Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 152

A Survey on Variational Optic Flow Methods for Small Displacements

Joachim Weickert, Andrés Bruhn,
Thomas Brox and Nils Papenberg

Saarbrücken 2005

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 152
submitted: September 28, 2005

A Survey on Variational Optic Flow Methods for Small Displacements

Joachim Weickert
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Building 27.1
66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Andrés Bruhn
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Building 27.1
66041 Saarbrücken, Germany
bruhn@mia.uni-saarland.de

Thomas Brox
Institute for Computer Science
Römerstraße 164
53113 Bonn, Germany
brox@mia.uni-saarland.de

Nils Papenberg
Institute of Mathematics
University of Lübeck, Wallstraße 40
23560 Lübeck, Germany
papenber@math.uni-luebeck.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Abstract

Optic flow describes the displacement field in an image sequence. Its reliable computation constitutes one of the main challenges in computer vision, and variational methods belong to the most successful techniques for achieving this goal. Variational methods recover the optic flow field as a minimiser of a suitable energy functional that involves data and smoothness terms. In this paper we present a survey on different model assumptions for each of these terms and illustrate their impact by experiments. We restrict ourselves to rotationally invariant convex functionals with a linearised data term. Such models are appropriate for small displacements. Regarding the data term, constancy assumptions on the brightness, the gradient, the Hessian, the gradient magnitude, the Laplacian, and the Hessian determinant are investigated. Local integration and nonquadratic penalisation are considered in order to improve robustness under noise. With respect to the smoothness term, we review a recent taxonomy that links regularisers to diffusion processes. It allows to distinguish five types of regularisation strategies: homogeneous, isotropic image-driven, anisotropic image-driven, isotropic flow-driven, and anisotropic flow-driven. All these regularisations can be performed either in the spatial or the spatiotemporal domain. After discussing well-posedness results for convex optic flow functionals, we sketch some numerical ideas in order to achieve real-time performance on a standard PC by means of multigrid methods, and we survey a simple and intuitive confidence measure.

Key Words: computer vision, optic flow, variational methods, partial differential equations.

Contents

1 Introduction

Finding the displacement field between subsequent frames of an image sequence has become a classical computer vision problem. This displacement field is called *optic flow*. Solving the optic flow problem does not only have an impact in fields like video coding or robot navigation, it is also a prototype for the entire class of correspondence problems, where one seeks a sufficiently smooth mapping that maps the features in one image to the structures in another one. Other applications where such problems appear include the fields of stereo reconstruction and medical image registration.
Already in 1981, Horn and Schunck introduced the first variational method for computing the optic flow field in an image sequence [44]. This method is based on two assumptions that are characteristic for many variational optic flow methods: a brightness constancy assumption and a smoothness assumption. These assumptions enter a continuous energy functional whose minimiser yields the desired optic flow field.
Performance evaluations such as [9, 35] showed that variational methods belong to the best performing techniques for computing the optic flow field. It is thus not surprising that a lot of research has been carried out in order to improve these techniques even further: These amendments include refined model assumptions with discontinuity-preserving constraints [2, 28, 42, 62, 65, 73, 91] or spatiotemporal regularisation [11, 61, 92], improved data terms with modified constraints [3, 26, 62, 74] or nonquadratic penalisation [11, 43, 56, 26], and efficient multigrid algorithms [15, 22, 39, 38, 78, 95] for minimising these energy functionals.
The goal of the present chapter is to analyse the data term and the smoothness term in detail and to survey some of our recent results on variational optic flow computation. For the sake of simplicity we focus on small displacements, where Taylor linearisations of the data term are valid approximations. This restriction allows to consider convex functionals where many theoretical and practical aspects become significantly easier and more transparent.
Our chapter is organised as follows: In Section 2 we sketch the general structure of these techniques. While Section 3 analyses the data term in more detail, a discussion of the different possibilities for smoothness constraints is given in Section 4. Suitable combinations of data and smoothness tersm are investigated in Section 5, well-posedness results are presented in Section 6, and algorithmic aspects are sketched in Section 7. A simple but general confidence measure for energy-based optic flow methods in discussed in Section 8. Our chapter is concluded with a summary in Section 9. A significantly shorter early version of the present chapter has been presented at a workshop [90].

2 General Structure

Let $f(x_1, x_2, x_3)$ denote some scalar-valued image sequence, where (x_1, x_2) is the location and x_3 denotes time. Often f is obtained by preprocessing some initial image sequence f_0 by convolving it with a Gaussian K_σ of standard deviation σ:

$$f = K_\sigma * f_0. \tag{1}$$

Let us assume that $D^k f$ describes the set of all partial (spatial and temporal) derivatives of f of order k, and that the optic flow field $u(x_1, x_2, x_3) = (u_1(x_1, x_2, x_3), u_2(x_1, x_2, x_3), 1)^\top$ gives the displacement rate between subsequent frames with temporal frame distance 1. In the present paper we consider variational methods that are based on the minimisation of the continuous energy functional

$$E(u) = \int_\Omega (\underbrace{M(D^k f, u)}_{\text{data term}} + \alpha \underbrace{S(\nabla f, \nabla u)}_{\text{regulariser}})\, dx \tag{2}$$

where the integration domain Ω is either a spatial or a spatiotemporal domain. In the spatial case we have $x := (x_1, x_2)^\top$ and $\nabla := \nabla_2 := (\partial_{x_1}, \partial_{x_2})^\top$, and in the spatiotemporal case we use the notations $x := (x_1, x_2, x_3)^\top$ and $\nabla := \nabla_3 := (\partial_{x_1}, \partial_{x_2}, \partial_{x_3})^\top$. The optic flow field $u(x_1, x_2, x_3)$ is obtained as a function that minimises $E(u)$. The energy functional $E(u)$ penalises all deviations from model assumptions. Typically is consists of a *data term* $M(D^k f, u)$ which expresses e.g. a brightness constancy assumption, and a *regulariser* $S(\nabla f, \nabla u)$ with $\nabla u := (\nabla u_1, \nabla u_2)^\top$ that penalises deviations from (piecewise) smoothness. The weight $\alpha > 0$ serves as *regularisation parameter:* Larger values correspond to more simplified flow fields.
The simplest and oldest representative of the class (2) is given by the method of Horn and Schunck [44]. It is based on the minimisation of the spatial functional

$$E(u) = \int_\Omega \Big((u^\top \nabla_3 f)^2 + \alpha \sum_{i=1}^{2} |\nabla u_i|^2 \Big)\, dx. \tag{3}$$

As will be detailed in the forthcoming sections, the Horn–Schunck functional combines a data term that describes the brighness constancy of moving patterns with a smoothness term which involves homogeneous (Tikhonov [79]) regularisation.
It should be noted that continuous energy functionals of type (2) may be formulated in a rotationally invariant way: Apart from very few exceptions such as [6, 28, 52], almost all continuous optic flow functionals that have been proposed are rotationally invariant. Results from numerical analysis then show that consistent discretisations approximate this invariance under rotations arbitrarily well if the sampling is sufficiently fine. Moreover, if the energy functional is convex, a unique minimiser exists that can be found in a relatively simple way by globally convergent algorithms. Variational optic flow methods are *global* methods: If there is not sufficient local information, the data term $M(D^k f, u)$ is so small that it is dominated by the smoothness term $\alpha S(\nabla f, \nabla u)$ which fills in information from more reliable surrounding locations. Thus, in contrast to local methods, the *filling-in effect* of global variational approaches always yields dense flow fields and no subsequent interpolation steps are necessary: Everything is automatically accomplished within a single variational framework.

3 Data Terms

In the design of data terms for optic flow methods prior knowledge plays an important role. This knowledge may include information on the imaging device (e.g. the quality of the images with respect to noise), on the conditions during the acquisition of the video material (e.g. the occurrence of frequent illumination changes) as well as information on the expected type of motion (e.g. mainly translational motion of cars in traffic sequences). For a specific problem, this information may allow to select a data term that is especially appropriate and thus improves the quality of the estimation significantly. For this reason, the following section gives an overview on data terms that are frequently used in literature. Moreover, a detailed discussion on their advantages and shortcomings should guide the reader to select an appropriate data term for a specific problem.

3.1 Constancy Assumptions

In order to analyse motion within subsequent frames of an image sequence, temporal constancy has to be imposed on certain image features. The most frequently used feature in this context is the image brightness. Many differential methods are based on the assumption that this brightness is constant, i.e. that the grey value of objects does not change over time. If we denote the motion of some image structure by $(x_1(x_3), x_2(x_3))^\top$ this assumptions can be formulated as

$$\frac{df(x_1(x_3), x_2(x_3), x_3)}{dx_3} = 0. \tag{4}$$

By applying the chain rule and defining $f_{x_i} := \partial_{x_i} f$ the following *optic flow constraint (OFC)* is obtained:

$$f_{x_1} u_1 + f_{x_2} u_2 + f_{x_3} = 0. \tag{5}$$

Note that the optic flow field satisfies $(u_1, u_2, 1)^\top = (\partial_{x_3} x_1, \partial_{x_3} x_2, 1)^\top$.

It also is instructive to derive this constraint in a second way: Assuming a frame distance of 1, the brightness constancy constraint between two subsequent frames at time x_3 and $x_3 + 1$ can be expressed as

$$0 \;=\; f(x_1{+}u_1,\, x_2{+}u_2,\, x_3{+}1) - f(x_1, x_2, x_3) \tag{6}$$

such that (5) follows from a Taylor linearisation in the point $(x_1, x_2, x_3)^\top$. However, this Taylor linearisation is only a reasonable approximation if the flow field varies sufficiently smooth and the displacement rates are small, i.e. in the order of one pixel or below. In the following we assume that this is the case, because it would be much more burdensome to deal with the unlinearised constraint (6) than its linearised counterpart (5).

In order to use equation (5) within the energy functional (2), we penalise all deviations from zero by considering the quadratic data term [44]

$$M_1(D^1 f, u) := (u^\top \nabla_3 f)^2. \tag{7}$$

As long as the image data does not violate the brightness constancy assumption, the use of M_1 can give good results. In particular with regard to image data with non-constant brightness, however, constancy assumptions should be based on image features that are less sensitive to illumination changes. A simple and efficient strategy in this context is the consideration of derivatives. Instead of imposing constancy to the image brightness f along the path $(x_1(x_3), x_2(x_3))^\top$, one may e.g. assume that the spatial brightness gradient $(f_{x_1}, f_{x_2})^\top$ does not change along the same path [83]:

$$\frac{df_{x_1}(x_1(x_3), x_2(x_3), x_3)}{dx_3} = 0, \tag{8}$$

$$\frac{df_{x_2}(x_1(x_3), x_2(x_3), x_3)}{dx_3} = 0. \tag{9}$$

This gives the two equations

$$u^\top \nabla_3 f_{x_1} \;=\; 0, \tag{10}$$

$$u^\top \nabla_3 f_{x_2} \;=\; 0. \tag{11}$$

Squaring and adding them produces the data term

$$M_2(D^2 f, u) := \sum_{i=1}^{2} (u^\top \nabla_3 f_{x_i})^2. \tag{12}$$

In a straightforward way, constancy assumptions can also be imposed on higher-order derivatives, e.g. on the (spatial) Hessian $\mathcal{H}_2 f$. Squaring and adding the corresponding equations we obtain the following data term:

$$M_3(D^3 f, u) := \sum_{i=1}^{2} \sum_{j=1}^{2} (u^\top \nabla_3 f_{x_i x_j})^2. \tag{13}$$

With M_2 and M_3 we have proposed data terms that are designed for sequences with illumination changes. However, one should note that their performance depends significantly on the occurring type of motion. This has the following reason: In contrast to the image brightness both gradient and Hessian contain directional information. As a consequence, any constancy assumption on these expressions implies a constancy assumption on their orientation. On one hand, this property may be useful if it comes to the estimation of translational, divergent or slow rotational motion. In this case the orientation of the features does hardly change and the combination of two or three constraints in one data term may improve the results. On the other hand, poor results have to be expected if fast rotations are dominating and the implied orientation constancy does not hold.

A way to overcome this limitation is to create motion invariant image features from these "oriented" derivatives. Instead of imposing constancy on the (spatial) brightness gradient and therewith on its orientation, one may e.g. assume that only its magnitude is constant over time. Then, the following data term is obtained:

$$M_4(D^2 f, u) := (u^\top \nabla_3 |\nabla f|)^2. \tag{14}$$

This idea can also be extended to higher-order derivatives. As an example, let us consider the (spatial) Hessian $\mathcal{H}_2 f$. In this case, one may either think of imposing constancy on the (spatial) Laplacian $\Delta_2 f$ or on the determinant of the (spatial) Hessian $\mathcal{H}_2 f$. While the data term associated to the Laplacian is given by

$$M_5(D^3 f, u) := (u^\top \nabla_3 (\Delta_2 f))^2, \tag{15}$$

the data term based on the constancy of the determinant of the Hessian reads

$$M_6(D^3 f, u) := (u^\top \nabla_3 \det(\mathcal{H}_2 f))^2. \tag{16}$$

This example shows that in general multiple of such scalar valued expressions can be derived from the set of derivatives of a single order. However, there is no general rule which expression gives the best performance. An overview of all data terms presented so far is given in Table 1. It may also be useful to combine multiple of these terms by means of a linear combination. Moreover, one should note that M_2–M_6 can be more sensitive to noise than M_1, since they involve higher orders of derivatives of the image sequence.

In Figure 1 we illustrate the impact of different constancy assumptions on the computed flow field. To this end we use the data terms M_1–M_6 within a spatial energy functional based on homogeneous regularisation of Horn–Schunck type, i.e. we minimise

$$E(u) = \int_\Omega \left(M_j + \alpha \sum_{i=1}^{2} |\nabla u_i|^2 \right) dx. \tag{17}$$

for $j = 1,...,6$. As test sequence we take the popular Yosemite sequence *with* clouds. It consists of 15 frames of size 316×252 and combines divergent and translational motion under varying illumination. Both the sequence and its ground truth flow field are available from `ftp://csd.uwo.ca`

Table 1: Comparison of the data terms M_1–M_6.

	data term	constancy assumption	illum. changes	motion type
M_1	$(u^\top \nabla_3 f)^2$	brightness	no	any
M_2	$\sum_{i=1}^{2} (u^\top \nabla_3 f_{x_i})^2$	gradient	yes	translational divergent slow rotational
M_3	$\sum_{i=1}^{2}\sum_{j=1}^{2} (u^\top \nabla_3 f_{x_i x_j})^2$	Hessian	yes	translational divergent slow rotational
M_4	$(u^\top \nabla_3 \lvert\nabla f\rvert)^2$	gradient magnitude	yes	any
M_5	$(u^\top \nabla_3 (\Delta_2 f))^2$	Laplacian	yes	any
M_6	$(u^\top \nabla_3 \det(\mathcal{H}_2 f))^2$	Hessian determinant	yes	any

under the directory `pub/vision`. In order to allow for a quantitative comparison of the different data terms we computed the so-called *average angular error (AAE)* as proposed in [9] :

$$\mathrm{AAE}(\mathrm{u_c}, \mathrm{u_e}) = \frac{1}{|\Omega|} \int_\Omega \arccos\left(\frac{u_c^\top u_e}{|u_c||u_e|}\right) dx. \tag{18}$$

In this context the subscripts c and e denote the correct respectively the estimated spatiotemporal optic flow vectors $u_c = (u_{c1}, u_{c2}, 1)^\top$ and $u_e = (u_{e1}, u_{e2}, 1)^\top$. Moreover, $|\Omega| = \int_\Omega dx$ is the integration domain, and $|u| = \sqrt{u_1^2 + u_2^2 + 1}$. The obtained results for optimised Gaussian presmoothing parameter σ (cf. equation (1)) and regularsiation parameter α are presented in Table 2. As one can see, the commonly used grey value constancy assumption is outperformed by almost all other constraints that involve higher derivatives. This quantitative impressions are also confirmed qualitatively by the corresponding flow fields shown in Figure 1. While M_1 gives slightly better results at the mountain site, the other data terms are significantly superior in estimating the sky region where illumination changes are present. This shows that it can be worthwhile to replace the brightness constancy constraint by constraints that involve higher derivatives, in particular when varying illumination has to be expected. We also observe that constancy assumptions based on higher order derivatives require a larger Gaussian width σ in order to give optimal results.

3.2 Increasing the Robustness of the Data Term

With M_1–M_6 we have proposed data terms for different illumination conditions and different types of motion. Let us now discuss by the example of M_1 how these data terms can be modified such that they become more robust. To this end we investigate three strategies: local least square fitting, adaptive averaging with nonlinear diffusion, and nonquadratic penalisation.

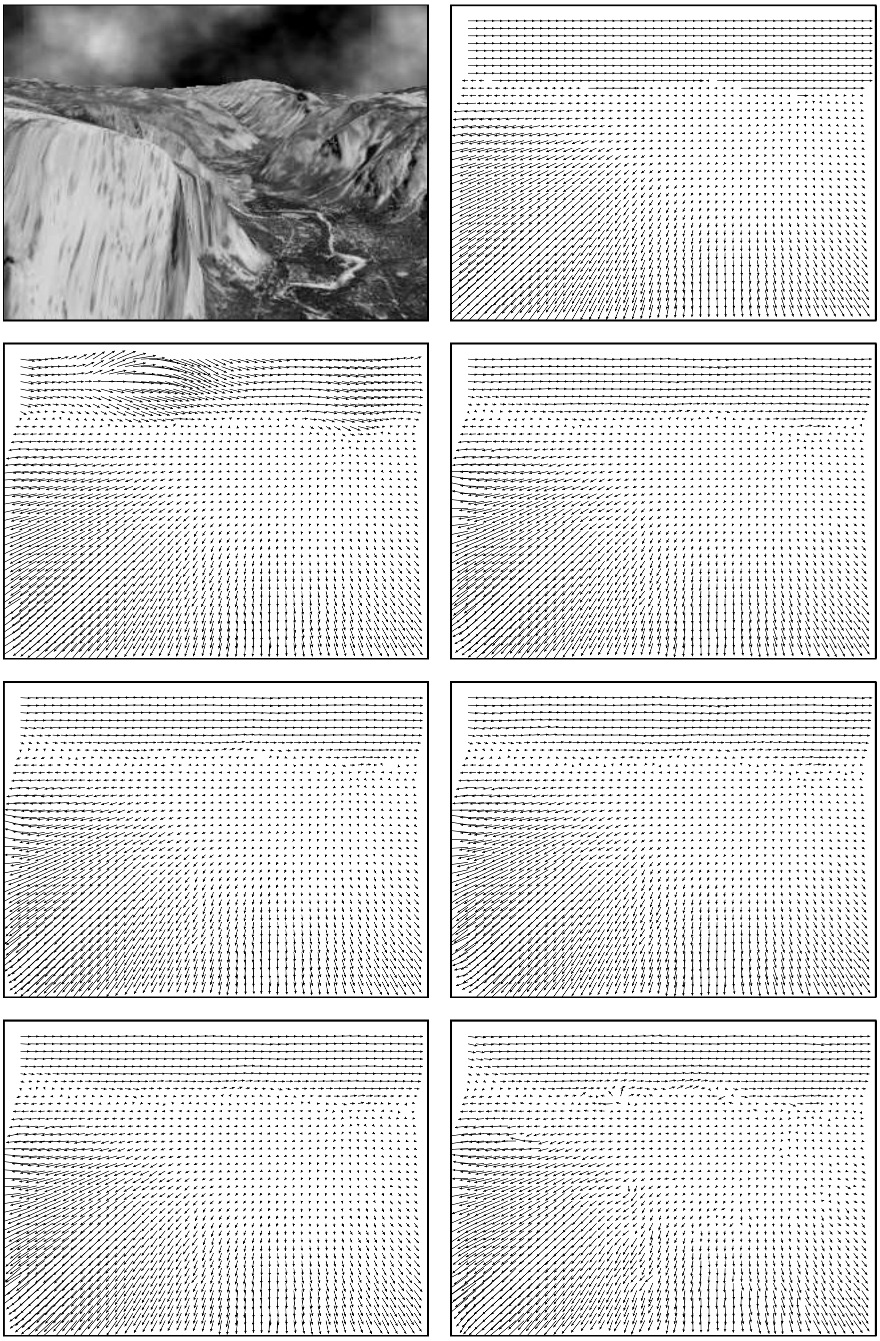

Figure 1: *From left to right, and from top to bottom: (a)* Frame 8 of the Yosemite sequence *with* clouds of size 316×256. *(b)* Ground truth. *(c)* Computed flow field for a spatial approach with data term M_1 (brightness constancy) and homogeneous regularisation as smoothness term. *(d)* Data term M_2 (gradient constancy). *(e)* Data term M_3 (constancy of Hessian). *(f)* Data term M_4 (gradient magnitude constancy). *(g)* Data term M_5 (constancy of Laplacian). *(h)* Data term M_6 (constancy of Hessian determinant).

Table 2: Impact of the constancy assumption on the quality of the optic flow field. We used a spatial energy functional with homogeneous regularisation, and computed the average angular error (AAE) for the Yosemite sequence with clouds. The parameters σ and α have been optimised.

constancy assumption	data term	σ	α	AAE
brightness	M_1	1.30	500	7.17°
gradient	M_2	2.10	20	5.91°
Hessian	M_3	2.70	1.8	6.46°
gradient magnitude	M_4	1.90	14	6.37°
Laplacian	M_5	2.50	3.0	6.18°
Hessian determinant	M_6	3.00	0.1	8.10°

3.2.1 Local Least Square Fitting

A useful strategy to make optic flow estimation more robust under noise is the consideration of neighbourhood information within the data term [26]. To this end one may e.g. assume that the optic flow is constant within some spatial or spatiotemporal neighbourhood of size ρ. Then, simple statistical methods such as least square regressions can be applied to estimate the flow vector from the considered neighbourhood [54]. In this context it is common to decrease the weight of neighbours with increasing distance to the center. Let us now apply such a Gaussian weighted least square fit to $M_1 = u^\top \nabla_3 f \, \nabla_3 f^\top u$. Then the corresponding data term reads

$$M_7(D^1 f, u) := u^\top J_\rho(\nabla_3 f)\, u \tag{19}$$

where the *structure tensor* (see e.g. [10, 33, 69]

$$J_\rho(\nabla_3 f) := K_\rho * (\nabla_3 f \, \nabla_3 f^\top) \tag{20}$$

results from componentwise Gaussian convolution of the tensor product $J_0 = \nabla_3 f \, \nabla_3 f^\top$. In this case the standard deviation ρ of the Gaussian K_ρ is called *integration scale*. One should note that for $\rho = 0$ this least square fit by minimising M_7 comes down to the original data term M_1.

3.2.2 Adaptive Averaging with Nonlinear Diffusion

Although the preceding integration of local information by means of a Gaussian convolution is a good concept for achieving robustness under noise, the integration relies on the underlying assumption that the optic flow field is constant within the local neighbourhood described by the Gaussian kernel. Especially in the area of discontinuities in the flow field this assumption is not valid, and thus the Gaussian convolution compromises the flow estimation. As a remedy, one can assume that the flow field is only *piecewise* constant. Then one replaces the (linear) structure tensor in (20) that is based on Gaussian convolution – or equivalently linear diffusion – by a nonlinear structure tensor [89, 20] that uses nonlinear tensor-valued diffusion for the local integration. Since nonlinear diffusion reduces the amount of smoothing at discontinuities, it avoids the integration of unrelated data beyond these discontinuities and therefore leads to less ambiguity in the least square regression.

Since the structure tensor is a matrix field, a matrix-valued scheme for nonlinear diffusion is needed. Such a scheme is proposed in [81] where the matrix channels are coupled by a joint

Table 3: Comparison of data terms M_1, M_2, M_7 and M_8 under noise. We added Gaussian noise with varying standard deviations σ_n to the Yosemite sequence with clouds and used a spatial energy functional with homogeneous regularisation to compute the average angular error (AAE). The parameters σ, α, ρ, and t have been optimised.

noise	data term	σ	α	integration parameter	AAE
$\sigma_n = 0$	M_1	1.30	500	-	7.17°
	M_2	2.10	20	-	5.91°
	M_7	1.30	500	$\rho = 1.80$	7.14°
	M_8	1.30	300	$t = 250$	6.97°
$\sigma_n = 20$	M_1	2.08	2200	-	12.17°
	M_2	3.60	35	-	12.26°
	M_7	2.09	1600	$\rho = 10.70$	11.71°
	M_8	2.10	1600	$t = 225$	11.76°
$\sigma_n = 40$	M_1	2.45	4100	-	16.80°
	M_2	4.20	55	-	18.00°
	M_7	2.38	2000	$\rho = 17.60$	15.82°
	M_8	2.40	2500	$t = 500$	16.29°

diffusivity. With $J_0 = \nabla_3 f \nabla_3 f^\top$ as initial value for the nonlinear diffusion process

$$\partial_t \hat{J}_{ij} = \operatorname{div}\Big(g\Big(\sum_{k,l=1}^{3} |\nabla \hat{J}_{kl}|^2\Big)\nabla \hat{J}_{ij}\Big) \qquad (i,j = 1,2,3) \tag{21}$$

the solution $\hat{J}_t$ constitutes a nonlinear structure tensor for a certain diffusion time t. The diffusion time is the scale parameter of the nonlinear structure tensor, similar to the standard deviation of the Gaussian kernel used in (20), and steers the size of the local neighbourhood. The so-called diffusivity function g is a decreasing function that reduces the amount of smoothing at discontinuities in the data. An appropriate choice is the regularised total variation (TV) diffusivity [5]

$$g(s^2) = \epsilon_1 + \frac{1}{\sqrt{s^2 + \epsilon_2^2}} \tag{22}$$

where the small positive constants ϵ_1 and ϵ_1 are introduced for theoretical reasons and in order to avoid unbounded diffusivities. In practice they can be set, for instance, to 0.001.
If we apply the nonlinear structure tensor to M_1, we obtain the data term

$$M_8(D^1 f, u) := u^\top \hat{J}_t(\nabla_3 f)\, u, \tag{23}$$

which is a nonlinear alternative to M_7.
Alternative ways of creating adaptive structure tensors are studied in [63] and [19]. It is also worth noting that if one chooses the diffusivity function

$$g(s^2) = 1 \tag{24}$$

one ends up with homogeneous diffusion, which does not adapt to the data. Homogeneous diffusion with diffusion time t is equivalent to Gaussian convolution with standard deviation

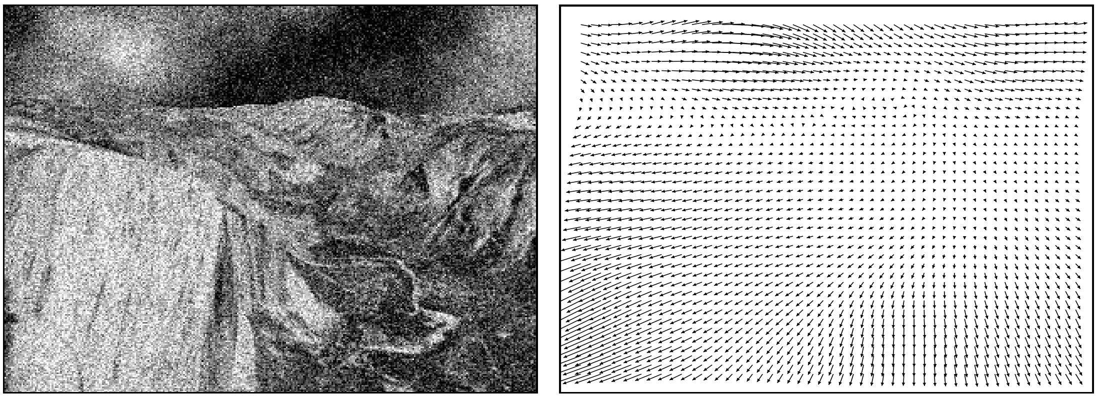

Figure 2: *(a) Left:* Frame 8 of the Yosemite sequence *with* clouds degraded by Gaussian of standard deviation $\sigma_n = 40$. *(b) Right:* Computed flow field for a spatial approach with data term M_7 (least squares) and homogeneous smoothness term.

$\rho = \sqrt{2t}$. This shows the direct relation between the employment of the structure tensor J_ρ and the nonlinear structure tensor $\hat{J}_t$.

In our second experiment we compare different data terms regarding their robustness under noise. To this end we have added Gaussian noise with zero mean and varying standard deviation σ_n to the Yosemite sequence *with* clouds. Apart from the data terms M_7 and M_8 that are based on the concept of local integration, we also considered the ordinary data terms M_1 and M_2. As expected, the results in Table 3 show a better performance of the data terms M_7 and M_8 when noise is present. Figure 2 depicts the corresponding flow field for the data term M_7 and $\sigma_n = 40$. Although the original sequence was degraded severely, the computed flow field still looks reasonable. In this context one should also note the worse performance of M_2. It shows that higher-order derivatives are more sensitive to noise.

3.2.3 Nonquadratic Penalisation

So far we have only considered data terms that penalise deviations from constancy assumptions in a quadratic way. From a statistical viewpoint, however, it seems desirable to penalise outliers less severely than in a quadratic setting. In particular with regard to the preservation of discontinuities in the data term, this concept from robust statistics [41, 45] proves to be very useful; see e.g. [11, 43, 56]. In order to guarantee well-posedness for the remaining problem and allow the construction of simple globally convergent algorithms it is advantageous to use penalisers $\Psi(s^2)$ that are convex in s. Such penalisers comprise e.g. the regularised TV penaliser [70, 64]

$$\Psi(s^2) = \epsilon_1^2\, s^2 + 2\sqrt{s^2 + \epsilon_2^2}, \tag{25}$$

where ϵ_1 and ϵ_2 are small positive constants.
In Figure 3 the graphs of the corresponding functions are depicted. Apart from TV penalisation also an example for a nonconvex function is shown. However, one should note that in the case of such nonconvex functions multiple minima have to be expected. As a consequence, minimisation strategies do usually not succeed in finding the global minimum. Let us now replace the quadratic

Table 4: Comparison of data terms M_7–M_{10} and their suitability for respecting discontinuities in the image sequence.

	data term	concept	discontinuities
M_7	$u^\top J_\rho(\nabla_3 f)\, u$	least squares	no
M_8	$u^\top \hat{J}_t(\nabla_3 f)\, u$	nonlinear diffusion	yes
M_9	$\Psi((u^\top \nabla_3 f)^2)$	nonquadratic penaliser	yes
M_{10}	$\Psi(u^\top J_\rho(\nabla_3 f)\, u)$	least squares and nonquadratic penaliser	yes

penaliser in M_1 and M_7 by one of the proposed convex functions. Then we obtain the data terms given by

$$\begin{aligned} M_9(D^1 f, u) &:= \Psi((u^\top \nabla_3 f)^2), \\ M_{10}(D^1 f, u) &:= \Psi(u^\top J_\rho(\nabla_3 f)\, u). \end{aligned}$$

An overview on the data terms M_7–M_{10} and their capability of handling discontinuities in the data is given in Table 4.
In our last experiment on the impact of data terms we investigate the advantages of nonquadratic penalisers. This is done in Table 5 where the terms M_1, M_9 and M_{10} are compared. Again, the listed results refer to the Yosemite sequence *with* clouds. Obviously, one can improve the average angular error by replacing the quadratic penaliser with a nonquadratic one. The reason for this improvement can be found in Figure 4. It depicts a zoom into the lower left corner of frame 8 and 9, the ground truth as well as the computed flow fields for the different data terms. As one can see, those boundary pixels from frame 8 that are not present in frame 9 have a large impact on the estimated flow field when penalised in a quadratic way. By using a nonquadratic approach, however, their influence is reduced significantly. As a consequence, the estimation at these locations becomes more precise and the average angular error decreases.

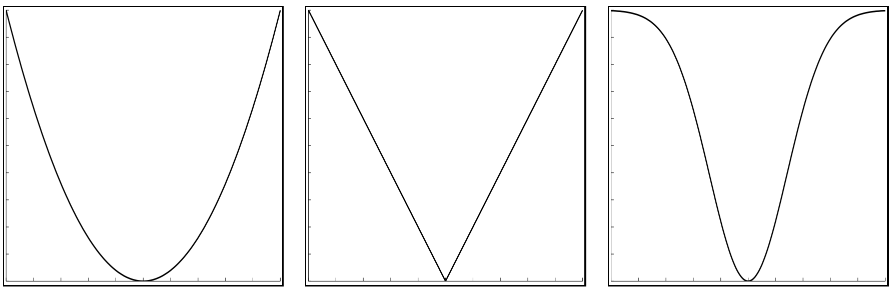

Figure 3: Comparison of different penalising functions. From left to right: (a) Tikhonov (quadratic). (b) Total variation (linear). (c) Example of a nonconvex function.

Table 5: Comparison of quadratic and nonquadratic penalisers for the data term M_1 (brightness constancy). We used a spatial energy functional with homogeneous regularisation, and computed the average angular error (AAE) for the Yosemite sequence with clouds. The parameters σ, α and ρ have been optimised.

penaliser	data term	σ	α	ρ	AAE
quadratic	M_1	1.30	500	-	7.17°
nonquadratic	M_9	1.40	190	-	7.08°
nonquadratic + least squares	M_{10}	1.40	200	2.0	6.76°

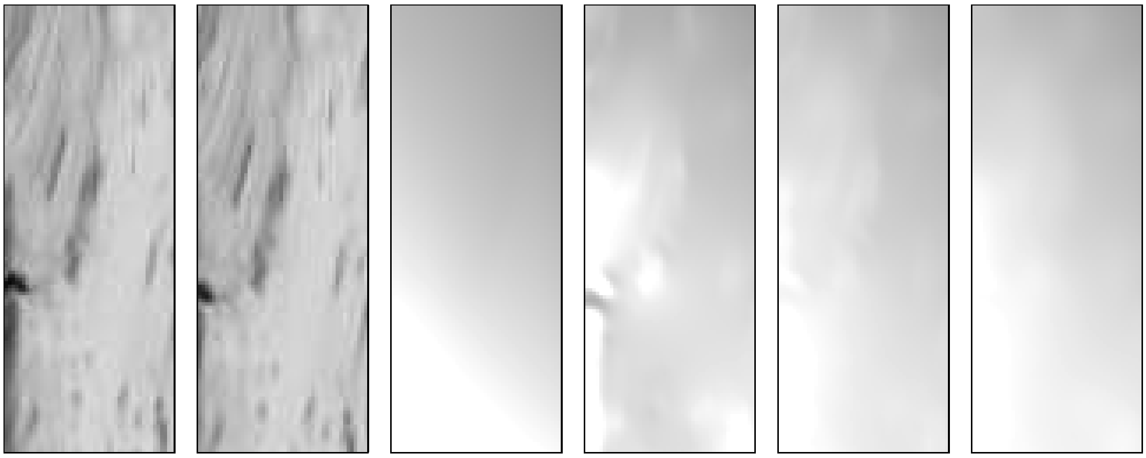

Figure 4: *From left to right: (a)* Detail from Frame 8 of the Yosemite sequence *with* clouds (48×128 pixels). *(b)* Frame 9. *(c)* Ground truth. *(d)* Computed flow field for a spatial approach with data term M_1 (quadratic penaliser) and homogeneous regularisation. *(e)* Data term M_9 (nonquadratic penaliser). *(f)* Data term M_{10} (nonquadratic penaliser and least squares).

4 Smoothness Terms

So far we have analysed different possibilities for modelling the data term. Let us now explore different models for the smoothness term. This is done in two steps: First we survey a taxonomy that links the regularisers in optic flow functionals to vector-valued diffusion processes. In a second step we investigate the impact of replacing a spatial smoothness assumption by a spatiotemporal one.

4.1 A Diffusion Taxonomy for Smoothness Terms

A taxonomy of the different possibilities to design smoothness constraints has been presented in [91]. It exploits the connection between regularisation methods and diffusion filtering. In order to describe this taxonomy we derive the steepest descent equations for the optic flow functionals. Since they come down to a diffusion–reaction system, we analyse diffusion filters for vector-valued images. Finally we transfer this classification into the optic flow setting.

4.1.1 From Energy Functionals to Diffusion–Reaction Systems

Minimising the energy functional (2) can be done in two ways:
One possibility is to compute the so-called Euler–Lagrange equations. They constitute necessary conditions a minimiser of $E(u)$ has to satisfy [29, 36]. In the specific case of a spatial energy functional (2) they are given by the two-dimensional system of partial differential equations (PDEs)

$$0 = \partial_{x_1} S_{u_{1x_1}} + \partial_{x_2} S_{u_{1,x_2}} - \tfrac{1}{\alpha}\partial_{u_1} M, \tag{26}$$

$$0 = \partial_{x_1} S_{u_{2x_1}} + \partial_{x_2} S_{u_{2,x_2}} - \tfrac{1}{\alpha}\partial_{u_2} M \tag{27}$$

equipped with homogeneous Neumann (reflecting) boundary conditions. The term $S_{u_{i,x_j}}$ denotes the partial derivative of S with respect to $\partial_{x_j} u_i$.
Alternatively we can minimise $E(u)$ by means of the steepest descent method. In the case of a spatial functional we obtain a system of two-dimensional diffusion–reaction equations, where the diffusion term results from the regulariser $S(\nabla f, \nabla u)$, and the reaction term is induced by the data term $M(D^k f, u)$:

$$\partial_t u_1 = \partial_{x_1} S_{u_{1,x_1}} + \partial_{x_2} S_{u_{1,x_2}} - \tfrac{1}{\alpha}\partial_{u_1} M, \tag{28}$$

$$\partial_t u_2 = \partial_{x_1} S_{u_{2,x_1}} + \partial_{x_2} S_{u_{2,x_2}} - \tfrac{1}{\alpha}\partial_{u_2} M \tag{29}$$

The parameter t is a pure numerical parameter that should not be confused with the time x_3 of the image sequence. If $E(u)$ is strictly convex, a unique minimiser exist and the steepest descent evolution is globally convergent, i.e. its steady–state does not depend on the initialisation. For $t \to \infty$, this steady–state of the diffusion–reaction system is given by the Euler–Lagrange equations (26)–(27).
Since we are interested in a taxonomy for optic flow regularisers, it it sufficient to restrict ourselves to the diffusion part of (28)–(29). This leads to the vector-valued diffusion process

$$\partial_t u_i = \partial_{x_i} S_{u_{i,x_1}} + \partial_{x_i} S_{u_{i,x_2}} \qquad (i = 1, 2). \tag{30}$$

In order to get a better understanding of such processes, it is instructive to make a little excursion to diffusion filters for multichannel images. This shall be done next, following the description in [89].

4.1.2 Diffusion of Vector-Valued Images

Vector-valued images arise for example as colour images, multispectral satellite images and multi-spin echo MR images. Diffusion filtering of some multichannel image $f = (f_1(x), ..., f_m(x))^\top$ with $x \in \mathbb{R}^2$ may be based on one of the following evolutions:

(a) *Homogeneous diffusion* (introduced in [46] in the scalar case):

$$\partial_t u_i = \Delta u_i \qquad (i = 1, ..., m) \tag{31}$$

(b) *Linear isotropic diffusion* (introduced in [34] in the scalar case):

$$\partial_t u_i = \operatorname{div}\Big(g\Big(\sum_j |\nabla f_j|^2\Big)\, \nabla u_i\Big) \qquad (i = 1, ..., m) \tag{32}$$

(c) *Linear anisotropic diffusion* (introduced in [47] in the scalar case):

$$\partial_t u_i \;=\; \operatorname{div}\left(D\Big(\sum_j \nabla f_j \nabla f_j^\top\Big)\, \nabla u_i\right) \qquad (i = 1, ..., m) \tag{33}$$

(d) *Nonlinear isotropic diffusion [37]:*

$$\partial_t u_i \;=\; \operatorname{div}\left(g\Big(\sum_j |\nabla u_j|^2\Big)\, \nabla u_i\right) \qquad (i = 1, ..., m) \tag{34}$$

(e) *Nonlinear anisotropic diffusion [87]:*

$$\partial_t u_i \;=\; \operatorname{div}\left(D\Big(\sum_j \nabla u_j \nabla u_j^\top\Big)\, \nabla u_i\right) \qquad (i = 1, ..., m) \tag{35}$$

where $f(x)$ acts as initial condition for the solution $u(x,t)$:

$$u_i(x,0) = f_i(x) \qquad (i = 1, ..., m). \tag{36}$$

Here, g denotes a scalar-valued diffusivity, and D is a positive definite diffusion matrix. The diffusivity $g(s^2)$ is a decreasing function in its argument. Moreover, we assume that the flux function $g(s^2)s$ is nondecreasing in s. One may e.g. use the regularised TV diffusivity (22). In the *linear* case this ensures that at edges of the *initial* image f, where $\sum_j |\nabla f_j|^2$ is large, the diffusivity $g(\sum_j |\nabla f_j|^2)$ is close to zero. Consequently, diffusion at edges is inhibited. In the *nonlinear* case one introduces a feedback by adapting the diffusivity g to the *evolving* image u. In physics, a diffusion process with a scalar-valued diffusivity is called *isotropic*, since its diffusive behavior does not depend on the direction. *Anisotropic* diffusion with a direction depending behavior may be realised by replacing the scalar-valued diffusivity g by some positive definite diffusion matrix D. One may design the diffusion matrix D such that diffusion along edges of f or u is preferred and diffusion across edges is inhibited. This may be very useful in cases when noisy edges are present.
How can edge directions in some vector-valued image f be measured? Di Zenzo [30] has proposed to consider the matrix $\sum_j \nabla f_j \nabla f_j^\top$. It serves as a structure tensor for vector-valued images since its eigenvectors v_1, v_2 describe the directions of highest and lowest contrast. This contrast is given by the corresponding eigenvalues μ_1 and μ_2.
A natural choice for the design of some diffusion matrix D as a function of a vector-valued image f would thus be to specify its eigenvectors as the eigenvectors v_1, v_2 of $\sum_j \nabla f_j \nabla f_j^\top$, and its eigenvalues λ_1, λ_2 via

$$\lambda_1 \;=\; g(\mu_1), \tag{37}$$
$$\lambda_2 \;=\; g(\mu_2), \tag{38}$$

with a diffusivity function g as e.g. in (22).
Three remarks are in order here:

1. The fact that in the preceding models the same diffusivity or diffusion matrix is used for all channels ensures that the evolutions between the channels are synchronised. This prevents e.g. that discontinuities evolve at different locations in each channel.

2. Let $J \in \mathbb{R}^{2\times 2}$ be symmetric with eigenvectors v_1, v_2 and eigenvalues μ_1, μ_2:

$$J = \mu_1 v_1 v_1^\top + \mu_2 v_2 v_2^\top . \tag{39}$$

A formal way to extend some scalar-valued function $g(s^2)$ to a matrix-valued function $g(J)$ is to define

$$g(J) := g(\mu_1) v_1 v_1^\top + g(\mu_2) v_2 v_2^\top . \tag{40}$$

With this notation we may characterise the linear and nonlinear isotropic models by their diffusivities $g(\sum_j \nabla f_j^\top \nabla f_j)$ and $g(\sum_j \nabla u_j^\top \nabla u_j)$, while their anisotropic counterparts are given by $g(\sum_j \nabla f_j \nabla f_j^\top)$ and $g(\sum_j \nabla u_j \nabla u_j^\top)$. Hence, isotropic and anisotropic models only differ by the location of the transposition.

3. The preceding models are not the only PDE methods that have been proposed for processing vector-valued images. For alternative approaches the reader is referred to [14, 50, 72, 82, 88] and the references therein. Our classification is based on diffusion processes in divergence form that can be derived as steepest descent methods for minimising suitable energy functionals.

Figure 5 illustrates the effect of the different smoothing strategies for a noisy color image with three channels corresponding to the red, green and blue components. We observe that homogeneous diffusion performs well with respect to denoising, but does not respect image edges. Space-variant linear isotropic diffusion, however, may suffer from noise sensitivity as strong noise may be misinterpreted as an important edge structure where the diffusivity is reduced. Anisotropic linear diffusion allows smoothing along edges, but reduces smoothing across them. This leads to a better performance than isotropic linear diffusion if images are noisy. We can also observe that nonlinear models give better results than their linear counterparts. This is not surprising, since the nonlinear models adapt the diffusion process to the evolving image instead of the initial one.

4.1.3 From Vector-Valued Diffusion to Optic Flow Regularisation

Having discussed a taxonomy for vector-valued diffusion, we can transfer it to the optic flow setting. The idea is to identify the optic flow regularisers $S(\nabla f, \nabla u)$ that produce homogeneous, linear isotropic, linear anisotropic, nonlinear isotropic, and nonlinear anistropic diffusion. It should be noted that now that we returned to the optic flow setting, f denotes the image sequence again, and u is the flow field.

The simplest optic flow regulariser is the *homogeneous* regularisation of Horn and Schunck [44]. This quadratic regulariser of type $S(\nabla u) = |\nabla u_1|^2 + |\nabla u_2|^2$ penalises all deviations from smoothness of the flow field. It can be related to linear diffusion with a constant diffusivity. Thus, the flow field is blurred in a homogeneous way such that motion discontinuities may loose sharpness and get dislocated. It is thus not surprising that people have tried to construct a variety of discontinuity-preserving regularisers. Depending on the structure of the resulting diffusion term, we can classify a regulariser $S(\nabla f, \nabla u)$ as image-driven or flow-driven, and isotropic or anisotropic.

For *image-driven* regularisers, S is not only a function of the flow gradient ∇u but also of the image gradient ∇f. This function is chosen in such a way that it respects discontinuities in the image data. If only the gradient *magnitude* $|\nabla f|$ matters, the method is called *isotropic*. It can avoid smoothing at image edges. An *anisotropic* technique depends also on the *direction* of ∇f.

Figure 5: Diffusion filtering of colour images. *(a) Top left:* Noisy color image. *(b) Top right:* Homogeneous diffusion. *(c) Middle left:* Linear isotropic diffusion. *(d) Middle right:* Linear anisotropic diffusion. *(e) Bottom left:* Nonlinear isotropic diffusion. *(f) Bottom right:* Nonlinear anisotropic diffusion. From [89].

Table 6: Vector-valued diffusion processes and their corresponding optic flow regularisers. In the diffusion context, f denotes the vector-valued initial image and u its evolution. In the optic flow setting, f is the scalar-valued image sequence and u describes the optic flow field.

vector-valued diffusion process $\partial_t u_i = \partial_{x_1} S_{u_{ix_1}} + \partial_{x_2} S_{u_{ix_2}}$	optic flow regulariser $S(\nabla f, \nabla u)$
homogeneous $\partial_t u_i = \Delta u_i$ (scalar case: Iijima 1959 [46])	homogeneous $S_1 = \sum_{i=1}^{2} \lvert\nabla u_i\rvert^2$ (Horn/Schunck 1981 [44])
linear isotropic $\partial_t u_i = \text{div}\left(g(\sum_j \lvert\nabla f_j\rvert^2)\,\nabla u_i\right)$ (scalar case: Fritsch 1992 [34])	image-driven, isotropic $S_2 = g(\lvert\nabla f\rvert^2) \sum_{i=1}^{2} \lvert\nabla u_i\rvert^2$ (Alvarez et al. 1999 [2])
linear anisotropic $\partial_t u_i = \text{div}\left(g(\sum_j \nabla f_j \nabla f_j^\top)\,\nabla u_i\right)$ (scalar case: Iijima 1962 [47])	image-driven, anisotropic $S_3 = \sum_{i=1}^{2} \nabla u_i^\top D(\nabla f)\,\nabla u_i$ (Nagel 1983 [60])
nonlinear isotropic $\partial_t u_i = \text{div}\left(\Psi'(\sum_j \lvert\nabla u_j\rvert^2)\,\nabla u_i\right)$ (Gerig et al. 1992 [37])	flow-driven, isotropic $S_4 = \Psi\left(\sum_{i=1}^{2} \lvert\nabla u_i\rvert^2\right)$ (Schnörr 1994 [73])
nonlinear anisotropic $\partial_t u_i = \text{div}\left(\Psi'(\sum_j \nabla u_j \nabla u_j^\top)\,\nabla u_i\right)$ (Weickert 1994 [87])	flow-driven, anisotropic $S_5 = \text{trace}\,\Psi\left(\sum_{i=1}^{2} \nabla u_i \nabla u_i^\top\right)$ (Weickert/Schnörr 2001 [91])

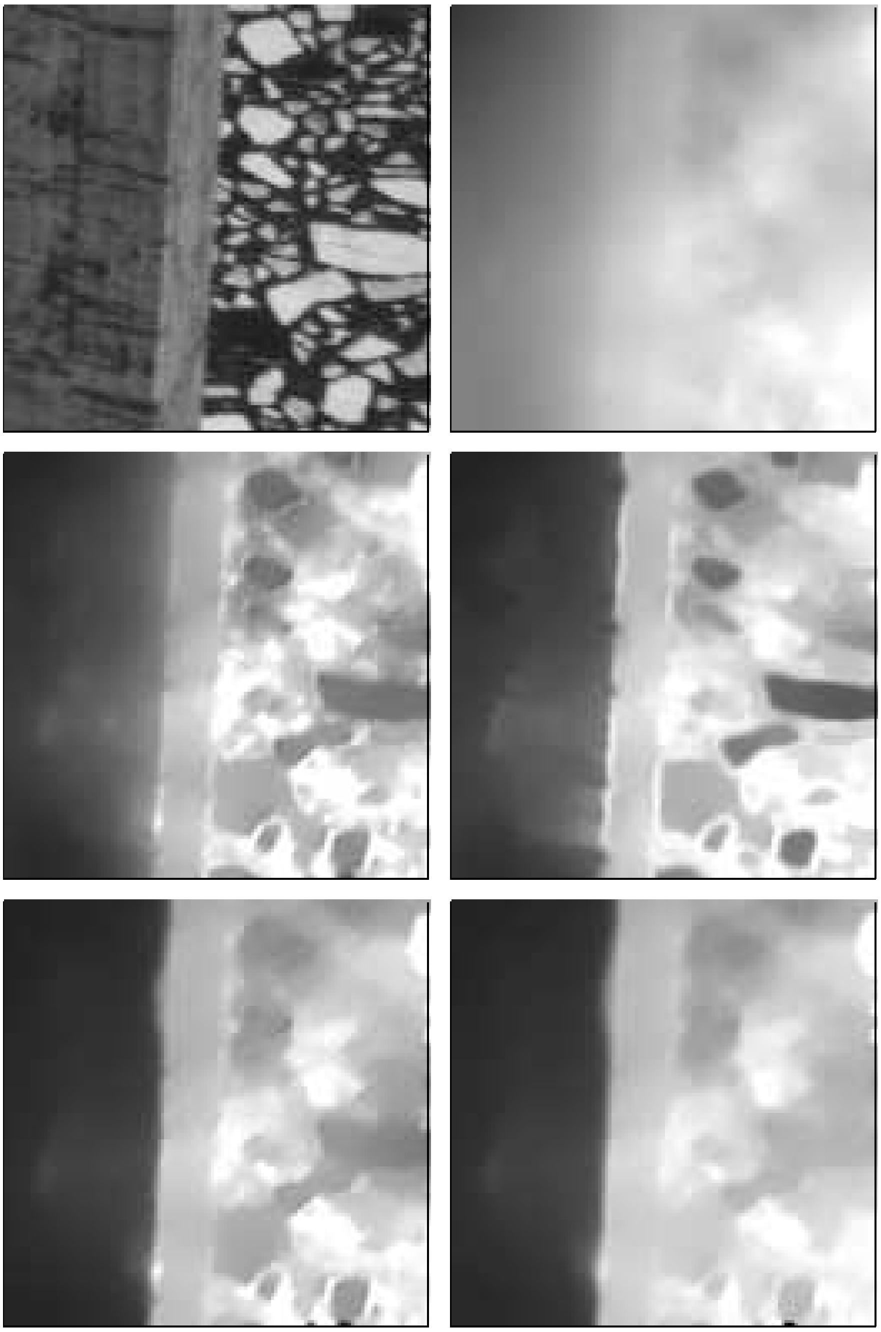

Figure 6: *(a) Top left:* Detail from Frame 16 of the *Marble* sequence (128×128 pixels). *(b) Top right:* Computed optic flow magnitude for a spatial approach with data term M_1 (brightness constancy) and smoothness term S_1 (homogeneous regularisation). *(c) Middle left:* Smoothness term S_2 (image-driven isotropic regularisation). *(d) Middle right:* Smoothness term S_3 (image-driven anisotropic regularisation). *(e) Bottom left:* Smoothness term S_4 (flow-driven isotropic regularisation) *(f) Bottom right:* Smoothness term S_5 (flow-driven anisotropic regularisation). From [91].

Typically it reduces smoothing across edges of f (i.e. along ∇f), while smoothing along edges of f is still permitted. Image-driven regularisers can be related to linear diffusion processes.
Flow-driven regularisers take into account discontinuities of the unknown flow field u by preventing smoothing at or across flow discontinuities. If the resulting diffusion process uses a scalar-valued diffusivity that only depends on $|\nabla u|^2 := |\nabla u_1|^2 + |\nabla u_2|^2$, it is an *isotropic* process. Cases where also the direction of ∇u_1 and ∇u_2 matters are named *anisotropic*. Flow-driven regularisers lead to nonlinear diffusion processes.
Table 6 gives an overview of the different regularisers and their corresponding diffusion filters. As a rule of thumb, one can expect that flow-driven regularisers offer advantages over image-driven ones for highly textured sequences, where the numerous texture edges create an oversegmentation of the flow field. Moreover, anisotropic methods may give somewhat better results than isotropic ones, since the latter ones are too "lazy" at noisy discontinuities.
Figure 6 presents an experiment that illustrates the impact of the smoothness terms we have discussed so far. We compare the regularisers S_1–S_5 from Table 6 within a spatial approach based on the brightness constancy assumption M_1. In order to illustrate their impact on the flow field, we use the 512×512 *Marble* scene by Otte and Nagel. This sequence that is available at `http://i21www.ira.uka.de/image-sequences` consists of 31 frames and requires the estimation of flow discontinuities within a globally translational motion. Figure 6 depicts a zoom into the computed flow fields, where one of these discontinuities is shown. The performance of the different regularisers is not surprising: Homogeneous regularisation is fairly blurry and cannot preserve the discontinuity. Flow-driven and image-driven regularisers perform better whereby the usage of flow information offers advantages in textured regions. And finally, one observes that anisotropic regularisation yields slightly more accurate results than the isotropic one.

4.2 Spatiotemporal Regularisation

While our general functional (2) allows either spatial or spatiotemporal models, the regularisers 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. Let us now investigate what happens if we consider such spatiotemporal models.
Going from spatial to spatiotemporal models is not very difficult in principle: All one has to do is to replace the spatial integration domain Ω in (2) by a spatiotemporal one, and to consider spatiotemporal instead of spatial derivatives. As a resulting steepest descent method, one obtains the three-dimensional diffusion–reaction system

$$\partial_t u_1 = \partial_{x_1} S_{u_{1,x_1}} + \partial_{x_2} S_{u_{1,x_2}} + \partial_{x_3} S_{u_{1,x_3}} - \tfrac{1}{\alpha}\partial_{u_1} M, \quad (41)$$

$$\partial_t u_2 = \partial_{x_1} S_{u_{2,x_1}} + \partial_{x_2} S_{u_{2,x_2}} + \partial_{x_3} S_{u_{2,x_3}} - \tfrac{1}{\alpha}\partial_{u_2} M \quad (42)$$

instead of its two-dimensional counterpart (28)–(29).
In practice, spatiotemporal models have not been used too often so far. An early suggestion for spatiotemporal anisotropic image-driven regularisers goes back to Nagel [61], followed by spatiotemporal flow-driven approaches such as [11, 92]. It appears that the limited memory of previous computer architectures prevented many researchers from studying approaches with spatiotemporal regularisers, since they require to keep the entire image stack in the computer memory. On contemporary PCs, however, these memory requirements are no longer a severe restriction in most cases. With respect to the computing time, the additional requirements are moderate if the entire sequence has to be analysed anyway. Often spatiotemporal models reward

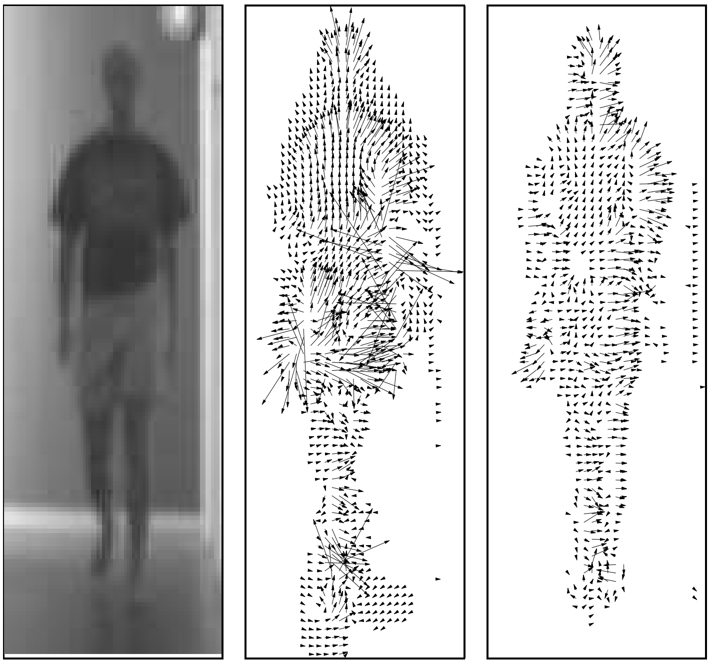

Figure 7: *(a) Left:* Detail of Frame 8 of the *Copenhagen hallway* sequence. *(b) Middle:* Computed flow field for the spatial approach with data term M_1 (brightness constancy) and smoothness term S_4 (isotropic flow-driven regularisation). *(c) Right:* Ditto for the spatiotemporal approach. From [92].

their users by significantly improved optic flow estimates. It is thus likely that spatiotemporal regularisers will become more important in the future.
In Figure 7 we study the effect of replacing spatial by spatiotemporal regularisation. This is done by the example of the 256×256 *Copenhagen hallway* sequence by Olsen and Nielsen. This real-world sequence consists of 16 frames and shows a person who walks along a hallway towards the camera. Comparing the quality of both flow fields, one sees that the additional assumption of temporal smoothness may lead to significantly improved results. In particular the displacements of fast moving body parts such as arms and legs are estimated with a much higher precision.

5 Experiments with Suitable Combinations

In the previous experiments we have focused either on the data or on the smoothness term. Let us now present experiments that illustrate how useful suitable combinations of these terms are. We start by considering a spatial approach with the least square regression data term M_7 and homogeneous regulariser S_1. Then we replace the quadratic penalisers in *both* the data and the smoothness term by nonquadratic penalising functions. Thus, a spatial approach with data term M_{10} and isotropic flow-driven regulariser S_4 is obtained. And finally, the energy functionals of both the original and the modified variant are extended to the spatiotemporal domain.
A comparison of these four approaches is performed in Table 7 where average angular errors for the *Marble* sequence are listed. The improvements of the results thereby clearly show that established concepts in data and smoothness term should be combined in order to obtain the

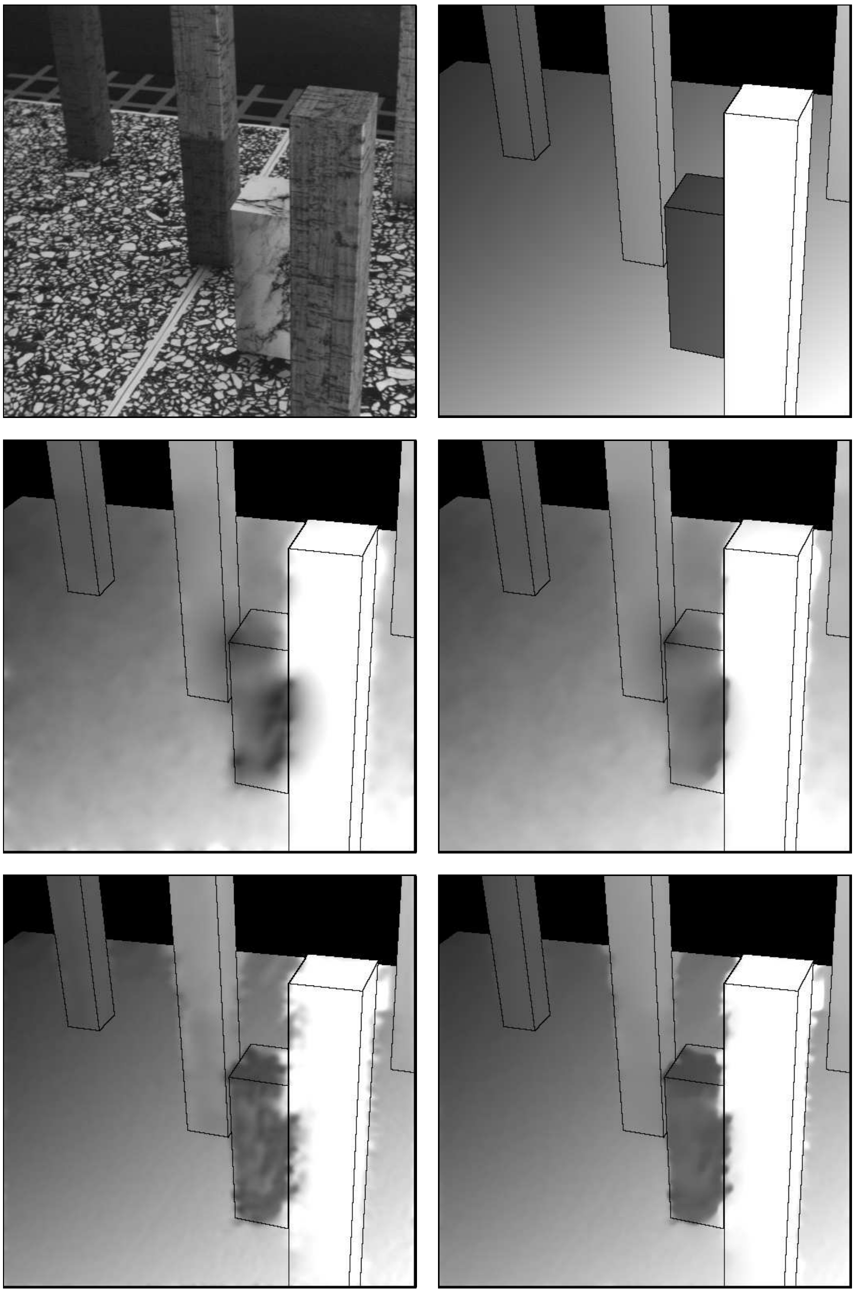

Figure 8: *(a) Top left:* Frame 16 of the *Marble* sequence. *(b) Top right:* Ground truth magnitude. *(c) Middle left:* Computed flow field for a spatial approach with data term M_7 (least squares) and smoothness term S_1 (homogeneous regularisation). *(d) Middle right:* Ditto with data term M_{10} (nonquadratic and least squares) and smoothness term S_4 (isotropic flow-driven regularisation). *(e) Bottom left:* Spatiotemporal approach with data term M_7 (least squares) and smoothness term S_1 (homogeneous regularisation). *(f) Bottom right:* Ditto with data term M_{10} (nonquadratic and least squares) and smoothness term S_4 (isotropic flow-driven regularisation). Adapted from [26].

Table 7: Results for different combinations based on local integration. The average angular error (AAE) has been computed for the *Marble* sequence. Adapted from [26].

approach	data term	smoothness term	AAE
2-D quadratic	M_7	S_1	5.30°
2-D nonquadratic	M_{10}	S_4	5.14°
3-D quadratic	M_7	S_1	2.06°
3-D nonquadratic	M_{10}	S_4	1.70°

Table 8: Comparison between results from the literature with 100 % density and our results using a 3-D functional with data term M_{11} (nonquadratic penalised gradient constancy) and smoothness term S_4 (isotropic flow-driven regulariser). All data refer to the *Yosemite* sequence with cloudy sky. Multiscale means that some focusing strategy using linear scale-space or pyramids has been applied. AAE = average angular error.

technique	multiscale	AAE
Horn/Schunck, original [9]	no	31.69°
Singh, step 1 [9]	no	15.28°
Anandan [9]	no	13.36°
Singh, step 2 [9]	no	10.44°
Nagel [9]	no	10.22°
Horn/Schunck, modified [9]	no	9.78°
Uras *et al.*, unthresholded [9]	no	8.94°
Alvarez/Weickert/Sánchez [3]	yes	5.53°
Mémin/Pérez (IEEE TIP) [56]	yes	5.38°
Bruhn/Weickert/Schnörr [26]	no	5.18°
Mémin/Pérez (ICCV '98) [57]	yes	4.69°
2-D nonquadratic / gradient constancy ($M_{11} + S_4$)	**no**	**3.50°**
3-D nonquadratic / gradient constancy ($M_{11} + S_4$)	**no**	**2.78°**

best performance. This is also confirmed by Figure 8, where we depict the computed flow fields. One can see that each component contributes to the overall improvement: The non-quadratic data term improves the estimation for outliers in the boundary region, the flow-driven isotropic regulariser allows a better preservation of the discontinuities at the marbled blocks and the temporal extension produces a more homogeneous estimation of the floor.

In a second experiment we replace the brightness constancy assumption within M_{10} by the gradient constancy assumption used in M_2. Let us denote this new data term by M_{11}. In Table 8 the resulting spatial and spatiotemporal approach are compared to other methods from the literature, when being applied to the Yosemite sequence with clouds. With 2.78° respectively 3.50° very low average angular errors are obtained[1]. The corresponding flow fields for the spatiotemporal method are depicted in Fig. 9. Obviously, they match the ground truth very well. This shows that sophisticated variational approaches belong to the qualitatively best performing optic flow methods.

[1]This method has been further modified in [18] where it yielded the best results in the literature so far.

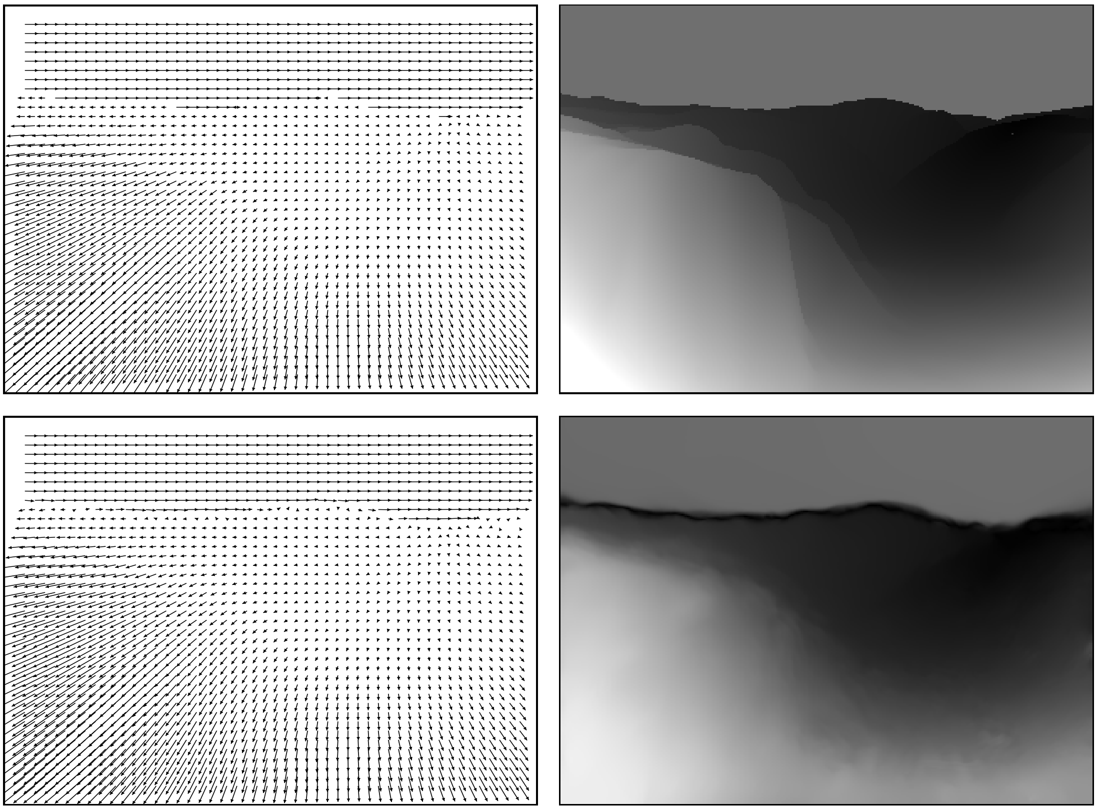

Figure 9: *(a) Top left:* Ground truth for the Yosemite sequence *with* clouds. *(b) Top right:* Magnitude of the ground truth. *(c) Bottom left:* Computed flow field for a spatiotemporal approach with data term M_{11} (nonquadratic gradient constancy) and smoothness term S_4 (isotropic flow-driven regularisation). *(d) Bottom right:* Magnitude of the computed flow field.

6 Well-Posedness Results

One specific advantage of convex variational methods for optic flow computations results from the fact that they allow a rigorous mathematical analysis. As an example, the following result has been proven in [91] for spatial or spatiotemporal energy functionals with the brightness constancy assumption as data term M_1 and any of the smoothness terms S_1,...,S_5:

Theorem (Well-Posedness of Optic Flow Functionals).
Assume that the following properties hold:

(a) The penalising function $\Psi(s^2)$ is differentiable and strictly convex in $s \in \mathbb{R}$.

(b) There exist $c_1, c_2 > 0$ such that $c_1 s^2 \leq \Psi(s^2) \leq c_2 s^2$ for all s.

(c) The initial data are sufficiently smooth: $f \in H^1(\Omega)$.

(d) f_{x_1} and f_{x_2} are linearly independent in $L^2(\Omega)$ and have finite $L^\infty(\Omega)$ norm.

Then the (spatial or spatiotemporal) energy functional

$$E(u) = \int_\Omega \Big(\langle u, \nabla_3 f \rangle^2 \; + \; \alpha \, S_j(\nabla f, \nabla u) \Big) \, dx \qquad (43)$$

with $j \in \{1,...,5\}$ has a unique minimiser $w := (u_1, u_2) \in H^1(\Omega) \times H^1(\Omega) =: \mathcal{H}$. It depends in a continuous way on the image sequence f.

The proof of this theorem combines methods from [75] and from [91] where two essential properties are required:

1. In order to guarantee strict convexity of the smoothness term, a convexity estimate for matrices is needed:
   Let $\Psi : \mathbb{R} \to \mathbb{R}$ be strictly convex, A and B two positive semidefinite symmetric $m \times m$ matrices with $A \neq B$, and $\beta \in (0,1)$. Then

$$\operatorname{trace} \Psi(\beta A + (1-\beta)B) \; < \; \beta \operatorname{trace} \Psi(A) + (1-\beta) \operatorname{trace} \Psi(B). \tag{44}$$

2. On the other hand, strict convexity of the data term requires to address degeneracies by showing that there exists a constant $c > 0$ such that

$$\int_\Omega \Big((\nabla f^\top w)^2 + \gamma |\nabla w|^2 \Big) dx \; \geq \; c \, \|w\|^2_{\mathcal{H}} \,, \qquad \forall w \in \mathcal{H}. \tag{45}$$

It should be noted that such a well-posedness proof is much more than a pure theoretical result: In practise it also guarantees e.g. stability of the optic flow field with respect to noise that perturbs the image data. In this sense it is the real reason behind the high robustness that distinguishes good variational approaches from a number of alternative ways to estimate the optic flow field. For alternative ways to obtain well-posedness results for optic flow functionals we refer to [6, 7, 43].

7 Algorithms

For the numerical minimisation of the energy functional (2), two strategies are used very frequently:
In the first strategy, one discretises the parabolic diffusion–reaction system (28), (29) and recovers the optic flow field as the steady–state solution for $t \to \infty$. The simplest numerical scheme would be an explicit (Euler forward) finite difference scheme [58, 59, 76]. More efficient methods include semi-implicit approaches that offer better stability properties at the expense of the need to solve linear systems of equations.
Alternatively, one can directly discretise the elliptic Euler-Lagrange equations (26), (27), either by finite differences [58, 59, 76] or finite elements [27, 85]. This also requires to solve large linear or nonlinear systems of equations. Efficient methods for this task include *successive overrelaxation (SOR)* methods [84, 94], *preconditioned conjugate gradient (PCG)* algorithms [55, 71] and *multigrid* techniques [16, 17, 40, 80, 93].
Figure 10 illustrates an example of a full multigrid cycle with 4 levels. Such strategies have been used in [22, 23] for finding the minimum of a variational approach with data term M_2 and a homogeneous regulariser. Thus, it was possible to compute up to 18 dense flow fields of size 316×252 pixels on 3.06 GHz Pentium 4 PC within a single second. Table 9 compares the performance of this numerical scheme to widely used iterative solvers like the Gauß-Seidel method or its extrapolated SOR variant. As one can see, the full multigrid cycle is almost three orders of magnitude more efficient than the Gauß-Seidel relaxation scheme and 13 times faster than the SOR method. Even frequently used coarse-to-fine strategies without error correction

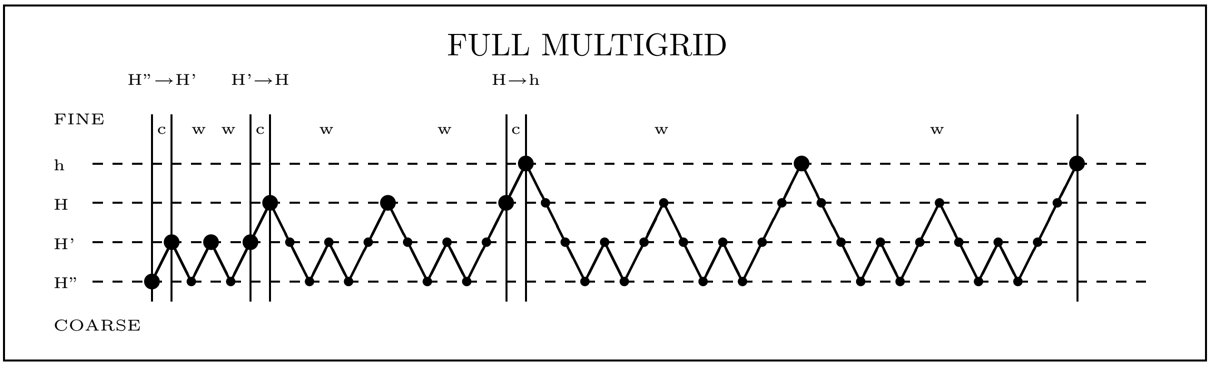

Figure 10: Example of a full multigrid implementation for four levels. Starting from a coarse scale the solution is refined step by step. From [22].

Table 9: Performance benchmark for the 316×252 Yosemite sequence with clouds. FPS = frames per second. Runtimes refer to the computation of all 14 frames with a numerical precision of 10^{-3}. The implementation was done in C on a 3.06 GHz Pentium 4 PC. The obtained average angular error is 7.17°. From [22].

solver	iterations/frame	runtime $[s]$	FPS $[s^{-1}]$	speedup
Gauß–Seidel	21931	543.799	0.026	1
SOR	286	10.140	1.381	54
Gauß–Seidel, coarse-to-fine	237	8.399	1.667	65
SOR, coarse-to-fine	25	1.723	8.125	316
full multigrid	1	0.768	18.229	708

steps are outperformed clearly. This shows that computational efficiency is no problem for variational optic flow methods, when state-of-the-art numerical methods are used.
While this example refers to a quadratic energy functional that leads to linear Euler–Lagrange equations, it is also possible to achieve real-time performance with nonquadratic functionals that give rise to nonlinear Euler–Lagrange equations. This is shown in [24] as well as in [25] where a larger variety of methods is studied.

8 A Simple and General Confidence Measure

While global, energy-based optic flow methods yield dense flow fields due to the filling-in effect, it is clear that the flow estimates cannot have the same reliability at all locations. It would thus be interesting to find a confidence measure that allows to assess the reliability of a dense optic flow field. In 1994 Barron *et al.* [9] have identified the absence of such good measure as one of the main drawbacks of energy-based global optic flow techniques: Simple heuristics such as using $|\nabla f|$ as a confidence measure did not work well. As a remedy, we present a confidence measure that is not only very simple, but also suited for any variational optic flow method. In our description we follow [26].

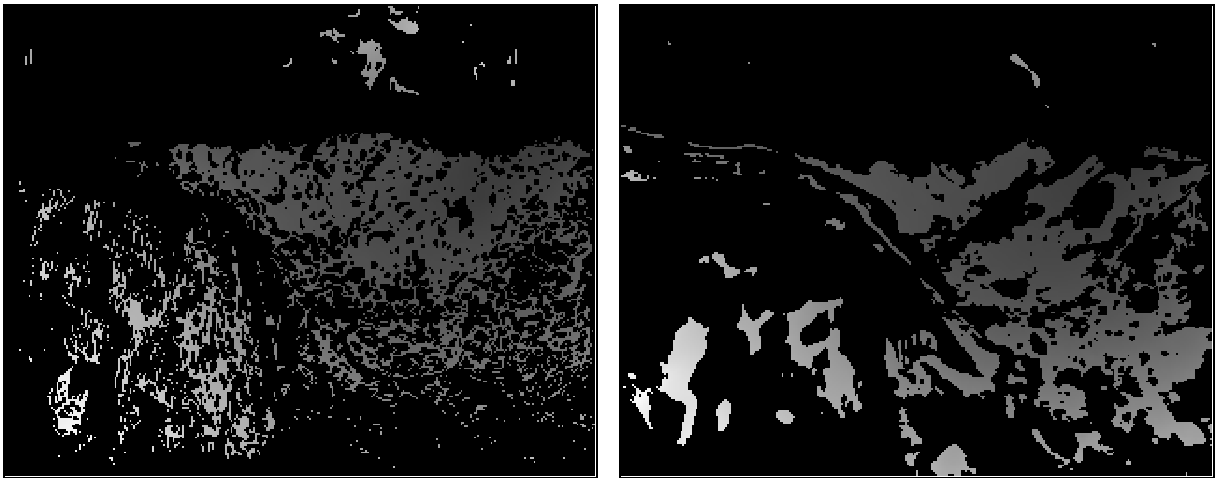

Figure 11: Confidence criterion for the *Yosemite* sequence with clouds. *(a) Left:* Locations with the lowest contributions to the energy (20 % quantile). The non-black grey values depict the optic flow magnitude. *(b) Right:* Locations where the angular error is lowest (20 % quantile).

Since the energy functional E penalises deviations from model assumptions by summing up the deviations E_i from all pixels i in the image domain, it appears natural to use E_i for assessing the local reliability of the computation. All we have to do is to consider the cumulative histogram of the contributions E_i of all pixels $i \in \{1,...,N\}$ in the image domain. As an approximation to the p percent locations with the highest reliability, we look for the p percent locations where the contribution E_i is lowest. There are very efficient algorithms available for this purpose; see e.g. [67, Section 8.5].

Let us now evaluate the quality of our energy-based confidence measure. To this end we consider the spatiotemporal energy functional with the local least square fit data term M_7 and the isotropic flow-driven regulariser S_4. In [26], this technique is named *3-D CLG (combined local–global) method*. Figure 11(a) depicts the 20 % quantile of locations where the 3-D CLG method has lowest contributions to the energy. A comparison with Figure 11(b) – which displays the result of a theoretical confidence measure that would be optimal with respect to the average angular error – demonstrates that the energy-based confidence method leads to a fairly realistic sparsification of flow fields. In particular, we observe that this confidence criterion is very successful in removing the cloudy sky regions. These locations are well-known to create large angular errors in many optic flow methods [9]. A number of authors have thus only used the modified *Yosemite* sequence without cloudy sky, or they have neglected the flow values from the sky region for their evaluations [8, 12, 13, 31, 32, 48, 49, 53, 77]. As we have seen one may get significantly lower angular errors than for the full sequence with cloudy sky.

A quantitative evaluation of our confidence measure is given in Table 10. Here we have used the energy-based confidence measure to sparsify the dense flow field such that the reduced density coincides with densities of well-known optic flow methods. Most of them have been evaluated by Barron *et al.* [9]. We observe that the sparsified 3-D CLG method performs very favourably: It has a far lower angular error than all corresponding methods with the same density. In several cases there is an order of magnitude between these approaches. At a flow density of 2.4 %, an average angular error of 0.76 ° is reached. To our knowledge, these are the best values that have been obtained for this sequence in the entire literature. It should be noted that these results

Table 10: Comparison between the "nondense" results from Barron *et al.* [9], Weber and Malik [86], Ong and Spann [66] and our results for the *Yosemite* sequence with cloudy sky. AAE = average angular error. CLG = average angular error of the 3-D CLG method with the same density. The sparse flow field has been created using our energy-based confidence criterion. The table shows that using this criterion clearly outperforms all results in the evaluation of Barron *et al.*

Technique	Density	AAE	CLG
Singh, step 2, $\lambda_1 \leq 0.1$	97.7 %	10.03°	6.04°
Ong/Spann	89.9 %	5.76°	5.26°
Heeger, level 0	64.2 %	22.82°	3.00°
Weber/Malik	64.2 %	4.31°	3.00°
Horn/Schunck, original, $\|\nabla f\| \geq 5$	59.6 %	25.33°	2.72°
Ong/Spann, tresholded	58.4 %	4.16°	2.66°
Heeger, combined	44.8 %	15.93°	2.07°
Lucas/Kanade, $\lambda_2 \geq 1.0$	35.1 %	4.28°	1.71°
Fleet/Jepson, $\tau = 2.5$	34.1 %	4.63°	1.67°
Horn/Schunck, modified, $\|\nabla f\| \geq 5$	32.9 %	5.59°	1.63°
Nagel, $\|\nabla f\| \geq 5$	32.9 %	6.06°	1.63°
Fleet/Jepson, $\tau = 1.25$	30.6 %	5.28°	1.55°
Heeger, level 1	15.2 %	9.87°	1.15°
Uras *et al.*, $\det(H) \geq 1$	14.7 %	7.55°	1.14°
Singh, step 1, $\lambda_1 \leq 6.5$	11.3 %	12.01°	1.07°
Waxman *et al.*, $\sigma_f = 2.0$	7.4 %	20.05°	0.95°
Heeger, level 2	2.4 %	12.93°	0.76°

have been computed from an image sequence that suffers from quantisation errors since its grey values have been stored in 8-bit precision only.

In Table 10 we also observe that the angular error decreases *monotonically* under sparsification over the entire range from 100 % down to 2.4 %. This in turn indicates an interesting finding that may seem counterintuitive at first glance: *Regions in which the filling-in effect dominates give particularly small angular errors.* In such flat regions, the data term vanishes such that a smoothly extended flow field may yield only a small local contribution to the energy functional. If there were large angular errors in regions with such low energy contributions, our confidence measure would not work well for low densities. This also confirms the observation that $|\nabla f|$ is not necessarily a good confidence measure [9]: Areas with large gradients may represent noise or occlusions, where reliable flow information is difficult to obtain. The filling-in effect, however, may create more reliable information in flat regions by averaging less reliable information that comes from all the surrounding high-gradient regions. A more extensive experimental evaluation of the energy based confidence measure is presented in [21].

9 Summary and Extensions

In this chapter we have outlined some basic design principles for variational optic flow methods and studied their performance in a number of experiments. For theoretical and practical reasons we have restricted ourselves to convex energy functionals that use linearised data terms. They are valid approximations when the temporal sampling is sufficiently fine such that the displacements between subsequent frames are small. We have seen that contemporary variational optic flow models have reached a high degree of sophistication that allows to achieve highly accurate computations of the displacement fields. Moreover, they are mathematically well-founded, they allow real-time computations on standard hardware, and it is possible to apply a simple and intuitive confidence measure.

There are several possibilities to improve the performance of these methods even further: One may for instance use data terms that renounce linearisations [3, 11, 62]. They create models that are better suitable for large displacements between subsequent frames. Unfortunately they lead to nonconvex functionals that may possess numerous local minimisres. In such a case one often uses multilevel strategies that encourage convergence towards a global minimiser [3, 4, 56]. Another extension that becomes relevant for large displacements consists of using modified functionals in order to deal with occlusion problems [1, 68]. On the numerical side, parallelisation strategies can be investigated, e.g. domain decomposition methods [51]. A detailed discussion of these extensions is beyond the scope of the present chapter.

It is our hope that the models we have described do not remain restricted to optic flow computation, but will also prove their use in related correspondence problems such as stereo reconstruction and image registration.

Acknowledgements. Our research has partly been funded by the *Deutsche Forschungsgemeinschaft (DFG)* and the *Graduiertenkolleg "Leistungsgarantien für Rechnersysteme"*. This is gratefully acknowledged.

References

[1] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez. Symmetrical dense optical flow estimation with occlusion detection. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 721–736. Springer, Berlin, 2002.

[2] 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, September 1999.

[3] 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, August 2000.

[4] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[5] F. Andreu–Vaillo, V. Caselles, and J. M. Mazon. *Parabolic Quasilinaer Equations Minimizing Linear Growth Functionals*, volume 223 of *Progress in Mathematics*. Birkhäuser, Basel, 2004.

[6] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal on Applied Mathematics*, 60(1):156–182, 1999.

[7] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer, New York, 2002.

[8] A. Bab-Hadiashar and D. Suter. Robust optic flow computation. *International Journal of Computer Vision*, 29(1):59–77, August 1998.

[9] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[10] 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(8):775–790, August 1991.

[11] M. J. 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.

[12] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[13] M. J. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):972–986, October 1996.

[14] 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, March 1998.

[15] A. Borzi, K. Ito, and K. Kunisch. Optimal control formulation for determining optical flow. *SIAM Journal on Scientific Computing*, 24(3):818–847, 2002.

[16] A. Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31(138):333–390, April 1977.

[17] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, Philadelphia, second edition, 2000.

[18] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.

[19] T. Brox, R. van den Boomgaard, F. Lauze, J. van de Weijer, J. Weickert, R. Mrázek, and P. Kornprobst. Adaptive structure tensors and their applications. In J. Weickert and H. Hagen, editors, *Visualization and Image Processing of Tensor Fields*. Springer, Berlin, 2005. To appear.

[20] T. Brox and J. Weickert. Nonlinear matrix diffusion for optic flow estimation. In L. Van Gool, editor, *Pattern Recognition*, volume 2449 of *Lecture Notes in Computer Science*, pages 446–453. Springer, Berlin, 2002.

[21] A. Bruhn and J. Weickert. Confidence measures for variational optic flow methods. In R. Klette, R. Kozera, L. Noakes, and J. Weickert, editors, *Geometric Properties from Incomplete Data*, Computational Imaging and Vision. Springer, Dordrecht, 2005. To appear.

[22] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr. Real-time optic flow computation with variational methods. In N. Petkov and M. A. Westenberg, editors, *Computer Analysis of Images and Patterns*, volume 2756 of *Lecture Notes in Computer Science*, pages 222–229. Springer, Berlin, 2003.

[23] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr. Variational optic flow computation in real-time. *IEEE Transactions on Image Processing*, 14(5):608–615, May 2005.

[24] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. Discontinuity-preserving computation of variational optic flow in real-time. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale-Space and PDE Methods in Computer Vision*, volume 3459 of *Lecture Notes in Computer Science*, pages 585–597, Berlin, 2005. Springer.

[25] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. Technical report, Dept. of Mathematics, Saarland University, Saarbrücken, Germany, May 2005. Submitted to *International Journal of Computer Vision*.

[26] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[27] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. SIAM, Philadelphia, 2002.

[28] 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.

[29] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience, New York, 1953.

[30] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics and Image Processing*, 33:116–125, 1986.

[31] G. Farnebäck. Fast and accurate motion estimation using orientation tensors and parametric motion models. In *Proc. 15th International Conference on Pattern Recognition*, volume 1, pages 135–139, Barcelona, Spain, September 2000.

[32] G. Farnebäck. Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field. In *Proc. Eighth International Conference on Computer Vision*, volume 1, pages 171–177, Vancouver, Canada, July 2001. IEEE Computer Society Press.

[33] 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.

[34] 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.

[35] 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, September 1998.

[36] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Dover, New York, 2000.

[37] 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.

[38] S. Ghosal and P. Č. Vaněk. Scalable algorithm for discontinuous optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2):181–194, February 1996.

[39] F. Glazer. Multilevel relaxation in low-level computer vision. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, pages 312–330. Springer, Berlin, 1984.

[40] W. Hackbusch. *Multigrid Methods and Applications*. Springer, New York, 1985.

[41] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. MIT Press, Cambridge, MA, 1986.

[42] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1217–1232, December 1993.

[43] W. Hinterberger, O. Scherzer, C. Schnörr, and J. Weickert. Analysis of optical flow models in the framework of calculus of variations. *Numerical Functional Analysis and Optimization*, 23(1/2):69–89, May 2002.

[44] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[45] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[46] T. Iijima. Basic theory of pattern observation. In *Papers of Technical Group on Automata and Automatic Control*. IECE, Japan, December 1959. In Japanese.

[47] T. Iijima. Observation theory of two-dimensional visual patterns. In *Papers of Technical Group on Automata and Automatic Control*. IECE, Japan, October 1962. In Japanese.

[48] S. Ju, M. Black, and A. Jepson. Skin and bones: multi-layer, locally affine, optical flow and regularization with transparency. In *Proc. 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 307–314, San Francisco, CA, June 1996. IEEE Computer Society Press.

[49] J. Karlholm. *Local Signal Models for Image Sequence Analysis*. PhD thesis, Linköping University, Sweden, 1998. Dissertation No. 536.

[50] 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, September 2000.

[51] T. Kohlberger, C. Schnörr, A. Bruhn, and J. Weickert. Parallel variational motion estimation by domain decomposition and cluster computing. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 205–216. Springer, Berlin, 2004.

[52] A. Kumar, A. R. Tannenbaum, and G. J. Balas. Optic flow: a curve evolution approach. *IEEE Transactions on Image Processing*, 5(4):598–610, April 1996.

[53] S.-H. Lai and B. C. Vemuri. Reliable and efficient computation of optical flow. *International Journal of Computer Vision*, 29(2):87–105, October 1998.

[54] 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, August 1981.

[55] A. Meister. *Numerik linearer Gleichungssysteme*. Vieweg, Braunschweig, 1999.

[56] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703–719, May 1998.

[57] E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proc. Sixth International Conference on Computer Vision*, pages 933–938, Bombay, India, January 1998. Narosa Publishing House.

[58] A. R. Mitchell and D. F. Griffiths. *The Finite Difference Method in Partial Differential Equations*. Wiley, Chichester, 1980.

[59] K. W. Morton and L. M. Mayers. *Numerical Solution of Partial Differential Equations*. Cambridge University Press, Cambridge, UK, 1994.

[60] 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.

[61] 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.

[62] 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.

[63] H.-H. Nagel and A. Gehrke. Spatiotemporally adaptive estimation and segmentation of OF-fields. In H. Burkhardt and B. Neumann, editors, *Computer Vision – ECCV '98*, volume 1407 of *Lecture Notes in Computer Science*, pages 86–102. Springer, Berlin, 1998.

[64] M. Z. Nashed and O. Scherzer. Least squares and bounded variation regularization with nondifferentiable functionals. *Numerical Functional Analysis and Optimization*, 19:873–901, 1998.

[65] P. Nesi. Variational approach to optical flow estimation managing discontinuities. *Image and Vision Computing*, 11(7):419–439, September 1993.

[66] E. P. Ong and M. Spann. Robust optical flow computation based on least-median-of-squares regression. *International Journal of Computer Vision*, 31(1):51–82, 1999.

[67] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, second edition, 1992.

[68] 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.

[69] A. R. Rao and B. G. Schunck. Computing oriented texture fields. *CVGIP: Graphical Models and Image Processing*, 53:157–185, 1991.

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

[71] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, second edition, 2003.

[72] G. Sapiro. *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press, Cambridge, UK, 2001.

[73] 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, October 1994. IEEE Computer Society Press.

[74] 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.

[75] 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.

[76] G. D. Smith. *Numerical Solution of Partial Differential Equations: Finite Difference Methods*. Clarendon Press, Oxford, third edition, 1985.

[77] R. Szeliski and J. Coughlan. Hierarchical spline-based image registration. In *Proc. 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 194–201, Seattle, WA, June 1994. IEEE Computer Society Press.

[78] D. Terzopoulos. Image analysis using multigrid relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):129–139, March 1986.

[79] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady*, 4:1035–1038, 1963.

[80] U. Trottenberg, C. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, San Diego, 2001.

[81] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 948–953, Kauai, HI, December 2001. IEEE Computer Society Press.

[82] D. Tschumperlé and R. Deriche. Diffusion PDE's on vector-valued images. *IEEE Signal Processing Magazine*, 19(5):16–25, 2002.

[83] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60:79–87, 1988.

[84] R. A. Varga. *Matrix Iterative Analysis*. Springer, New York, second edition, 2000.

[85] R. Wait and A. R. Mitchell. *Finite Element Analysis and Applications*. Wiley, Chichester, 1985.

[86] 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.

[87] J. Weickert. Scale-space properties of nonlinear diffusion filtering with a diffusion tensor. Technical Report 110, Laboratory of Technomathematics, University of Kaiserslautern, Germany, October 1994.

[88] J. Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17(3–4):199–210, March 1999.

[89] J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M. Z. Nashed and O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging*, volume 313 of *Contemporary Mathematics*, pages 251–268. AMS, Providence, 2002.

[90] J. Weickert, A. Bruhn, N. Papenberg, and T. Brox. Variational optic flow computation: From continuous models to algorithms. In L. Alvarez, editor, *IWCVIA '03: International Workshop on Computer Vision and Image Analysis*, volume 0026 of *Cuardernos del Instituto Universitario de Ciencias y Technologías Cibernéticas*, pages 1–6, Spain, February 2004.

[91] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, December 2001.

[92] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, May 2001.

[93] P. Wesseling. *An Introduction to Multigrid Methods*. R. T. Edwards, Flourtown, 2004.

[94] D. M. Young. *Iterative Solution of Large Linear Systems*. Dover, New York, 2003.

[95] G. Zini, A. Sarti, and C. Lamberti. Application of continuum theory and multi-grid methods to motion evaluation from 3D echocardiography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 44(2):297–308, March 1997.