In X.-C. Tai, E. Bae, T. F. Chan, M. Lysaker (Eds.):
Energy Minimization Methods in Computer Vision and Pattern Recognition.
Lecture Notes in Computer Science, Springer, Vol. 8932, 263-277, Berlin, 2015.
The final publication is available at link.springer.com

Justifying Tensor-Driven Diffusion from Structure-Adaptive Statistics of Natural Images

Pascal Peter[1], Joachim Weickert[1], Axel Munk[2],
Tatyana Krivobokova[3], and Housen Li[2]

[1] Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science,
Campus E1.7, Saarland University, 66041 Saarbrücken, Germany.
{peter, weickert}@mia.uni-saarland.de
[2] Felix-Bernstein-Chair for Mathematical Statistics, Institute of Mathematical Stochastics,
Goldschmidtstrasse 7, 37077 Göttingen, Germany
munk@math.uni-goettingen.de, housen.li@stud.uni-goettingen.de
[3] Statistical Methods Group, Courant Research Centre "Poverty, Equity and Growth",
Wilhelm-Weber-Str. 2, 37073 Göttingen, Germany
tkrivob@uni-goettingen.de

Abstract. Tensor-driven anisotropic diffusion and regularisation have been successfully applied to a wide range of image processing and computer vision tasks such as denoising, inpainting, and optical flow. Empirically it has been shown that anisotropic models with a diffusion tensor perform better than their isotropic counterparts with a scalar-valued diffusivity function. However, the reason for this superior performance is not well understood so far. Moreover, the specific modelling of the anisotropy has been carried out in a purely heuristic way. The goal of our paper is to address these problems. To this end, we use the statistics of natural images to derive a unifying framework for eight isotropic and anisotropic diffusion filters that have a corresponding variational formulation. In contrast to previous statistical models, we systematically investigate structure-adaptive statistics by analysing the eigenvalues of the structure tensor. With our findings, we justify existing successful models and assess the relationship between accurate statistical modelling and performance in the context of image denoising.

Keywords: diffusion, regularisation, anisotropy, diffusion tensor, statistics of natural images, image priors

1 Introduction

Anisotropic diffusion and regularisation models involve a positive definite 2×2 matrix called diffusion tensor. Its eigenvalues steer the amount of data propagation in the direction of the corresponding eigenvector. Throughout more than two decades of research,

such anisotropic methods have been successfully used for a large number of image processing and computer vision problems. These tasks include denoising [3], inpainting [24], image compression [6], optical flow computation [16], stereo reconstruction [31], and shape from shading [1]. Application domains cover e.g. computer aided quality control [25], medical image processing [14], and seismic image analysis [8].

To this date, modelling nonlinear diffusion filters is a heuristic, task-driven procedure, where images are processed towards a certain goal. It is a well-known fact that anisotropic models can be much more powerful in certain applications than isotropic diffusion approaches with a scalar-valued diffusivity function. Clearly, one reason for the success of anisotropic concepts are their additional degrees of freedom which can be adapted to a task at hand. However, another potential explanation for this success is still unexplored: Could it be that smoothness assumptions of anisotropic models reflect statistical properties of natural images more accurately than isotropic ones?

For specific *isotropic* diffusion models, there exists a well-known connection to probabilistic filter models based on the statistics of natural images [29]: There is a negative logarithmic correspondence between natural image priors and regularisation terms in variational models. However, in particular for *anisotropic* diffusion, previous investigations have focused on isolated, specific models in practical contexts such as parameter learning. In particular, there is a lack of a cohesive theory that systematically analyses the correspondence between probabilistic filters and diffusion filters that can be expressed by energy minimisation.

Our Contributions. The goal of our paper is to provide a justification of tensor-driven diffusion models via the statistics of natural images. We aim at systematically assessing the differences between isotropic and anisotropic approaches from a probabilistic perspective. To this end, we use natural image priors to derive a unifying framework that incorporates eight existing diffusion filters that have a corresponding variational formulation. In order to cover the full range of nonlinear models, these statistics have to reflect the local image structure and allow to involve directional information. The eigenvalue statistics of the structure tensor in databases of natural images provide not only such information, but also offer a lot of flexibility to generate a wide range of derivative-based priors. This allows us to construct probabilistic filters that represent existing isotropic and anisotropic filter classes and analyse the differences in the underlying priors. We discuss the implications of these differences on filter performance in the context of image denoising.

Related Work. At its core, our work relies on the non-Gaussian nature of the histograms that result from applying filters to natural images. For wavelet coefficients, these specific attributes were first reported by Huang and Mumford [5]. These observations were systematically investigated for both derivative filters and wavelet coefficients in [10]. Invariances of these statistics are vital for their practical relevance. Zhu and Mumford proposed that these statistical priors are invariant to scale and verified this empirically in [29]. Evaluations on databases containing different motives were conducted by Huang and Mumford in [9]. For more details on the statistics of natural images, we refer to the recent monograph of Pouli et al. [19].

General connections between diffusion processes and statistical image processing models have been pioneered by Zhu and Mumford [29] within a Gibbs diffusion–

reaction framework. Later on, Roth and Black [20] have found additional relations in the context of fields of experts. Works considering anisotropic diffusion models are, however, very rare. In the context of parameter learning, Scharr et al. [22] introduced an anisotropic model with Gaussian derivatives. A more recent parameter-free model goes back to Krajsek and Scharr [12]. They consider a two step procedure. In the first step, an isotropic diffusion process is derived. Afterwards, this is used to construct a *linear* anisotropic regularisation model. More recently, Kunisch and Pock [13] have analysed parameter learning for regularisation methods with a bilevel optimisation scheme.

Organisation of the Paper. We start with a brief overview of existing tensor-driven diffusion in Section 2. In Section 3, we investigate the properties of the structure tensor as an image feature and use it to derive a probabilistic denoising filter in Section 4. We show that this model is related to a unifying framework for diffusion filtering in Section 5. In Section 6 we investigate diffusion models that are learned from a database, evaluate their performance for denoising and interpret the results. Finally, we present our conclusions and outlook on future work in Section 7.

2 Tensor-Driven Diffusion Processes

Let us start by reviewing a number of isotropic and anisotropic diffusion filters which can be derived from a general energy functional that we present in Section 5.

General Structure. Let $\boldsymbol{f} = (f_1, ..., f_m)^\top$ represent a vector-valued image with m channels. Each of these channels is a function $f_k : \Omega \to \mathbb{R}$ that maps the rectangular image domain $\Omega \subset \mathbb{R}^2$ to the colour value range $\mathbb{R}$. A tensor-driven, vector-valued diffusion process computes filtered versions $\{\boldsymbol{u}(x,y,t) \,|\, (x,y) \in \Omega, t \geq 0\}$ of $\boldsymbol{f}(x,y)$ as solutions of the diffusion equation

$$\partial_t u_k = \boldsymbol{\nabla}^\top (\boldsymbol{D} \boldsymbol{\nabla} u_k) \quad \text{on} \quad \Omega \times (0,\infty), \quad k = 1, \ldots, m \tag{1}$$

with $\boldsymbol{u}(x,y,0) = \boldsymbol{f}(x,y)$ as initial condition on Ω, and reflecting boundary conditions:

$$\langle \boldsymbol{D} \boldsymbol{\nabla} u_k, \boldsymbol{n} \rangle = 0 \quad \text{on} \quad \partial\Omega \times (0,\infty), \quad k = 1, \ldots, m. \tag{2}$$

The diffusion time t serves as a scale parameter: Larger times yield simpler image representations. The nabla operator $\boldsymbol{\nabla}$ and the divergence operator $\boldsymbol{\nabla}^\top$ involve spatial derivatives only, and $\boldsymbol{n}$ denotes the outer normal vector to the image boundary $\partial\Omega$. The diffusion tensor $\boldsymbol{D}$ is a positive definite 2×2 matrix that steers the diffusion. Its eigenvalues specify the amount of diffusion in the direction of the eigenvectors.

Isotropic Models. The simplest diffusion process, *homogeneous* diffusion [11], is obtained for $\boldsymbol{D} := \boldsymbol{I}$ with a unit matrix $\boldsymbol{I}$. In this case, the diffusion does not depend on the image structure. For more sophisticated *nonlinear isotropic* diffusion models the diffusion tensor is of the form $\boldsymbol{D} := g(|\boldsymbol{\nabla} u|^2)\boldsymbol{I}$. If one wants to permit strong smoothing within homogeneous regions and inhibit smoothing across edges, one chooses the diffusivity $g(|\boldsymbol{\nabla} u|^2)$ as a decreasing positive function of its argument. Many diffusivity

functions have been proposed, e.g. the Perona/Malik diffusivity g_{PM} [18] or the Charbonnier diffusivity g_{C} [2]:

$$g_{\mathrm{PM}}(s^2) := \left(1 + \frac{s^2}{\lambda^2}\right)^{-1}, \qquad g_{\mathrm{C}}(s^2) := \left(1 + \frac{s^2}{\lambda^2}\right)^{-1/2}. \tag{3}$$

Note that locations where $|\nabla u| \gg \lambda$ are regarded as edges where the diffusivity is close to 0, while we have full diffusion in regions with $|\nabla u| \ll \lambda$. Therefore, $\lambda > 0$ acts as a contrast parameter. Isotropic models allow space-variant smoothing, but due to their scalar-valued diffusivity, the diffusion process acts in the same way in all directions. The first isotropic nonlinear model goes back to Perona and Malik [18] and is designed for greyscale images. Gerig et al. [7] have extended it to colour image processing by coupling the evolution of the individual channels through a diffusivity of the form $g(\sum_{k=1}^{m} |\nabla u_k|^2)$. Scherzer and Weickert [23] have investigated an isotropic nonlinear diffusion model where all spatial gradients ∇ are replaced by Gaussian-smoothed gradients $\nabla_\sigma := K_\sigma * \nabla$. Here K_σ denotes a Gaussian with standard deviation σ.

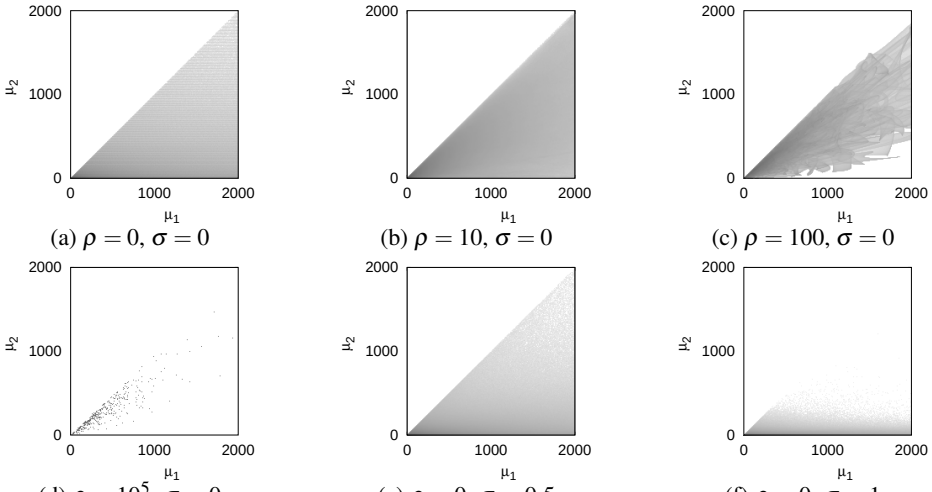
Anisotropic Models. In order to model direction-dependent diffusion processes, we need an anisotropic diffusion tensor $\boldsymbol{D}$ whose eigenvalues can differ significantly. These eigenvalues and their corresponding eigenvectors are adapted to the local image structure. A popular descriptor of the local image geometry is the structure tensor of Di Zenso [4]. In its most sophisticated form, it is given by the symmetric positive semidefinite matrix

$$\boldsymbol{J}_{m,\rho,\sigma} := K_\rho * \Big(\sum_{k=1}^{m} \nabla_\sigma u_k \nabla_\sigma u_k^\top \Big) \tag{4}$$

with eigenvalues $\mu_{1,\rho,\sigma} \geq \mu_{2,\rho,\sigma} \geq 0$. The corresponding diffusion tensor $\boldsymbol{D} := g(\boldsymbol{J}_{m,\rho,\sigma})$ uses the same set of eigenvectors and obtains its eigenvalues as functions of $\mu_{1,\rho,\sigma}$ and $\mu_{2,\rho,\sigma}$. The anisotropic models of Weickert/Brox [27] and Tschumperlé/Deriche [24] do not incorporate any smoothing in the structure tensor (i.e. $\sigma = \rho = 0$). However, such models degenerate to isotropic diffusion on greyscale images ($m = 1$). The methods of Roussos/Maragos [21] and Scharr et al. [22] involve a smoothing scale $\rho > 0$ and remain also anisotropic for $m = 1$. While Roussos/Maragos use $\sigma = 0$, Scharr et al. consider the case $\sigma > 0$ and replace all gradients ∇ by their Gaussian-smoothed counterparts ∇_σ.

3 Structure-Adaptive Analysis of the Berkeley Database

Interpretation of the Structure Tensor. The local image structure of a vector-valued image $\boldsymbol{u}$ with m channels can be characterised by the joint structure tensor from Eq. (4). Its eigenvalues $\mu_1 \geq \mu_2$ represent the local contrast in the direction of the corresponding eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. For $\mu_1 \gg \mu_2$, the eigenvector $\boldsymbol{v}_2$ describes the direction of coherent structures while $\boldsymbol{v}_1$ points across these structures. Locally isotropic image content is characterised by $\mu_1 \approx \mu_2$. Thus, the eigenvalues of the structure tensor are image features that describe local geometry.



(a) $\rho = 0$, $\sigma = 0$ (b) $\rho = 10$, $\sigma = 0$ (c) $\rho = 100$, $\sigma = 0$

(d) $\rho = 10^5$, $\sigma = 0$ (e) $\rho = 0$, $\sigma = 0.5$ (f) $\rho = 0$, $\sigma = 1$

Fig. 1. Evolution of the negative logarithmic histogram of the eigenvalue pairs $(\mu_{1,\rho,\sigma}, \mu_{2,\rho,\sigma})$ of the structure tensor $\boldsymbol{J}_{m,\rho,\sigma}$ over different scales ρ and σ. Dark values indicate high occurrences and bright values low occurrences.

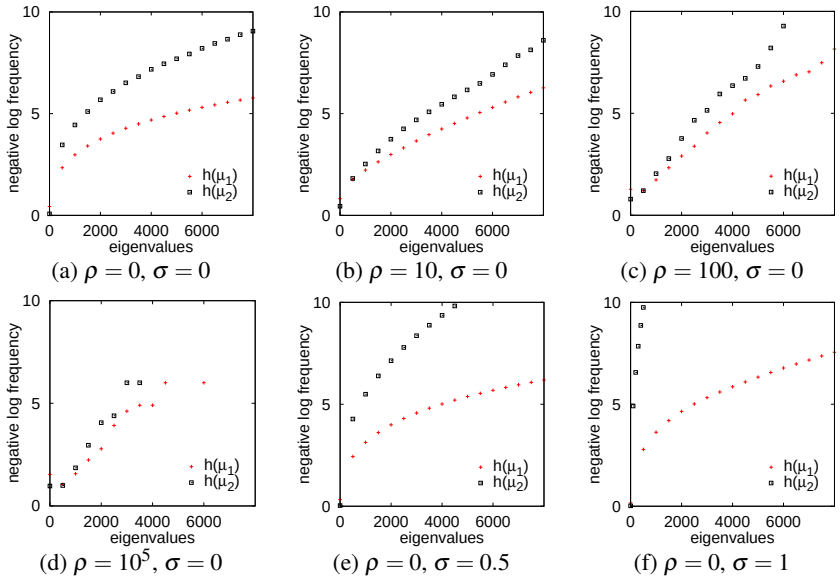
The Gaussian smoothing scales σ and ρ play distinct roles for the analysis of local image structure: Smoothing with K_σ removes noise and small-scale details. Thus, it should be chosen as small as possible. The smoothing scale ρ is usually chosen to be larger since its task is to accumulate neighbourhood information in the structure tensor.

In our implementation of the structure tensor, we use the finite difference discretisation from [28] with a parameter $\alpha = 1/6$. Its leading error term is rotationally invariant.

Anisotropic Statistics of Colour Images. Let us now use the aforementioned structure tensor for a statistical analysis of the Berkeley database [15]. The histogram of the eigenvalue pairs (μ_1, μ_2) with $\sigma = \rho = 0$ is displayed in Fig. 1(a). The fact that the eigenvalue μ_1 clearly dominates and there are many structure tensors where μ_2 is significantly smaller confirms two things: Firstly, colour images contain many strongly oriented structures which legitimates the use of anisotropic filters. Secondly, these structures have some correlations over the colour channels. Fig. 2(a) reveals that both eigenvalues have the heavy-tailed distributions that are characteristic for filter results on natural images. Such kurtotic distributions are captured well by the function

$$\psi(x^2) = \frac{\lambda^2}{1-\gamma}\left(1 + \frac{x^2}{\lambda^2}\right)^{1-\gamma}. \tag{5}$$

The free parameters λ and γ can be adapted to fit ψ to the discrete histograms. A related model with one more degree of freedom was also proposed in [12]. Similar statistics have been shown to be nearly identical on many databases of natural images such as the Berkeley [15] or McGill [17] test sets. In particular, they are also invariant for image content on different scales. Therefore, they form a good prior for natural images. This scale invariance implies that the statistics do hardly change under subsampling.



(a) $\rho = 0$, $\sigma = 0$ (b) $\rho = 10$, $\sigma = 0$ (c) $\rho = 100$, $\sigma = 0$

(d) $\rho = 10^5$, $\sigma = 0$ (e) $\rho = 0$, $\sigma = 0.5$ (f) $\rho = 0$, $\sigma = 1$

Fig. 2. Evolution of the negative logarithmic histograms $h(\mu_1)$, $h(\mu_2)$ of the eigenvalues $\mu_{1,\rho,\sigma}$, $\mu_{2,\rho,\sigma}$ of the structure tensor $\boldsymbol{J}_{m,\rho,\sigma}$ over different scales ρ and σ.

Behaviour under Smoothing. If one averages with overlapping neighbourhoods, the statistics depend significantly on the neighbourhood size. This happens for the Gaussian-smoothed structure tensor $\boldsymbol{J}_{m,\rho,\sigma}$, where the tensor entries are embedded in a Gaussian scale-space. Let us first fix σ and consider the scale-space behaviour with respect to ρ. Fig. 1(a)–(d) shows the evolution of the histogram for the eigenvalue pairs $(\mu_{1,\rho,\sigma},\ \mu_{2,\rho,\sigma})$. We observe that for increasing ρ, the joint histogram clusters towards the diagonal. This shows that $\mu_{1,\rho,\sigma}$ and $\mu_{2,\rho,\sigma}$ approach each other, i.e. the structure tensor becomes more isotropic. This is plausible, since one smoothes over structures with different orientations. For $\rho \to \infty$, all tensors $\boldsymbol{J}_{m,\rho,\sigma}$ converge to the average structure tensor of the whole image. If all directions were equally prominent over the database, this average tensor would be purely isotropic . However, the steady state of the statistics ($\rho = 10^5$ in Fig. 1(d) and Fig. 2(d)) reveals some anisotropy. Thus, the average eigenvalue histograms show the inherent directional bias of the image database.

Now we fix ρ and investigate the evolution under σ. For $\sigma \to \infty$, the local contrast given by $\mu_{1,\rho,\sigma}$ and $\mu_{2,\rho,\sigma}$ approaches 0 and the corresponding diffusion tensor $\boldsymbol{D}$ converges to the unit matrix $\boldsymbol{I}$. Interestingly, Figs. 1(e)–(f) and 2(e)–(f) show that for small σ, the presmoothing increases the difference between the histograms of $\mu_{1,\rho,\sigma}$ and $\mu_{2,\rho,\sigma}$. This fosters anisotropy of the image prior. We conjecture that Gaussian convolution effectivly removes high-frequent isotropic perturbations, such that anisotropic image structures become more dominant. For larger σ their dominance decreases again.

In conclusion, we observe that natural images contain pronounced anisotropies and their statistics strongly depend on the smoothing scales ρ and σ. This suggest to design filters that take into account such anisotropic phenomena as priors.

4 Probabilistic Denoising with a Structure Tensor Prior

We can use the statistics from Section 3 as a prior for a Bayesian denoising approach. Let a discrete, noisy image $\boldsymbol{f}$ of size $M \times N$ with m channels be given. The goal is to compute an approximation $\boldsymbol{u}$ to the original image $\boldsymbol{v}$ under two assumptions: $\boldsymbol{v}$ belongs to the class of natural images and is degraded by Gaussian noise. For any image $\boldsymbol{u}$, let $p(\boldsymbol{u})$ be the natural image prior. It describes the probability that $\boldsymbol{u}$ is a natural image and is derived from the statistics of image features on a suitable database. Furthermore, an assumption on the distribution of the noise yields the noise prior $p(\boldsymbol{f}|\boldsymbol{u})$. According to Bayes' rule, the posterior probability for a candidate image $\boldsymbol{u}$ to be the ground truth to an observed noisy image $\boldsymbol{f}$ obeys

$$p(\boldsymbol{u}|\boldsymbol{f}) \sim p(\boldsymbol{f}|\boldsymbol{u}) \cdot p(\boldsymbol{u}). \tag{6}$$

Thus, the optimally denoised image $\hat{\boldsymbol{u}}$ can be obtained by maximising the posterior probability $p(\boldsymbol{u}|\boldsymbol{f})$ over all candidates $\boldsymbol{u}$:

$$\hat{\boldsymbol{u}} = \operatorname*{argmax}_{\boldsymbol{u}} p(\boldsymbol{u}|\boldsymbol{f}). \tag{7}$$

Since we assume independent identically distributed Gaussian noise for each channel k with $k \in \{1,...,m\}$, the noise prior is given by

$$p(\boldsymbol{f}|\boldsymbol{u}) \sim \prod_{k=1}^{m} \prod_{i=1}^{M} \prod_{j=1}^{N} \exp\Big(-\frac{1}{2\sigma^2}(u_{k,i,j} - f_{k,i,j})^2\Big). \tag{8}$$

In order to formulate a natural image prior, we follow the minimax entropy model that has been used to model texture [30] and whole images [29]. For a set of given linear or nonlinear filters $\{\boldsymbol{F}_1, \ldots, \boldsymbol{F}_L\}$ the distribution of natural images is modelled as

$$p(\boldsymbol{u}) = \prod_{\ell=1}^{L} \prod_{i=1}^{N} \prod_{j=1}^{M} \phi_\ell(\boldsymbol{F}_\ell(\boldsymbol{u})_{i,j}). \tag{9}$$

Here the potential functions ϕ_ℓ model the distribution of the corresponding filter $\boldsymbol{F}_\ell$. Current state-of-the-art models like the fields of experts approach [20] use specifically learned linear filters as a feature set. Interestingly many of these learned filters resemble derivative filters as was shown in [20].

Let $\phi(\mu_1, \mu_2)$ define the distribution of an arbitrary image feature that is derived from the eigenvalues μ_1 and μ_2 of $\boldsymbol{J}_{m,\rho,\sigma}$. In particular, this formulation also includes separate statistics for both eigenvalues, i.e. $\phi(\mu_1, \mu_2) := \phi_1(\mu_1) \cdot \phi_2(\mu_2)$. Such image features can be interpreted as *second-level priors* in the terminology of [29], since they model the local geometry of image structures. In particular, these priors adapt to dominant directions in the image in contrast to linear filters that approximate derivatives in a fixed, global direction. By specifying the natural image prior (9) with a feature based on μ_1 and μ_2 and including the noise prior (8) we obtain the following energy:

$$\boxed{E_P(\boldsymbol{u}) = \prod_{i=1}^{M} \prod_{j=1}^{N} \left(\prod_{k=1}^{m} \left(\exp\Big(-\frac{(u_{k,i,j} - f_{k,i,j})^2}{2\sigma^2}\Big)\right) \cdot \phi(\mu_{1,i,j}, \mu_{2,i,j})\right).} \tag{10}$$

Maximising E_P gives the denoised image $\hat{\boldsymbol{u}}$.

Model	m	σ	ρ	PDE	Penaliser	Prior
Homogeneous Diffusion [11]	1	0	0	$\partial_t u = \nabla^\top \nabla u$	$\psi(\mu_1+\mu_2) = \lvert\nabla u\rvert^2$	$\phi(\mu_1+\mu_2) = e^{-\lvert\nabla u\rvert^2}$
Perona/ Malik [18]	1	0	0	$\partial_t u = \nabla^\top \left(\psi'(\lvert\nabla u\rvert^2)\right)\nabla u)$	$\psi(\mu_1+\mu_2) = -\ln\phi(\lvert\nabla u\rvert^2)$	$\phi(\mu_1+\mu_2) = \phi(\lvert\nabla u\rvert^2)$
Gerig et al. [7]	≥ 1	0	0	$\partial_t u_k = \nabla^\top(\psi'(\sum_{\ell=1}^m \lvert\nabla u_\ell\rvert^2)\nabla u_k)$	$\psi(\mu_1+\mu_2) = -\ln\phi(\sum_{\ell=1}^m \lvert\nabla u_\ell\rvert^2)$	$\phi(\mu_1+\mu_2) = \phi\left(\sum_{\ell=1}^m \lvert\nabla u_\ell\rvert^2\right)$
Scherzer/ Weickert [23]	1	≥ 0	0	$\partial_t u = \nabla_\sigma^\top\left(\psi'(\lvert\nabla_\sigma u\rvert^2)\nabla_\sigma u\right)$	$\psi(\mu_1+\mu_2) = -\ln\phi(\lvert\nabla_\sigma u\rvert^2)$	$\phi(\mu_1+\mu_2) = \phi(\lvert\nabla_\sigma u\rvert^2)$
Weickert/ Brox [27]	>1	0	0	$\partial_t u_k = \nabla^\top\left(\left(\psi'(\mu_1)v_1v_1^\top + \psi'(\mu_2)v_2v_2^\top\right)\nabla u_k\right)$	$\psi(\mu_{1,2}) = -\ln\phi(\mu_{1,2})$	$\phi(\mu_1)\cdot\phi(\mu_2)$
Tschumperlé/ Deriche [24]	>1	0	0	$\partial_t u_k = \nabla^\top\left(\left(\frac{\partial\psi(\mu_1,\mu_2)}{\partial\mu_1}v_1v_1^\top + \frac{\partial\psi(\mu_1,\mu_2)}{\partial\mu_2}v_2v_2^\top\right)\nabla u_k\right)$	$\psi(\mu_1,\mu_2) = -\ln\phi(\mu_1,\mu_2)$	$\phi(\mu_1,\mu_2)$
Roussos/ Maragos [21]	≥ 1	0	>0	$\partial_t u_k = \nabla^\top\left(K_\rho * \left(\frac{\partial\psi(\mu_1,\mu_2)}{\partial\mu_1}v_1v_1^\top + \frac{\partial\psi(\mu_1,\mu_2)}{\partial\mu_2}v_2v_2^\top\right)\nabla u_k\right)$	$\psi(\mu_1,\mu_2) = -\ln\phi(\mu_1,\mu_2)$	$\phi(\mu_1,\mu_2)$
Scharr et al. [22]	1	≥ 0	>0	$\partial_t u = \nabla_\sigma^\top\left(K_\rho * \left(\psi_1'(\mu_1)v_1v_1^\top + \psi_2'(\mu_2)v_2v_2^\top\right)\nabla_\sigma u\right)$	$\psi_{1,2}(\mu_{1,2}) = -\ln\phi_{1,2}(\mu_{1,2})$	$\phi_1(\mu_1)\cdot\phi_2(\mu_2)$

Table 1. Existing diffusion models and their relation to the unifying framework. The models primarily differ in the number of image channels m, the smoothing scale ρ for the structure tensor and the presmoothing scale σ. Additionally, most models impose certain restrictions to the general prior $\phi(\mu_1,\mu_2)$. The priors and penalisers are always given in the most general form, that the corresponding model allows. Note that the Roussos/Maragos model has the same prior structure as Tschumperlé/Deriche, but yields a different PDEs due to the nonzero smoothing scale ρ.

5 The Unifying Prior-Based Diffusion Framework

Let us now show that the probabilistic denoising model (10) is the discrete counterpart to a unifying diffusion framework that incorporates a large family of existing diffusion approaches. Instead of maximising the energy E_P, we consider the minimisation of its negative logarithm

$$E_{\log}(\boldsymbol{u}) := \frac{1}{2}\sum_{i=1}^{M}\sum_{i=1}^{N}\Big(\sum_{k=1}^{m}\frac{1}{\tau}(u_{k,i,j}-f_{k,i,j})^2+\psi(\mu_{1,i,j},\mu_{2,i,j})\Big). \tag{11}$$

Here, we define the penaliser ψ as $\psi(\mu_1,\mu_2) = -\log\phi(\mu_1,\mu_2)$, and we choose $\tau\sim\sigma^2$. A variational regularisation approach is obtained by the minimisation of the continuous counterpart to $E_{\log}$:

$$\boxed{E(\boldsymbol{u}) = \frac{1}{2}\int_\Omega\Big(\frac{1}{\tau}|\boldsymbol{u}-\boldsymbol{f}|^2+\psi(\mu_1,\mu_2)\Big)\,dxdy} \tag{12}$$

where $|\cdot|$ denotes the Euclidean norm. Interestingly, this energy provides a unifying framework for the eight diffusion models from Section 2. The key result for understanding this connection is given by the following proposition.

Proposition 1 (Euler-Lagrange Equations of the General Energy Functional). *The energy functional $E(\boldsymbol{u})$ from Eq.* (12) *gives rise to the Euler–Lagrange equations*

$$\frac{u_k-f_k}{\tau} \;=\; \boldsymbol{\nabla}_\sigma^\top\big((K_\rho * \boldsymbol{D})\;\boldsymbol{\nabla}_\sigma u_k\big), \qquad k=1,...,m, \tag{13}$$

with natural boundary conditions $\boldsymbol{n}^\top(K_\sigma * K_\rho * \boldsymbol{D}\boldsymbol{\nabla}_\sigma u_k) = 0$ on $\partial\Omega$. Here, $\boldsymbol{n}$ is the outer image normal and $\boldsymbol{D}$ is given in terms of the eigenvectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and eigenvalues μ_1, μ_2 of the structure tensor $\boldsymbol{J}_{m,\sigma,\rho}$:

$$\boldsymbol{D} \;:=\; \frac{\partial\psi(\mu_1,\mu_2)}{\partial\mu_1}\,\boldsymbol{v}_1\boldsymbol{v}_1^\top \;+\; \frac{\partial\psi(\mu_1,\mu_2)}{\partial\mu_2}\,\boldsymbol{v}_2\boldsymbol{v}_2^\top. \tag{14}$$

Proof. The Euler-Lagrange equations are obtained from the Gâteaux derivatives of $E(\boldsymbol{u})$. We focus on the derivative of the penaliser ψ. with $d_{\varepsilon_k}(f) := \frac{\partial}{\partial\varepsilon_k}f|_{\varepsilon_k=0}$, $k\in\{1,...,m\}$, a test function $\boldsymbol{h}:\mathbb{R}^2\mapsto\mathbb{R}^m$, $\boldsymbol{H} := \mathbf{diag}(\boldsymbol{h})$, and $\boldsymbol{\epsilon}\in\mathbb{R}^m$ we calculate:

$$d_{\varepsilon_k}\big(\psi(\mu_1(\boldsymbol{u}+\boldsymbol{H}\boldsymbol{\epsilon}),\mu_2(\boldsymbol{u}+\boldsymbol{H}\boldsymbol{\epsilon}))\big) \;=\; \frac{\partial\psi}{\partial\mu_1}d_{\varepsilon_k}(\mu_1) \;+\; \frac{\partial\psi}{\partial\mu_2}d_{\varepsilon_k}(\mu_2). \tag{15}$$

Therefore, the derivatives of the eigenvalues μ_1 and μ_2 of $\boldsymbol{J}_{m,\rho,\sigma}$ must be computed. In terms of the matrix elements $J_{1,1}$, $J_{1,2}$, $J_{2,2}$, the eigenvalue μ_1 is given by

$$\mu_1 = \frac{1}{2}\Big(J_{1,1}+J_{2,2}+\sqrt{(J_{1,1}-J_{2,2})^2+4J_{1,2}^2}\Big). \tag{16}$$

By writing the derivatives $d_{\varepsilon_k}(J_{1,2})$, $d_{\varepsilon_k}(J_{1,1}+J_{2,2})$, and $d_{\varepsilon_k}(J_{1,1}-J_{2,2})$ as dot products with $\boldsymbol{\nabla}_\sigma h_k$, we can simplify $d_{\varepsilon_k}(\mu_1)$ to

$$d_{\varepsilon_k}(\mu_1) \;=\; K_\rho * \Big((\boldsymbol{M}\boldsymbol{\nabla}_\sigma u_k)^\top \boldsymbol{\nabla}_\sigma h_k \Big), \tag{17}$$

$$\boldsymbol{M} \;:=\; \frac{2}{\mu_1-\mu_2} \begin{pmatrix} \mu_1-\mu_2+J_{1,1}-J_{2,2} & 2J_{1,2} \\ 2J_{1,2} & \mu_1-\mu_2-J_{1,1}+J_{2,2} \end{pmatrix}. \tag{18}$$

Algebraic computations similar to [21] lead to $\boldsymbol{M} = 2\boldsymbol{v}_1\boldsymbol{v}_1^\top$. With analogous results for $d_{\varepsilon_k}(\mu_2)$, we obtain $d_{\varepsilon_k}(\psi(\mu_1,\mu_2)) = (K_\rho * \boldsymbol{D}\boldsymbol{\nabla}_\sigma u_k)^\top \boldsymbol{\nabla}_\sigma h_k$. Plugging these results into the Gâteaux derivative $d_{\varepsilon_k}E$ of the energy and applying partial integration yields

$$d_{\varepsilon_k}E = \sum_{\ell=1}^{2} \Big[\Big(K_\sigma * K_\rho * \boldsymbol{D}\boldsymbol{\nabla}_\sigma u_k \Big)_\ell h_k \Big]_{a_\ell}^{b_\ell} - \int_\Omega \boldsymbol{\nabla}_\sigma^\top \Big((K_\rho * \boldsymbol{D}) \boldsymbol{\nabla}_\sigma u_k \Big) h_k \, dxdy \tag{19}$$

with $\Omega = [a_1,b_1]\times[a_2,b_2]$. Variational calculus yields Eq. (13) and the natural boundary conditions $\boldsymbol{n}^\top (K_\sigma * K_\rho * \boldsymbol{D}\boldsymbol{\nabla}_\sigma u_k) = 0$ on $\partial\Omega$. □

According to [23], Eq. (13) can be interpreted as an implicit time discretisation with one time step of size τ of the general diffusion equation

$$\boxed{\partial_t u_k \;=\; \boldsymbol{\nabla}_\sigma^\top \Big((K_\rho * \boldsymbol{D}) \, \boldsymbol{\nabla}_\sigma u_k \Big), \quad k = 1,\dots,m} \tag{20}$$

with initial condition $\boldsymbol{u}(t=0) = \boldsymbol{f}$. In Table 1 we demonstrate that a large number of existing diffusion models can be considered as special cases of this unifying partial differential equation. To see this, note that the isotropic models use $\rho = 0$ and the prior

$$\phi(\mu_1+\mu_2) \;=\; \phi\left(\mathrm{tr}\boldsymbol{J}_{m,0,\sigma}\right) \;=\; \phi\Big(\sum_{\ell=1}^{m} |\boldsymbol{\nabla}_\sigma u_\ell|^2 \Big). \tag{21}$$

Moreover, for greyscale images ($m=1$) and smoothing scale $\rho = 0$, the structure tensor $\boldsymbol{J}_{1,0,\sigma} = \boldsymbol{\nabla}_\sigma u \boldsymbol{\nabla}_\sigma u^\top$ has the normalised eigenvectors $\boldsymbol{v}_1 = \frac{\boldsymbol{\nabla}_\sigma u}{|\boldsymbol{\nabla}_\sigma u|}$ and $\boldsymbol{v}_2 = \boldsymbol{v}_1^\perp$. As a consequence, the diffusion process from Eq. (20) degenerates to isotropic diffusion with a scalar diffusivity: Using (14) we get

$$\begin{aligned} \boldsymbol{D}\boldsymbol{\nabla}_\sigma u \;&=\; \Big(\frac{\partial\psi}{\partial\mu_1} \frac{\boldsymbol{\nabla}_\sigma u \boldsymbol{\nabla}_\sigma u^\top}{|\boldsymbol{\nabla}_\sigma u|^2} \;+\; \frac{\partial\psi}{\partial\mu_2} \frac{\boldsymbol{\nabla}_\sigma u^\perp \boldsymbol{\nabla}_\sigma u^{\perp\top}}{|\boldsymbol{\nabla}_\sigma u|^2} \Big) \boldsymbol{\nabla}_\sigma u \\ &=\; \frac{\partial\psi}{\partial\mu_1} \boldsymbol{\nabla}_\sigma u \;=\; \psi'\left(|\boldsymbol{\nabla}_\sigma u|^2\right) \boldsymbol{\nabla}_\sigma u. \end{aligned} \tag{22}$$

Homogeneous diffusion is also captured by the model (20), if one chooses $\phi(|\boldsymbol{\nabla}u|^2) := \exp(-|\boldsymbol{\nabla}u|^2)$ as prior distribution. The four anisotropic models are covered as follows: Weickert/Brox [27] and Scharr et al. [22] use the factorised prior $\phi_1(\mu_1)\cdot\phi_2(\mu_2)$, in the case of Weickert/Brox with identical functions ϕ_1 and ϕ_2 and $\sigma = \rho = 0$. The models of Tschumperlé/Deriche [24] and Roussos/Maragos [21] allow general priors $\phi(\mu_1,\mu_2)$, but specify $\sigma := 0$. Moreover, Tschumperlé/Deriche also set $\rho := 0$.

The whole framework was derived from a common natural image prior, the directional statistics of the structure tensor. This shows that the ad hoc choices that were made for diffusion models during decades of research in fact reflect inherent properties of natural images. This observation can be even extended to the choice of diffusivities: If we consider the special case $\phi(\mu_1,\mu_2) = \phi_1(\mu_1)\cdot\phi_2(\mu_2)$, we are able to decompose $\psi(\mu_1,\mu_2) := \psi_1(\mu_1)+\psi_2(\mu_2)$ into two separate penalisers $\psi_\ell = -\ln\phi_\ell$ with $\ell\in\{1,2\}$. The kurtotic distribution model (5) gives rise to the following family of diffusivities:

$$\psi'(x^2) \;=\; \left(1+\frac{x^2}{\lambda^2}\right)^{-\gamma}. \tag{23}$$

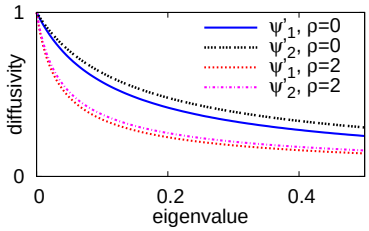
Comparing this to (3) shows that the Perona/Malik diffusivity [18] is covered for $\gamma = 1$ and the Charbonnier diffusivity [2] results for $\gamma = 0.5$. To the best of our knowledge, our framework covers all relevant diffusion models that offer a variational interpretation. Since it is a variational framework, it is natural that it cannot be applied to models for which no variational formulation is known, e.g. edge- and coherence-enhancing diffusion filters [26].

6 Denoising Experiments

In the following, we compare the performance of different diffusion models in the context of image denoising. We focus on those models from Table 1 that are designed for colour images and apply small modifications where necessary: In analogy to [7], we extend the Scherzer/Weickert model to colour images by coupling the gradient within a joint diffusivity. Furthermore, we use separate penalisers $\psi_1(\mu_1)$ and $\psi_2(\mu_2)$ for the anisotropic models. This extends the Weickert/Brox model with individual diffusivities for both eigenvalues, which is a special case of the Tschumperlé/Deriche model. In the accompanying figures we use the abbreviations H for homogeneous diffusion [11], GKKJ for Gerig et al. [7], SW for Scherzer/Weickert [23], WBTD for the hybrid model of Weickert/Brox [27] and Tschumperlé/Deriche [24], RM for Roussos/Maragos [21], and SBH for a vector-valued extension of Scharr et al. [22].

For our experiments, we first determine the parameters λ and γ of the prior distribution (5) and the corresponding diffusivity (23). To this end, we compute the discrete histograms of μ_1 and μ_2 on the 200 training images of the Berkeley database [15]. For a nonlinear least squares fit to these histograms, we have chosen the Matlab implementation of the Levenberg–Marquardt algorithm (version 3.2.1 of the Matlab curve fitting toolbox). In Fig. 3(a) we see that the resulting diffusivities decrease more rapidly for μ_1 than for μ_2. Thus, they inhibit diffusion across coherent structures more than along them. For increasing smoothing scales σ and ρ this anisotropic behaviour is reduced, since the difference between the diffusivities ψ_1' and ψ_2' is less pronounced. In the following, we use $\rho = 0.5$ and $\sigma = 0.2$.

For our denoising experiments, we consider the partial differential formulation of the statistically-derived diffusion filters and apply them to the 100 images of the Berkeley test set [15] with added Gaussian noise. The average peak signal to noise (PSNR) values for different standard deviations of the noise are given in Fig. 3(b). We observe that for all noise levels, homogenous diffusion H yields the worst results, and



Model	γ_1	λ_1	γ_2	λ_2	$n=20$	$n=40$	$n=60$
H	-	-	-	-	27.26	24.40	22.67
GKKJ	0.756	0.340	-	-	29.08	25.59	23.45
SW	0.754	0.336	-	-	29.09	25.64	23.52
WBTD	0.752	0.334	0.694	0.326	29.58	25.88	23.63
RM	0.733	0.304	0.644	0.231	29.56	25.99	23.76
SBH	0.739	0.315	0.646	0.234	29.66	26.05	23.80

Fig. 3. (a) Left: Diffusivities estimated for the eigenvalues μ_1 and μ_2 on the Berkeley database for different smoothing scales. **(b) Right:** Diffusivity parameters and denoising results for different diffusion models on the Berkeley test set. See Section 6 for the abbreviations. In the last three columns, the average PSNR for Gaussian noise with standard deviation n is given.

the isotropic methods GKKJ and SW perform consistently below the anisotropic models WBTD, RM and SBH. With increasing noise levels, the Gaussian smoothing scales σ and ρ within the models SW, RM and SBH offer a slight PSNR advantage over their counterparts GKKJ and WBTD that have to cope without Gaussian smoothing. Visually, the most distinct difference is the severe blurring of edges in homogenous diffusion that sets it apart from the other models.

Let us now interpret these findings from a probabilistic modelling perspective. The performance ranking according to the PSNR mirrors the accuracy of the underlying natural image priors. In particular, the large gap between homogenous diffusion and the rest of the models is caused by the wrongly assumed Gaussian-like distribution of the underlying image prior $\mu_1 + \mu_2 = |\nabla u|^2$ in model H (see Tab. 1). Since all of the remaining filters accurately reproduce the kurtotic shape of the prior distributions, they perform much better. Finally, the inherent directional bias in natural image models is only respected by the anisotropic models WBTD, RM and SBH, which gives them a consistent advantage over the isotropic models GKKJ and SW.

7 Conclusion and Outlook

We have presented a unifying framework for eight diffusion filters that have a corresponding variational formulation. It enabled us to derive these models from probabilistic filters with a structure tensor prior. We have verified experimentally that those filters which model the structure adaptive statistics of natural images more accurately also offer a better performance in practice. This justifies their use in digital image processing and computer vision, and it establishes a hitherto unknown reason for the success of anisotropic filters. From a statistical viewpoint, we have emphasised the importance of directional statistics that take into account the local image structure and its scale dependency. Interestingly, our statistical foundation of tensor-driven diffusion gives also additional insights that go beyond a pure statistical foundation of existing models: For instance, it sheds light on how the decay function of each eigenvalue should be adapted to the smoothing scales of the structure tensor.

Our results give rise to a number of ongoing and future activities. We are focussing our current research on anisotropic models that are tailored optimally to the statistics of

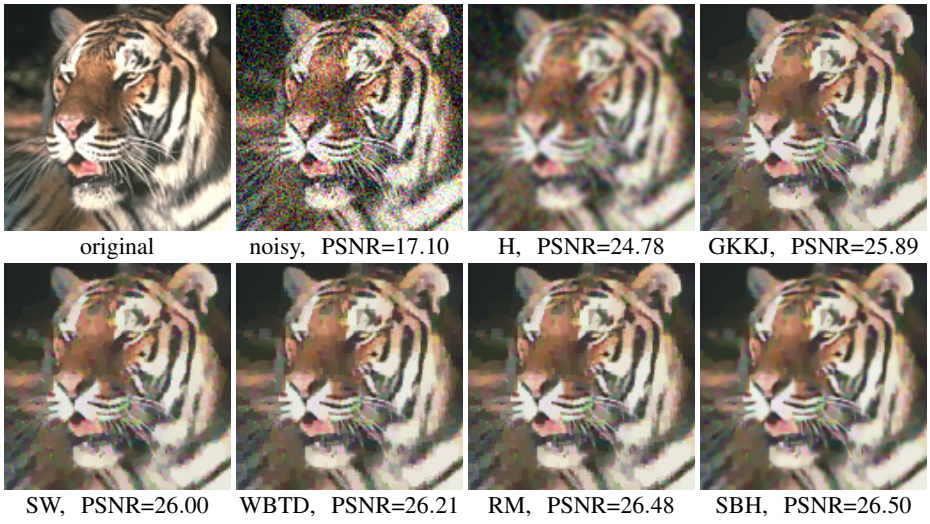


Fig. 4. Denoising results for image *108082* of the Berkeley test set with Gaussian noise of standard deviation $n = 40$ for the models L, GKKJ, SW, WBTD, RM and SBH. The PSNR is given for the whole image, but only a zoom of size 128×128 is shown.

natural images in a specific application context. Moreover, we expect that our framework can also be extended to novel energy functionals once they will be discovered for other important classes of anisotropic diffusion filters.

References

1. Agrawal, A., Raskar, R., Chellappa, R.: What is the range of surface reconstructions from a gradient field? In: Bischof, H., Leonardis, A., Pinz, A. (eds.) Computer Vision – ECCV 2006, Part I, Lecture Notes in Computer Science, vol. 3951, pp. 578–591. Springer, Berlin (2006)
2. Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proc. 1994 IEEE International Conference on Image Processing. vol. 2, pp. 168–172. IEEE Computer Society Press, Austin, TX (Nov 1994)
3. Cottet, G.H.: Diffusion approximation on neural networks and applications for image processing. In: Hodnett, F. (ed.) Proc. Sixth European Conference on Mathematics in Industry, pp. 3–9. Teubner, Stuttgart (1992)
4. Di Zenzo, S.: A note on the gradient of a multi-image. Computer Vision, Graphics and Image Processing 33, 116–125 (1986)
5. Field, D.J.: Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America A 4(12), 2379–2394 (1987)
6. Galić, I., Weickert, J., Welk, M., Bruhn, A., Belyaev, A., Seidel, H.P.: Image compression with anisotropic diffusion. Journal of Mathematical Imaging and Vision 31(2–3), 255–269 (2008)
7. Gerig, G., Kübler, O., Kikinis, R., Jolesz, F.A.: Nonlinear anisotropic filtering of MRI data. IEEE Transactions on Medical Imaging 11, 221–232 (1992)

8. Höcker, C., Fehmers, G.: Fast structural interpretation with structure-oriented filtering. The Leading Edge 21(3), 238–243 (Mar 2002)
9. Huang, J., Mumford, D.: Image statistics for the British Aerospace segmented database. Tech. rep., Divison of Applied Math, Brown University, Providence (1999)
10. Huang, J., Mumford, D.: Statistics of natural images and models. In: Proc. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 1541–1547. IEEE Computer Society Press, Ft. Collins, CO (Jun 1999)
11. Iijima, T.: Basic theory on normalization of pattern (in case of typical one-dimensional pattern). Bulletin of the Electrotechnical Laboratory 26, 368–388 (1962), in Japanese
12. Krajsek, K., Scharr, H.: Diffusion filtering without parameter tuning: Models and inference tools. In: Proc. 2010 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2536–2543. IEEE Computer Society Press, San Francisco (CA) (Jun 2010)
13. Kunisch, K., Pock, T.: A bilevel optimization approach for parameter learning in variational models. SIAM Journal on Imaging Sciences 6(2), 938–983 (2013)
14. Manniesing, R., Viergever, M.A., Niessen, W.J.: Vessel enhancing diffusion: A scale space representation of vessel structures. Medical Image Analysis 10, 815–825 (2006)
15. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. Eigth International Conference on Computer Vision. pp. 416–423. Vancouver, Canada (Jul 2001)
16. Nagel, H.H., Enkelmann, W.: 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)
17. Olmos, A., Kingdom, F.: A biologically inspired algorithm for the recovery of shading and reflectance images. Perception 33(12), 1463–1473 (2004)
18. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 629–639 (1990)
19. Pouli, T., Reinhard, E., Cunningham, D.W.: Image Statistics in Visual Computing. CRC Press, Boca Raton (2013)
20. Roth, S., Black, M.J.: Fields of experts: A framework for learning image priors. In: Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 2, pp. 860–867. IEEE Computer Society Press, San Diego, CA (Jun 2005)
21. Roussos, A., Maragos, P.: Tensor-based image diffusions derived from generelizations of the total variation and Beltrami functionals. In: Proc. 17th IEEE International Conference on Image Processing. pp. 4141–4144. Hong Kong (Sep 2010)
22. Scharr, H., Black, M.J., Haussecker, H.W.: Image statistics and anisotropic diffusion. In: Proc. Ninth International Conferance on Computer Vision. vol. 2, pp. 840–847. IEEE Computer Society Press, Nice, France (2003)
23. Scherzer, O., Weickert, J.: Relations between regularization and diffusion filtering. Journal of Mathematical Imaging and Vision 12(1), 43–63 (Feb 2000)
24. Tschumperlé, D., Deriche, R.: Vector-valued image regularization with PDEs: A common framework for different applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(4), 506–516 (Apr 2005)
25. Weickert, J.: Anisotropic diffusion filters for image processing based quality control. In: Fasano, A., Primicerio, M. (eds.) Proc. Seventh European Conference on Mathematics in Industry, pp. 355–362. Teubner, Stuttgart (1994)
26. Weickert, J.: Anisotropic Diffusion in Image Processing. Teubner, Stuttgart (1998)
27. Weickert, J., Brox, T.: Diffusion and regularization of vector- and matrix-valued images. In: Nashed, M.Z., Scherzer, O. (eds.) Inverse Problems, Image Analysis, and Medical Imaging, Contemporary Mathematics, vol. 313, pp. 251–268. AMS, Providence (2002)

28. Welk, M., Steidl, G., Weickert, J.: Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. Applied and Computational Harmonic Analysis 24, 195–224 (2008)
29. Zhu, S.C., Mumford, D.: Prior learning and Gibbs reaction-diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(11), 1236–1250 (1997)
30. Zhu, S.C., Wu, Y., Mumford, D.: Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. International Journal of Computer Vision 27(2), 107–126 (1998)
31. Zimmer, H., Valgaerts, L., Bruhn, A., Breuß, M., Weickert, J., Rosenhahn, B., Seidel, H.P.: PDE-based anisotropic disparity-driven stereo vision. In: Deussen, O., Keim, D., Saupe, D. (eds.) Vision, Modelling, and Visualization 2008, pp. 263–272. AKA, Heidelberg (2008)