. Morel. Numerical analysis of the fundamental equation of image processing. Technical Report 9254, CEREMADE, Université Paris IX – Dauphine, France, 1992.

[87] D. Colton. *Partial Differential Equations*. Random House, New York, 1998.

[88] N. Cornelius and T. Kanade. Adapting optical flow to measure object motion in reflectance and X-ray image sequences. In *Proc. ACM SIGRAPH/SIGART Interdisciplinary Workshop on Motion: Representation and Perception*, pages 50–58, Toronto, Canada, Apr. 1983.

[89] G.-H. Cottet. Diffusion approximation on neural networks and applications for image processing. In F. Hodnett, editor, *Proc. Sixth European Conference on Mathematics in Industry*, pages 3–9. Teubner, Stuttgart, 1992.

[90] G.-H. Cottet and M. El Ayyadi. Nonlinear PDE operators with memory terms for image processing. In *Proc. 1996 IEEE International Conference on Image Processing*, volume 1, pages 481–483, Lausanne, Switzerland, Sept. 1996.

[91] G.-H. Cottet and L. Germain. Image processing through reaction combined with nonlinear diffusion. *Mathematics of Computation*, 61:659–673, 1993.

[92] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience, New York, 1953.

[93] D. Cremers, C. Schnörr, J. Weickert, and C. Schellewald. Diffusion-snakes using statistical shape knowledge. In G. Sommer and Y. Y. Zeevi, editors, *Algebraic Frames for the Perception-Action Cycle*, volume 1888 of *Lecture Notes in Computer Science*, pages 164–174. Springer, Berlin, 2000.

[94] D. Cremers, C. Schnörr, J. Weickert, and C. Schellewald. Learning of translation invariant shape knowledge for steering diffusion-snakes. In G. Baratoff and H. Neumann, editors, *Dynamische Perzeption*, volume 9 of *Proceedings in Artificial Intelligence*, pages 117–122. Akademische Verlagsgesellschaft Aka, Berlin, 2000.

[95] B. Dacorogna. *Weak Continuity and Weak Lower Semicontinuity of Non-Linear Functionals.* Lecture Notes in Mathematics. Springer, Berlin, 1982.

[96] B. Dacorogna. *Direct Methods in the Calculus of Variations.* Springer, Berlin, 1989.

[97] M. Dæhlen and A. Tveito. *Numerical Methods and Software Tools in Industrial Mathematics.* Birkhäuser, Boston, 1997.

[98] J. Damon. Local Morse theory for solutions to the heat equation and Gaussian blurring. *Journal of Differential Equations*, 115:368–401, 1995.

[99] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.

[100] D. De Vleeschauer, F. A. Cheikh, R. Hamila, and M. Gabbouj. Watershed segmentation of an image enhanced by Teager energy driven diffusion. In *Proc. Sixth International Conference on Image Processing and its Applications*, pages 254–258, Dublin, Ireland, July 1997.

[101] G. Demoment. Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2024–2036, Dec. 1989.

[102] R. Deriche. Fast algorithms for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:78–87, 1990.

[103] R. Deriche. Recursively implementing the Gaussian and its derivatives. In V. Srinivasan, S. H. Ong, and Y. H. Ang, editors, *Proc. Second International Singapore Conference on Image Processing*, pages 263–267, Singapore, Sept. 1992.

[104] R. Deriche and O. Faugeras. Les EDP en traitement des images et vision par ordinateur. *Traitement du Signal*, 13(6):551–577, 1996. Numéro Special.

[105] R. Deriche, P. Kornprobst, and G. Aubert. Optical-flow estimation while preserving its discontinuities: a variational approach. In *Proc. Second Asian Conference on Computer Vision*, volume 2, pages 290–295, Singapore, Dec. 1995.

[106] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr. Anisotropic feature-preserving denoising of height fields and bivariate data. In *Proc. Graphics Interface 2000*, Montreal, Canada, May 2000.

[107] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics and Image Processing*, 33:116–125, 1986.

[108] F. Dibos. Projective analysis of 2-D images. *IEEE Transactions on Image Processing*, 7(3):274–279, Mar. 1998.

[109] D. Dobson and O. Scherzer. Analysis of regularized total variation penalty methods for denoising. *Inverse Problems*, 12:601–617, 1996.

[110] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM Journal on Numerical Analysis*, 34:1779–1791, 1997.

[111] J. Douglas and J. E. Gunn. A general formulation of alternating direction methods. Part I. Parabolic and hyperbolic problems. *Numerische Mathematik*, 6:428–453, 1964.

[112] A. Dressel. Nichtlineare Diffusion in der Bildverarbeitung. Master's thesis, Faculty of Mathematics, University of Heidelberg, Germany, 1999.

[113] L. Elsgolc. *Calculus of Variations.* Pergamon, Oxford, 1961.

[114] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer, Dordrecht, 1996.

[115] W. Enkelmann. Investigation of multigrid algorithms for the estimation of optical flow fields in image sequences. *Computer Vision, Graphics and Image Processing*, 43:150–177, 1987.

[116] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, 1998.

[117] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions.* CRC Press, Boca Raton, 1992.

[118] J. Fairfield. Toboggan contrast enhancement for contrast segmentation. In *Proc. Tenth International Conference on Pattern Recognition*, volume 1, pages 712–716, Atlantic City, NJ, June 1990. IEEE Computer Society Press.

[119] O. Faugeras. Sur l'évolution de courbes simples du plan projectif réel. *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, 317:565–570, 1993.

[120] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint.* MIT Press, Cambridge, MA, 1993.

[121] O. Faugeras. Cartan's moving frame method and its application to the geometry and evolution of curves in the Euclidean, affine and projective planes. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 11–46. Springer, Berlin, 1994.

[122] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, Mar. 1998.

[123] O. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images.* MIT Press, Cambridge, MA, 2001. To appear.

[124] M. Ferraro, G. Boccignone, and T. Caelli. On the representation of image structures via scale space entropy conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1199–1203, Nov. 1999.

[125] D. T. Finkbeiner. *Introduction to Matrices and Linear Transformations.* Freeman, San Francisco, 1966.

[126] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5:77–104, 1990.

[127] L. Florack. *Image Structure*, volume 10 of *Computational Imaging and Vision.* Kluwer, Dordrecht, 1997.

[128] L. M. J. Florack. Data, models, and images. In *Proc. 1996 IEEE International Conference on Image Processing*, volume 1, pages 469–472, Lausanne, Switzerland, Sept. 1996.

[129] L. M. J. Florack, W. J. Niessen, and M. Nielsen. The intrinsic structure of the optic flow incorporating measurement duality. *International Journal of Computer Vision*, 27:263–286, 1998.

[130] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, 1992.

[131] G. E. Ford, R. R. Estes, and H. Chen. Scale-space analysis for image sampling and interpolation. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 165–168, San Francisco, CA, Mar. 1992.

[132] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[133] J. Fourier. *Théorie analytique de la chaleur.* Didot, Paris, 1822.

[134] A. S. Frangakis and R. Hegerl. Nonlinear anisotropic diffusion in three-dimensional electron microscopy. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 386–397. Springer, Berlin, 1999.

[135] D. S. Fritsch. A medial description of greyscale image structure by gradient-limited diffusion. In R. A. Robb, editor, *Visualization in Biomedical Computing '92*, volume 1808 of *Proceedings of SPIE*, pages 105–117. SPIE Press, Bellingham, 1992.

[136] J. Fröhlich and J. Weickert. Image processing using a wavelet algorithm for nonlinear diffusion. Technical Report 104, Laboratory of Technomathematics, University of Kaiserslautern, Germany, Mar. 1994.

[137] S. Fučik, A. Kratochvil, and J. Nečas. Kačanov–Galerkin method. *Commentationes Mathematicae Universitatis Carolinae*, 14(4):651–659, 1973.

[138] D. Gabor. Information theory in electron microscopy. *Laboratory Investigation*, 14:801–807, 1965.

[139] B. Galvin, B. McCane, K. Novins, D. Mason, and S. Mills. Recovering motion fields: an analysis of eight optical flow algorithms. In *Proc. 1998 British Machine Vision Conference*, Southampton, England, Sept. 1998.

[140] D. Geiger and A. Yuille. A common framework for image segmentation. *International Journal of Computer Vision*, 6(3):227–243, 1991.

[141] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:367–383, 1992.

[142] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4:932–945, 1995.

[143] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[144] G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11:221–232, 1992.

[145] T. Gijbels, P. Six, L. Van Gool, F. Catthoor, H. De Man, and A. Oosterlinck. A VLSI-architecture for parallel non-linear diffusion with applications in vision. In *Proc. IEEE Workshop on VLSI Signal Processing*, volume 7, pages 398–407, 1994.

[146] D. Gilbarg and N. Trudinger. *Elliptic Partial Differential Equations of Second Order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften.* Springer, Berlin, second edition, 1994.

[147] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation.* Birkhäuser, Basel, 1984.

[148] R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky. Fast geodesic active contours. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 34–45. Springer, Berlin, 1999.

[149] E. González and A. Trujillo. *XMegaWave. A Window Oriented Image Processing Software.* Departamento de Informática y Sistemas, Universidad de Las Palmas de Gran Canaria, Spain. http://amiserver.dis.ulpgc.es/xmwgus/.

[150] R. C. Gonzalez and P. Wintz. *Digital Image Processing.* Addison–Wesley, Reading, 1987.

[151] A. R. Gourlay. Implicit convolution. *Image and Vision Computing*, 3:15–23, 1985.

[152] T. Grahs, A. Meister, and T. Sonar. Image processing for numerical approximations of conservation laws: nonlinear anisotropic artificial dissipation. Technical Report F8, Institute for Applied Mathematics, University of Hamburg, Germany, Dec. 1998.

[153] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision.* Kluwer, Dordrecht, 1995.

[154] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9:84–93, 1990.

[155] L. D. Griffin, A. C. F. Colchester, and G. P. Robinson. Scale and segmentation of grey-level images using maximum gradient paths. *Image and Vision Computing*, 10:389–402, 1992.

[156] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind.* Pitman, Boston, 1984.

[157] C. W. Groetsch. Spectral methods for linear inverse problems with unbounded operators. *Journal of Approximation Theory*, 70:16–28, 1992.

[158] F. Guichard and J.-M. Morel. Partial differential equations and image iterative filtering. In I. S. Duff and G. A. Watson, editors, *The State of the Art in Numerical Analysis*, number 63 in IMA Conference Series (New Series), pages 525–562. Clarendon Press, Oxford, 1997.

[159] F. Guichard and L. Rudin. Accurate estimation of discontinuous optical flow by minimizing divergence related functionals. In *Proc. Third International Conference on Image Processing*, volume 1, pages 497–500, Lausanne, Switzerland, Sept. 1996.

[160] M. Hanke and C. W. Groetsch. Nonstationary iterated Tikhonov regularization. *Journal of Optimization Theory and Applications*, 98:37–53, 1998.

[161] C. G. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fouth Alvey Vision Conference*, pages 147–152, Manchester, England, Aug. 1988.

[162] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, Cambridge, UK, 2000.

[163] T. C. Hasley, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Physical Review*, 33:1141–1151, 1986.

[164] J. Heers, C. Schnörr, and H.-S. Stiehl. Investigation of parallel and globally convergent iterative schemes for nonlinear variational image smoothing and segmentation. In *Proc. 1998 IEEE International Conference on Image Processing*, Chicago, IL, Oct. 1998.

[165] J. Heers, C. Schnörr, and H. S. Stiehl. Globally-convergent iterative numerical schemes for non-linear variational image smoothing and segmentation on a multi-processor machine. *IEEE Transactions on Image Processing*, 2001. To appear.

[166] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:1217–1232, 1993.

[167] H. G. E. Hentschel and I. Procaccia. The infinite number of generated dimensions of fractals and strange attractors. *Physica D*, 8:435–444, 1983.

[168] C. Herbe. Numerical methods for nonlinear diffusion models of sandpile growth. Master's thesis, Dept. of Mathematics, University of Kaiserslautern, Germany, Jan. 1999.

[169] E. Hille and R. S. Philips. *Functional Analysis and Semi-Groups*. American Mathematical Society, Providence, 1957.

[170] W. Hinterberger. Generierung eines Films zwischen zwei Bildern mit Hilfe des optischen Flusses. Master's thesis, Industrial Mathematics Institute, University of Linz, Austria, 1999.

[171] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[172] B. K. P. Horn and M. J. Brooks, editors. *Shape from Shading*. MIT Press, Cambridge, MA, 1989.

[173] R. A. Hummel. Representations based on zero-crossings in scale space. In *Proc. 1986 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 204–209, Miami Beach, FL, June 1986. IEEE Computer Society Press.

[174] T. Iijima. Basic theory of pattern observation. In *Papers of Technical Group on Automata and Automatic Control*. IECE, Japan, Dec. 1959. In Japanese.

[175] T. Iijima. Basic theory on normalization of pattern (in case of typical one-dimensional pattern). *Bulletin of the Electrotechnical Laboratory*, 26:368–388, 1962. In Japanese.

[176] T. Iijima. Observation theory of two-dimensional visual patterns. In *Papers of Technical Group on Automata and Automatic Control*. IECE, Japan, Oct. 1962. In Japanese.

[177] T. Iijima. Basic theory on normalization of two-dimensional visual pattern. *Studies on Information and Control (IECE, Japan)*, (1):15–22, 1963. Pattern Recognition Issue. In Japanese.

[178] T. Iijima. Theory of pattern recognition. *Electronics and Communications in Japan*, pages 123–134, Nov. 1963. In English.

[179] T. Iijima. Basic equation of figure and observational transformation. *Systems, Computers, Controls*, 2(4):70–77, 1971. In English.

[180] T. Iijima. Basic theory on normalization of figures. *Electronics and Communications in Japan*, 54–C(12):106–112, 1971. In English.

[181] T. Iijima. Basic theory on the construction of figure space. *Systems, Computers, Controls*, 2(5):51–57, 1971. In English.

[182] T. Iijima. A suppression kernel of a figure and its mathematical characteristics. *Systems, Computers, Controls*, 2(6):16–23, 1971. In English.

[183] T. Iijima. A system of fundamental functions in an abstract figure space. *Systems, Computers, Controls*, 2(6):96–103, 1971. In English.

[184] T. Iijima. Basic theory of feature extraction for figures. *Systems, Computers, Controls*, 3(1):32–39, 1972. In English.

[185] T. Iijima. Basic theory on the structural recognition of figures. *Systems, Computers, Controls*, 3(4):30–36, 1972. In English.

[186] T. Iijima. Theoretical studies on the figure identification by pattern matching. *Systems, Computers, Controls*, 3(4):37–44, 1972. In English.

[187] T. Iijima. *Pattern Recognition.* Corona-sha, 1973. In Japanese.

[188] T. Iijima. *Theory of Pattern Recognition*, volume 6 of *Series of Basic Information Technology.* Morishita Publishing, 1989. In Japanese.

[189] T. Iijima, H. Genchi, and K. Mori. A theoretical study of pattern identification by matching method. In *Proc. First USA–Japan Computer Conference*, pages 42–48, Tokyo, Japan, Oct. 1972. In English.

[190] T. Iijima, H. Genchi, and K. Mori. A theory of character recognition by pattern matching method. In *Proc. First International Joint Conference on Pattern Recognition*, pages 50–56, Washington, DC, Oct. 1973. In English.

[191] R. Illner and H. Neunzert. Relative entropy maximization and directed diffusion equations. *Mathematical Methods in the Applied Sciences*, 16:545–554, 1993.

[192] A. Imiya and R. Katsuta. Extraction of a structure feature from three-dimensional objects by scale-space analysis. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 345–348. Springer, Berlin, 1997.

[193] K. Ito and K. Kunisch. An active set strategy based on the augmented Lagrangian formulation for image restoration. *RAIRO Mathematical Models and Numerical Analysis*, 33:1–21, 1999.

[194] P. T. Jackway. Gradient watersheds in morphological scale-space. *IEEE Transactions on Image Processing*, 5:913–921, 1996.

[195] P. T. Jackway and M. Deriche. Scale-space properties of the multiscale morphological dilation–erosion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:38–51, 1996.

[196] M. Jägersand. Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach. In *Proc. Fifth International Conference on Computer Vision*, pages 195–202, Cambridge, MA, June 1995. IEEE Computer Society Press.

[197] B. Jähne. *Spatio-Temporal Image Processing*, volume 751 of *Lecture Notes in Computer Science.* Springer, Berlin, 1993.

[198] B. Jähne and H. Haussecker. Performance characteristics of low-level motion estimators in spatiotemporal images. In R. Klette, H. Stiehl, M. Viergever, and K. Vincken, editors, *Performance Characterization in Computer Vision*, Computational Imaging and Vision. Kluwer, Dordrecht, 2000.

[199] F. G. Jensen, J. Sporring, M. Nielsen, and P. G. Sørensen. Analysing the dynamics of target and spiral waves by image processing techniques. Technical Report DIKU-98/16, Dept. of Computer Science, University of Copenhagen, Denmark, 1998.

[200] M. Kass and A. Witkin. Analyzing oriented patterns. *Computer Vision, Graphics and Image Processing*, 37:362–385, 1987.

[201] J. Kačur, J. Nečas, J. Polák, and J. Souček. Convergence of a method for solving the magnetostatic field in nonlinear media. *Aplikace Matematiky*, 13:456–465, 1968.

[202] B. Kawohl and N. Kutev. Maximum and comparison principle for one-dimensional anisotropic diffusion. *Mathematische Annalen*, 311:107–123, 1998.

[203] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations.* SIAM, Philadelphia, 1995.

[204] S. Kichenassamy. The Perona–Malik paradox. *SIAM Journal on Applied Mathematics*, 57:1343–1372, 1997.

[205] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *Proc. Fifth International Conference on Computer Vision*, pages 810–815, Cambridge, MA, June 1995. IEEE Computer Society Press.

[206] B. B. Kimia and K. Siddiqi. Geometric heat equation and non-linear diffusion of shapes and images. *Computer Vision and Image Understanding*, 64:305–322, 1996.

[207] B. B. Kimia, A. Tannenbaum, and S. W. Zucker. Toward a computational theory of shape: an overview. In O. Faugeras, editor, *Computer Vision – ECCV '90*, volume 427 of *Lecture Notes in Computer Science*, pages 402–407. Springer, Berlin, 1990.

[208] R. Kimmel and A. M. Bruckstein. Tracking level sets by level sets: A method for solving the shape-from-shading problem. *Computer Vision and Image Understanding*, 62(2):47–58, July 1995.

[209] R. Kimmel, R. Malladi, and N. Sochen. Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129, Sept. 2000.

[210] R. Kimmel, N. Sochen, and R. Malladi. Images as embedding maps and minimal surfaces: movies, color, and volumetric medical images. In *Proc. 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 350–355, San Juan, Puerto Rico, June 1997. IEEE Computer Society Press.

[211] G. Kin and M. Sato. Scale space filtering on spherical pattern. In *Proc. Eleventh International Conference on Pattern Recognition*, volume C, pages 638–641, The Hague, The Netherlands, Aug. 1992.

[212] J. J. Koenderink. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22:773–791, 1975.

[213] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[214] J. J. Koenderink. Scale in perspective. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*, pages xv–xx. Kluwer, Dordrecht, 1997.

[215] P. Kornprobst, R. Deriche, and G. Aubert. Nonlinear operators in image restoration. In *Proc. 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 325–330, San Juan, Puerto Rico, June 1997. IEEE Computer Society Press.

[216] E. Kreyszig. *Differential Geometry.* University of Toronto Press, Toronto, 1959.

[217] H. Krim and D. H. Brooks. Feature-based segmentation of ECG signals. In *Proc. IEEE International Symposion on Time Frequency/Scale Analysis*, Paris, France, 1996.

[218] K. Krissian, G. Malandain, and N. Ayache. Directional anisotropic diffusion applied to segmentation of vessels in 3D images. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 345–348. Springer, Berlin, 1997.

[219] G. Kühne, J. Weickert, O. Schuster, and S. Richter. A tensor-driven active contour model for moving object segmentation. Submitted to Proc. 2001 IEEE International Conference on Image Processing (Thessaloniki, Greece, Oct. 2001), Jan. 2001.

[220] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[221] A. Kumar, A. R. Tannenbaum, and G. J. Balas. Optic flow: a curve evolution approach. *IEEE Transactions on Image Processing*, 5:598–610, 1996.

[222] C. Lamberti, M. Sitta, and F. Sgallari. Improvements to the anisotropic diffusion model for 2-D echo image processing. In *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 14, pages 1872–1873, 1992.

[223] J. Lasenby, W. J. Fitzgerald, C. J. L. Doran, and A. N. Lasenby. New geometric methods for computer vision. *International Journal of Computer Vision*, 36(3):191–213, 1998.

[224] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape infuence in geodesic active contours. In *Proc. 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 316–323, Hilton Head, SC, June 2000.

[225] S. Z. Li. *Markov Random Field Modeling in Computer Vision.* Springer, Tokyo, 1995.

[226] X. Li and T. Chen. Nonlinear diffusion with multiple edginess thresholds. *Pattern Recognition*, 27:1029–1037, 1994.

[227] Y. Li and F. Santosa. A computational algorithm for minimizing total variation in image enhancement. *IEEE Transactions on Image Processing*, 5:987–995, 1996.

[228] L. M. Lifshitz and S. M. Pizer. A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:529–540, 1990.

[229] T. Lindeberg. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:234–254, 1990.

[230] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer, Boston, 1994.

[231] T. Lindeberg. On the axiomatic formulations of linear scale-space. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*, pages 75–97. Kluwer, Dordrecht, 1997.

[232] T. Lindeberg and J. Garding. Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D brightness structure. *Image and Vision Computing*, 15:415–434, 1997.

[233] M. Lindenbaum, M. Fischer, and A. Bruckstein. On Gabor's contribution to image enhancement. *Pattern Recognition*, 27:1–8, 1994.

[234] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, Aug. 1981.

[235] D.-S. Luo, M. A. King, and S. Glick. Local geometry variable conductance diffusion for post-reconstruction filtering. *IEEE Transactions on Nuclear Science*, 41:2800–2806, 1994.

[236] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–75, Jan. 1996.

[237] F. Maes, D. Vandermeulen, P. Suetens, and G. Marchal. Computer-aided interactive object delineation using an intelligent paintbrush technique. In N. Ayache, editor, *Computer Vision, Virtual Reality and Robotics in Medicine*, volume 905 of *Lecture Notes in Computer Science*, pages 77–83. Springer, Berlin, 1995.

[238] R. Malladi, J. A. Sethian, and B. C. Vemuri. A topology independent shape modeling scheme. In B. Vemuri, editor, *Geometric Methods in Computer Vision*, volume 2031 of *Proceedings of SPIE*, pages 246–258. SPIE Press, Bellingham, 1993.

[239] R. Malladi, J. A. Sethian, and B. C. Vemuri. A fast level set based algorithm for topology-independent shape modeling. *Journal of Mathematical Imaging and Vision*, 6(2/3):269–289, 1996.

[240] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.

[241] G. I. Marchuk. Splitting and alternating direction methods. In P. G. Ciarlet and J.-L. Lions, editors, *Handbook of Numerical Analysis*, volume I, pages 197–462. North Holland, Amsterdam, 1990.

[242] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 207:187–217, 1980.

[243] M. D. McCool. Anisotropic diffusion for Monte Carlo noise reduction. *ACM Transactions on Graphics*, 18(2):171–194, 1999.

[244] E. Meijering. *Image Enhancement in Digital X-Ray Angiography*. PhD thesis, Image Sciences Institute, Utrecht University, The Netherlands, Oct. 2000.

[245] T. Meis and U. Marcowitz. *Numerische Behandlung partieller Differentialgleichungen*. Springer, Berlin, 1978.

[246] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7:703–719, 1998.

[247] Message Passing Interface Forum. MPI: A message-passing interface standard. Technical report, University of Tennessee, Knoxville, TN, 1995.

[248] F. Meyer and S. Beucher. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1:21–46, 1990.

[249] H. Meyer and M. Culman. Die Architektur der Spongiosa. *Archiv für Anatomie und Physiologie*, 47:615–628, 1867.

[250] A. R. Mitchell and D. F. Griffiths. *The Finite Difference Method in Partial Differential Equations*. Wiley, Chichester, 1980.

[251] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *International Journal of Computer Vision*, 19:29–55, 1996.

[252] C. R. Moloney and E. Dubois. Estimation of motion fields from image sequences with illumination variation. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 2425–2428, Toronto, Canada, May 1991.

[253] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation.* Birkhäuser, Basel, 1994.

[254] S. Mori, C. Y. Suen, and K. Yanamoto. Historical review of OCR research and development. *Proceedings of the IEEE*, 80:1029–1058, 1992.

[255] V. A. Morozov. *Methods for Solving Incorrectly Posed Problems.* Springer, New York, 1984.

[256] K. W. Morton and L. M. Mayers. *Numerical Solution of Partial Differential Equations.* Cambridge University Press, Cambridge, UK, 1994.

[257] D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 22–26, San Francisco, CA, June 1985. IEEE Computer Society Press.

[258] J. L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision.* MIT Press, Cambridge, MA, 1992.

[259] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:220–228, 1987.

[260] H.-H. Nagel. Constraints for the estimation of displacement vector fields from image sequences. In *Proc. Eighth International Joint Conference on Artificial Intelligence*, volume 2, pages 945–951, Karlsruhe, West Germany, August 1983.

[261] H.-H. Nagel. On the estimation of optical flow: relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.

[262] H.-H. Nagel. Image sequences—ten (octal) years—from phenomenology towards a theoretical foundation. *International Journal of Pattern Recognition and Artificial Intelligence*, 2:459–483, 1988.

[263] H.-H. Nagel. On a constraint equation for the estimation of displacement rates in image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):13–30, 1989.

[264] H.-H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In O. Faugeras, editor, *Computer Vision – ECCV '90*, volume 427 of *Lecture Notes in Computer Science*, pages 139–148. Springer, Berlin, 1990.

[265] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[266] L. Najman and M. Schmitt. Watershed of a continuous function. *Signal Processing*, 38:99–112, 1994.

[267] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:1163–1173, 1996.

[268] M. Z. Nashed and O. Scherzer. Least squares and bounded variation regularization with nondifferentiable functionals. *Numerical Functional Analysis and Optimization*, 19:873–901, 1998.

[269] S. Negahdaripour and C. M. Yu. A generalized brightness change model for computing optical flow. In *Proc. Fourth International Conference on Computer Vision*, pages 2–11, Berlin, May 1993. IEEE Computer Society Press.

[270] P. Nesi. Variational approach to optical flow estimation managing discontinuities. *Image and Vision Computing*, 11:419–439, 1993.

[271] E. Nicolet and M. Spühler. Charakterisierung von unscharfen blobförmigen 3-D Strukturen. Master's thesis, Image Science Division, ETH Zurich, Switzerland, 1996.

[272] M. Nielsen, L. Florack, and R. Deriche. Regularization, scale-space, and edge detection filters. In B. Buxton and R. Cipolla, editors, *Computer Vision – ECCV '96*, volume 1065 of *Lecture Notes in Computer Science*, pages 70–81. Springer, Berlin, 1996.

[273] M. Nielsen, L. Florack, and R. Deriche. Regularization, scale-space and edge detection filters. *Journal of Mathematical Imaging and Vision*, 7:291–307, 1997.

[274] M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors. *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*. Springer, Berlin, 1999.

[275] W. J. Niessen, A. M. López, W. J. van Enk, P. M. van Roermund, B. M. ter Haar Romeny, and M. A. Viergever. In vivo analysis of trabecular bone architecture. In J. S. Duncan and G. Gindi, editors, *Information Processing in Medical Imaging*, volume 1230 of *Lecture Notes in Computer Science*, pages 435–440. Springer, Berlin, 1997.

[276] W. J. Niessen, K. L. Vincken, J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Multiscale segmentation of three-dimensional MR brain images. *International Journal of Computer Vision*, 31(2/3):185–202, May 1999.

[277] W. J. Niessen, K. L. Vincken, J. Weickert, and M. A. Viergever. Nonlinear multiscale representations for image segmentation. *Computer Vision and Image Understanding*, 66(2):233–245, May 1997.

[278] W. J. Niessen, K. L. Vincken, J. Weickert, and M. A. Viergever. Three-dimensional MR brain segmentation. In *Proc. Sixth International Conference on Computer Vision*, pages 53–58, Bombay, India, Jan. 1998.

[279] H. Nishino. PIPS (pattern information processing system) project – background and outline. In *Proc. Fourth International Joint Conference on Pattern Recognition*, pages 1152–1161, Kyoto, Japan, Nov. 1978.

[280] M. Nitzberg and T. Shiota. Nonlinear image filtering with edge and corner enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:826–833, 1992.

[281] N. Nordström. Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection. *Image and Vision Computing*, 8:318–327, 1990.

[282] E. Norman. A discrete analogue of the Weierstrass transform. *Proceedings of the American Mathematical Society*, 11(596–604), 1960.

[283] O. F. Olsen. Multiscale watershed segmentation. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*, pages 191–200. Kluwer, Dordrecht, 1997.

[284] O. F. Olsen and J. Sporring. Segmenting by compression using linear scale-space and the watersheds. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 513–518. Springer, Berlin, 1999.

[285] H. Oltmans. A watershed algorithm. Laboratory of Technomathematics, University of Kaiserslautern, Germany, 1993. Project Work.

[286] P. J. Olver, G. Sapiro, and A. Tannenbaum. Classification and uniqueness of invariant geometric flows. *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, 319:339–344, 1994.

[287] A. Oomes and P. R. Snoeren. Structural information in scale-space. In P. Johansen, editor, *Proc. Copenhagen Workshop on Gaussian Scale-Space Theory*, pages 48–57, Copenhagen, Denmark, May 1996. Technical Report DIKU-96/19, Dept. of Computer Science, University of Copenhagen, Denmark.

[288] S. C. Orphanoudakis, G. Tziritas, and K. Haris. A hybrid algorithm for the segmentation of 2D and 3D medical images. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*, volume 3 of *Computational Imaging and Vision*, pages 385–386. Kluwer, Dordrecht, 1995.

[289] J. M. Ortega. *Matrix Theory*. Plenum Press, New York, 1987.

[290] S. Osher and L. I. Rudin. Feature-oriented image enhencement using shock filters. *SIAM Journal on Numerical Analysis*, 27:919–940, 1990.

[291] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[292] N. Otsu. Mathematical studies on feature extraction in pattern recognition. Technical Report 818 (PhD Thesis), Electrotechnical Laboratory, Tsukuba, Japan, July 1981. In Japanese.

[293] M. Otte and H.-H. Nagel. Estimation of optical flow based on higher-order spatiotemporal derivatives in interlaced and non-interlaced image sequences. *Artificial Intelligence*, 78:5–43, 1995.

[294] N. Paragios and R. Deriche. A PDE-based level-set approach for detection and tracking of moving objects. In *Proc. Sixth International Conference on Computer Vision*, pages 1139–1145, Bombay, India, Jan. 1998.

[295] E. J. Pauwels, L. J. Van Gool, P. Fiddelaers, and T. Moons. An extended class of scale-invariant and recursive scale space filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:691–701, 1995.

[296] E. Payot, R. Guillemaud, Y. Trousset, and F. Preteux. An adaptive and constrained model for 3D X-ray vascular reconstruction. In P. Grangeat and J.-L. Amans, editors, *Three-Dimensional Image Reconstruction in Radiation and Nuclear Medicine*, volume 4 of *Computational Imaging and Vision*, pages 47–57. Kluwer, Dordrecht, 1996.

[297] S.-C. Pei, C.-C. Tseng, and C.-Y. Lin. Wavelet transform and scale space filtering of fractal images. *IEEE Transactions on Image Processing*, 4(5):682–687, May 1995.

[298] S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:518–523, 1984.

[299] M. Pelillo and E. R. Hancock, editors. *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 1223 of *Lecture Notes in Computer Science*. Springer, Berlin, 1997.

[300] M. Pelillo and E. R. Hancock, editors. *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 1654 of *Lecture Notes in Computer Science*. Springer, Berlin, 1999.

[301] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. In *Proc. IEEE Computer Society Workshop on Computer Vision*, pages 16–22, Miami Beach, FL, Nov. 1987. IEEE Computer Society Press.

[302] P. Perona and J. Malik. A network for multiscale image segmentation. In *Proc. IEEE International Symposium on Circuits and Systems*, pages 2565–2568, Espoo, Finland, June 1988.

[303] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[304] R. Picard, C. Graczyk, S. Mann, J. Wachman, L. Picard, and L. Campbell. *Vistex.* Copyright 1995 Massachusetts Institute of Technology, Cambridge, MA, 1995. ftp:whitechapel.media.mit.edu.

[305] T. Preußer and M. Rumpf. An adaptive finite element method for large scale image processing. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 223–234. Springer, Berlin, 1999.

[306] T. Preußer and M. Rumpf. Anisotropic nonlinear diffusion in flow visualization. In *Proc. 1999 IEEE Visualization Conference*, pages 223–232, San Francisco, CA, Oct. 1999.

[307] C. B. Price, P. Wambacq, and A. Oosterlinck. Applications of reaction–diffusion equations to image processing. In *Proc. IEE Third International Conference on Image Processing and its Applications*, pages 49–53, Warwick, UK, July 1989.

[308] M. Proesmans, E. Pauwels, and L. Van Gool. Coupled geometry-driven diffusion equations for low-level vision. In B. M. ter Haar Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, volume 1 of *Computational Imaging and Vision*, pages 191–228. Kluwer, Dordrecht, 1994.

[309] M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck. Determination of optical flow and its discontinuities using non-linear diffusion. In J.-O. Eklundh, editor, *Computer Vision – ECCV '94*, volume 801 of *Lecture Notes in Computer Science*, pages 295–304. Springer, Berlin, 1994.

[310] E. Radmoser, O. Scherzer, and J. Weickert. Scale-space properties of regularization methods. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 211–222. Springer, Berlin, 1999.

[311] E. Radmoser, O. Scherzer, and J. Weickert. Scale-space properties of nonstationary iterative regularization methods. *Journal of Visual Communication and Image Representation*, 11(2):96–114, June 2000.

[312] I. Rambaux and P. Garçon. Nonlinear anisotropic diffusion filtering of 3D images. Département Génie Mathématique, INSA de Rouen, France, and Laboratory of Technomathematics, University of Kaiserslautern, Germany, 1994. Project Work.

[313] A. R. Rao and B. G. Schunck. Computing oriented texture fields. *CVGIP: Graphical Models and Image Processing*, 53:157–185, 1991.

[314] A. Rényi. On measures of entropy and information. In P. Turán, editor, *Selected Papers of Alfréd Rényi*, pages 565–579. Akadémiai Kiadó, Budapest, 1976. Originally: Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1961, pp. 547–561, University of California Press.

[315] A. Rényi. On the foundations of information theory. In P. Turán, editor, *Selected Papers of Alfréd Rényi*, pages 304–318. Akadémiai Kiadó, Budapest, 1976. Originally: Rev. Inst. Internat. Stat., 33, 1965, pp. 1–14.

[316] A. Rényi. Some fundamental questions of information theory. In P. Turán, editor, *Selected Papers of Alfréd Rényi*, pages 526–552. Akadémiai Kiadó, Budapest, 1976. Originally: MTA III. Oszt. Közl., 10, 1960, pp. 251–282.

[317] J. Rexilius. Anisotrope nichtlineare Diffusion für die Bildverarbeitung. Technical Report B-99-07, Institute of Mathematics and Computer Science, Medical University of Lübeck, Germany, Dec. 1999.

[318] T. Richardson and S. Mitter. Approximation, computation, and distortion in the variational formulation. In B. M. ter Haar Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, volume 1 of *Computational Imaging and Vision*, pages 169–190. Kluwer, Dordrecht, 1994.

[319] K. O. Riedel. Corner-preserving anisotropic diffusion and junction detection using the structure tensor. In W. Förstner, J. Buhmann, A. Faber, and P. Faber, editors, *Mustererkennung 1999*, pages 164–171. Springer, Berlin, 1999.

[320] J. H. Rieger. Generic evolution of edges on families of diffused greyvalue surfaces. *Journal of Mathematical Imaging and Vision*, 5:207–217, 1995.

[321] L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In B. Buxton and R. Cipolla, editors, *Computer Vision – ECCV '96*, volume 1064 of *Lecture Notes in Computer Science*, pages 439–451. Springer, Berlin, 1996.

[322] L. Robert, R. Deriche, and O. Faugeras. Dense depth map reconstruction using multiscale regularization. In V. Srinivasan, S. H. Ong, and Y. H. Ang, editors, *Proc. Second International Singapore Conference on Image Processing*, pages 123–127, Singapore, Sept. 1992.

[323] E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM Journal on Numerical Analysis*, 29:867–884, 1992.

[324] Y. Rubner and C. Tomasi. Coalescing texture descriptors. In *Proc. ARPA Image Understanding Workshop*, pages 927–936, Palm Springs, CA, Feb. 1996.

[325] L. I. Rudin. *Images, Numerical Analysis of Singularities and Shock Filters.* PhD thesis, California Institute of Technology, Pasadena, CA, 1987.

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

[327] P. Sahoo, C. Wilkins, and J. Yeager. Threshold selection using Renyi's entropy. *Pattern Recognition*, 30:71–84, 1997.

[328] A. H. Salden. *Dynamic Scale-Space Paradigms*. PhD thesis, Faculty of Medicine, Utrecht University, The Netherlands, Nov. 1996.

[329] A. H. Salden, B. M. ter Haar Romeny, and M. A. Viergever. Linear scale-space theory from physical principles. *Journal of Mathematical Imaging and Vision*, 9:103–139, 1998.

[330] G. Sapiro. *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press, Cambridge, UK, 2001.

[331] G. Sapiro and V. Caselles. Histogram modification via differential equations. *Journal of Differential Equations*, 135:238–268, 1997.

[332] G. Sapiro and D. L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Transactions on Image Processing*, 5(11):1582–1586, 1996.

[333] G. Sapiro and A. Tannenbaum. Affine invariant scale-space. *International Journal of Computer Vision*, 11:25–44, 1993.

[334] A. Sarti, K. Mikula, and F. Sgallari. Nonlinear multiscale analysis of 3D echocardiographic sequences. *IEEE Transactions on Medical Imaging*, 18(6):453–466, 1999.

[335] M. Sato and T. Wada. A hierarchical representation of random waveforms by scale-space filtering. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 273–276, Dallas, TX, Apr. 1987.

[336] M. Sato, T. Wada, and H. Kawarada. A morphological study on structure line. In *Proc. Ninth International Conference on Pattern Recognition*, pages 559–562, Rome, Italy, Nov. 1988.

[337] H. Scharr, B. Jähne, S. Böckle, J. Kazenwandel, T. Kunzelmann, and C. Schulte. Flame front analysis in turbulent combustion. In G. Sommer, N. Krüger, and C. Perwass, editors, *Mustererkennung 2000*, pages 325–332. Springer, Berlin, 2000.

[338] H. Scharr and J. Weickert. An anisotropic diffusion algorithm with optimized rotation invariance. In G. Sommer, N. Krüger, and C. Perwass, editors, *Mustererkennung 2000*, pages 460–467. Springer, Berlin, 2000.

[339] O. Scherzer. Stable evaluation of differential operators and linear and nonlinear multi-scale filtering. *Electronic Journal of Differential Equations*, (15):1–12, 1997. http://ejde.math.unt.edu.

[340] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60:1–27, 1998.

[341] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63, Feb. 2000.

[342] C. Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *International Journal of Computer Vision*, 6(1):25–38, 1991.

[343] C. Schnörr. On functionals with greyvalue-controlled smoothness terms for determining optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:1074–1079, 1993.

[344] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *Proc. Twelfth International Conference on Pattern Recognition*, volume A, pages 661–663, Jerusalem, Israel, Oct. 1994. IEEE Computer Society Press.

[345] C. Schnörr. Unique reconstruction of piecewise smooth images by minimizing strictly convex non-quadratic functionals. *Journal of Mathematical Imaging and Vision*, 4:189–198, 1994.

[346] C. Schnörr. Convex variational segmentation of multi-channel images. In M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, and J.-M. Morel, editors, *ICAOS '96: Images, Wavelets and PDEs*, volume 219 of *Lecture Notes in Control and Information Sciences*, pages 201–207. Springer, London, 1996.

[347] C. Schnörr. A study of a convex variational diffusion approach for image segmentation and feature extraction. *Journal of Mathematical Imaging and Vision*, 8(3):271–292, 1998.

[348] C. Schnörr. Variational methods for adaptive image smoothing and segmentation. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook on Computer Vision and Applications. Vol. 2: Signal Processing and Pattern Recognition*, pages 451–484. Academic Press, San Diego, 1999.

[349] C. Schnörr and R. Sprengel. A nonlinear regularization approach to early vision. *Biological Cybernetics*, 72:141–149, 1994.

[350] C. Schnörr, R. Sprengel, and B. Neumann. A variational approach to the design of early vision algorithms. *Computing Suppl.*, 11:149–165, 1996.

[351] C. Schnörr and J. Weickert. Variational motion computation: theoretical framework, problems and perspectives. In G. Sommer, N. Krüger, and C. Perwass, editors, *Mustererkennung 2000*, pages 476–487. Springer, Berlin, 2000.

[352] H. R. Schwarz. *Numerische Mathematik*. Teubner, Stuttgart, 1988.

[353] J. A. Sethian. *Level Set Methods*. Cambridge University Press, Cambridge, UK, 1996.

[354] J. Shah. A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. In *Proc. 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 34–40, New York, NY, June 1993. IEEE Computer Society Press.

[355] J. Shah. Curve evolution and segmentation functionals: application to color images. In *Proc. 1996 IEEE International Conference on Image Processing*, volume 1, pages 461–464, Lausanne, Switzerland, Sept. 1996.

[356] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1949.

[357] K. Siddiqi, B. B. Kimia, and C.-W. Shu. Geometric shock-capturing ENO schemes for subpixel interpolation, computation and curve evolution. *Graphical Models and Image Processing*, 59:278–301, 1997.

[358] J. Sijbers, P. Scheunders, M. Verhoye, A. Van der Linden, D. Van Dyck, and E. Raman. Watershed-based segmentation of 3D MR data for volume quantization. *Magnetic Resonance Imaging*, 15:679–688, 1997.

[359] J. S. Simonoff. *Smoothing Methods in Statistics*. Springer Series in Statistics. Springer, New York, 1996.

[360] B. Smith, P. Bjørstad, and W. Gropp. *Domain Decomposition: Parallel Multilevel Methods for the Solution of Elliptic Partial Differential Equations*. Cambridge University Press, Cambridge, UK, 1996.

[361] M. A. Snyder. On the mathematical foundations of smoothness constraints for the determination of optical flow and for surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:1105–1114, 1991.

[362] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Transactions on Image Processing*, 7(3):310–318, Mar. 1998.

[363] A. F. Solé, A. López, C. Cañero, P. Radeva, and J. Saludes. Crease enhancement diffusion. In M. I. Torres and A. Sanfeliu, editors, *Proc. Eighth National Symposium on Pattern Recognition and Image Analysis*, volume 1, pages 279–286, Bilbao, Spain, May 1999.

[364] G. Sommer. The global algebraic frame of the perception–action cycle. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook on Computer Vision and Applications, Vol. 3: Systems and Applications*, pages 221–264. Academic Press, San Diego, 1999.

[365] G. Sommer, editor. *Geometric Computing with Clifford Algebra: Theory and Applications in Vision and Robotics*. Springer, Berlin, 2001. To appear.

[366] G. Sommer and J. J. Koenderink, editors. *Algebraic Frames for the Perception–Action Cycle*, volume 1315 of *Lecture Notes in Computer Science*. Springer, Berlin, 1997.

[367] G. Sommer and Y. Y. Zeevi, editors. *Algebraic Frames for the Perception–Action Cycle*, volume 1888 of *Lecture Notes in Computer Science*. Springer, Berlin, 2000.

[368] J. Sporring. The entropy of scale-space. In *Proc. 13th International Conference on Pattern Recognition*, volume A, pages 900–904, Vienna, Austria, Aug. 1996.

[369] J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors. *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*. Kluwer, Dordrecht, 1997.

[370] J. Sporring, M. Nielsen, O. F. Olsen, and J. Weickert. Smoothing images creates corners. *Image and Vision Computing*, 18(3):261–266, Feb. 2000.

[371] J. Sporring, M. Nielsen, J. Weickert, and O. F. Olsen. A note on differential corner measures. In *Proc. 14th International Conference on Pattern Recognition*, volume 1, pages 652–654, Brisbane, Australia, Aug. 1998. IEEE Computer Society Press.

[372] J. Sporring and J. Weickert. On generalized entropies and scale-space. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 53–64. Springer, Berlin, 1997.

[373] J. Sporring and J. Weickert. Information measures in scale-spaces. *IEEE Transactions on Information Theory*, 45(3):1051–1058, Apr. 1999.

[374] M. Sramek and T. Wrbka. Watershed based image segmentation – an effective tool for detecting landscape structure. In E. Wenger and L. I. Dimitrov, editors, *Sixth International Workshop on Digital Image Processing and Computer Graphics: Applications in Humanities and Natural Sciences*, volume 3346 of *Proceedings of SPIE*, pages 227–235. SPIE Press, Bellingham, 1997.

[375] J. L. Stansfield. Conclusions from the commodity expert project. Technical Report 671 (A.I. Memo), Artificial Intelligence Lab., Massachusetts Inst. of Technolgy, Cambridge, MA, U.S.A., 1980.

[376] R. L. Stevenson, B. E. Schmitz, and E. J. Delp. Discontinuity preserving regularization of inverse visual problems. *IEEE Transactions on Systems, Man and Cybernetics*, 24:455–469, 1994.

[377] C. Stiller and J. Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, 16:70–91, 1999.

[378] D. M. Strong and T. F. Chan. Relation of regularization parameter and scale in total variation based image denoising. Technical Report CAM-96-7, Dept. of Mathematics, Univ. of California at Los Angeles, CA, U.S.A., 1996.

[379] E. P. Stuijfzand, M. D. Abràmoff, K. J. Zuiderveld, L. M. P. Ramos, J. Weickert, M. P. Mourits, F. W. Zonneveld, and W. P. T. H. Mali. Fast kinematic MR imaging of the eye and orbit. *RSNA Electronic Journal*, 1, 1997. http://ej.rsna.org/.

[380] M. Tanaka, T. Watanabe, and T. Mishima. Tsallis entropy in scale-spaces. In L. J. Latecki, R. A. Melter, D. M. Mount, and A. Y. Wu, editors, *Vision Geometry VIII*, volume 3811 of *Proceedings of SPIE*. SPIE Press, Bellingham, 1999.

[381] B. ter Haar Romeny, L. Florack, and J. Koenderink, editors. *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*. Springer, Berlin, 1997.

[382] B. M. ter Haar Romeny, editor. *Geometry-Driven Diffusion in Computer Vision*, volume 1 of *Computational Imaging and Vision*. Kluwer, Dordrecht, 1994.

[383] B. M. ter Haar Romeny and L. M. J. Florack. Front-end vision, a multiscale geometry engine. In S.-W. Lee, H. H. Bülthoff, and T. Poggio, editors, *Biologically Motivated Computer Vision*, volume 1811 of *Lecture Notes in Computer Science*. Springer, Berlin, 2000.

[384] B. M. ter Haar Romeny, W. J. Niessen, J. Weickert, P. van Roermund, W. J. van Enk, A. Lopez, and R. Maas. Orientation detection of trabecular bone. In *Progress in Biophysics and Molecular Biology (Proc. Twelfth International Biophysics Congress)*, volume 65, Amsterdam, The Netherlands, Aug. 1996. Poster P-H5-43.

[385] B. M. ter Haar Romeny, K. J. Zuiderveld, P. F. G. M. van Waes, T. van Walsum, R. van der Weijden, J. Weickert, R. Stokking, O. Wink, S. Kalitzin, T. Maintz, F. Zonneveld, and M. A. Viergever. Advances in three-dimensional diagnostic radiology. *Journal of Anatomy*, 193(3):363–371, 1998.

[386] A. Thompson and D. M. Titterington. A study of choosing the smoothing parameters in image reconstruction by regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:326–339, 1991.

[387] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill–Posed Problems*. Wiley, Washington, DC, 1977.

[388] V. Torre and T. A. Poggio. On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:148–163, 1986.

[389] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60:79–87, 1988.

[390] R. van den Boomgaard. The morphological equivalent of the Gauss convolution. *Nieuw Archief Voor Wiskunde*, 10(3):219–236, Nov. 1992.

[391] V. van Gogh. Selfportrait, 1889. Saint-Rémy (Paris, Museé d'Orsay).

[392] V. van Gogh. Starry night, 1889. Saint-Rémy (New York, The Museum of Modern Art).

[393] V. van Gogh. Road with cypress and star, 1890. Auvers-sur-Oise (Otterlo, Rijksmuseum Kröller–Müller).

[394] P. J. M. van Laarhoven and E. H. L. Aarts. *Simulated Annealing: Theory and Applications.* Reidel, Dordrecht, 1988.

[395] R. A. Varga. *Matrix Iterative Analysis.* Prentice Hall, Englewood Cliffs, 1962.

[396] J. L. Véhel and P. Mignot. Multifractal segmentation of images. *Fractals*, 2(3):379–382, June 1994.

[397] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:583–589, 1991.

[398] K. L. Vincken, A. S. E. Koster, and M. A. Viergever. Probabilistic multiscale image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:109–120, 1997.

[399] C. R. Vogel and M. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Transactions on Image Processing*, 7:813–824, 1998.

[400] T. Wada, Y. H. Gu, and M. Sato. Scale-space filtering for periodic waveforms. *Systems and Computers in Japan*, 22:45–54, 1991.

[401] T. Wada and M. Sato. Scale-space tree and its hierarchy. In *Proc. Tenth International Conference on Pattern Recognition*, volume 3, pages 103–108, Atlantic City, NJ, June 1990. IEEE Computer Society Press.

[402] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. *International Journal of Computer Vision*, 14:67–81, 1995.

[403] J. Weickert. Zwischenbericht zum Projekt "Nichtlineare Diffusionsfilter". In *Bericht über die wissenschaftliche Tätigkeit Januar 1991 – Dezember 1991*, pages 133–142. Center for Applied Mathematics, Darmstadt and Kaiserslautern, 1992.

[404] J. Weickert. Anisotropic diffusion filters for image processing based quality control. In A. Fasano and M. Primicerio, editors, *Proc. Seventh European Conference on Mathematics in Industry*, pages 355–362. Teubner, Stuttgart, 1994.

[405] J. Weickert. Scale-space properties of nonlinear diffusion filtering with a diffusion tensor. Technical Report 110, Laboratory of Technomathematics, University of Kaiserslautern, Germany, Oct. 1994.

[406] J. Weickert. Multiscale texture enhancement. In V. Hlaváč and R. Šára, editors, *Computer Analysis of Images and Patterns*, volume 970 of *Lecture Notes in Computer Science*, pages 230–237. Springer, Berlin, 1995.

[407] J. Weickert. Nonlinear diffusion scale-spaces: From the continuous to the discrete setting. In M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, and J.-M. Morel, editors, *ICAOS '96: Images, Wavelets and PDEs*, volume 219 of *Lecture Notes in Control and Information Sciences*, pages 111–118. Springer, London, 1996.

[408] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing Suppl.*, 11:221–236, 1996.

[409] J. Weickert. Coherence-enhancing diffusion of colour images. In A. Sanfeliu, J. J. Villanueva, and J. Vitrià, editors, *Proc. Seventh National Symposium on Pattern Recognition and Image Analysis*, volume 1, pages 239–244, Barcelona, Spain, Apr. 1997.

[410] J. Weickert. Recursive separable schemes for nonlinear diffusion filters. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 133–142. Springer, Berlin, 1997.

[411] J. Weickert. A review of nonlinear diffusion filtering. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 3–28. Springer, Berlin, 1997.

[412] J. Weickert. *Anisotropic Diffusion in Image Processing.* Teubner, Stuttgart, 1998.

[413] J. Weickert. Fast segmentation methods based on partial differential equations and the watershed transformation. In P. Levi, R.-J. Ahlers, F. May, and M. Schanz, editors, *Mustererkennung 1998*, pages 93–100. Springer, Berlin, 1998.

[414] J. Weickert. On discontinuity-preserving optic flow. In S. Orphanoudakis, P. Trahanias, J. Crowley, and N. Katevas, editors, *Proc. Computer Vision and Mobile Robotics Workshop*, pages 115–122, Santorini, Greece, Sept. 1998.

[415] J. Weickert. Coherence-enhancing diffusion filtering. *International Journal of Computer Vision*, 31(2/3):111–127, Apr. 1999.

[416] J. Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17(3–4):199–210, Mar. 1999.

[417] J. Weickert. Applications of nonlinear diffusion in image processing and computer vision. Technical Report 18/2000, Computer Science Series, University of Mannheim, Germany, Nov. 2000.

[418] J. Weickert and B. Benhamouda. A semidiscrete nonlinear scale-space theory and its relation to the Perona–Malik paradox. In F. Solina, W. G. Kropatsch, R. Klette, and R. Bajcsy, editors, *Advances in Computer Vision*, pages 1–10. Springer, Wien, 1997.

[419] J. Weickert, J. Heers, C. Schnörr, K. J. Zuiderveld, O. Scherzer, and H. S. Stiehl. Fast parallel algorithms for a broad class of nonlinear variational diffusion approaches. *Real-Time Imaging*, 2001. In press.

[420] J. Weickert, S. Ishikawa, and A. Imiya. On the history of Gaussian scale-space axiomatics. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*, pages 45–59. Kluwer, Dordrecht, 1997.

[421] J. Weickert, S. Ishikawa, and A. Imiya. Linear scale-space has first been proposed in Japan. *Journal of Mathematical Imaging and Vision*, 10(3):237–252, May 1999.

[422] J. Weickert and H. Scharr. A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. Technical Report 4/2000, Computer Science Series, University of Mannheim, Germany, Feb. 2000. To appear in Journal of Visual Communication and Image Representation.

[423] J. Weickert and C. Schnörr. Räumlich–zeitliche Berechnung des optischen Flusses mit nichtlinearen flußabhängigen Glattheitstermen. In W. Förstner, J. Buhmann, A. Faber, and P. Faber, editors, *Mustererkennung 1999*, pages 317–324. Springer, Berlin, 1999.

[424] J. Weickert and C. Schnörr. PDE-based preprocessing of medical images. *Künstliche Intelligenz*, (3):5–10, July 2000.

[425] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. Technical Report 13/2000, Computer Science Series, University of Mannheim, Germany, June 2000.

[426] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. Technical Report 15/2000, Computer Science Series, University of Mannheim, Germany, July 2000. To appear in Journal of Mathematical Imaging and Vision.

[427] J. Weickert, B. M. ter Haar Romeny, A. Lopez, and W. J. van Enk. Orientation analysis by coherence-enhancing diffusion. In *Proc. 1997 Real World Computing Symposium*, pages 96–103, Tokyo, Japan, Jan. 1997.

[428] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Conservative image transformations with restoration and scale-space properties. In *Proc. 1996 IEEE International Conference on Image Processing*, volume 1, pages 465–468, Lausanne, Switzerland, Sept. 1996.

[429] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, Mar. 1998.

[430] J. Weickert, K. J. Zuiderveld, B. M. ter Haar Romeny, and W. J. Niessen. Parallel implementations of AOS schemes: A fast way of nonlinear diffusion filtering. In *Proc. 1997 IEEE International Conference on Image Processing*, volume 3, pages 396–399, Santa Barbara, CA, Oct. 1997.

[431] J. Weidmann. *Lineare Operatoren in Hilberträumen*. Teubner, Stuttgart, 1976.

[432] R. Whitaker and G. Gerig. Vector-valued diffusion. In B. M. ter Haar Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, volume 1 of *Computational Imaging and Vision*, pages 93–134. Kluwer, Dordrecht, 1994.

[433] R. T. Whitaker and S. M. Pizer. A multi-scale approach to nonuniform diffusion. *CVGIP: Image Understanding*, 57:99–110, 1993.

[434] N. Wiener. *Cybernetics*. Wiley, New York, 1948.

[435] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, volume 27 of *Applications of Mathematics*. Springer, Berlin, 1995.

[436] A. P. Witkin. Scale-space filtering. In *Proc. Eighth International Joint Conference on Artificial Intelligence*, volume 2, pages 945–951, Karlsruhe, West Germany, August 1983.

[437] J. Wloka. *Partielle Differentialgleichungen*. Teubner, Stuttgart, 1982.

[438] J. Wolff. Über die innere Architektur der Knochen. *Virchows Archiv*, 50:389–453, 1870.

[439] N. N. Yanenko. *The Method of Fractional Steps: the Solution of Problems of Mathematical Physics in Several Variables*. Springer, New York, 1971.

[440] G. Z. Yang, P. Burger, D. N. Firmin, and S. R. Underwood. Structure adaptive anisotropic filtering. *Image and Vision Computing*, 14:135–145, 1996.

[441] Y.-L. You and M. Kaveh. Anisotropic blind image restoration. In *Proc. 1996 IEEE International Conference on Image Processing*, volume 2, pages 461–464, Lausanne, Switzerland, Sept. 1996.

[442] A. L. Yuille and T. A. Poggio. Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:15–25, 1986.

[443] E. Zeidler. *Nonlinear Functional Analysis and its Applications*, volume IIb. Springer, Berlin, 1990.

[444] D. Zhao and B. Li. A new implementation of discrete multiscale filtering. In *Proc. 1996 IEEE International Conference on Image Processing*, volume 1, pages 383–386, Lausanne, Switzerland, Sept. 1996.

[445] N.-Y. Zhao. *Feature Extraction of Images by Stable Gaze Tree.* PhD thesis, Dept. of Computer Science, Tokyo Inst. of Technology, Japan, 1985. In Japanese.

[446] N.-Y. Zhao and T. Iijima. A theory of feature extraction by the tree of stable view-points. *IECE Transactions on Information and Systems*, J68-D:1125–1132, 1985. In Japanese.

[447] N.-Y. Zhao and T. Iijima. Theory on the method of determination of view-point and field of vision during observation and measurement of figure. *IECE Transactions on Information and Systems*, J68-D:508–514, 1985. In Japanese.

[448] W. Ziemer. *Weakly Differentiable Functions.* Springer, New York, 1989.

Publications

Books

1. J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner, Stuttgart, 1998. ISBN 3-519-02606-6.

2. M. Nielsen, P. Johansen, O.F. Olsen, J. Weickert (Eds.), *Scale-Space Theories in Computer Vision*, Lecture Notes in Computer Science, Vol. 1682, Springer, Berlin, 1999. ISBN 3-540-66498-X.

Book Chapters

3. J. Weickert, Design of nonlinear diffusion filters, in B. Jähne, H. Haußecker (Eds.), *Computer Vision and Applications*, Academic Press, San Diego, 439–458, 2000.

4. J. Weickert, Nonlinear diffusion filtering, in B. Jähne, H. Haußecker, P. Geißler (Eds.), *Handbook of Computer Vision and Applications, Vol. 2: Signal Processing and Pattern Recognition*, Academic Press, San Diego, 423–450, 1999.

5. J. Weickert, S. Ishikawa, A. Imiya, On the history of Gaussian scale-space axiomatics, in J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.), *Gaussian Scale-Space Theory*, Kluwer, Dordrecht, 45–59, 1997.

6. J. Weickert, Nonlinear diffusion scale-spaces, in J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.), *Gaussian Scale-Space Theory*, Kluwer, Dordrecht, 221–234, 1997.

Journal Articles

7. J. Weickert, C. Schnörr, Variational optic flow computation with a spatio-temporal smoothness constraint, *Journal of Mathematical Imaging and Vision*, in press.

8. J. Weickert, Efficient image segmentation using partial differential equations and morphology, *Pattern Recognition*, in press.

9. L. Alvarez, R. Deriche, J. Sánchez, J. Weickert, Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach, *Journal of Visual Communication and Image Representation*, in press.

10. J. Weickert, H. Scharr, A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance, *Journal of Visual Communication and Image Representation*, in press.

11. L. Alvarez, J. Weickert, J. Sánchez, Reliable estimation of dense optical flow fields with large displacements, *International Journal of Computer Vision*, Vol. 39, 41–56, 2000.

12. J. Weickert, C. Schnörr, PDE-based preprocessing of medical images, *Künstliche Intelligenz*, No. 3, 5–10, 2000.

13. E. Radmoser, O. Scherzer, J. Weickert, Scale-space properties of nonstationary iterative regularization methods, *Journal of Visual Communication and Image Representation*, Vol. 8, 96–114, 2000. Invited paper.

14. J. Weickert, J. Heers, C. Schnörr, K.J. Zuiderveld, O. Scherzer, H.S. Stiehl, Fast parallel algorithms for a broad class of nonlinear variational diffusion approaches, *Real-Time Imaging*, in press.

15. O. Scherzer, J. Weickert, Relations between regularization and diffusion filtering, *Journal of Mathematical Imaging and Vision*, Vol. 12, 43–63, 2000.

16. J. Sporring, M. Nielsen, O.F. Olsen, J. Weickert, Smoothing images creates corners, *Image and Vision Computing*, Vol. 18, 261–266, 2000.

17. J. Weickert, Coherence-enhancing diffusion filtering, *International Journal of Computer Vision*, Vol. 31, 111–127, 1999.

18. W.J. Niessen, K.L. Vincken, J. Weickert, B.M. ter Haar Romeny, M.A. Viergever, Multiscale segmentation of three-dimensional MR brain images, *International Journal of Computer Vision*, Vol. 31, 185–202, 1999.

19. J. Weickert, S. Ishikawa, A. Imiya, Linear scale-space has first been proposed in Japan, *Journal of Mathematical Imaging and Vision*, Vol. 10, 237–252, 1999.

20. J. Sporring, J. Weickert, Information measures in scale-spaces, *IEEE Transactions on Information Theory*, Vol. 45, 1051–1058, 1999.

21. J. Weickert, Coherence-enhancing diffusion of colour images, *Image and Vision Computing*, Vol. 17, 199–210, 1999.

22. J. Weickert, A real-time algorithm for assessing inhomogeneities in fabrics, *Real-Time Imaging*, Vol. 5, 15–22, 1999.

23. B.M. ter Haar Romeny, K.J. Zuiderveld, P.F.G.M. van Waes, T. van Walsum, R. van der Weijden, J. Weickert, R. Stokking, O. Wink, S. Kalitzin, T. Maintz, F. Zonneveld, M.A. Viergever, Advances in three-dimensional diagnostic radiology, *Journal of Anatomy*, Vol. 193, 363–371, 1998.

24. J. Weickert, B.M. ter Haar Romeny, M.A. Viergever, Efficient and reliable schemes for nonlinear diffusion filtering, *IEEE Transactions on Image Processing*, Vol. 7, 398–410, 1998.

25. E.P. Stuijfzand, M.D. Abràmoff, K.J. Zuiderveld, L.M.P. Ramos, J. Weickert, M.P. Mourits, F.W. Zonneveld, W.P.T.H. Mali, Fast kinematic MR imaging of the eye and orbit, *RSNA Electronic Journal*, Vol. 1, 1997.

26. W.J. Niessen, K.L. Vincken, J. Weickert, M.A. Viergever, Nonlinear multiscale representations for image segmentation, *Computer Vision and Image Understanding*, Vol. 66, 233–245, 1997.

27. J. Weickert, Foundations and applications of nonlinear anisotropic diffusion filtering, *Zeitschrift für Angewandte Mathematik und Mechanik*, Vol. 76, Suppl. 1, 283–286, 1996.

28. J. Weickert, Theoretical foundations of anisotropic diffusion in image processing, *Computing*, Suppl. 11, 221–236, 1996.

29. J. Weickert, A mathematical model for diffusion and exchange phenomena in Ultra napkins, *Mathematical Methods in the Applied Sciences*, Vol. 16, 759–777, 1993.

Conference Articles

30. L. Alvarez, R. Deriche, J. Sánchez, J. Weickert, Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach, *Proc. IAPR Workshop on Machine Vision* (Tokyo, Nov. 28–30, 2000), 423–427, 2000.

31. D. Cremers, C. Schnörr, J. Weickert, C. Schellewald, Learning of translation invariant shape knowledge for steering diffusion snakes, in G. Baratoff, H. Neumann (Eds.), *Dynamische Perzeption*, Akademische Verlagsgesellschaft Aka, 117–122, 2000.

32. H. Scharr, J. Weickert, An anisotropic diffusion algorithm with optimized rotation invariance, in G. Sommer, N. Krüger, C. Perlwass (Eds.), *Mustererkennung 2000*, Springer, Berlin, 460–467, 2000.

33. C. Schnörr, J. Weickert, Variational image motion computation: theoretical framework, problems and perspectives, in G. Sommer, N. Krüger, C. Perlwass (Eds.), *Mustererkennung 2000*, Springer, Berlin, 476–487, 2000. Invited paper.

34. D. Cremers, C. Schnörr, J. Weickert, C. Schellewald, Diffusion-snakes using statistical shape knowledge, in G. Sommer, Y.Y. Zeevi (Eds.), *Algebraic Frames for the Perception–Action Cycle*, Lecture Notes in Computer Science, Springer, Vol. 1888, Berlin, 164–174, 2000.

35. E. Radmoser, O. Scherzer, J. Weickert, Scale-space properties of regularization methods, in M. Nielsen, P. Johansen, O.F. Olsen, J. Weickert (Eds.), *Scale-Space Theories in Computer Vision*, Lecture Notes in Computer Science, Vol. 1682, Springer, Berlin, 211–222, 1999.

36. L. Alvarez, J. Weickert, J. Sánchez, A scale-space approach to nonlocal optical flow calculations, in M. Nielsen, P. Johansen, O.F. Olsen, J. Weickert (Eds.), *Scale-Space Theories in Computer Vision*, Lecture Notes in Computer Science, Vol. 1682, Springer, Berlin, 235–246, 1999.

37. J. Weickert, C. Schnörr, Räumlich-zeitliche Berechnung des optischen Flusses mit nichtlinearen flußabhängigen Glattheitstermen, in W. Förstner, J.M. Buhmann, A. Faber, P. Faber (Eds.), *Mustererkennung 1999*, Springer, Berlin, 317–324, 1999.

38. J. Weickert, Fast segmentation methods based on partial differential equations and the watershed transformation, in P. Levi, R.-J. Ahlers, F. May, M. Schanz (Eds.), *Mustererkennung 1998*, Springer, Berlin, 93–100, 1998.

39. J. Weickert, On discontinuity-preserving optic flow, in S. Orphanoudakis, P. Trahanias, J. Crowley, N. Katevas (Eds.), *Proc. Computer Vision and Mobile Robotics Workshop* (CVMR '98, Santorini, Sept. 17–18, 1998), 115–122, 1998.

40. J. Sporring, M. Nielsen, J. Weickert, O.F. Olsen, A note on differential corner measures, *Proc. 14th Int. Conf. Pattern Recognition* (ICPR 14, Brisbane, Aug. 17–20, 1998), IEEE Computer Society Press, Los Alamitos, Vol. 1, 652–654, 1998.

41. W.J. Niessen, K.L. Vincken, J. Weickert, M.A. Viergever, Three-dimensional MR brain segmentation, *Proc. Sixth Int. Conf. on Computer Vision* (ICCV '98, Bombay, Jan. 4–7, 1998), 53–58, 1998.

42. J. Weickert, B. Benhamouda, A semidiscrete nonlinear scale-space theory and its relation to the Perona–Malik paradox, in F. Solina, W.G. Kropatsch, R. Klette, R. Bajcsy (Eds.), *Advances in Computer Vision*, Springer, Wien, 1–10, 1997.

43. J. Weickert, K.J. Zuiderveld, B.M. ter Haar Romeny, W.J. Niessen, Parallel implementations of AOS schemes: a fast way of nonlinear diffusion filtering, *Proc. 1997 IEEE International Conference on Image Processing* (ICIP–97, Santa Barbara, Oct. 26–29, 1997), Vol. 3, 396–399, 1997.

44. J. Weickert, A review of nonlinear diffusion filtering, in B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), *Scale-Space Theory in Computer Vision*, Lecture Notes in Computer Science, Vol. 1252, Springer, Berlin, 3–28, 1997. Invited paper.

45. J. Sporring, J. Weickert, On generalized entropies and scale-space, in B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), *Scale-Space Theory in Computer Vision*, Lecture Notes in Computer Science, Vol. 1252, Springer, Berlin, 53–64, 1997.

46. J. Weickert, Recursive separable schemes for nonlinear diffusion filters, in B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), *Scale-Space Theory in Computer Vision*, Lecture Notes in Computer Science, Vol. 1252, Springer, Berlin, 260–271, 1997.

47. J. Weickert, Coherence-enhancing diffusion of colour images, in A. Sanfeliu, J.J. Villanueva, J. Vitrià (Eds.), *Proc. VII National Symposium on Pattern Recognition and Image Analysis* (VII NSPRIA, Barcelona, April 21–25, 1997), Vol. 1, 239–244, 1997.

48. J. Weickert, B.M. ter Haar Romeny, A. Lopez, W.J. van Enk, Orientation analysis by coherence-enhancing diffusion, *Proc. Symposium on Real World Computing* (RWC '97, Tokyo, Jan. 29–31, 1997), 96–103, 1997.

49. H. Neunzert, B. Claus, K. Rjasanowa, R. Rösch, J. Weickert, Mathematische Werkzeuge in der Bildverarbeitung zur Qualitätsbestimmung von Oberflächen, in K.-H. Hoffmann, W. Jäger, T. Lohmann, H. Schunck (Eds.), *Mathematik – Schlüsseltechnologie für die Zukunft*, Springer, Berlin, 449–462, 1997.

50. J. Weickert, B.M. ter Haar Romeny, M.A. Viergever, Conservative image transformations with restoration and scale-space properties, *Proc. 1996 IEEE International Conference on Image Processing* (ICIP–96, Lausanne, Sept. 16–19, 1996), Vol. 1, 465–468, 1996.

51. J. Weickert, Nonlinear diffusion scale-spaces: from the continuous to the discrete setting, in M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, J.-M. Morel (Eds.), *ICAOS '96: Images, Wavelets and PDEs*, Lecture Notes in Control and Information Sciences, Vol. 219, Springer, London, 111–118, 1996.

52. J. Weickert, A model for the cloudiness of fabrics, in H. Neunzert (Ed.), *Progress in Industrial Mathematics at ECMI 94*, Wiley–Teubner, Chichester, 258–265, 1996.

53. J. Weickert, Multiscale texture enhancement, in V. Hlaváč, R. Šára (Eds.), *Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, Vol. 970, Springer, Berlin, 230–237, 1995.

54. J. Weickert, Ein mathematisches Modell für Ultra-Windeln, *Proc. Workshop "Mathematische Modelle und Methoden in den Anwendungen"* (Mainz, March 19–20, 1992), Vol. ALEF–15, 91–110, 1995.

55. J. Weickert, Anisotropic diffusion filters for image processing based quality control, in A. Fasano, M. Primicerio (Eds.), *Proc. Seventh European Conference on Mathematics in Industry*, Teubner, Stuttgart, 355–362, 1994.

56. J. Weickert, The modelling of napkins, in F. Hodnett (Ed.), *Proc. Sixth European Conference on Mathematics in Industry*, Teubner, Stuttgart, 297–300, 1992.

Theses

57. J. Weickert, *Anisotropic Diffusion in Image Processing*, Ph.D. thesis, Dept. of Mathematics, University of Kaiserslautern, Germany, 1996.

58. J. Weickert, *Mathematische Modelle eines Diffusions- und Austauschprozesses*, M.Sc. thesis, Dept. of Mathematics, University of Kaiserslautern, Germany, 1990.

Manuscripts Submitted to Journals

59. J. Weickert, Applications of nonlinear diffusion in image processing and computer vision, Technical Report 18/2000, Computer Science Series, University of Mannheim, Germany, Nov. 2000. Submitted to *Acta Mathematica Universitatis Comenianae*.

60. J. Weickert, C. Schnörr, A theoretical framework for convex regularizers in PDE-based computation of image motion, Technical Report 13/2000, Computer Science Series, University of Mannheim, Germany, June 2000. Submitted to *International Journal of Computer Vision.*

Manuscripts Submitted to Conferences

61. G. Kühne, J. Weickert, O. Schuster, S. Richter, A tensor-driven active contour model for moving object segmentation, Jan. 2001. Submitted to *Proc. 2001 IEEE International Conference on Image Processing* (Thessaloniki, Greece, Oct. 2001).

62. D. Cremers, C. Schnörr, J. Weickert, Diffusion-snakes: combining statistical shape knowledge and image information in a variational framework, Dec. 2000. Submitted to *Proc. IEEE Workshop on Variational and Level Set Methods in Computer Vision* (Vancouver, Canada, July 2001).

Technical Reports and Non-Reviewed Work

63. J. Weickert, C. Schnörr, Adaptive Signalverarbeitung zur Auswertung digitaler Bilddaten, *Future*, in press.

64. J. Weickert, Anisotrope Diffusionsfilter: Grundlagen und Anwendungsmöglichkeiten, *Tagungsband zum Workshop "Initiative Industrielle Bildverarbeitung"* (Kaiserslautern, 19. 7. 1995).

65. J. Weickert, *Scale-Space Properties of Nonlinear Diffusion Filtering with a Diffusion Tensor*, Technical Report 110, Laboratory of Technomathematics, University of Kaiserslautern, Germany, Okt. 1994.

66. J. Fröhlich, J. Weickert, *Image Processing Using a Wavelet Algorithm for Nonlinear Diffusion*, Technical Report 104, Laboratory of Technomathematics, University of Kaiserslautern, Germany, March 1994.

67. J. Weickert, Abschlußbericht zum Projekt "Nichtlineare Diffusionsfilter", *Abschlussbericht und Bericht über die wissenschaftliche Tätigkeit Januar 1992 – Dezember 1993*, Zentrum für Praktische Mathematik Darmstadt–Kaiserslautern, 191–209, 1994.

68. J. Weickert, Zwischenbericht zum Projekt "Nichtlineare Diffusionsfilter", *Bericht über die wissenschaftliche Tätigkeit Januar 1991 – Dezember 1991*, Zentrum für Praktische Mathematik Darmstadt–Kaiserslautern, 133–142, 1992.