CrossMark

Image Denoising Using Directional Adaptive Variable Exponents Model

Juha Tiirola[1]

Abstract In this paper, a new variational image denoising model is proposed. The new model could be seen to be a two-step method. In the first step, structure tensor analysis is used to infer something about the local geometry. The eigenvectors and the eigenvalues of the structure tensor are used in the construction of the denoising energy. In the second step, the actual variational denoising takes place. The steps are coupled in the sense that the energy expression is built using the underlying image, not the data. Two variable exponents are incorporated into the regularizer in order to reduce the staircasing effect, which is often present in the methods based on the first-order partial derivatives, and to increase smoothing along the image boundaries. In addition, two pointwise weight functions try to help to preserve small-scale details. In the theoretical part, the existence of a minimizer of a weak form of the original energy is considered. In the numerical part, an algorithm based on iterative minimization is presented and the numerical experiments demonstrate the possible advantages of the new model over some existing variational and partial differential equations methods.

Keywords Image denoising · Anisotropic total variation · Variable exponent

Mathematics Subject Classification 49N45 · 94A08 · 49J40

1 Introduction

Variational image denoising methods have attracted much attention since the model based on the minimization of the total variation was proposed by Rudin, Osher and Fatemi (ROF) [27]. Despite its success, some drawbacks of the ROF method are its tendency to produce piecewise constant images and loss of contrast. Since the introduction of the model, several modifications have been proposed.

In this paper, it is assumed that the noise is additive and white. Mathematically,

$$f = f_0 + n$$

where $f_0 : \Omega \rightarrow \mathbb{R}$ is the original noise-free image, $n : \Omega \rightarrow \mathbb{R}$ is the white noise and $f : \Omega \rightarrow \mathbb{R}$ is the noisy image or the measurement which is available. Variational models to find approximation of f_0 are formulated as energy minimization problems.

Methods, which are based on some kind of orientation matching, could be seen to form one group which tries to improve the ROF method. In these methods, some smoothing process is applied on $\nabla f / |\nabla f|$ in order to get an approximation v of $\nabla f_0 / |\nabla f_0|$. In the second step, this vector field v is used to guide the variational denoising in such a way that the energy minimization favours images u such that ∇u is approximately pointwise parallel to v. For methods belonging to this group, see e.g. the Lysaker–Osher–Tai model [25], the TV-Stokes model [26] and the model proposed in [18].

One group of denoising methods which avoids the staircasing effect is based on the use of the higher-order partial derivatives. For instance in [5], the concept of total generalized variation was introduced where the regularizer automatically balances between the first- and higher-order derivatives. In [23], an adaptive, directionally dependent

✉ Juha Tiirola
juha.tiirola@oulu.fi

[1] Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland

regularizer based on the first- and second-order derivatives was proposed. Article [11] presents a general framework for higher integer-order variational methods and non-linear diffusion filtering for image denoising. Higher-order methods are not limited to integer-order differentiability, e.g. in [10] fractional-order derivatives appear in the proposed regularizer to handle one-dimensional signals. Earlier, the use of fractional derivatives had appeared in differential equations form, see e.g. [12]. The work [19] on higher degree total variation contains also an interesting discussion regarding the separability of the total variation which is relevant also for our work in this paper.

The methods based on the use of variable exponents are also intended to reduce the staircasing. Edge-preserving regularization is used in the areas where the gradient of the noise-free image is probably large and in the areas where the gradient is probably small, isotropic smoothing is used. The first method on this area was [2] where the variable exponent in the regularizer was based on the underlying image. Due to the non-convexity of the method, in [8] a convex variant of [2] was proposed where the exponent is based solely on the noisy image. See also the subsequent papers [3,16,24] on denoising which are based on the use of variable exponents and first-order partial derivatives.

Studying image structures can be challenging due to noise. It could be preferable to use operations which are based on local neighbourhoods, since they could be more robust to noise than operations which are just pointwise based. For instance, structure tensor (see e.g. [14] for a pioneering work on the concept) provides a way to study the dominant orientation of the gradient in a neighbourhood. In the partial differential equations and calculus of variations approaches to image processing, the early works which utilize the structure tensor, are formulated in the partial differential equations framework, see Weickert [32] and Tschumperlé [30]. Currently, there are also variational methods which utilize the structure tensor.

In [15], Grasmair and Lenzen considered an anisotropic total variation model based on the minimization of the energy

$$||f - u||^2_{L^2(\Omega)} + \lambda \int_\Omega \left(\nabla u^\mathrm{T} A(u) \nabla u \right)^{\frac{1}{2}} \mathrm{d}x \qquad (1)$$

where the pointwise anisotropy matrix $A(u)$ is constructed from the linear structure tensor of u. In this model, the approximation of the unit vector field parallel to ∇u and the actual denoising step are coupled in the sense that the unit vector field is estimated from the underlying image u, not from the data.

In [28], a two-step method was proposed for image denoising. In the first step, the orientations of the image structures are estimated from the data. In the second step, this directional information is used to build an energy functional to be minimized. One of the regularizers in that paper has the form

$$\int_\Omega |r_1 \cdot \nabla u| + |r_2 \cdot \nabla u| \, \mathrm{d}x \qquad (2)$$

in the continuous domain where r_1 and r_2 are the pointwise directions derived in the step 1.

In [13], an iterative two-step method is proposed for image reconstruction where in the first step the structure tensor of the original image is estimated. In the second step, this information is used to construct an energy expression whose minimization favours solutions respecting the local geometry obtained by the structure tensor analysis. In [21], the so-called structure tensor total variation is proposed to solve inverse problems in imaging. The regularizer is based on penalizing the eigenvalues of the structure tensor. The work [20] extends [21] in the sense that in [20] the regularizer is based on the eigenvalues of a non-local version of the structure tensor. In [22], solution adaptive variants of the total variation were considered where the adaptivity is modelled as a fixed point problem. In [36], a gradient energy total variation is proposed using the gradient energy tensor.

The linear structure tensor $K_\rho * (\nabla u_\sigma \nabla u_\sigma^\mathrm{T})$, where $\rho > 0$ and $\sigma > 0$ are fixed and K is the Gaussian kernel, can also be obtained by applying the linear isotropic heat equation to the elements of the matrix $\nabla u_\sigma \nabla u_\sigma^\mathrm{T}$ where the stopping time depends on ρ. There are also non-linear methods to smooth the tensor field $\nabla u_\sigma \nabla u_\sigma^\mathrm{T}$ which aim at better edge preservation, see e.g. Brox et al. [6] and Hahn and Lee [17].

Regarding the works which have some resemblance to our work we also mention [1] where a directional total variation was considered and [35] where an adaptive directional total variation model was proposed for latent fingerprint segmentation.

In this paper, we consider a new adaptive variational image denoising model which is mostly inspired by the works [3,15]. The regularizer has a formulation which should help to preserve the edges in the estimated directions given by the structure tensor but the regularizer should also encourage smoothing along the edges. On the other hand, in the estimated smooth areas, the regularizer simplifies to isotropic smoothing. The pointwise positive weights in the regularizer also try to help to preserve small-scale details.

In [3], the authors considered variants of [2] for denoising and the regularizer was an interpolation between isotropic total variation and Tikhonov regularization. On the other hand, we consider in this paper a regularizer which in some sense is an interpolation between directional varying anisotropic total variation and Tikhonov regularization.

In the new model we consider in this paper, similar to [15], the approximation of the unit vector field parallel to ∇u and the actual denoising step are coupled in the sense that

the directional vector field is computed from the underlying image. We utilize both the eigenvectors and the eigenvalues of the structure tensor in the construction of the regularizer. Compared to the energy in (1), our regularizer has a pointwise separable formulation regarding the smoothing along the level lines and in the direction orthogonal to them. This separability is similar to the regularizer in (2). This separability property combined with the local coordinate system, whose axes are the eigenvectors of the structure tensor, is used to design the smoothing behaviour along the level lines and in the direction orthogonal to them. Roughly speaking, one can use locally linear or non-linear smoothing along the edges, whereas in the direction orthogonal to the edges, the smoothing is minor and of non-linear character in order to prevent the edges from getting blurred.

2 The New Model

Let $\Omega \subset \mathbb{R}^2$ be an open and bounded rectangle and $f \in L^2(\Omega)$. In this paper, we study variational image denoising of an image f based on the minimization of the energy

$$\lambda||f - u||^2_{L^2(\Omega)} + R(u) \tag{3}$$

where the regularizer R is

$$R(u) := \int_{\Omega} \alpha_1(u)|\nabla u \cdot \xi(u)|^{p_1(u)} + \alpha_2(u)|\nabla u \cdot \xi(u)^{\perp}|^{p_2(u)} \, dx$$

and $\lambda > 0$ is the global tuning parameter. In the regularizer, $(\xi(u)(x), \xi(u)^{\perp}(x))$ is a pointwise, smoothly varying orthonormal coordinate system computed from the image u such that $\xi(u)(x)$ is approximately parallel to $\nabla u(x)$ provided $\nabla u(x)$ is non-vanishing. The choices of the pointwise weights $\alpha_1(u)(x)$ and $\alpha_2(u)(x)$ and the varying integrabilities $p_1(u)(x)$ and $p_2(u)(x)$ aim at preserving the edges, reducing noise also along the edges of u and preserving small-scale details.

For simplicity, at times we may omit the explicit dependency of $\alpha_1, \alpha_2, p_1, p_2, \xi$ and $\xi^{\perp}$ on u.

The key features regarding the choices of the parameter functions are:

1. $0 < \varepsilon \leq \alpha_1, \alpha_2 \leq 1, 1 \leq p_1, p_2 \leq 2$. Here $\varepsilon > 0$ is small. This assumption is needed when the existence of a minimizer of the model is considered.
2. In the nearly constant, homogeneous areas, we have for the pointwise weights $\alpha_1(x) \approx \alpha_2(x) \approx 1$ and for the integrabilities $p_1(x) \approx p_2(x) \approx 2$, so $\alpha_1(u)|\nabla u \cdot \xi(u)|^{p_1(u)} + \alpha_2(u)|\nabla u \cdot \xi(u)^{\perp}|^{p_2(u)} \approx |\nabla u|^2$ by the Pythagorean theorem. This aims at reducing the staircasing effect.

We mention that if $\alpha_1 = \alpha_2 = 1$, $p_1 = p_2 = 2$, then one can actually solve the minimization of the energy (3) in the case $\Omega = \mathbb{R}^2$. The solution is obtained by applying a linear low-pass filter on f. This could be considered to correspond to averaging f locally in a linear way. The derivation of the low-pass filter in the Fourier domain is made as e.g. in [10].

3. In the areas where the structure is one-dimensional, the choices $\alpha_1(x) \approx 0$, $p_1(x) \approx 1$ and $p_2(x) \approx 2$ try to preserve the edges and at the same time encourage smoothing along the edges. Another possibility is to use also $p_2(x) \approx 1$.

   Again, if $\alpha_1 = 0$, $p_2 = 2$ and α_2, ξ are constant, then one could solve the minimization of the energy (3) in the case $\Omega = \mathbb{R}^2$. In the Fourier domain, the solution is obtained by applying the linear filter

$$\frac{2\lambda}{2\lambda + 8\pi^2 \alpha_2 |\xi^{\perp} \cdot w|^2}$$

on $\hat{f}(w)$. From the expression of the filter, we see that the filter attenuates all sinusoidal components except those oscillations which in the spatial domain are constant along $\xi^{\perp}$. Components which in the spatial domain oscillate in the direction $\xi^{\perp}$ are attenuated the most. In the spatial domain, along $\xi^{\perp}$, the filter has a low-pass character.

4. In the textured areas, where the structure is clearly not one-dimensional, or at the corners, $p_1(x) \approx p_2(x) \approx 1$ which tries to preserve the small-scale details along several directions.

To prove the existence of a minimizer of the new model we make the following assumptions on the pointwise varying coordinate system, the pointwise weights and the integrabilities:

If $u \in L^1(\Omega)$, then $\xi(u) \in C^1(\overline{\Omega}; \mathbb{R}^2)$, $\alpha_1(u)$, $\alpha_2(u)$, $\quad$ (A) $p_1(u)$, $p_2(u) \in C^1(\overline{\Omega})$, $p_1(u)$, $p_2(u) \in [1, 2]$, $\alpha_1(u)$, $\alpha_2(u) \in [\varepsilon, 1]$, where $\varepsilon \in (0, 1)$.

Continuity: If $u_n \to u$ in $L^1(\Omega)$, then $\xi(u_n) \to \xi(u)$ in $\quad$ (B)

$C^1(\overline{\Omega}; \mathbb{R}^2)$, $\alpha_1(u_n) \to \alpha_1(u)$, $\alpha_2(u_n) \to \alpha_2(u)$, $p_1(u_n) \to p_1(u)$, $p_2(u_n) \to p_2(u)$ in $C^1(\overline{\Omega})$.

3 Strong and Weak Forms

We denote the regularizer R in (3) by E_{strong},

$$E_{\text{strong}}(u) := \int_{\Omega} \alpha_1(u) |\nabla u \cdot \xi(u)|^{p_1(u)} \tag{4}$$
$$+ \alpha_2(u) |\nabla u \cdot \xi^{\perp}(u)|^{p_2(u)} \, dx$$

which is defined at least if u is differentiable.

We aim to define a suitable weak form E_{weak} such that $E_{\text{weak}}(u) = E_{\text{strong}}(u)$ if u is smooth enough and if $f \in L^2(\Omega)$ and $\lambda > 0$, then the energy $\lambda ||f - u||_{L^2(\Omega)} + E_{\text{weak}}(u)$ has a minimizer in the space of functions of bounded variation. The choice of the weak form is motivated by the analysis made in [3] where a weak form of the regularizer of the form $\int_{\Omega} |\nabla u|^{p(u)} \, dx$ or $\int_{\Omega} |\nabla u|^{p(f)} \, dx$ was considered. We use the results from [3] in many places in this section.

We aim to express the energy E_{strong} in such a way that in the energy, the dependency on u does not involve the derivatives of u. In the same way, the total variation can be obtained starting with the expression $\int_{\Omega} |\nabla u| \, dx$. First, we proceed formally to show how E_{weak} is obtained. Since for $t \in \mathbb{R}$ and $1 \le p \le 2$, we have

$$|t|^p = \max_{v \ge 0} \left(p v^{(p-1)} |t| - (p-1) v^p \right),$$

then we get the expression (5) for $|\nabla u \cdot \xi(u)|^{p_1(u)}$.

$$|\nabla u \cdot \xi(u)|^{p_1(u)}$$
$$= \max_{v_1 \ge 0} \left(p_1(u) v_1^{(p_1(u)-1)} |\nabla u \cdot \xi(u)| \right.$$
$$\left. - (p_1(u) - 1) v_1^{p_1(u)} \right) \tag{5}$$
$$= \max_{v_1 \ge 0, \sigma_1 \in \mathbb{R}, |\sigma_1| \le 1} \left(p_1(u) v_1^{(p_1(u)-1)} \sigma_1 \nabla u \cdot \xi(u) \right.$$
$$\left. - (p_1(u) - 1) v_1^{p_1(u)} \right)$$

Analogously, we get (6) for $|\nabla u \cdot \xi(u)^{\perp}|^{p_2(u)}$.

$$|\nabla u \cdot \xi(u)^{\perp}|^{p_2(u)} = \max_{v_2 \ge 0, \sigma_2 \in \mathbb{R}, |\sigma_2| \le 1}$$
$$\left(p_2(u) v_2^{(p_2(u)-1)} \sigma_2 \nabla u \cdot \xi(u)^{\perp} - (p_2(u) - 1) v_2^{p_2(u)} \right) \tag{6}$$

If it is possible to take the supremum outside of the integral, then using the integration by parts, if $\xi^{\perp} = (-(\xi)_2, (\xi)_1)$, and if

$$A(u) := \begin{pmatrix} (\xi(u))_1 & (\xi(u)^{\perp})_1 \\ (\xi(u))_2 & (\xi(u)^{\perp})_2 \end{pmatrix} = \begin{pmatrix} \xi(u)_1 & -\xi(u)_2 \\ \xi(u)_2 & \xi(u)_1 \end{pmatrix}, \tag{7}$$

we see that we have at least the approximate expression (8) for $E_{\text{strong}}(u)$ where the occurrences of ∇u have vanished. We can now define

$$E_{\text{weak}} : L^1(\Omega) \to \mathbb{R} \cup \{\pm \infty\}$$

as in (8) where we use the notation

$$||||(\sigma_1, \sigma_2)||_{\infty}||_{\infty} := \operatorname*{esssup}_{x \in \Omega} \max\{|\sigma_1(x)|, |\sigma_2(x)|\}.$$

$$E_{\text{strong}}(u) \approx \sup \left\{ \int_{\Omega} \operatorname{div} \left(A(u) \begin{bmatrix} \alpha_1(u) p_1(u) v_1^{p_1(u)-1} \sigma_1 \\ \alpha_2(u) p_2(u) v_2^{p_2(u)-1} \sigma_2 \end{bmatrix} \right) u \right.$$
$$- \alpha_1(u)(p_1(u) - 1) v_1^{p_1(u)} - \alpha_2(u)(p_2(u) - 1) v_2^{p_2(u)} \, dx \Big|$$
$$v_1, v_2 \in C^1(\overline{\Omega}), \sigma_1, \sigma_2 \in C_c^1(\Omega), \inf_{x \in \Omega} v_1(x), \tag{8}$$
$$\left. \inf_{x \in \Omega} v_2(x) > 0, \, || ||(\sigma_1, \sigma_2)||_{\infty}||_{\infty} \le 1 \right\} =: E_{\text{weak}}(u)$$

Next we intend to show that E_{weak} is minorized by a multiple of the total variation plus a constant. This result is similar to [3, p. 75] and the result is needed later when the existence of a minimizer is considered. We begin with a short technical lemma.

Lemma 1 *The function*

$$y \mapsto \begin{cases} \left(\frac{1}{y} \right)^{\frac{1}{y-1}}, & 1 < y \le 2, \\ e^{-1}, & y = 1 \end{cases}$$

is a restriction of a C^1 function g defined in $(0, +\infty)$. Also, g is increasing on $(1, 2)$ and thus $g \ge e^{-1}$ on $[1, 2]$.

Proof Let $g : (0, +\infty) \to \mathbb{R}$ be defined by

$$g(y) := \begin{cases} e^{\frac{1}{y-1} \log \left(\frac{1}{y} \right)}, & y \ne 1, \\ e^{-1}, & y = 1. \end{cases}$$

It is clear that g is a C^1 function on $(0, 1) \cup (1, +\infty)$. It is also true that

$$\lim_{y \to 1} g(y) = \frac{1}{e} \text{ and } \lim_{h \to 0} \frac{g(1 + h) - g(1)}{h} = \frac{1}{2e}$$

and

$$\lim_{y \to 1} \frac{d}{dy} \left(e^{\frac{1}{y-1} \log \left(\frac{1}{y} \right)} \right) = \frac{1}{2e}.$$

The claim follows. $\qquad \square$

Next we prove that E_{weak} is minorized by a multiple of the total variation plus a constant.

Proposition 1 *Let $\varepsilon \in (0, 1)$ be such that $\alpha_1(u), \alpha_2(u) \in [\varepsilon, 1]$ for all $u \in L^1(\Omega)$. Then*

$$TV(u) \le \frac{1}{\varepsilon} E_{\text{weak}}(u) + C$$

for all $u \in L^1(\Omega)$ where the constant C does not depend on u.

Proof Now

$$E_{\text{weak}}(u)$$

$$\geq \sup_{v_1, v_2, \sigma_1, \sigma_2} \int_\Omega \text{div}\left(A(u) \begin{bmatrix} \alpha_1(u) p_1(u) v_1^{p_1(u)-1} \sigma_1 \\ \alpha_2(u) p_2(u) v_2^{p_2(u)-1} \sigma_2 \end{bmatrix} \right) u$$

$$- \alpha_1(u)(p_1(u) - 1) v_1^{p_1(u)}$$

$$- \alpha_2(u)(p_2(u) - 1) v_2^{p_2(u)} \, dx$$

where $v_1, v_2 \in C^1(\overline{\Omega})$, $\inf v_1$, $\inf v_2 > 0$, $\sigma_1, \sigma_2 \in C_c^1(\Omega)$, $\sigma_1^2 + \sigma_2^2 \leq 1$. Since the columns of A are linearly independent, A is of full rank and invertible. Actually,

$$A^{-1} = \begin{pmatrix} \xi_1 & \xi_2 \\ -\xi_2 & \xi_1 \end{pmatrix}.$$

Let $v_1 := g \circ p_1(u)$ where g is as in Lemma 1. Then $v_1 \in C^1(\overline{\Omega})$, $p_1 v_1^{p_1-1} = 1$, $\inf_{x \in \Omega} v_1(x) \geq e^{-1} > 0$ and v_1 is bounded. Similarly, we select $v_2 := g \circ p_2(u)$. Then

$$E_{\text{weak}}(u) \geq \sup_{\sigma_1^2 + \sigma_2^2 \leq 1} \int_\Omega \text{div}\left(A(u) \begin{bmatrix} \alpha_1(u) \sigma_1 \\ \alpha_2(u) \sigma_2 \end{bmatrix} \right) u$$

$$- \alpha_1(u)(p_1(u) - 1) v_1^{p_1(u)}$$

$$- \alpha_2(u)(p_2(u) - 1) v_2^{p_2(u)} \, dx.$$

Let $(\sigma_1', \sigma_2') \in C_c^1(\Omega; \mathbb{R}^2)$, $(\sigma_1')^2 + (\sigma_2')^2 \leq 1$, be given. We select (σ_1, σ_2) such that

$$A(u) \begin{bmatrix} \alpha_1(u) & 0 \\ 0 & \alpha_2(u) \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} \sigma_1' \\ \sigma_2' \end{bmatrix}.$$

Then $(\sigma_1, \sigma_2) \in C_c^1(\Omega; \mathbb{R}^2)$ and since $\varepsilon \leq \alpha_1, \alpha_2 \leq 1$ and A^{-1} is length-preserving, we see that $\sigma_1^2 + \sigma_2^2 \leq \frac{1}{\varepsilon^2}$ and thus $(\varepsilon \sigma_1)^2 + (\varepsilon \sigma_2)^2 \leq 1$. Now

$$\int_\Omega \text{div}\left(\begin{bmatrix} \sigma_1' \\ \sigma_2' \end{bmatrix} \right) u \, dx = \frac{1}{\varepsilon} \int_\Omega \text{div}\left(A \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} \varepsilon \sigma_1 \\ \varepsilon \sigma_2 \end{bmatrix} \right) u \, dx$$

and thus

$$\sup_{(\sigma_1')^2 + (\sigma_2')^2 \leq 1} \int_\Omega \text{div}\left(\begin{bmatrix} \sigma_1' \\ \sigma_2' \end{bmatrix} \right) u \, dx$$

$$\leq \frac{1}{\varepsilon} \sup_{\sigma_1^2 + \sigma_2^2 \leq 1} \int_\Omega \text{div}\left(A \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} \right) u \, dx.$$

We mention that by a similar reasoning, it also follows that

$$\sup_{(\sigma_1')^2 + (\sigma_2')^2 \leq 1} \int_\Omega \text{div}\left(A \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} \sigma_1' \\ \sigma_2' \end{bmatrix} \right) u \, dx$$

$$\leq \sup_{\sigma_1^2 + \sigma_2^2 \leq 1} \int_\Omega \text{div}\left(\begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} \right) u \, dx.$$

Since $0 < \varepsilon \leq \alpha_1, \alpha_2 \leq 1$, $1 \leq p_1, p_2 \leq 2$, $v_i = g \circ p_i(u)$ where g is bounded on $[1, 2]$ by Lemma 1, we also get an estimate for the constant C. □

4 Equivalence of the Strong and Weak Forms for Smooth Functions

In this section, we show that if $u \in C^1(\overline{\Omega})$, then $E_{\text{strong}}(u) = E_{\text{weak}}(u)$.

Lemma 2 *If $u \in C^1(\overline{\Omega})$, then*

$$E_{\text{strong}}(u) \geq E_{\text{weak}}(u).$$

Proof Let $v_1, v_2 \in C^1(\overline{\Omega})$, $(\sigma_1, \sigma_2) \in C_c^1(\Omega; \mathbb{R}^2)$ be fixed, where $\inf v_1$, $\inf v_2 > 0$ and $||\sigma_1||_{L^\infty}, ||\sigma_2||_{L^\infty} \leq 1$. If $u \in C^1(\overline{\Omega})$, then using the fact that $|t|^{p_1} = \max_{v \geq 0}(p_1 v^{p_1-1}|t| - (p_1 - 1)v^{p_1})$, we see that $|\nabla u \cdot \xi|^{p_1(u)} \geq p_1 v_1^{p_1-1}|\nabla u \cdot \xi| - (p_1 - 1)v_1^{p_1} \geq p_1 v_1^{p_1-1} \sigma_1 \nabla u \cdot \xi - (p_1 - 1)v_1^{p_1}$. Using this and the integration by parts we see that

$$E_{\text{strong}}(u) \geq$$

$$- \int_\Omega \text{div}\left(\alpha_1 p_1 v_1^{p_1-1} \sigma_1 \xi + \alpha_2 p_2 v_2^{p_2-1} \sigma_2 \xi^\perp \right) u \, dx$$

$$- \int_\Omega \alpha_1(p_1 - 1)v_1^{p_1} + \alpha_2(p_2 - 1)v_2^{p_2} \, dx$$

and since $(\sigma_1, \sigma_2) \in C_c^1(\Omega; \mathbb{R}^2)$, $||\sigma_1||_{L^\infty}, ||\sigma_2||_{L^\infty} \leq 1$ and $v_1 \in C^1(\overline{\Omega})$, $v_2 \in C^1(\overline{\Omega})$, where $\inf v_1$, $\inf v_2 > 0$, are arbitrary, we get $E_{\text{strong}}(u) \geq E_{\text{weak}}(u)$. □

Lemma 3 *Let $u \in C^1(\overline{\Omega})$ and $\varepsilon > 0$. Then there exists $\sigma_1 \in C_c^1(\Omega)$ such that $||\sigma_1||_\infty \leq 1$ and*

$$(*) := \left| \int_\Omega \alpha_1 |\nabla u \cdot \xi|^{p_1} \, dx \right.$$

$$- \int_\Omega \alpha_1 p_1 |\nabla u \cdot \xi|^{p_1-1} \sigma_1 \nabla u \cdot \xi$$

$$\left. - \alpha_1(p_1 - 1)|\nabla u \cdot \xi|^{p_1} \, dx \right| < \varepsilon.$$

An analogous estimate holds for the triplet α_2, p_2 and $\xi^\perp$.

Proof Set $w_1(x) := \frac{\nabla u \cdot \xi}{|\nabla u \cdot \xi|}$ if $|\nabla u \cdot \xi| \neq 0$ and $w_1(x) = 0$ otherwise. Then $\alpha_1 |\nabla u \cdot \xi|^{p_1} = \alpha_1 |\nabla u \cdot \xi|^{p_1-1} w_1 \nabla u \cdot \xi$.

Select $\sigma_1 \in C_c^1(\Omega)$ such that

$$||w_1 - \sigma_1||_{L^1(\Omega)} < \frac{1}{n}. \tag{9}$$

Since σ_1 can be obtained by convolving with the standard mollifier a function whose absolute value is less than or equal to one, it follows that σ_1 can be chosen such that $||\sigma_1||_\infty \leq 1$. Then

$$(*) = \left| \int_\Omega \alpha_1 p_1 |\nabla u \cdot \xi|^{p_1-1} [w_1 \nabla u \cdot \xi - \sigma_1 \nabla u \cdot \xi] \, dx \right|$$

$$\leq \int_\Omega \alpha_1 p_1 |\nabla u \cdot \xi|^{p_1} |w_1 - \sigma_1| \, dx.$$

Since $u \in C^1(\overline{\Omega})$, then $|\nabla u| \in L^\infty(\Omega)$. By choosing n in (9) large enough, we see that $(*) < \varepsilon$. □

Lemma 4 *Let $u \in C^1(\overline{\Omega})$ and $\varepsilon > 0$. Then there exist $\sigma_1 \in C_c^1(\Omega)$ and $v_1 \in C^1(\overline{\Omega})$ such that $|\sigma_1| \leq 1$, $\inf v_1 > 0$ and*

$$(**) := \left| \int_\Omega \alpha_1 |\nabla u \cdot \xi|^{p_1} \, dx - \int_\Omega \alpha_1 p_1 v_1^{p_1-1} \sigma_1 \xi \cdot \nabla u \right.$$

$$\left. - \alpha_1(p_1 - 1)v_1^{p_1} \, dx \right| < \varepsilon.$$

An analogous estimate holds for the triplet α_2, p_2 and $\xi^\perp$.

Proof Select σ_1 by Lemma 3 corresponding to the tolerance $\frac{\varepsilon}{2}$. Since $u \in C^1(\overline{\Omega})$, then $|\nabla u| \in L^\infty(\Omega)$. Let $\rho_n := \eta_{1/n} * |\nabla u \cdot \xi|$ where it is assumed that $|\nabla u \cdot \xi| = 0$ outside Ω and η is the standard mollifier. Let $x \in \Omega$. Choose n so large that $B(x, 1/n) \subset \Omega$. Then

$$|\rho_n(x) - |\nabla u \cdot \xi(x)||$$

$$= \left| \int_{B(x,1/n)} \eta_{1/n}(x-y)[|(\nabla u \cdot \xi)(y)| - |\nabla u \cdot \xi(x)|] \, dy \right|$$

$$\leq \int_{B(x,1/n)} \eta_{1/n}(x-y)|(\nabla u \cdot \xi)(y) - (\nabla u \cdot \xi)(x)| \, dy.$$

It follows from the uniform continuity of $\nabla u \cdot \xi$ that $\rho_n(x) \to |\nabla u \cdot \xi(x)|$ as $n \to \infty$.

Let $v_n := \rho_n + \frac{1}{n}$. Then $v_n \in C^1(\overline{\Omega})$, $\inf v_n > 0$ for each n and $v_n \to |\nabla u \cdot \xi|$ pointwise as $n \to +\infty$. The dominated convergence theorem implies that

$$\int_\Omega \alpha_1 p_1 v_n^{p_1-1} \sigma_1 \xi \cdot \nabla u - \alpha_1(p_1 - 1)v_n^{p_1} \, dx \to$$

$$\int_\Omega \alpha_1 p_1 |\nabla u \cdot \xi|^{p_1-1} \sigma_1 \xi \cdot \nabla u - \alpha_1(p_1 - 1)|\nabla u \cdot \xi|^{p_1} \, dx$$

as $n \to \infty$. We set v_n in the place of v_1 in $(**)$. By adding and subtracting the term $\int_\Omega \alpha_1 p_1 |\nabla u \cdot \xi|^{p_1-1} \sigma_1 \nabla u \cdot \xi - \alpha_1(p_1 - 1)|\nabla u \cdot \xi|^{p_1} \, dx$ and by using Lemma 3 we see that

$$(**) \leq \left| \int_\Omega \alpha_1 |\nabla u \cdot \xi|^{p_1} \, dx \right.$$

$$- \int_\Omega \alpha_1 p_1 |\nabla u \cdot \xi|^{p_1-1} \sigma_1 \nabla u \cdot \xi$$

$$\left. - \alpha_1(p_1 - 1)|\nabla u \cdot \xi|^{p_1} \, dx \right|$$

$$+ \left| \int_\Omega \alpha_1 p_1 |\nabla u \cdot \xi|^{p_1-1} \sigma_1 \nabla u \cdot \xi \right.$$

$$- \alpha_1(p_1 - 1)|\nabla u \cdot \xi|^{p_1} \, dx$$

$$\left. - \int_\Omega \alpha_1 p_1 v_n^{p_1-1} \sigma_1 \xi \cdot \nabla u - \alpha_1(p_1 - 1)v_n^{p_1} \, dx \right|$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$

provided n is large enough. □

The next corollary follows directly from Lemmas 2 and 4.

Corollary 1 *If $u \in C^1(\overline{\Omega})$, then $E_{\text{weak}}(u) = E_{\text{strong}}(u)$.*

5 Existence of a Minimizer

The goal of this section is to show that the energy $E_{\text{weak}}(u) + \lambda ||f - u||_{L^2(\Omega)}^2$, $u \in L^2(\Omega)$, has a minimizer in $BV(\Omega)$. We begin with a technical lemma.

Lemma 5 *Let v_1, $v_2 \in C^1(\overline{\Omega})$, $\inf v_1$, $\inf v_2 > 0$ and σ_1, $\sigma_2 \in C_c^1(\Omega)$ be fixed. Let A be as in (7). Then the map*

$$u \mapsto A(u) \begin{bmatrix} \alpha_1(u)p_1(u)v_1^{p_1(u)-1}\sigma_1 \\ \alpha_2(u)p_2(u)v_2^{p_2(u)-1}\sigma_2 \end{bmatrix} =: G(u)$$

is continuous from $L^1(\Omega)$ to $C_c^1(\Omega; \mathbb{R}^2)$.

Proof Now for $G(u)$ we have the expression (10).

Since $v_1 \in C^1(\overline{\Omega})$, $\inf v_1 > 0$ and $p_1(u) \in C^1(\overline{\Omega})$, it follows that the function $x \mapsto v_1(x)^{p_1(u)(x)-1}$ is a $C^1(\overline{\Omega})$ function. This can be seen by writing $v^{p-1} = e^{(p-1)\log v}$ and applying the mean value theorem. Similarly for v_2 and p_2. Then using Assumption (A) on page 3, we see that $G(u) \in C^1(\overline{\Omega}; \mathbb{R}^2)$. By using [3, Lemma 5] and Assumption (B) it follows that $u \mapsto v_1^{p_1(u)-1}$ and $u \mapsto v_2^{p_2(u)-1}$ are continuous maps from $L^1(\Omega)$ to $C^1(\overline{\Omega})$.

By Assumption (B) on page 3, $u \mapsto \xi_i(u)$, $u \mapsto p_i(u)$ and $u \mapsto \alpha_i(u)$ are continuous from $L^1(\Omega)$ to $C^1(\overline{\Omega})$. The claim then follows using the Leibniz rule as in [3, p. 80]. □

Theorem 1 *Let v_1, v_2, σ_1, σ_2 and G be as in Lemma 5. Let A be as in (7). Then*

$$u \mapsto \int_\Omega \operatorname{div}(G(u))\, u - \alpha_1(u)(p_1(u) - 1)v_1^{p_1(u)}$$
$$- \alpha_2(u)(p_2(u) - 1)v_2^{p_2(u)}\, dx$$

is continuous from $L^1(\Omega)$ to $\mathbb{R}$.

Proof Let $u_n \to u$ in $L^1(\Omega)$. Using [3, Lemma 7] and Assumption (B) on page 3 we see that $(p_1(u_n) - 1)v_1^{p_1(u_n)} \to (p_1(u) - 1)v_1^{p_1(u)}$ in $L^1(\Omega)$ as $n \to \infty$ and an analogous result holds for p_2 and v_2. By Assumption (B) on page 3, $\alpha_1(u_n) \to \alpha_1(u)$ and $\alpha_2(u_n) \to \alpha_2(u)$ in $C^1(\overline{\Omega})$. We get

$$\int_\Omega \alpha_1(u_n)(p_1(u_n) - 1)v_1^{p_1(u_n)}$$
$$+ \alpha_2(u_n)(p_2(u_n) - 1)v_2^{p_2(u_n)}\, dx$$
$$\to \int_\Omega \alpha_1(u)(p_1(u) - 1)v_1^{p_1(u)}$$
$$+ \alpha_2(u)(p_2(u) - 1)v_2^{p_2(u)}\, dx$$

as $n \to \infty$. Next we show that

$$\int_\Omega \operatorname{div}(G(u_n))\, u_n \, dx \to \int_\Omega \operatorname{div}(G(u))\, u \, dx.$$

Now

$$\left| \int \operatorname{div}(G(u))\, u - \operatorname{div}(G(u_n))\, u_n \, dx \right|$$
$$\le \left| \int \operatorname{div}(G(u))\, (u - u_n)\, dx \right|$$
$$+ \left| \int [\operatorname{div}(G(u)) - \operatorname{div}(G(u_n))]\, u_n \, dx \right|$$
$$=: (*) + (**).$$

Here $(*) \to 0$ as $n \to \infty$, since $||u_n - u||_{L^1(\Omega)} \to 0$ as $n \to \infty$ and $G(u)$ is a $C_c^1(\Omega; \mathbb{R}^2)$ function by Lemma 5.

Since $u_n \to u$ in $L^1(\Omega)$, then u_n is bounded in $L^1(\Omega)$. On the other hand, by Lemma 5,

$$||\operatorname{div}(G(u)) - \operatorname{div}(G(u_n))||_{L^\infty} \to 0$$

as $n \to \infty$ and thus $(**) \to 0$ as $n \to \infty$. $\qquad\square$

$$G(u) = \begin{bmatrix} \xi_1(u)\alpha_1(u)p_1(u)v_1^{p_1(u)-1}\sigma_1 - \xi_2(u)\alpha_2(u)p_2(u)v_2^{p_2(u)-1}\sigma_2 \\ \xi_2(u)\alpha_1(u)p_1(u)v_1^{p_1(u)-1}\sigma_1 + \xi_1(u)\alpha_2(u)p_2(u)v_2^{p_2(u)-1}\sigma_2 \end{bmatrix}$$
$$(10)$$

Corollary 2 *The functional E_{weak} is lower-semicontinuous on $L^1(\Omega)$.*

Proof By the definition of E_{weak} and using Theorem 1 we see that E_{weak} is a supremum of continuous functions so then E_{weak} is lower-semicontinuous. $\qquad\square$

Now we come to the main result of this section.

Theorem 2 *Let $\Omega \subset \mathbb{R}^2$ be an open rectangle, $f \in L^2(\Omega)$, $\lambda > 0$ and set*

$$E(u) := E_{\text{weak}}(u) + \lambda ||f - u||_{L^2(\Omega)}^2$$

for $u \in L^2(\Omega)$. Then there exists $u \in \text{BV}(\Omega)$ which is a minimizer of E.

Proof Let $(u_n) \subset L^2(\Omega)$ be a minimizing sequence for E. Since $E(0) = \lambda ||f||_{L^2(\Omega)}^2 < \infty$, we see that $E(u_n)$ is bounded from above. Then by Proposition 1, $\text{TV}(u_n) \le c$ for each n where $c > 0$ does not depend on n. Since $L^2(\Omega) \hookrightarrow L^1(\Omega)$, we see that $||u_n||_{L^1(\Omega)} \le C||u_n||_{L^2(\Omega)} \le C(||f - u_n||_{L^2(\Omega)} + ||f||_{L^2(\Omega)})$. We see that $u_n \in \text{BV}(\Omega)$ and

$$||u_n||_{\text{BV}(\Omega)} = ||u_n||_{L^1(\Omega)} + \text{TV}(u_n) \le C$$

for each n. Thus there exists $u \in \text{BV}(\Omega)$ and a subsequence n_k such that $u_{n_k} \rightharpoonup u$ weakly* in $\text{BV}(\Omega)$ as $k \to \infty$. Thus $u_{n_k} \to u$ in $L^1(\Omega)$.

Using Corollary 2 we see that

$$E_{\text{weak}}(u) \le \liminf_{k \to \infty} E_{\text{weak}}(u_{n_k}).$$

Since u_n is bounded in $L^2(\Omega)$ we can also assume by selecting again a subsequence if necessary that $u_{n_k} \rightharpoonup u$ in $L^2(\Omega)$ and thus $||f - u||_{L^2(\Omega)} \le \liminf_{k \to \infty} ||f - u_{n_k}||_{L^2(\Omega)}$ which combined with the previous gives that

$$E(u) \le \liminf_{k \to \infty} E(u_{n_k})$$

and the proof is complete. $\qquad\square$

6 Parameter Functions

In this section, we discuss a possible choice of the pointwise varying coordinate system satisfying the assumptions (A) and (B) on page 3. The choice is based on smoothing of a matrix representation of the image gradient field. We also discuss how the pointwise weights and the varying integrabilities can be chosen in general.

We begin by discussing the concept of the structure tensor. In general, structure tensor can be used to infer something about the local geometry of an image. By a tensor field we understand a function from Ω to the space of symmetric, positive semidefinite matrices of size 2×2.

Let $u \in L^1(\Omega)$ be an arbitrary image. First, the Gaussian kernel G_σ, where $\sigma > 0$ is fixed, is applied on u in order to make the pointwise partial derivatives well defined. Let

$$S_\sigma(u) := \begin{bmatrix} [\partial_x G_\sigma * u]^2 & [\partial_x G_\sigma * u] \cdot [\partial_y G_\sigma * u] \\ [\partial_x G_\sigma * u] \cdot [\partial_y G_\sigma * u] & [\partial_y G_\sigma * u]^2 \end{bmatrix}. \tag{11}$$

Let $\rho > 0$ denote the size of the neighbourhood where the average of the tensor field $S_\sigma(u)$ is considered and set

$$S(u)(x) := G_\rho * S_\sigma(u)(x) =: \begin{bmatrix} a(x) \; b(x) \\ b(x) \; c(x) \end{bmatrix}. \tag{12}$$

The obtained tensor field is called the structure tensor of u. The eigenvectors and eigenvalues of $S(u)(x)$ can be used to infer something about the local geometry of u. This is based on a variational characterization of the eigenvalues of a positive semidefinite matrix. We follow [32]. Let

$$S(u)(x) = \lambda_1 \theta_1 \theta_1^T + \lambda_2 \theta_2 \theta_2^T$$

be an eigenvalue decomposition of $S(u)(x)$. Here $\lambda_1 \geq \lambda_2 \geq 0$ since $S(u)(x)$ is positive semidefinite. If $\lambda_1 \approx \lambda_2 \approx 0$, it is assumed that x is in the nearly constant intensity area. If $\lambda_1 \gg \lambda_2 \approx 0$, then around x the image structure is nearly one-dimensional. If $\lambda_1 \geq \lambda_2 \gg 0$, then x is in the corner area or in the texture area where the texture is not locally one-dimensional.

In (12), $S(u)$ can also be obtained by applying the linear, isotropic heat equation to the tensor field $S_\sigma(u)$. There are also non-linear smoothing processes which can be applied to $S_\sigma(u)$. Let us discuss the smoothing proposed in [17]. Let $A(t) : \Omega \to \mathbb{R}^{2\times2}$, $A(t) := \begin{bmatrix} a_{11}(t) \; a_{12}(t) \\ a_{21}(t) \; a_{22}(t) \end{bmatrix}$ such that $A(0)(x) = S_\sigma(u)(x)$ and

$$\begin{cases} \dfrac{\partial a_{ij}}{\partial t} = \mathrm{div}(g(G_\varsigma * A)\nabla a_{ij}) \text{ in } \Omega \times (0, T], \\ (g(G_\varsigma * A)\nabla a_{ij}) \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega \times (0, T], \\ a_{ij}(x, 0) = (S_\sigma(u))_{ij}(x). \end{cases} \tag{13}$$

Here $g(M) = \tilde{g}(\Lambda)v_\Lambda v_\Lambda^T + \tilde{g}(\lambda)v_\lambda v_\lambda^T$ where it is assumed that $\Lambda v_\Lambda v_\Lambda^T + \lambda v_\lambda v_\lambda^T$ is an eigenvalue decomposition of M and $\tilde{g}(s) = \frac{1}{\sqrt{s+1}}$, $s \geq 0$. In [17] it is proved that if $A(0)$ is pointwise positive semidefinite, then also $A(t)$ is pointwise positive semidefinite for $t \in (0, T]$.

In (12), if $u_n \to u$ in $L^1(\Omega)$, then quite similarly as in [15], it follows from Young's inequality for convolution that we have $S_\sigma(u_n) \to S_\sigma(u)$ in $L^2(\Omega; \mathbb{R}^{2\times2})$. Since $S(u)$ is obtained from $S_\sigma(u)$ by convolution where the convolution kernel is smooth, then again by Young's inequality for convolution we see that $S(u_n) \to S(u)$ in $C^1(\overline{\Omega}, \mathbb{R}^{2\times2})$. If the matrix elements in (12) are considered, we have shown that $u_n \to u$ in $L^1(\Omega)$ implies that $a(u_n) \to a(u), b(u_n) \to b(u)$ and $c(u_n) \to c(u)$ in $C^1(\overline{\Omega})$.

In (13), by working as in [33, Theorem 1] and using [17, Lemmas 1-2] one could prove an L^2 stability result: If $A(t)$

and $A_n(t)$ are tensor fields obtained by the diffusion such that $A_n(0) \to A(0)$ in $L^2(\Omega; \mathbb{R}^{2\times2})$ as $n \to \infty$, and if $\rho > 0$ is a fixed time instant, then $A_n(\rho) \to A(\rho)$ in $L^2(\Omega; \mathbb{R}^{2\times2})$ as $n \to \infty$. As in the linear case, $u_n \to u$ in $L^1(\Omega)$ as $n \to \infty$ implies that $S_\sigma(u_n) \to S_\sigma(u)$ in $L^2(\Omega; \mathbb{R}^{2\times2})$ which in turn implies that $A_n(\rho) \to A(\rho)$ in $L^2(\Omega; \mathbb{R}^{2\times2})$ where $A_n(0) = S_\sigma(u_n)$ and $A(0) = S_\sigma(u)$. If in the non-linear case, we call the structure tensor $S(u)$ the tensor field $G_a * A(\rho)$, where $a > 0$, then as in the linear case we see that $S(u_n) \to S(u)$ in $C^1(\overline{\Omega}, \mathbb{R}^{2\times2})$. In the numerical experiments, we assume that the amount of this smoothing is negligible and we actually use $a = 0$ and in the non-linear case we write

$$S(u) = A(\rho) := \begin{bmatrix} a(x) \; b(x) \\ b(x) \; c(x) \end{bmatrix}. \tag{14}$$

If u is differentiable, then in the limiting case $\sigma = \rho = 0$ we have e.g. in (12) that $S(u) = \nabla u (\nabla u)^T$ and

$$\begin{bmatrix} (\partial_x u)^2 & (\partial_x u)(\partial_y u) \\ (\partial_x u)(\partial_y u) & (\partial_y u)^2 \end{bmatrix} \frac{\nabla u}{|\nabla u|} = |\nabla u|^2 \frac{\nabla u}{|\nabla u|}.$$

We can then consider that the largest eigenvalue of the structure tensor $S(u)(x)$ is approximately equal to $|\nabla u|^2$.

The matrix $S(u)(x)$ in (12) or in (14) has eigenvalues

$$\lambda_1 := \frac{1}{2}\left(a + c + \sqrt{(a-c)^2 + 4b^2}\right) \tag{15}$$

and

$$\lambda_2 := \frac{1}{2}\left(a + c - \sqrt{(a-c)^2 + 4b^2}\right). \tag{16}$$

Since $S(u)(x)$ is positive semidefinite, $\lambda_1(x), \lambda_2(x) \geq 0$. The unit eigenvector corresponding to the eigenvalue λ_1 is parallel to the vector

$$(2b, c - a + \sqrt{(a-c)^2 + 4b^2}) \tag{17}$$

provided $b \neq 0$.

Next we consider how $(\xi, \xi^\perp)$ can be chosen in order to fulfil the assumptions (A) and (B) on page 3. We begin with a short technical result.

Proposition 2 *Let $h_n \to h$ in $C^1(\overline{\Omega}, \mathbb{R}^n)$ and $S > 0$ such that $\|h\|_{C^1(\Omega)}, \|h_n\|_{C^1(\Omega)} < \frac{S}{2}$ for all n. Let $G \in C^2(\overline{B(0, S)}, \mathbb{R})$. Then $G \circ h_n \to G \circ h$ in $C^1(\overline{\Omega}, \mathbb{R})$ as $n \to +\infty$.*

Proof Using the mean value theorem we see that

$$\sup_{x \in \Omega} |G(h_n(x)) - G(h(x))|$$

$$\leq \sup_{|y| \leq \frac{S}{2}} |\nabla G(y)| \sup_{x \in \Omega} |h_n(x) - h(x)|$$

which tends to zero as $n \to \infty$. Now

$$\partial_j (G \circ h)(x) = \sum_k (\partial_k G)(h(x)) \partial_j h^k(x).$$

Using this and the mean value theorem we get

$$\sup_{x \in \Omega} |\partial_j (G \circ h(x)) - \partial_j (G \circ h_n(x))|$$

$$= \sup_{x \in \Omega} \left| \sum_k [(\partial_k G)(h(x))[\partial_j h^k(x) - \partial_j h_n^k(x)]\right.$$

$$\left. - [(\partial_k G)(h_n(x)) - (\partial_k G)(h(x))] \partial_j h_n^k(x) \right|$$

$$\leq \sup_{|y| \leq \frac{S}{2}} |\nabla G(y)| \sum_k \sup_{x \in \Omega} |\partial_j h^k(x) - \partial_j h_n^k(x)|$$

$$+ \sum_k \sup_{|y| \leq \frac{S}{2}} |\nabla \partial_k G(y)| \sup_{x \in \Omega} |h_n(x) - h(x)| \cdot \frac{S}{2}$$

which tends to zero as $n \to \infty$. Then the claim follows. □

Now we come to the choice of $(\xi, \xi^\perp)$. Our choice is based, as in anisotropic diffusion filtering [32], on the eigenvectors of the structure tensor. We use in addition some regularization in order to get smooth dependencies. We can replace in (17) the square root by its regularization $s \mapsto \sqrt{s + \delta^2}$ where $\delta > 0$ is small. Let $G : \mathbb{R}^3 \to \mathbb{R}^2$,

$$G(x, y, z) = \frac{(2y, z - x + \sqrt{(x-z)^2 + 4y^2 + \delta^2})}{\sqrt{4y^2 + (z - x + \sqrt{(x-z)^2 + 4y^2 + \delta^2})^2}}$$

and set

$$\xi(u)(x) := G(a(u)(x), b(u)(x), c(u)(x)).$$

Let $\xi^\perp(u)(x) := (-(\xi(u)(x))_2, (\xi(u)(x))_1)$. Then if $u_n \to u$ in $L^1(\Omega)$ implies that $a(u_n) \to a(u)$, $b(u_n) \to b(u)$ and $c(u_n) \to c(u)$ in $C^1(\overline{\Omega})$, it follows by applying Proposition 2 componentwise on the function G that we have $\xi(u_n) \to \xi(u)$ in $C^1(\overline{\Omega}; \mathbb{R}^2)$.

We can do analogous regularization in (15) and (16), and obtain regularized eigenvalues λ_1 and λ_2. If also the pointwise weights α_1 and α_2 and the varying integrabilities p_1 and p_2 are obtained by applying a smooth function on the regularized eigenvalues of $S(u)(x)$ or on $|\nabla G_\sigma * u|^2$, then the assumptions (A) and (B) on page 3 are valid.

7 Discretization

7.1 General Algorithm

In this section, we consider the numerical aspects of the new model. Since in general the energy (3) is not convex with

respect to u, the minimization is challenging. As in [15], we resort to iterative minimization. In [31], a lagged diffusivity fixed point iteration, where the diffusivity is based on the previous iterate, was proposed for the numerical solution of the total variation denoising.

We use two nested loops in the algorithm. In the outer loop we update the pointwise coordinate system $(\xi, \xi^\perp)$, the pointwise weights α_1, α_2 and the pointwise integrabilities p_1, p_2 using the previous iterate of u. We use the inner loop to solve a non-quadratic minimization problem with respect to u. There is no guarantee that the algorithm we use finds an actual local minimizer of the total energy.

First, we regularize the energy (3) and instead consider the energy

$$\lambda ||f - u||_{L^2(\Omega)}^2 + \int_\Omega \alpha_1(u) \left[(\nabla u \cdot \xi(u))^2 + \delta^2 \right]^{\frac{p_1(u)}{2}}$$

$$+ \alpha_2(u) \left[(\nabla u \cdot \xi^\perp(u))^2 + \delta^2 \right]^{\frac{p_2(u)}{2}} \, dx,$$

where $\delta > 0$ is small. If ξ, $\xi^\perp$, α_1, α_2, p_1 and p_2 are fixed, let

$$E^{\xi, \alpha, p}(u) := \lambda ||f - u||_{L^2(\Omega)}^2 + \int_\Omega \alpha_1 \left[(\nabla u \cdot \xi)^2 + \delta^2 \right]^{\frac{p_1}{2}}$$

$$+ \alpha_2 \left[(\nabla u \cdot \xi^\perp)^2 + \delta^2 \right]^{\frac{p_2}{2}} \, dx.$$

The functional gradient of $E^{\xi, \alpha, p}$ is

$$\nabla E^{\xi, \alpha, p}(u) = 2\lambda(u - f)$$

$$- \operatorname{div} \left(\alpha_1 p_1 \left[(\nabla u \cdot \xi)^2 + \delta^2 \right]^{\frac{p_1}{2} - 1} (\nabla u \cdot \xi) \xi \right)$$

$$- \operatorname{div} \left(\alpha_2 p_2 \left[(\nabla u \cdot \xi^\perp)^2 + \delta^2 \right]^{\frac{p_2}{2} - 1} (\nabla u \cdot \xi^\perp) \xi^\perp \right).$$

Let

$$\lambda_1(u) := \frac{\alpha_1 p_1}{[(\nabla u \cdot \xi)^2 + \delta^2]^{1 - \frac{p_1}{2}}}, \tag{18}$$

$$\lambda_2(u) := \frac{\alpha_2 p_2}{[(\nabla u \cdot \xi^\perp)^2 + \delta^2]^{1 - \frac{p_2}{2}}} \tag{19}$$

and

$$A(u)$$

$$:= \begin{bmatrix} \lambda_1(u)\xi_1^2 + \lambda_2(u)(\xi_1^\perp)^2 & \lambda_1(u)\xi_1\xi_2 + \lambda_2(u)\xi_2^\perp \xi_1^\perp \\ \lambda_1(u)\xi_1\xi_2 + \lambda_2(u)\xi_1^\perp \xi_2^\perp & \lambda_1(u)\xi_2^2 + \lambda_2(u)(\xi_2^\perp)^2 \end{bmatrix}.$$

Then

$$\nabla E^{\xi, \alpha, p}(u) = 2\lambda(u - f) - \operatorname{div}(A(u)\nabla u).$$

We summarize the numerical steps in the algorithm below. In the inner part of the algorithm, a fixed point iteration is used to solve the equation $2\lambda(u - f) - \mathrm{div}(A(u)\nabla u) = 0$ for u.

Algorithm:
    $i = 0$;
    $u_i = f$;
    do
        $i = i + 1$;
        $\xi = \xi(u_{i-1}), \xi^\perp = \xi^\perp(u_{i-1}), \alpha_1 = \alpha_1(u_{i-1}),$
        $\alpha_2 = \alpha_2(u_{i-1}), p_1 = p_1(u_{i-1}), p_2 = p_2(u_{i-1})$;
        $v = u_{i-1}$;
        for $j = 1 : N$
            $A = A(v)$;
            Update v by solving
            $2\lambda(v - f) - \mathrm{div}(A\nabla v) = 0$;
        end
        $u_i = v$;
    while $\|u_i - u_{i-1}\|_2 >$ tolerance or the max. number of iterations is not reached

7.2 Update of u

If h denotes the grid size, we assume that we know the noisy image f at the points (ih, jh) where $i, j \in \mathbb{Z}$. Let $f_{ij} := f(ih, jh)$ and similarly for u and the other functions. When we update u, we assume that $\xi, \xi^\perp, \alpha_1, \alpha_2, p_1$ and p_2 are available at the locations (ih, jh). We discuss in section 7.3 how they are obtained. In λ_1 of (18), the product $\nabla u \cdot \xi = \partial_x u \, \xi_1 + \partial_y u \, \xi_2$ at (ih, jh) is computed as

$$(\nabla u \cdot \xi)_{ij} = m\left((D_x^+ u)_{ij}, (D_x^- u)_{ij}\right)(\xi_1)_{ij}$$
$$+ m\left((D_y^+ u)_{ij}, (D_y^- u)_{ij}\right)(\xi_2)_{ij}$$

where m denotes the minmod function. Analogously $(\nabla u \cdot \xi^\perp)_{ij}$ is computed in λ_2 of (19).

In the fixed-point algorithm, A is based on the previous iterate of u. When A is fixed, we solve the equation of the form $2\lambda(u - f) - \mathrm{div}(A\nabla u) = 0$ as an evolution equation by applying the AOS (additive operator splitting) scheme on the version involving the time. For more on the AOS scheme in image processing, see [34].

So we assume that A is fixed, $A := \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. For $\mathrm{div}(A\nabla u)$ we use the discretization described at the end of [32, Chapter 3.4]. Then we can write

$$(\mathrm{div}(A\nabla u))_{ij} = \sum_{k=-1}^{1} A_{i,j,k}\, u_{i+k,j} + \sum_{l=-1}^{1} C_{i,j,l}\, u_{i,j+l}$$

$$+ \sum_{k,l=-1}^{1} B_{i,j,k,l}\, u_{i+k,j+l}$$

where $A_{i,j,k}$ is constructed from a, $C_{i,j,l}$ is constructed from c and $B_{i,j,k,l}$ is constructed from b. Referring to [32, Chapter 3.4], we have e.g. $A_{i,j,-1} = \frac{a_{i-1,j}+a_{i,j}}{2h_1^2}$, $C_{i,j,1} = \frac{c_{i,j+1}+c_{i,j}}{2h_2^2}$ and $B_{i,j,1,1} = \frac{|b_{i+1,j+1}|+b_{i+1,j+1}}{4h_1h_2}$. To solve the equation $2\lambda(u - f) - \mathrm{div}(A\nabla u) = 0$ we introduce an artificial time variable and find the steady state of $\partial_t u + 2\lambda(u - f) - \mathrm{div}(A\nabla u) = 0$. If we discretize the previous equation we get

$$\frac{1}{2}\partial_t u + \lambda u - \sum_{k=-1}^{1} A_{i,j,k} u_{i+k,j} +$$

$$\frac{1}{2}\partial_t u + \lambda u - \sum_{l=-1}^{1} C_{i,j,l} u_{i,j+l} = W$$

where

$$W := 2\lambda f + \sum_{k,l=-1}^{1} B_{i,j,k,l} u_{i+k,j+l}.$$

If τ denotes the timestep, let $u_{ij}^{t+1} = \frac{v_{ij}^{t+1}+w_{ij}^{t+1}}{2}$ where

$$\frac{v_{ij}^{t+1} - u_{ij}^t}{2\tau} + \lambda v_{ij}^{t+1} - \sum_{k=-1}^{1} A_{i,j,k} v_{i+k,j}^{t+1} = \frac{1}{2}W_{ij}^t$$

and

$$\frac{w_{ij}^{t+1} - u_{ij}^t}{2\tau} + \lambda w_{ij}^{t+1} - \sum_{l=-1}^{1} C_{i,j,l} w_{i,j+l}^{t+1} = \frac{1}{2}W_{ij}^t$$

which can be solved rapidly by using the tridiagonal matrix algorithm as in [34].

7.3 Structure Tensor Computations

In (11), the partial derivatives ∂_x and ∂_y of $v := G_\sigma * u$ are computed at the locations $((i + 1/2)h, (j + 1/2)h)$ using the formulas $\frac{v_{i+1,j}-v_{ij}+v_{i+1,j+1}-v_{i,j+1}}{2h}$ and $\frac{v_{i,j+1}-v_{ij}+v_{i+1,j+1}-v_{i+1,j}}{2h}$. After the smoothing of $S_\sigma(u)$ by G_ρ or by (13), we have $S(u)$ defined at the locations $((i + 1/2)h, (j + 1/2)h)$. We use $\nabla(G_\sigma * u)$ and the eigenvalues and eigenvectors of $S(u)$ to compute $\xi, \xi^\perp, \alpha_1, \alpha_2, p_1$ and p_2 at the locations $((i + 1/2)h, (j + 1/2)h)$. These functions can then be computed at the locations (ih, jh) as the averages at the four locations $((i \pm 1/2)h, (j \pm 1/2)h)$.

8 Numerical Experiments

We begin this section by describing the choices of the weight functions α_1 and α_2 and the integrabilities p_1 and p_2 in (3) used in the experiments. The choices are motivated by the key features mentioned in section 2.

We begin by constructing an auxiliary function which changes its value smoothly from one to zero. Let $\gamma(x) = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$ where the constants $a, b, c, d, e, f \in \mathbb{R}$. Let $A > 0$ and suppose we require that $\gamma(0) = 1$, $\gamma(A) = 0$, $\gamma'(0) = \gamma''(0) = \gamma'(A) = \gamma''(A) = 0$. Then after some calculations we see that $a = -6/A^5$, $b = 15/A^4$, $c = -10/A^3$, $d = e = 0$ and $f = 1$. Then γ changes its value smoothly from 1 at $x = 0$ to 0 at $x = A$.

Let $M, N \in \mathbb{R}^+$, $0 < M < N$ such that if $|\nabla u(x)| < M$ we assume that x is in the smooth area and if $|\nabla u(x)| > N$, x is near an edge. Let $A := N - M$ and $\rho : [0, \infty) \to [0, 1]$ where $\rho(s) = 0$ if $0 \le s \le M$, $\rho(s) = 1$ if $s \ge N$ and $\rho(s) = 1 - \gamma(s - M)$ if $M < s < N$. Now $\rho'(M) = \rho''(M) = \rho'(N) = \rho''(N) = 0$. Then ρ changes its value smoothly from 0 at $s = M$ to 1 at $s = N$.

Let

$$g := \rho \circ \sqrt{\lambda_1} \tag{20}$$

where λ_1 denotes the largest eigenvalue of $S(u)$. Then g works as an edge detector such that $g \approx 0$ in the near constant areas and $g \approx 1$ in the non-smooth areas. We use the ratio of the eigenvalues of $S(u)$ to split the edges found by g into one-dimensional edges and the rest, which consists of corners and texture which is not one-dimensional. Let

$$\psi := \sqrt{\frac{\lambda_2}{\lambda_1 + 0.01}} \tag{21}$$

which measures the pointwise anisotropy such that $\psi \approx 1$ if $\lambda_1 \approx \lambda_2$ and $\psi \approx 0$ if $\lambda_1 >> \lambda_2$. See also [29] for discussions on measuring the anisotropy in kernel regression approach to denoising. The function ψ is used to split the edge areas detected by g. Let

$$d_a := (1 - \psi)g$$

and

$$d_r := \psi g.$$

Then $d_a \approx 1$ in the one-dimensional edge areas and $d_a \approx 0$ otherwise. Analogously, $d_r \approx 1$ in the corners and texture areas where the texture is not one-dimensional, and $d_r \approx 0$ otherwise.

In the numerical experiments of this paper, the pointwise weights α_1 and α_2 are of the form

$$\alpha_1 = \frac{d_a}{1 + k_1 \lambda_1^2} + \frac{d_r}{1 + k_2 \lambda_1^2} + (1 - d_a - d_r) \tag{22}$$

and

$$\alpha_2 = \frac{d_a}{1 + k_3 \lambda_2^2} + \frac{d_r}{1 + k_2 \lambda_2^2} + (1 - d_a - d_r) \tag{23}$$

where $k_1, k_2, k_3 > 0$ are constants.

Next we consider the choices of the integrabilities. Let again $A = N - M$. Recalling how the function γ was constructed earlier in this section, let $\varphi : [0, \infty) \to [1, 2]$ such that $\varphi(s) = 2$ if $s \le M$, $\varphi(s) = 1$ if $s \ge N$ and $\varphi(s) = 1 + \gamma(s - M)$ for $M < s < N$. Then $\varphi'(M) = \varphi''(M) = \varphi'(N) = \varphi''(N) = 0$ and φ changes its value smoothly from 2 at $s = M$ to 1 at $s = N$. In the experiments of this paper, we use

$$p_1 = \varphi \circ |\nabla u_t(\cdot)| \tag{24}$$

where t corresponds to the time needed to blur u using the linear, isotropic heat equation such that the result equals to the convolution with the Gaussian kernel G_σ. For the second integrability we use

$$p_2 = 2 - d_r.$$

In Algorithm of Sect. 7.1, the maximum number of outer iterations is set to be 25 in each experiment. We use 5 inner iterations to solve $2\lambda(u - f) - \text{div}(A(u)\nabla u) = 0$ for u. We use 10 AOS iterations with the timestep 0.1 to find a candidate for the solution of the equation of the form $2\lambda(v - f) - \text{div}(A\nabla v) = 0$.

We use the non-linear structure tensor obtained by (13) in the experiments. We did not see much difference in terms of the PSNR values when the linear structure tensor was used. But visually, the non-linear version seemed to give somewhat better results near the edges and reduced rounding of the corners. In G_ς of (13), we always use $\varsigma = 0.3$. In the discretization of (13), we use the AOS scheme with the timestep 1.0.

In the numerical experiments, M and N refer to the values used in the construction of g in (20) and p_1 in (24). The parameters k_1, k_2 and k_3 refer to α_1 and α_2 in (22) and (23); t_1 refers to the time in (24) and L refers to the number of the AOS iterations in (13).

In the experiments, we also consider the classical total variation regularization which is based on the minimization of the energy

$$\int_\Omega |\nabla u| \, dx + \mu \|f - u\|_{L^2(\Omega)}^2. \tag{25}$$

We solve the corresponding Euler–Lagrange equation using the minmod scheme with explicit time.

For each numerical comparison with the total variation regularization, we minimized the energy (25) for several values of μ and then we selected the μ for which the best PSNR value was obtained. The PSNR (peak signal-to-noise ratio) of an $M \times N$ pixel image u against the ground truth image f_0 is defined by

$$\text{PSNR} = 10 \log_{10} \left(\frac{(\max(f_0))^2 MN}{\sum_{i,j}(u_{ij} - f_{0_{ij}})^2} \right).$$

The noise-free images used in the experiments have their intensity values in the range [0, 255].

We also consider denoising results obtained by the edge-enhancing diffusion method proposed by Weickert [32]. The method is based on the diffusion equation

$$\begin{cases} \dfrac{\partial u}{\partial t} = \text{div}(D\nabla u) & \text{in } \Omega \times (0, T], \\ u(\cdot, 0) = f(\cdot), \\ n \cdot (D\nabla u) = 0. \end{cases} \quad (26)$$

Here $D = g(|\nabla u_\sigma|^2)\theta_1\theta_1^T + \theta_2\theta_2^T$, $g(s) = 1 - \exp\left(\frac{-C_4}{(s/\lambda^2)^4}\right)$ for $s > 0$, $g(s) = 1$ for $s \le 0$, $\theta_2||\nabla u_\sigma, |\theta_2| = 1, \theta_2 \perp \theta_1$, $\lambda > 0$ and $C_4 = 3.31488$. We use t_1 to denote the time needed to blur u using the linear heat equation such that the result corresponds to the convolution of u with the Gaussian kernel G_σ. We used the explicit time scheme and the discretization [32, Chapter 3.4] for $\text{div}(A\nabla u)$. We ran the diffusion (26) for several values of λ and t_1 and then selected the stopping time T such that the best PSNR value was obtained. We then selected λ, t_1 and T corresponding to the overall best PSNR value.

In the experiments, we also used the second-order total generalized variation regularizer [5] which is defined as

$$\text{TGV}_\alpha^2(u) := \sup \left\{ \int_\Omega u \, \text{div}^2 v \, dx \mid v \in C_c^2(\Omega, S^{2\times2}), \right.$$

$$\left. ||v||_\infty \le \alpha_0, || \text{div } v||_\infty \le \alpha_1 \right\}$$

where $S^{2\times2}$ denotes the space of symmetric 2×2 matrices. To minimize the total energy, we used the numerical implementation described in [4].

For comparisons, we also consider results obtained by the non-local means filter [7] and the BM3D method [9]. Regarding the BM3D, we used the matlab code available at http://www.cs.tut.fi/~foi/GCF-BM3D/. We ran this algorithm for several parameter values and selected the one for which the best PSNR value was obtained.

Regarding the implementation of the non-local means filter, we used

$$\hat{u}_{i,j} = \frac{1}{C_{i,j}} \sum_{|r-i|\le a, |s-j|\le a} C_{i,j}(r, s) f(r, s) \quad (27)$$

where

$$C_{i,j}(r, s) = e^{-\frac{||P((i,j),b)-P((r,s),b)||_{L^2}^2}{h(2b+1)^2\sigma^2}} \quad (28)$$

for $(r, s) \ne (i, j)$. For the centre pixel of a patch, $C_{i,j}(i, j) = \max_{(r,s)\ne(i,j)} C_{i,j}(r, s)$ is used. In (27) and (28), σ is the standard deviation of the noise, f is the noisy image, $a \in \mathbb{Z}^+$ is such that $(2a+1)^2$ is the size of the search window, $b \in \mathbb{Z}^+$ is such that $(2b+1)^2$ is the patch size, $P((i, j), b)$ is a patch of f such that (i, j) is the centre of the patch and $(2b+1)^2$ is the size of the patch, and $h > 0$ is a regularization parameter. In the denominator of (27), $C_{i,j}$ is a normalization factor,

$$C_{i,j} = \sum_{|r-i|\le a, |s-j|\le a} C_{i,j}(r, s).$$

In the experiments, we ran the algorithm for several values of a, b and h and then we selected the parameter values for which the best PSNR value was obtained.

First, we consider the denoising of a synthetic image. In Fig. 1, we see typical differences between the total variation regularization and the new method. In the solution obtained by the total variation regularization, some staircasing and loss of contrast are present. The new method aims at reducing these artefacts. In the new method, we used $\lambda = 0.05$, $M = 4$,
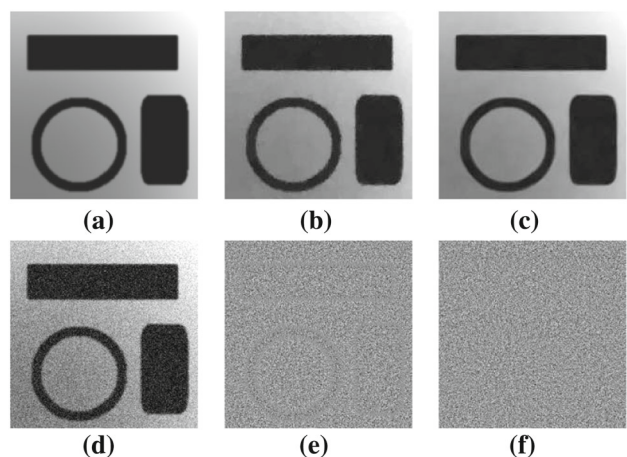

Fig. 1 *First column* the original noise-free image and the noisy image. *Second column* the restored image and the corresponding residual obtained by the TV denoising using $\mu = 0.03$. *Third column* the restored image and the corresponding residual obtained by the new method using $\lambda = 0.05$

Table 1 The parameter values in the results obtained by the edge-enhancing diffusion in Figs. 2, 3, 4, 5, 6, 7, 8 and 9

Image	PSNR	t_1	λ	T
Cameraman	28.2	0.1	16	0.57
Cameraman	24.3	0.2	16	2.85
Boat	29.4	0.2	11	0.89
Boat	26.5	0.4	11	2.45
Lena	27.9	0.2	10	2.38
Lena	25.7	0.4	11	4.24
Lighthouse	26.2	0.2	12	1.14
Lighthouse	23.7	0.2	14	2.82

Table 2 The parameter values in the results obtained by the non-local means algorithm in Figs. 2, 3, 4, 5, 6, 7, 8 and 9

Image	PSNR	$2b + 1$	$2a + 1$	h
Cameraman	29.1	3	31	0.8
Cameraman	25.0	13	15	0.3
Boat	29.8	11	7	0.5
Boat	26.6	21	11	0.2
Lena	28.4	19	11	0.3
Lena	25.9	25	11	0.2
Lighthouse	28.3	11	21	0.3
Lighthouse	26.1	19	23	0.2

Table 3 The parameter values used in the results obtained by the new method in Figs. 2, 3, 4, 5, 6, 7, 8 and 9

λ	M	N	k_1	k_2	k_3	t_1	L
0.07	2	3	10	0.0001	1.0	0.1	5
0.015	1	2	10	0.0001	1.0	0.2	15
0.18	4	5	1.0	0.0001	0.00001	0.1	25
0.09	4	5	1.0	0.0001	0.00001	0.2	25
0.1	4	5	1.0	0.0001	0.001	0.2	10
0.06	4	5	1.0	0.0001	0.001	0.4	15
0.1	4	5	1.0	0.0001	0.0001	0.1	25
0.04	4	5	1.0	0.0001	0.0001	0.2	25

The order of the images is as in Tables 1 and 2

$N = 5$, $k_1 = 1.0$, $k_2 = 0.0001$, $k_3 = 0.0001$, $t_1 = 0.1$ and $L = 25$.

Next we consider the denoising of four standard test images: cameraman, boat, Lena and lighthouse, each with two different levels of noise. The parameter values used in the edge-enhancing diffusion, the non-local means and the new method are seen in Tables 1, 2 and 3.

First, the cameraman image. The solutions corresponding to different noise levels are seen in Figs. 2 and 3. In Fig. 2c we used $\mu = 0.044$, whereas in Fig. 3c, $\mu = 0.014$. In the total

generalized variation results, we used $\alpha_0 = 27$, $\alpha_1 = 13.5$ in Fig. 2e and $\alpha_0 = 92$, $\alpha_1 = 41.4$ in Fig. 3e.

Next we consider the denoising of the boat test image in Figs. 4 and 5. We used $\mu = 0.041$ in Fig. 4c and $\mu = 0.018$ in Fig. 5c. Regarding the total generalized variation results we used $\alpha_0 = 19$, $\alpha_1 = 14.25$ in Fig. 4e and $\alpha_0 = 50$, $\alpha_1 = 32.5$ in Fig. 5e.

Next we consider the denoising of the Lena test image in Figs. 6 and 7. We see two noisy versions of this test image in Figs. 6b and 7b. In the total variation regularization results we used $\mu = 0.026$ in Fig. 6c and $\mu = 0.014$ in Fig. 7c. In the total generalized variation results we used $\alpha_0 = 38$, $\alpha_1 = 22.8$ in Fig. 6e and $\alpha_0 = 100$, $\alpha_1 = 60$ in Fig. 7e.

Finally, we consider the denoising of the lighthouse test image in Figs. 8 and 9. In the solutions obtained by the total variation regularization we used $\mu = 0.027$ in Fig. 8c and $\mu = 0.015$ in Fig. 9c. In the total generalized variation results, we used $\alpha_0 = 35$, $\alpha_1 = 21$ in Fig. 8e and $\alpha_0 = 87$, $\alpha_1 = 39.15$ in Fig. 9e.

We see that in all the experiments in Figs. 2, 3, 4, 5, 6, 7, 8 and 9, in terms of the PSNR values, the new method gives better results than the total variation regularization, edge-enhancing diffusion and total generalized variation. Nevertheless, just higher PSNR values do not necessarily mean visually better. It also seems that the criterion we used to select the 'best' parameter values in the methods proposed earlier in the literature may leave some noise especially in the results obtained by the edge-enhancing diffusion method. Somewhat the same is with the results obtained by the second-order total generalized variation.

From the experiments, it also seems that using the minimization of the energy (3), the edges are still sharp but are not as grainy as in the results produced by the total variation regularization or the total generalized variation regularization. In this aspect, the new method is quite similar to the edge-enhancing diffusion method. The observation that the new method can reduce noise efficiently along the boundary is also quite expected from the key feature 3 in Sect. 2 since provided the edge direction is estimated correctly, the linear averaging along the edge reduces noise efficiently. This is clearly seen in Fig. 7.

Compared to the total variation regularization, the new method hardly suffers from the staircasing. The new method is also quite as good as total variation at preserving edges except very weak edges which our method tends to blur, since in those areas our regularizer may simplify to isotropic smoothing, but on the other hand, the total variation may not place the weak edges correctly as can be seen with the cameraman experiments. Also, the new method gives higher or the same PSNR values than the implementation we used of the NL means algorithm, except with the last image with strong noise, which is not a surprise considering that the lighthouse image has lot of repetitive components. In all
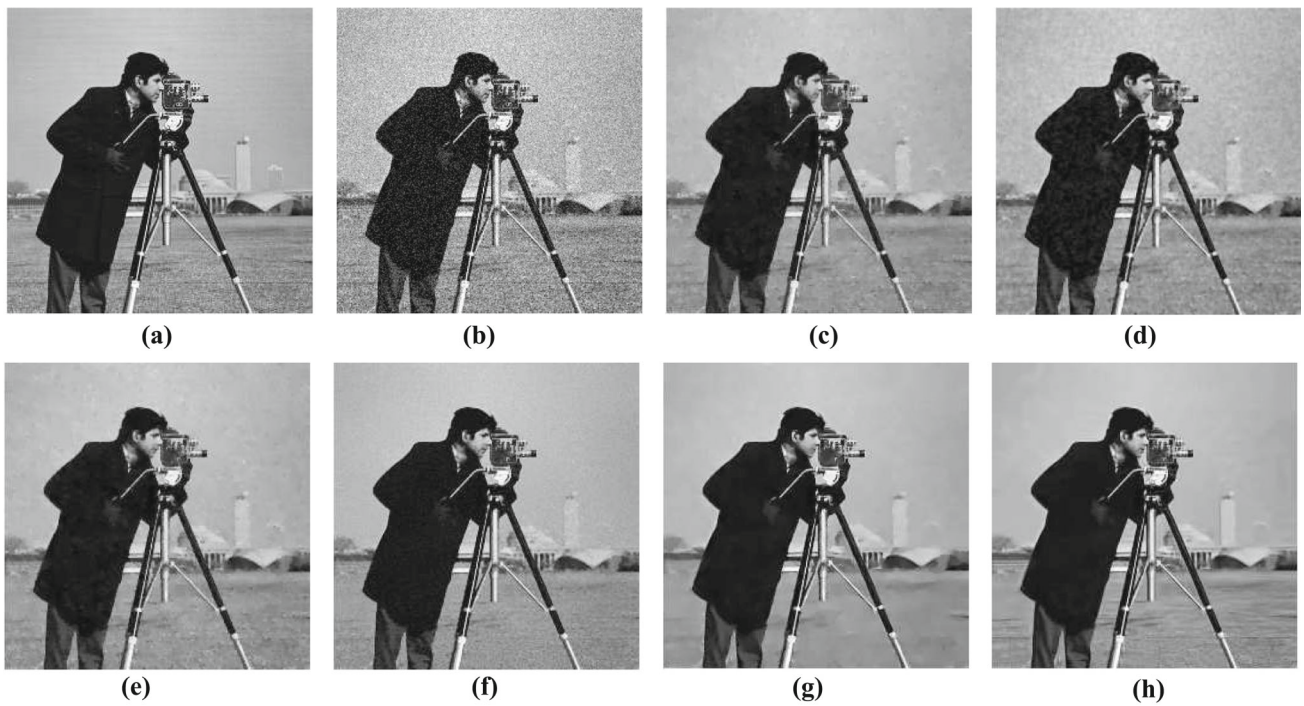
Fig. 2 **a** Noise-free. **b** Noisy, PSNR $= 22.1$. **c** TV, PSNR $= 28.9$. **d** EED, PSNR $= 28.2$. **e** TGV, PSNR $= 28.8$. **f** NL means, PSNR $= 29.1$. **g** New method, PSNR $= 29.3$. **h** B3MD, PSNR $= 30.4$
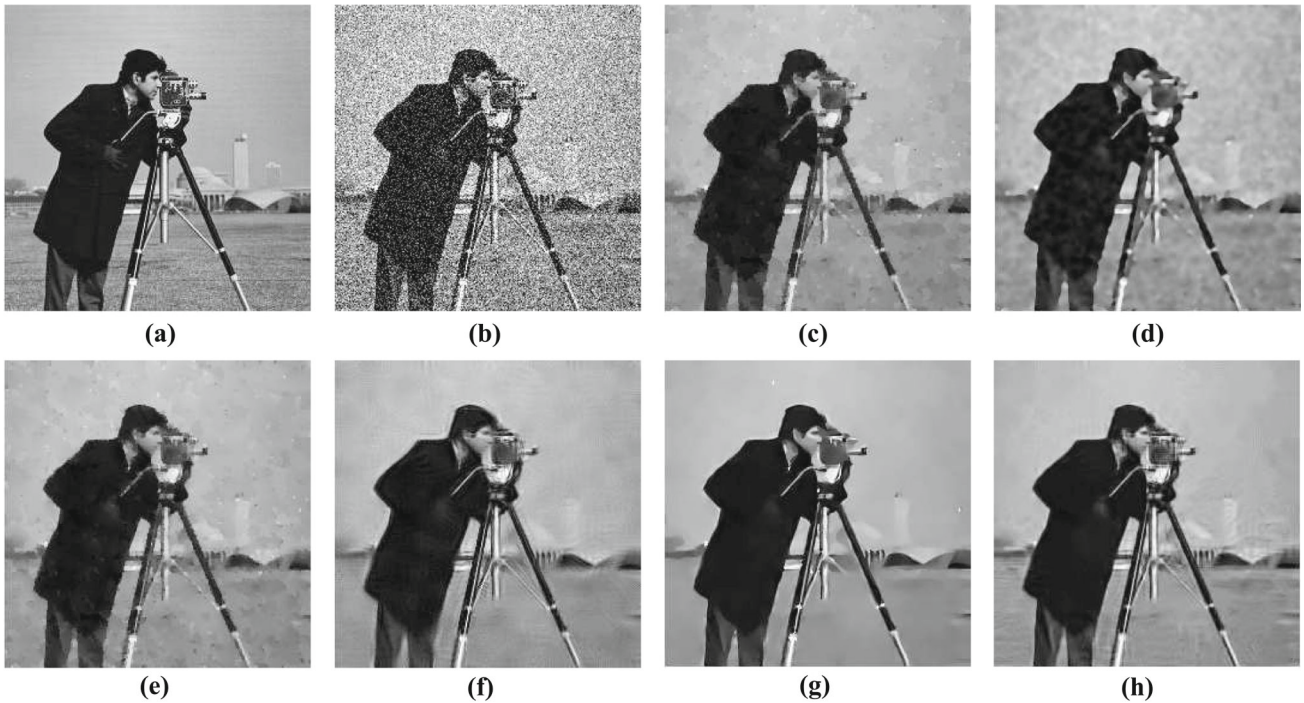

Fig. 3 **a** Noise-free. **b** Noisy, PSNR $= 14.2$. **c** TV, PSNR $= 24.6$. **d** EED, PSNR $= 24.3$. **e** TGV, PSNR $= 24.6$. **f** NL means, PSNR $= 25.0$. **g** New method, PSNR $= 25.7$. **h** B3MD, PSNR $= 26.3$
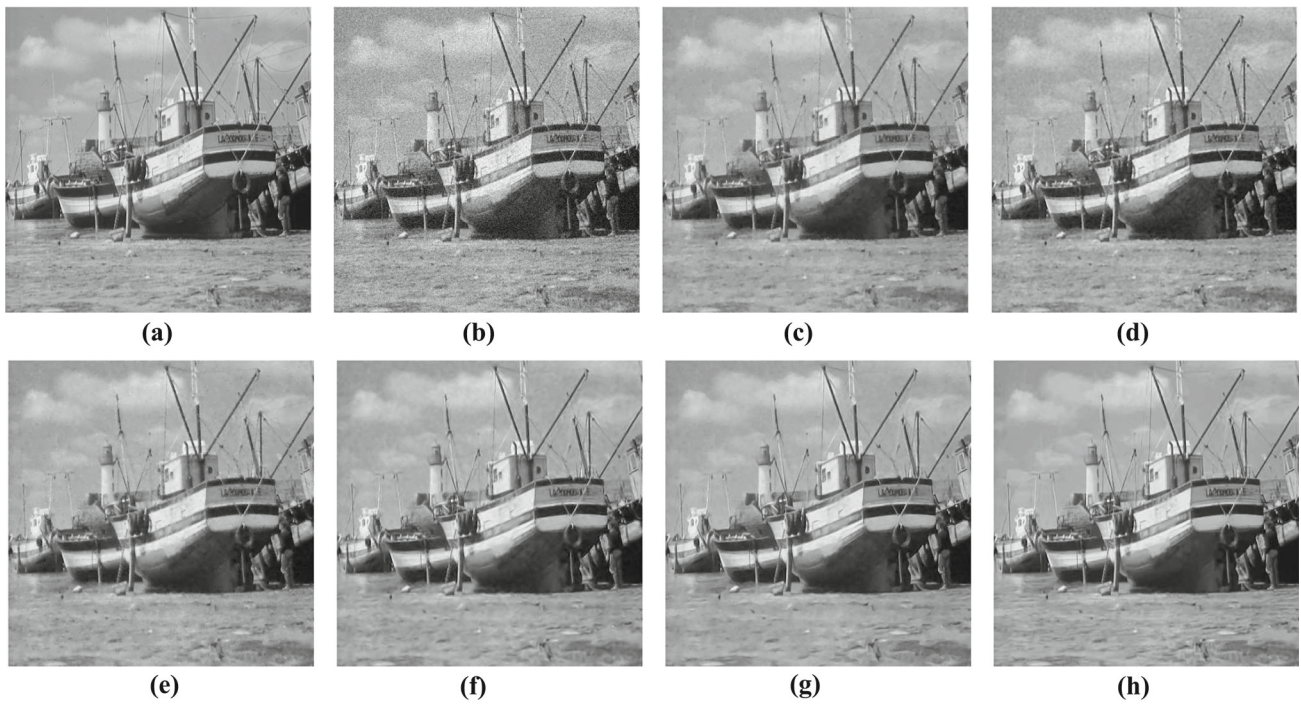
Fig. 4 **a** Noise-free. **b** Noisy, PSNR = 22.1. **c** TV, PSNR = 29.2. **d** EED, PSNR = 29.4. **e** TGV, PSNR = 29.3. **f** NL means, PSNR = 29.8. **g** New method, PSNR = 30.0. **h** B3MD, PSNR = 30.9
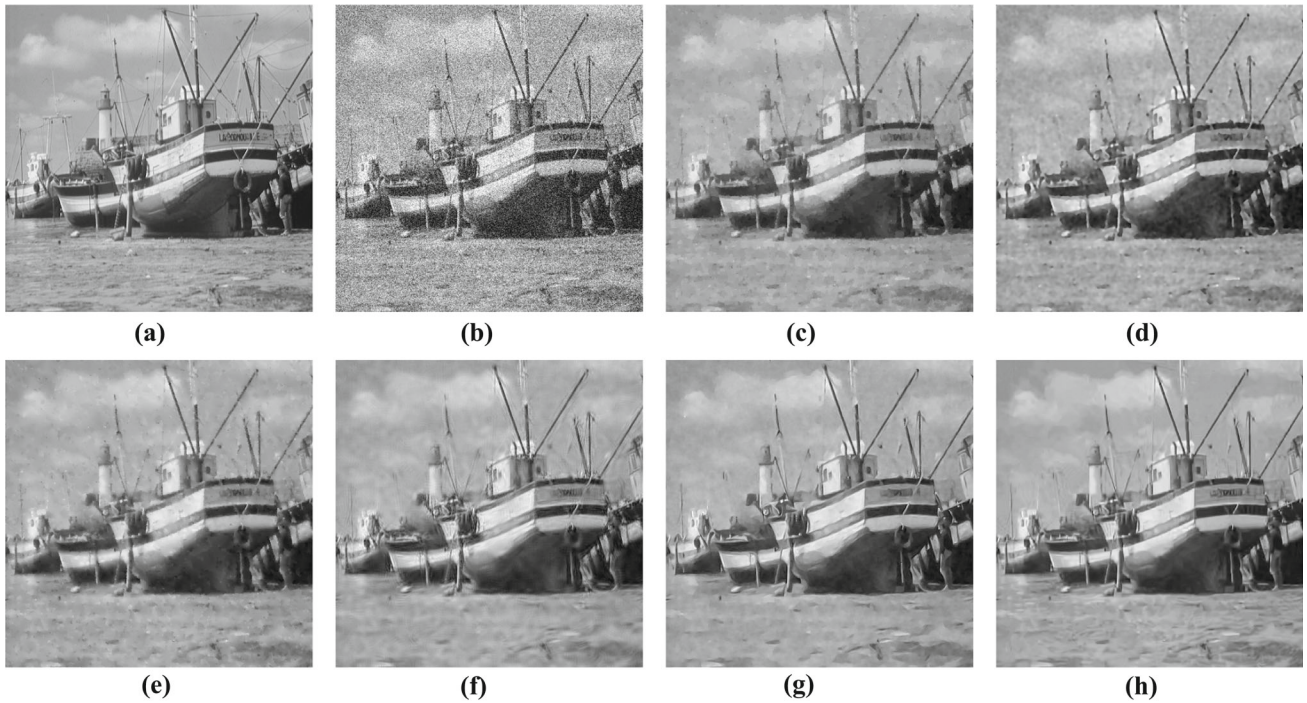

Fig. 5 **a** Noise-free. **b** Noisy, PSNR = 16.1. **c** TV, PSNR = 26.3. **d** EED, PSNR = 26.5. **e** TGV, PSNR = 26.4. **f** NL means, PSNR = 26.6. **g** New method, PSNR = 27.1. **h** B3MD, PSNR = 27.7

Fig. 6 **a** Noise-free. **b** Noisy, PSNR = 18.6. **c** TV, PSNR = 27.2. **d** EED, PSNR = 27.9. **e** TGV, PSNR = 27.4. **f** NL means, PSNR = 28.4. **g** New method, PSNR = 28.4. **h** B3MD, PSNR = 29.2
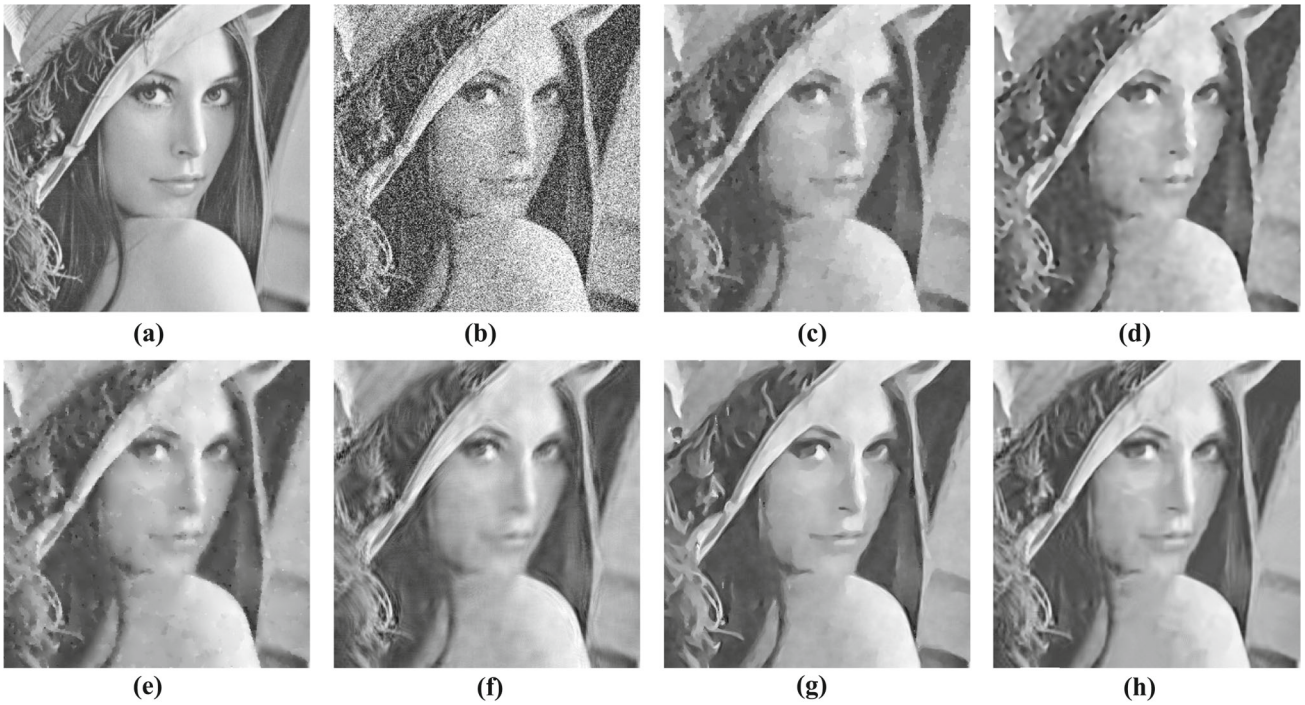

Fig. 7 **a** Noise-free. **b** Noisy, PSNR = 14.2. **c** TV, PSNR = 25.1. **d** EED, PSNR = 25.7. **e** TGV, PSNR = 25.2. **f** NL means, PSNR = 25.9. **g** New method, PSNR = 26.4. **h** B3MD, PSNR = 26.9

Fig. 8 **a** Noise-free. **b** Noisy, PSNR = 18.6. **c** TV, PSNR = 26.5. **d** EED, PSNR = 26.2. **e** TGV, PSNR = 26.3. **f** NL means, PSNR = 28.3. **g** New method, PSNR = 28.4. **h** B3MD, PSNR = 29.9

Fig. 9 **a** Noise-free. **b** Noisy, PSNR = 14.2. **c** TV, PSNR = 24.0. **d** EED, PSNR = 23.7. **e** TGV, PSNR = 24.0. **f** NL means, PSNR = 26.1. **g** New method, PSNR = 25.9. **h** B3MD, PSNR = 27.4

the experiments, the best results are obtained by the B3MD method, which is also a non-local method, contrary to our method which is local.

9 Conclusion

In this paper, a new variational image denoising method is proposed. The new method could be seen to be a two-step method in which the steps are in some sense coupled. Structure tensor analysis is used to guide the denoising in such a way that the energy minimization favours solutions whose gradient fields respect the approximation obtained by the structure tensor analysis. In the theoretical part, the existence of a minimizer of a weak form was considered. In the numerical part, an algorithm was proposed based on iterative minimization. The numerical experiments demonstrated the possible advantages of the new method over some existing variational and partial differential equations methods. The future considerations involve extending the method to the non-local framework.

References

1. Bayram, I., Kamasak, M.E.: Directional total variation. IEEE Signal Process. Lett. **19**, 781–784 (2012)
2. Blomgren, P., Mulet, P., Chan, T., Wong, C.: Total variation image restoration: Numerical methods and extensions. In: Proceedings of the IEEE International Conference on Image Processing, vol. 3, pp. 384–387. IEEE Computer Society Press, Piscataway (1997)
3. Bollt, E.M., Chartrand, R., Esedoglu, S., Schultz, P., Vixie, K.R.: Graduated, adaptive image denoising: local compromise between total-variation and isotropic diffusion. Adv. Comput. Math. **31**, 61–85 (2009)
4. Bredies, K.: Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty. In: Efficient Algorithms for Global Optimization Methods in Computer Vision, pp. 44–77 (2014)
5. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM J. Imaging Sci. **3**, 492–526 (2010)
6. Brox, T., Weickert, J., Burgeth, B., Mrazek, P.: Nonlinear structure tensors. Image Vis. Comput. **24**, 41–55 (2006)
7. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. **4**, 490–530 (2005)
8. Chen, Y., Levine, S., Rao, M.: Variable exponent, linear growth functionals in image restoration. SIAM J. Appl. Math. **66**, 1383–1406 (2006)
9. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**, 2080–2095 (2007)
10. Didas, S., Burgeth, B., Imiya, A., Weickert, J.: Regularity and scale-space properties of fractional high order linear filtering. In: Kimmel, R., Sochen, N.A., Weickert, J. (eds.) Scale Space and PDE Methods in Computer Vision, 5th International Conference, Scale-Space, pp. 13–25. Hofgeismar, Germany (2005)
11. Didas, S., Weickert, J., Burgeth, B.: Properties of higher order nonlinear diffusion filtering. J. Math. Imaging Vision **35**, 208–226 (2009)
12. Duits, R., Florack, L., de Graaf, J., ter Haar Romeny, B.: On the axioms of scale space theory. J. Math. Imaging Vision **20**, 267–298 (2004)
13. Estellers, V., Soatto, S., Bresson, X.: Adaptive regularization with the structure tensor. IEEE Trans. Image Process. **24**, 1777–1790 (2015)
14. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305 (1987)
15. Grasmair, M., Lenzen, F.: Anisotropic total variation filtering. Appl. Math. Optim. **62**, 323–339 (2010)
16. Guo, Z., Sun, J., Zhang, D., Wu, B.: Adaptive Perona–Malik model based on the variable exponent for image denoising. IEEE Trans. Image Process. **21**, 958–967 (2012)
17. Hahn, J., Lee, C.-O.: A nonlinear structure tensor with the diffusivity matrix composed of the image gradient. J. Math. Imaging Vis. **34**, 137–151 (2009)
18. Hahn, J., Tai, X.-C., Borok, S., Bruckstein, A.M.: Orientation-matching minimization for image denoising and inpainting. Int. J. Comput. Vis. **92**, 308–324 (2011)
19. Hu, Y., Jacob, M.: Higher degree total variation (HDTV) regularization for image recovery. IEEE Trans. Image Process. **21**, 2559–2571 (2012)
20. Lefkimmiatis, S., Osher, S.: Non-local structure tensor functionals for image regularization. IEEE Trans. Comput. Imaging **1**, 16–29 (2015)
21. Lefkimmiatis, S., Roussos, A., Maragos, P., Unser, M.: Structure tensor total variation. SIAM J. Imaging Sci. **8**, 1090–1122 (2015)
22. Lenzen, F., Berger, J.: Solution-driven adaptive total variation regularization. In: SSVM Proceedings, pp. 203–215 (2015)
23. Lenzen, F., Becker, F., Lellmann, J.: Adaptive second-order total variation: an approach aware of slope discontinuities. In: Proceedings of the Fourth International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), pp. 61–73 (2013)
24. Li, F., Li, Z., Pi, L.: Variable exponent functionals in image restoration. Appl. Math. Comput. **216**, 870–882 (2010)
25. Lysaker, M., Osher, S., Tai, X.-C.: Noise removal using smoothed normals and surface fitting. IEEE Trans. Image Process. **13**, 1345–1357 (2004)
26. Rahman, T., Tai, X-C., Osher, S.: A TV-Stokes denoising algorithm. In: Proceedings of Scale Space and Variational Methods in Computer Vision (SSVM), pp. 473–482 (2007)
27. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
28. Steidl, G., Teuber, T.: Anisotropic smoothing using double orientations. In: Proceedings of the Second International Conference Scale Space and Variational Methods in Computer Vision, pp. 477–489 (2009)
29. Takeda, H., Farsiu, S., Milanfar, P.: Kernel regression for image processing and reconstruction. IEEE Trans. Image Process. **16**, 349–366 (2007)
30. Tschumperlé, D.: PDE's based regularization of multivalued images and applications, PhD thesis, University of Nice Sophia Antipolis (2002)
31. Vogel, C.R., Oman, M.E.: Iterative methods for total variation denoising. SIAM J. Sci. Comput. **17**, 227–238 (1996)
32. Weickert, J.: Anisotropic Diffusion in Image Processing. Teubner-Verlag, Stuttgart (1998)

33. Weickert, J.: Coherence-enhancing diffusion filtering. Int. J. Comput. Vis. **31**, 111–127 (1999)
34. Weickert, J., Romeny, B.M.T.H., Viergever, M.A.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans. Image Process. **7**, 398–410 (1998)
35. Zhang, J., Lai, R., Kuo, C.J.: Adaptive directional total-variation model for latent fingerprint segmentation. IEEE Trans. Inf. Forensics Secur. **8**, 1261–1273 (2013)
36. Åstöm, F., Baravdish, G., Felsberg, M.: A Tensor variational formulation of gradient energy total variation. In: Proceedings of the 10th International Conference EMMCVPR, pp. 307–320 (2015)

Juha Tiirola received a Ph.D. degree in applied mathematics from the University of Oulu, Finland, in 2015. His research interests include inverse problems in imaging.