

Universität des Saarlandes



Fachrichtung 6.1 – Mathematik

Preprint Nr. 248

**Generalised Nonlocal Image Smoothing**

Luis Pizarro, Pavel Mrázek, Stephan Didas,  
Sven Grewenig and Joachim Weickert

Saarbrücken 2009



## Generalised Nonlocal Image Smoothing

**Luis Pizarro**

Saarland University  
Department of Mathematics  
P.O. Box 15 11 50  
D-66041 Saarbrücken  
Germany  
pizarro@mia.uni-saarland.de

**Pavel Mrázek**

UPEK Prague R&D Center  
Husinecká 7  
13000 Praha 3  
Czech Republic  
pavel.mrazek@upek.com

**Stephan Didas**

Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM  
Abteilung Bildverarbeitung  
Fraunhofer-Platz 1  
67663 Kaiserslautern  
Germany  
stephan.didas@itwm.fraunhofer.de

**Sven Grewenig**

Saarland University  
Department of Mathematics  
P.O. Box 15 11 50  
66041 Saarbrücken  
Germany  
grewenig@mia.uni-saarland.de

**Joachim Weickert**

Saarland University  
Department of Mathematics  
P.O. Box 15 11 50  
66041 Saarbrücken  
Germany  
weickert@mia.uni-saarland.de

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

Fax: + 49 681 302 4443  
e-Mail: [preprint@math.uni-sb.de](mailto:preprint@math.uni-sb.de)  
WWW: <http://www.math.uni-sb.de/>

## Abstract

We propose a discrete variational approach for image smoothing consisting of nonlocal data and smoothness constraints that penalise general dissimilarity measures defined on image patches. One of such dissimilarity measures is the weighted  $L_2$  distance between patches. In such a case we derive an iterative neighbourhood filter that induces a new similarity measure in the photometric domain. It can be regarded as an extended patch similarity measure that evaluates not only the patch similarity of two chosen pixels, but also the similarity of their corresponding neighbours. This leads to a more robust smoothing process since the pixels selected for averaging are more coherent with the local image structure. The suggested approach includes two recently proposed filters as special cases: The NL-means filter of Buades *et al.* and the NDS filter of Mrázek *et al.* In fact, the approach introduced here can be considered as a generalisation of the latter filter. We evaluate our method for the task of denoising greyscale and colour images degraded by Gaussian and impulse noise, demonstrating that it compares very well to other more sophisticated patch-based approaches.

**Keywords:** Discrete variational methods, Nonlocal image smoothing, Neighbourhood filters, Nonlinear filtering, Image denoising.

## 1 Introduction

Image smoothing is a fundamental task in image processing. It serves as a noise removal tool for improving the visual quality of noisy images taken from digital cameras or scanners, as well as for providing simplified input images that are further processed in tasks such as segmentation, feature extraction and texture analysis. There exist numerous approaches to image smoothing emerging from statistical methods, information theory, transforms in the frequency domain, partial differential equations (PDEs) and variational methods [82, 1, 84, 18]. Establishing equivalences and relations between the different approaches has been focus of intense research in recent years [6, 31, 32, 56, 67, 68, 77, 71, 85].

Mrázek *et al.* [56] pointed out the relations between several nonlinear smoothing methods such as M-estimators [23, 85], bilateral filtering [75], diffusion filters [61, 81], and regularisation/Bayesian techniques [7, 35, 57, 85]. Although these methods seem very different at the first glance and originate in different mathematical theories, Mrázek *et al.* showed that they lead to highly similar discrete algorithms, and that all these methods can be cast in a single unified framework of discrete regularisation theory. The unifying model is formulated as an energy functional with *nonlocal data* and *smoothness* (NDS) terms – hence called NDS model. The data term rewards similarity of the filtered image to the input (noisy)

image, while the smoothness term penalises high deviations from regularity on the solution. These terms can consider not only information from a small region around a pixel but also make it possible to involve large neighbourhoods. Pizarro *et al.* [64] showed that the NDS approach can outperform the methods obtained as special cases mainly by adjusting the spatial extent where the nonlocal pixel interactions occur. These interactions take the form of nonlinear differences of intensity measuring pixel similarity. However, single differences do not carry reliable information about the local image structure/geometry too far away from a chosen pixel. Thus, truly nonlocal interactions in the NDS model are rather limited in practice. This is actually the main drawback of *single differences*-based approaches.

Two equivalent and simultaneously proposed methods, namely the *non-local means* (NL-means) filter [14, 15] and the *unsupervised, information-theoretic, adaptive* (UINTA) filter [3, 4] are able to cope with such a problem. Both methods consider a whole neighbourhood (or patch) around a pixel to measure similarity. In this way, if the corresponding neighbourhoods of two pixels are similar, the pixels themselves will be considered alike even if they are spatially distant from each other. This simple idea allows a real incorporation of nonlocal pixel interactions in the smoothing process, providing impressive denoising results. The NL-means filter belongs to the class of neighbourhood filters [51, 86, 70, 75, 16] that average similar pixels based on their photometric and spatial proximities – where the spatial distance does not play a role in NL-means. In particular, it can be seen as a bilateral filter [75] with a patch-based photometric similarity measure. Several variational formulations of the NL-means filter have been proposed [47, 38, 5, 13, 52] together with acceleration techniques [55, 8, 26, 28, 13, 60] and invariant patch similarity measures [79, 48, 91, 53]. This method has inspired the appearance of numerous so-called *patch-based* approaches for image smoothing, deblurring, segmentation, inpainting, super-resolution, and texture synthesis, among others.

In this paper we propose the *Generalised NDS* (GNDS) framework for image smoothing as an extension of the NDS model of Mrázek *et al.* [56]. Instead of penalising deviations from similarity considering only single pixel differences, as in the NDS model, we introduce a discrete variational approach with nonlocal constraints that penalise general dissimilarity measures defined on image patches. As an example of such dissimilarity measures we consider the weighted  $L_2$  distance between patches used in the NL-means filter. In such a case the resulting *GNDS filter* can be considered as an iterative neighbourhood filter consisting of two terms, one prescribing the solution to be nonlocally similar to the input image and the other imposing nonlocal regularity on the solution. Another characteristic of this filtering model is that it induces a new similarity measure in the photometric domain. We regard it as an *extended* patch similarity measure that evaluates not only the patch similarity of two chosen pixels, but also the similarity of their corresponding neighbours. This makes the selection of the most similar pixels

in the averaging (filtering) process more robust. The new similarity measure includes three special cases: (i) similarity of single pixels, in which case we get the NDS filter of Mrázek *et al.* [56], (ii) *isotropic* patch similarity, which leads to the NL-means filter of Buades *et al.* [15], and (iii) *anisotropic* patch similarity, which results in a novel filter for removal of impulse noise.

This article is structured as follows: Section 2 describes the NDS model proposed by Mrázek *et al.*, and the most important filters that can be obtained from it as special cases are summarised in Section 3. In Section 4, we report new relations between the NDS framework and recently proposed graph regularisation techniques. In Section 5, we introduce the *Generalised NDS* model, discuss relations to other patch-based approaches, its extension to multichannel images and the use of other similarity measures. In Section 6, we evaluate both the NDS and the GNDS approaches for the task of denoising images degraded by Gaussian and salt-and-pepper noise. We show that the NDS model outperforms other classical non-patch-based approaches and that our GNDS model auspiciously compares to other more sophisticated patch-based methods. We conclude the paper in Section 7.

## 2 The Nonlocal Data and Smoothness (NDS) Model

Let  $f, u : \Omega \rightarrow \mathbb{R}$  be scalar images defined on the discrete image domain  $\Omega$ .  $f$  stands for the (noisy) original image while  $u$  represents a processed version of it. Let  $J = \{1, \dots, N\}$  be the index set of all pixels in the images. The pixel position in the bi-dimensional grid is indicated by  $x_i$  ( $i \in J$ ). The discrete energy function  $E$  of the NDS filter presented in [56] is a convex combination of a nonlocal data (or similarity) term  $E_D$  and a nonlocal smoothness term  $E_S$ :

$$E_D(u) = \sum_{i,j \in J} \Psi_D(|u_i - f_j|^2) w_D(|x_i - x_j|^2), \quad (1)$$

$$E_S(u) = \sum_{i,j \in J} \Psi_S(|u_i - u_j|^2) w_S(|x_i - x_j|^2). \quad (2)$$

Here  $\Psi_\circ : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$  are increasing functions that penalise large (greyvalue) *tonal distances*, e.g., the Cauchy function  $\Psi(s^2) = \lambda^2 \log(1 + s^2/\lambda^2)$  [42, 61]. The weights  $w_\circ : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$  are nonnegative functions downweighting large *spatial distances*, e.g., the hard window  $w(x^2) = \{1 \text{ for } x^2 < r^2, 0 \text{ otherwise}\}$  [49] or the soft window  $w(x^2) = \exp(-x^2/(2r^2))$  [23]. For a more comprehensive list of penalisers, see [58].

The complete NDS model can be regarded as a discrete nonlocal variational

method combining both the data (1) and the smoothness (2) terms:

$$\begin{aligned}
E(u) &= (1 - \alpha) E_D(u) + \alpha E_S(u) \\
&= (1 - \alpha) \sum_{i,j \in J} \Psi_D (|u_i - f_j|^2) w_D (|x_i - x_j|^2) \\
&\quad + \alpha \sum_{i,j \in J} \Psi_S (|u_i - u_j|^2) w_S (|x_i - x_j|^2)
\end{aligned} \tag{3}$$

with regularisation parameter  $\alpha \in [0, 1]$ .

## 2.1 Numerical Implementation

After introducing the NDS model in the previous section, we now consider a robust and stable iterative procedure for minimising the energy functional. Even if the presented iterative fixed point approach is very simple, we will see that it satisfies a maximum-minimum principle for a general set of penaliser functions, and we will prove the existence of a fixed point.

Taking the partial derivatives of the data term (1) yields

$$\frac{\partial E_D}{\partial u_k} = 2 \sum_{j \in J} \Psi'_D (|u_k - f_j|^2) (u_k - f_j) w_D (|x_k - x_j|^2) , \tag{4}$$

where  $\Psi'$  denotes the derivative of  $\Psi$  w.r.t. its argument. In a similar way we calculate the derivatives of the smoothness term (2) which leads to

$$\frac{\partial E_S}{\partial u_k} = 4 \sum_{j \in J} \Psi'_S (|u_k - u_j|^2) (u_k - u_j) w_S (|x_k - x_j|^2) . \tag{5}$$

It is clear that the complete derivatives then have the form

$$\frac{\partial E}{\partial u_i} = (1 - \alpha) \frac{\partial E_D}{\partial u_i} + \alpha \frac{\partial E_S}{\partial u_i} . \tag{6}$$

For a critical point  $u$  of the energy functional  $E$  we have

$$\nabla E(u) = 0 \iff \frac{\partial E}{\partial u_i} = 0 \text{ for all } i \in J . \tag{7}$$

We define the abbreviations

$$d_{i,j} := \Psi'_D (|u_i - f_j|^2) w_D (|x_i - x_j|^2) , \tag{8}$$

$$s_{i,j} := 2\Psi'_S (|u_i - u_j|^2) w_S (|x_i - x_j|^2) \tag{9}$$

which help us to rewrite (7) as

$$0 = (1 - \alpha) \sum_{j \in J} d_{i,j} (u_i - f_j) + \alpha \sum_{j \in J} s_{i,j} (u_i - u_j) \tag{10}$$



where we use the partial derivatives shown in (4) and (5). This can be transformed into fixed point form

$$u_i = \frac{(1 - \alpha) \sum_{j \in J} d_{i,j} f_j + \alpha \sum_{j \in J} s_{i,j} u_j}{(1 - \alpha) \sum_{j \in J} d_{i,j} + \alpha \sum_{j \in J} s_{i,j}}. \quad (11)$$

To have a positive denominator we assume that  $\Psi'_D(s^2) > 0$  and  $\Psi'_S(s^2) > 0$ , i. e., the penalisers are monotonically increasing. Furthermore we assume that  $w_D(s^2) \geq 0$ ,  $w_S(s^2) \geq 0$  as well as  $w_D(0) > 0$  and  $w_S(0) > 0$  for the spatial weights. We use this equation to build up a first iterative method to minimise the value of  $E$  where the upper index  $k$  denotes the iteration number. Note that  $d_{i,j}$  and  $s_{i,j}$  also depend on the evolving image  $u^k$  and thus also get a superscript to denote the iteration level involved. The corresponding fixed point iteration then reads as

$$u_i^0 := f_i, \quad (12)$$

$$u_i^{k+1} := \frac{(1 - \alpha) \sum_{j \in J} d_{i,j}^k f_j + \alpha \sum_{j \in J} s_{i,j}^k u_j^k}{(1 - \alpha) \sum_{j \in J} d_{i,j}^k + \alpha \sum_{j \in J} s_{i,j}^k}. \quad (13)$$

In the following we will write this scheme (13) in the form  $u^{k+1} = F(u^k)$  with  $F : \mathbb{R}^N \rightarrow \mathbb{R}^N$ . We note that we calculate  $u^{k+1}$  using only components of the vector  $u^k$  of the old iteration level:

$$u_i^{k+1} := F^i(u^k) \quad \text{for all } i \in J, k \in \mathbb{N}. \quad (14)$$

Such a method can also be called a *nonlinear Jacobi method*.

Let us now state two important results.

**Proposition 2.1 (Maximum-Minimum Principle)**

*With the assumptions on  $\Psi_D$ ,  $\Psi_S$ ,  $w_D$ , and  $w_S$  as above, the scheme (13) satisfies a maximum-minimum principle:*

$$\min_{j \in J} f_j \leq u_i^k \leq \max_{j \in J} f_j \quad \text{for all } i \in J, k \in \mathbb{N}. \quad (15)$$

**Proof:** With our assumptions on the tonal and spatial weights from above we know that  $d_{i,j}^k \geq 0$  and  $s_{i,j}^k \geq 0$  for all  $i, j, k$ . That means in (13),  $u_i^{k+1}$  is calculated as a convex combination of grey values of the initial image  $f$  and of the last iteration step  $u^k$ . Thus we have

$$\min_{j \in J} \{u_j^k, f_j\} \leq u_i^{k+1} \leq \max_{j \in J} \{u_j^k, f_j\} \quad \text{for all } i \in J, k \in \mathbb{N}. \quad (16)$$

Induction shows that the fixed point scheme (13) satisfies a maximum-minimum principle, i. e.

$$\min_{j \in J} f_j \leq u_i^k \leq \max_{j \in J} f_j \quad \text{for all } i \in J, k \in \mathbb{N}. \quad (17)$$

□

In the next proposition, we see that this property is not only useful from a practical point of view: Together with continuity, it gives us the existence of a fixed point.

**Proposition 2.2 (Existence of a Fixed Point)**

*The fixed point equation (11) has a solution.*

**Proof:** Let us consider the set  $M := \{u \in \mathbb{R}^N \mid \|u\|_\infty \leq \|f\|_\infty\}$  with the norm  $\|u\|_\infty := \max_{j \in J} |u_j|$ .  $M$  is nonempty, compact and convex. Then the maximum-minimum stability implies that  $F(M) \subseteq M$ . With our requirements on the tonal and spatial weights, the denominator in (13) is always larger than zero. This means that each component  $F_i : \mathbb{R}^N \rightarrow \mathbb{R}$  is continuous with respect to the norm  $\|\cdot\|_\infty$ . Since this holds for all  $i$ , we know that  $F : (\mathbb{R}^N, \|\cdot\|_\infty) \rightarrow (\mathbb{R}^N, \|\cdot\|_\infty)$  is continuous. Then Brouwer’s fixed point theorem (see [11] or [87, page 51], for example) shows that  $F$  has a fixed point in  $M$ . □

From the derivation it is clear that a fixed point corresponds to a critical point of  $E$ . If we have chosen our penaliser functions such that the energy functional is strictly convex, this is equivalent to the unique minimum of  $E$ .

Alternatively, the solution of the NDS energy (3) can be obtained by gradient descent optimisation:

$$\frac{u_i^{k+1} - u_i^k}{\tau} = -\frac{\partial E}{\partial u_i^k} \quad \text{for all } i \in J, \tag{18}$$

with step size  $\tau > 0$ . Considering (7)–(11), the energy minimiser is computed as

$$\begin{aligned} u_i^0 &:= f_i, & (19) \\ u_i^{k+1} &:= (1 - \tau)u_i^k + \tau \frac{(1 - \alpha) \sum_{j \in J} d_{i,j}^k f_j + \alpha \sum_{j \in J} s_{i,j}^k u_j^k}{(1 - \alpha) \sum_{j \in J} d_{i,j}^k + \alpha \sum_{j \in J} s_{i,j}^k}. & (20) \end{aligned}$$

Note that by setting  $\tau = 1$  one obtains the fixed point iteration (13).

### 3 Important Special Cases

Recall that the NDS functional (3) can be optimised using the fixed point iterations (11). Let us introduce the following notation for the tonal weights,

$$g_{i,j}^D := \Psi'_D (|u_i - f_j|^2), \tag{21}$$

$$g_{i,j}^S := 2 \Psi'_S (|u_i - u_j|^2), \tag{22}$$

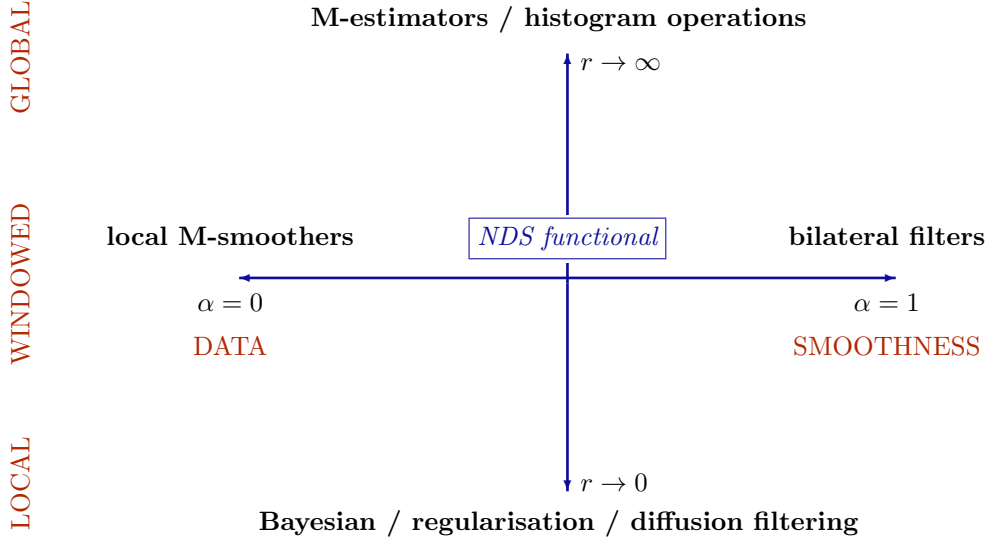


Figure 1: Overview of methods covered by the NDS functional. Left to right, the parameter  $\alpha$  balances the data and smoothness terms. Bottom to top, the window size  $r$  determines the spatial extent of the methods, from local to global.

and for the spatial weights,

$$w_{i,j}^{D,r} := w_D (|x_i - x_j|^2), \quad (23)$$

$$w_{i,j}^{S,r} := w_S (|x_i - x_j|^2), \quad (24)$$

where the spatial weights  $w$  implicitly contain a scale parameter  $r$  specifying how quickly the weight decreases: Small  $r$  means a local operation (or a smaller window), larger  $r$  leads to operations with large-scale effects. The window sizes for the data and smoothness terms may differ.

Using this notation, equation (11) can be written as

$$u_i = \frac{(1 - \alpha) \sum_{j \in J} g_{i,j}^D w_{i,j}^{D,r} f_j + \alpha \sum_{j \in J} g_{i,j}^S w_{i,j}^{S,r} u_j}{(1 - \alpha) \sum_{j \in J} g_{i,j}^D w_{i,j}^{D,r} + \alpha \sum_{j \in J} g_{i,j}^S w_{i,j}^{S,r}}. \quad (25)$$

In the following sections we show that many well known filtering and estimation methods can be derived from equation (25) (and thus from the NDS functional) by a simple choice of the parameter  $\alpha$  which balances the smoothness and data terms, the window size  $r$ , and by an appropriate selection of the weighting functions  $g^D$ ,  $g^S$ ,  $w^D$ , and  $w^S$ . Fig. 1 gives an overview of the NDS landscape and the methods covered below.

### 3.1 M-Estimators and Local M-Smothers

When estimating the underlying constant signal from noisy samples, the selected method should depend on the type of noise present in the data. For Gaussian noise, taking the sample mean is a good choice, providing the maximum a posteriori (MAP) estimate. For noise with heavier tails (caused either by the noise properties themselves, or because e.g. the samples were mixed from two distributions due to signal discontinuity), one has to use methods from robust statistics that are less effected by outliers, such as an M-estimator [43, 40]. An M-estimate of a constant value  $u$  from noisy data  $f_j$  is found by minimising

$$E(u) = \sum_{j=1}^K \Psi(|u - f_j|^2) \quad (26)$$

where the error norm  $\Psi$  can attain for example one of the forms presented in Table 1.

The right column of Table 1 gives an overview of what element minimises the functional (26) with the given error penaliser  $\Psi$ . For the  $L_2$  norm (a), the solution is the mean of the noisy samples. The  $L_1$  norm (b) is minimised by the median. For the robust error norms (c) and (d), the influence of outliers is very much reduced, and the solution  $u$  minimising (26) approximates a mode (maximum) of the probability density underlying the noisy samples. The mode ideally corresponds to the most frequent value present in the data. For the discrete noisy samples, the maximum of the density can be only estimated e.g. using suitable smoothing kernels; see [22] for some examples and a connection to iterative solvers. Note that while the  $L_2$  and  $L_1$  norms lead to a convex functional minimisation, the robust error norms (c) and (d) in Table 1 are nonconvex, and their corresponding functionals  $E(u)$  may exhibit multiple local minima.

The M-estimators were introduced to robustly estimate a single value from noisy samples. For images, we have to consider also the spatial distribution of the data. Such a generalization is known as *local M-smothers*, and the functional to minimise has the following structure [23, 85]:

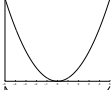
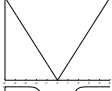
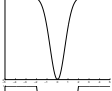
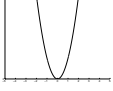
$$E(u) = \sum_{i=1}^N \sum_{j \in \mathcal{B}(i)} \Psi(|u_i - f_j|^2) w(|x_i - x_j|^2) \quad (27)$$

where  $w$  represents the spatial weight depending on sample distance. The local window  $\mathcal{B}(i)$  is introduced in (27) for computational convenience only, to make the index  $j$  run through the neighbourhood of  $x_i$  where  $w(|x_i - x_j|^2)$  exceeds some threshold of contribution importance.

The energy functional (27) can be minimized using an iterative scheme called *W-estimator* [85],

$$u_i^{k+1} = \frac{\sum_{j \in \mathcal{B}(i)} g(|u_i^k - f_j|^2) w(|x_i - x_j|^2) f_j}{\sum_{j \in \mathcal{B}(i)} g(|u_i^k - f_j|^2) w(|x_i - x_j|^2)} \quad (28)$$

Table 1: Examples of error norms for M-estimators. The parameter  $\lambda$  serves as contrast parameter.

<i>error norm</i>		<i>estimation result</i>
(a) $\Psi(s^2) = s^2$		→ mean
(b) $\Psi(s^2) =  s $		→ median
(c) $\Psi(s^2) = 1 - e^{-s^2/\lambda^2}$		→ mode approximation
(d) $\Psi(s^2) = \min(s^2, \lambda^2)$		→ mode approximation

where the process is initialised with  $u_i^0 := f_i$ . This iterative scheme converges to a local minimum of (27) close to the input data. Depending on the penaliser  $\Psi$ , the iterations may lead e.g. to a local mode approximation [78, 76, 77, 24], or to an approximation of a windowed median filter or Gaussian smoothing.

Comparing equation (28) with our scheme (25), we observe that the local M-smoothers and the W-estimator correspond to the data term of the NDS model. To obtain the W-estimator from (25), simply set the smoothness parameter  $\alpha = 0$ . The spatial weight  $w$  will be chosen so that it covers some area around the current pixel, typically larger than the immediate neighbourhood.

### 3.2 Bilateral Filtering

Contrary to the previous section, let us analyse the situation for the maximum smoothness parameter,  $\alpha = 1$ . Then, the data term from equations (3) and (25) vanishes, and the full scheme consists of the smoothness term only. The resulting energy functional

$$E_S(u) = \sum_{i,j \in J} \Psi_S(|u_i - u_j|^2) w_S(|x_i - x_j|^2) \quad (29)$$

can be minimized by the fixed point iterations

$$u_i^{k+1} = \frac{\sum_{j \in J} g_S(|u_i^k - u_j^k|^2) w_S(|x_i - x_j|^2) u_j^k}{\sum_{j \in J} g_S(|u_i^k - u_j^k|^2) w_S(|x_i - x_j|^2)}. \quad (30)$$

Equation (30) is known as *bilateral filter* [2, 70, 75]. While bilateral filtering was originally proposed as a heuristic algorithm, we have shown that it can be derived as a special case from the NDS energy functional (3) where only the smoothness term is considered, and the local smoothness of the signal  $u$  is evaluated in a nonlocal window  $w_S$ .

### 3.3 Regularisation Methods

Consider the optimality condition  $\frac{\partial E_S}{\partial u_k} = 0$  for (2) written as

$$0 = \sum_{j \in \mathcal{B}(k)} \Psi_S(|u_k - u_{k+j}|^2) (u_k - u_{k+j}) w_S(|x_k - x_{k+j}|^2), \quad (31)$$

with  $\mathcal{B}(k) = \{j \in \mathbb{Z} : |x_k - x_{k+j}| \leq r_S\}$ , and the hard window

$$w_S(|x_k - x_{k+j}|^2) = \begin{cases} 1 & \text{if } |x_k - x_{k+j}| \leq r_S \\ 0 & \text{otherwise} \end{cases}.$$

Then, following [80, 30], equation (31) can be regarded as a crude approximation of the steady state of the rotationally invariant PDE

$$\partial_t u = \frac{2}{\pi} \int_0^\pi \partial_{e_\varphi} \left( g_S(|\partial_{e_\varphi} u_\delta|^2) \partial_{e_\varphi} u \right) d\varphi \quad (32)$$

when the kernel size  $\delta$  in  $u_\delta := G_\delta * u$  vanishes,  $g_S := \Psi'_S$ , and  $e_\varphi = (\cos \varphi, \sin \varphi)^\top$ . Furthermore, equation (32) is equivalent to the anisotropic model

$$\partial_t u = \operatorname{div}(D \cdot \nabla u) \quad (33)$$

with the diffusion tensor

$$D := \frac{2}{\pi} \int_0^\pi e_\varphi e_\varphi^\top g_S(|\partial_{e_\varphi} u_\delta|^2) d\varphi.$$

In [80] it is shown that the eigenvectors and eigenvalues of  $D$  are given by

$$\begin{aligned} v_\parallel(\varphi) &= \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix}, & \lambda_\parallel(r^2) &= \frac{4}{\pi} \int_0^{\pi/2} \sin^2 \varphi g_S(|r \cos \varphi|^2) d\varphi; \\ v_\perp(\varphi) &= \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix}, & \lambda_\perp(r^2) &= \frac{4}{\pi} \int_0^{\pi/2} \cos^2 \varphi g_S(|r \cos \varphi|^2) d\varphi, \end{aligned}$$

where  $(r, \varphi)$  are the polar coordinates of  $\nabla u$ . In our case, i.e.  $\delta \rightarrow 0$ , the process (33) becomes isotropic with scalar diffusivity  $\tilde{g} := \lambda_\perp$ :

$$\partial_t u = \operatorname{div}(\tilde{g} \cdot \nabla u). \quad (34)$$

This means that the solution of the smoothness term (2) approximates a Perona-Malik filter that diffuses in direction  $\nabla u^\perp$  perpendicular to the gradient, i.e. along edges.

If we now include the data term (1) with a local spatial window

$$w_D(|x_k - x_{k+j}|^2) = \begin{cases} 1 & \text{if } x_k = x_{k+j} \\ 0 & \text{otherwise} \end{cases},$$

the resulting process approximates

$$\frac{u - f}{\frac{2\alpha}{1-\alpha}} = \operatorname{div}(\tilde{g} \cdot \nabla u), \quad (35)$$

which can be regarded as a fully implicit time discretisation of the diffusion process (34) with a single time step of size  $\frac{2\alpha}{1-\alpha} > 0$ . Following Scherzer and Weickert [68], it can be shown that (35) corresponds to the Euler-Lagrange equation of the continuous functional

$$E(u) = \int_{\Omega} \left( (1 - \alpha) \Psi_D(|u - f|^2) + \alpha \tilde{\Psi}_S(|\nabla u|^2) \right) dx, \quad (36)$$

where  $\tilde{\Psi}'_S := \tilde{g}$ . The continuous functional is the classical energy functional from regularisation or Bayesian frameworks; see e.g. [7, 35, 57, 85]. As an example, the continuous Mumford–Shah functional fits into this framework if we choose  $\Psi_D(s^2) := s^2$  and  $\Psi_S(s^2) := \min(s^2, \lambda^2)$ . Also, the diffusion filters [61, 81] and diffusion-reaction processes [59, 69, 72, 19] can be derived from equation (36).

### 3.4 Histogram Quantisation

For the sake of completeness, let us consider the case when the spatial support window grows to ‘infinite’ size, and all the pixels are connected with the same weight regardless of their position in the image,  $w_{D,S} \equiv 1$ . Then, the NDS functional simplifies to

$$E(u) = (1 - \alpha) \sum_{i,j \in J} \Psi_D(|u_i - f_j|^2) + \alpha \sum_{i,j \in J} \Psi_S(|u_i - u_j|^2). \quad (37)$$

Because the spatial information does not appear in the formula, the solution can be equivalently found in a space where the spatial information was omitted and only the tonal information remains: the image histogram. For example, minimizing the functional for the robust penaliser  $\Psi$  from Table 1 (c) or (d) corresponds to replacing each pixel value with the local mode of the corresponding image histogram. The resulting image will have a smaller number of gray values, adaptively quantised. The data and smoothness terms in this context correspond to the non-blurring or blurring mean shift process, respectively [22].

## 4 NDS and Graph Regularisation

In this section we show that the discrete NDS framework is closely related to graph regularisation techniques and that it extends recent developments in the context of image and manifold regularisation on weighted graphs.

A discrete image is usually defined on a regular domain, e.g. on a rectangular grid. However, for more general image domains it is more appropriate to represent an image as a graph with arbitrary topology. Every vertex (pixel)  $i$  of the graph encodes both the pixel location  $x_i$  and the pixel intensity  $f_i$ . The edge connecting two vertices  $i$  and  $j$  represents the similarity between both pixels, expressed as a weight function  $w(i, j) > 0$ . Employing such graph representation and special calculus on graphs [89, 90], several regularisation models for general data living on discrete spaces have been recently proposed. In the context of image denoising Weickert [82] developed a space-discrete theory for diffusion filtering that is directly applicable to functions defined on graphs, and Chan *et al.* [17] introduced the digital TV filter as a discrete version of the continuous ROF model [66]. In the context of semi-supervised learning Zhou and Schölkopf [89, 90] proposed a discrete analogue of classical regularisation [74] with a  $p$ -Dirichlet regulariser; and Zhou and Burges [88] introduced a discrete analogue of the Laplace-de Rham operator as a regulariser.

Following the ideas from graph theory presented in [89, 90], Gilboa and Osher [38] proposed the use of nonlocal operators to extend some known PDEs and variational techniques in image processing to a nonlocal framework. In particular, they use discretised differential operators such as gradient and divergence. The discretisations involve pixel differences that are weighted by a patch-based similarity between pixels as in [15]. Bogleux *et al.* [9, 33, 10] designed a discrete graph regularisation framework that can be seen as a digital extension of the continuous framework [38] employing a  $p$ -Dirichlet regulariser. The same discrete framework has been applied in image segmentation tasks [73]. Furthermore, nonlocal differential operators have been used to derive nonlocal morphological PDEs [34].

We now show that the discrete variational NDS model (3) can be regarded as a common regularisation method for general data defined on discrete spaces. Let us consider the smoothness term (2) of the NDS model using  $\Psi(s^2) = \frac{1}{p}|s|^p$ ,  $p > 0$ , as penaliser:

$$E_S(u) = \sum_{i \in J} \sum_{j \in J} \Psi(|u_i - u_j|^2) w(i, j) \quad (38)$$

$$= \frac{1}{p} \sum_{i \in J} \sum_{j \in J} |u_i - u_j|^p w(i, j) \quad (39)$$

$$= \frac{1}{p} \sum_{i \in J} \|\nabla_w u_i\|_p^p, \quad (40)$$

where  $\|\nabla_w u_i\|_p = \left( \sum_{j \in J} |u_i - u_j|^p w(i, j) \right)^{\frac{1}{p}}$  is the weighted  $L_p$  norm. Other definitions of the weighted gradient norm are possible using alternative weighted difference operators (see [41] and references therein). This regulariser has been used in [90, 9, 33, 10] for regularisation on arbitrary graphs. In particular, the



following energy functionals have been proposed in [10]:

$$E_{iso}(u) = \sum_{i \in J} \left( \frac{\lambda}{2} (u_i - f_i)^2 + \frac{1}{p} \|\nabla_w u_i\|_2^p \right), \quad (41)$$

$$E_{ani}(u) = \sum_{i \in J} \left( \frac{\lambda}{2} (u_i - f_i)^2 + \frac{1}{2p} \|\nabla_w u_i\|_p^p \right). \quad (42)$$

The functional (41) corresponds to an isotropic model whose minimiser is obtained by solving a linear system, whereas (42) is an anisotropic model leading to a nonlinear system. The nonlocal interactions between graph nodes are introduced via the weight function  $w$ . In the general case the weight  $w(i, j) := w(\mathcal{F}_i, \mathcal{F}_j)$  measures the similarity between the nodes  $i$  and  $j$  with respect to a certain feature vector  $\mathcal{F}$ . For instance, a weighted  $L_2$  norm between image patches [15] can be used for the task of image smoothing. This and other similarity measures are discussed later in this paper.

There exist three main differences between the NDS framework and the graph regularisation (GR) approaches previously reviewed: (i) in the NDS we allow the use of any penaliser for both the data similarity and the smoothness term, whereas GR only considers penalisers of the form  $\Psi(s^2) = \frac{1}{p}|s|^p$  for  $p \in ]0, 2]$ ; (ii) in the NDS model nonlocal interactions are present in both the data and the smoothness term, while in the GR techniques the non-localities are only considered in the regularisation term; and (iii) in the NDS framework the functions  $w$  only depend on the spatial node/pixel locations, whereas in the GR approaches  $w$  can be defined in terms of several node characteristics. This last issue suggests that the NDS model (3) can be generalised by extending the definition of the weighting functions  $w$ . In this way, we obtain a general framework for processing data sets defined on arbitrary discrete domains.

Despite the interesting extension of the NDS model described above, we do not further develop this idea here. That will be part of future work. However, we shall consider another generalisation of the NDS framework in the next section. We will rather concentrate on the penalisers  $\Psi$ , which we allow to act on more general constraints.

## 5 Generalised NDS Model

The NDS model of the previous section was termed *nonlocal* data and smoothness (NDS) because of the interactions between more distant pixels than the immediate neighbourhood. However, the tonal weights in (3) depend on the single differences between pairs of connected pixels. These single differences have a limited ability to express local image structure and geometry, and for practical purposes, the pixel interactions have to be kept to a relatively small neighbourhood.

Many recent approaches for image denoising make use of self-similarity of the whole image, or similarity between several images. For filtering, pixels from very distant locations could also contribute to the result. To distinguish which pixels are compatible, a more powerful measure is needed to evaluate the similarity: Not just pixel difference, but the similarity of a whole region of interest, or *image patch* around the central pixel, is considered. The NL-means filter [14, 15] is a typical example of this class of filters.

In this section, we combine the idea of patch similarity with the NDS functional, which leads to a *Generalized Nonlocal Data and Smoothness*, or GNDS model. We keep the discrete variational framework involving both data and smoothness terms, and allow for different ways to calculate the distance of the image patches. We will show which iterative filter can be derived as a minimizer of the GNDS energy functional. Inspired by its form, we will relax a constraint and present a new family of patch-based GNDS filters.

## 5.1 GNDS Functional and its Minimisation

First, let us introduce the tonal distance functions  $d_D, d_S : \mathbb{R}^{2p} \rightarrow \mathbb{R}_0^+$  in the data and the smoothness term. For example, in the data term, such a function calculates the distance between two image patches  $u(\mathcal{P}_i)$  of the evolving image and  $f(\mathcal{P}_j)$  of the initial image. The index sets  $\mathcal{P}_i$  and  $\mathcal{P}_j$  define image patches as neighbourhoods of the pixels  $i$  and  $j$ , respectively. Both patches are assumed to have the same size  $p \in \mathbb{N}$  and the same shape.

As distance function, for example the weighted  $L_2$  norm can be used, i. e.

$$|d(u(\mathcal{P}_i), f(\mathcal{P}_j))|^2 = \sum_p G_\sigma(p) (u_{i+p} - f_{j+p})^2, \quad (43)$$

where  $G_\sigma(p) := \exp(-p^2/(2\sigma^2))$ . This has also been used as a patch distance in the nonlocal means algorithm.

With these definitions, the *Generalised Nonlocal Data and Smoothness* (GNDS) model reads

$$\begin{aligned} E_G(u) &= (1 - \alpha) E_{GD}(u) + \alpha E_{GS}(u) \\ &= (1 - \alpha) \sum_{i,j \in J} \Psi_D \left( |d_D(u(\mathcal{P}_i), f(\mathcal{P}_j))|^2 \right) w_D (|x_i - x_j|^2) \\ &\quad + \alpha \sum_{i,j \in J} \Psi_S \left( |d_S(u(\mathcal{P}_i), u(\mathcal{P}_j))|^2 \right) w_S (|x_i - x_j|^2). \end{aligned} \quad (44)$$

As we did in Section 2.1 for the NDS model, we now obtain the corresponding fixed point form for (44). The minimiser  $u$  of (44) necessarily satisfies

$$\frac{\partial E_G}{\partial u_i} = (1 - \alpha) \frac{\partial E_{GD}}{\partial u_i} + \alpha \frac{\partial E_{GS}}{\partial u_i} = 0 \quad \text{for all } i \in J. \quad (45)$$

Using the distance function  $d_{D;i,j}^2 := |d_D(u(\mathcal{P}_i), f(\mathcal{P}_j))|^2$  as in (43), we have for the data term:

$$\begin{aligned}
\frac{\partial E_{GD}}{\partial u_k} &= \frac{\partial}{\partial u_k} \sum_{i,j \in J} \Psi_D(d_{D;i,j}^2) w_D(|x_i - x_j|^2) \\
&= \sum_{i,j \in J} \Psi'_D(d_{D;i,j}^2) \frac{\partial}{\partial u_k} \left( \sum_p G_\sigma(p) (u_{i+p} - f_{j+p})^2 \right) w_D(|x_i - x_j|^2) \\
&\stackrel{(i=k-p)}{=} \sum_{j \in J} \Psi'_D(d_{D;k-p,j}^2) \frac{\partial}{\partial u_k} \left( \sum_p G_\sigma(p) (u_{i+p} - f_{j+p})^2 \right) w_D(|x_i - x_j|^2) \\
&\stackrel{(j=l-p)}{=} 2 \sum_{l \in J} \sum_p G_\sigma(p) \Psi'_D(d_{D;k-p,l-p}^2) (u_k - f_l) w_D(|x_{k-p} - x_{l-p}|^2) \\
&= 2 \sum_{j \in J} G_\sigma * \Psi'_D(d_{D;k-\cdot,j-\cdot}^2)(0) (u_k - f_j) w_D(|x_k - x_j|^2) \quad (46)
\end{aligned}$$

where the operator ‘\*’ stands for convolution. A more detailed derivation, and the corresponding counterpart for the smoothness term, can be found in Appendix A.1.

Then, with help of the abbreviations

$$g_{i,j}^{GD} := G_\sigma * \Psi'_D \left( |d_D(u(\mathcal{P}_{i-\cdot}), f(\mathcal{P}_{j-\cdot}))|^2 \right) (0), \quad (47)$$

$$g_{i,j}^{GS} := 2 G_\sigma * \Psi'_S \left( |d_S(u(\mathcal{P}_{i-\cdot}), u(\mathcal{P}_{j-\cdot}))|^2 \right) (0), \quad (48)$$

and with the spatial weights defined as in (23)–(24), the fixed point for the GNDS model becomes

$$u_i = \frac{(1 - \alpha) \sum_{j \in J} g_{i,j}^{GD} w_{i,j}^{D,r} f_j + \alpha \sum_{j \in J} g_{i,j}^{GS} w_{i,j}^{S,r} u_j}{(1 - \alpha) \sum_{j \in J} g_{i,j}^{GD} w_{i,j}^{D,r} + \alpha \sum_{j \in J} g_{i,j}^{GS} w_{i,j}^{S,r}}, \quad (49)$$

for all  $i \in J$ . This equation can be embedded in a fixed point iteration scheme similar to (12)–(14). A maximum-minimum principle and the existence of a fixed point can be proven following Propositions 2.1 and 2.2. Analogously to (18), the energy minimiser can be obtained via gradient descent.

The data similarity and smoothness constraints in our generalised model (44) penalise tonal distances between patches rather than between single pixels as in the original NDS approach (3). Comparing (49) with the fixed point form of the NDS model (25) we note that the *patch distances induce convolutions with the neighbouring tonal weights*. In Sections 5.2 and 5.3 we discuss the implications of this fact and how it inspires the modelling of new filters.

## 5.2 Double Weighting

Considering the data term of eq. (49) only (the situation for the smoothness term is analogous), and expanding the convolution (47), the fixed point equation for the filtered pixel  $u_i$  becomes

$$u_i = \frac{1}{M_{i,j}} \sum_j \sum_p G_\sigma(p) \cdot \Psi' \left( \sum_q G_\sigma(q) |u_{i+p+q} - f_{j+p+q}|^2 \right) \cdot w_{i,j} \cdot f_j \quad (50)$$

where  $M_{i,j}$  is the usual normalisation by the sum of all applied weights,

$$M_{i,j} = \sum_j \sum_p G_\sigma(p) \cdot \Psi' \left( \sum_q G_\sigma(q) |u_{i+p+q} - f_{j+p+q}|^2 \right) \cdot w_{i,j}.$$

In (50),  $G_\sigma$  is the Gaussian of radius  $r_\sigma$  which represents the patch size in the patch similarity computation (43). Note that this weighting appears twice in formula (50): Once during the patch similarity calculation (summed over  $q$ ) before the nonlinearity  $\Psi'$  is applied. We call this  $G_\sigma$  the *inner* weighting of patch pixels. Moreover,  $G_\sigma$  appears also for a second time in (50), in the sum over  $p$ . We call this the *outer* weighting which is applied when summing the results of the function  $\Psi'$  after it is applied to individual patch distances. Figure 2 demonstrates this: the tonal weight (47) entering in (50) not only involves the comparison of the patches about the pixels  $i$  and  $j$ , but also the patch similarity between their corresponding neighbours is considered.

Equation (50), and particularly this double weighting, deserve a detailed discussion. The estimated pixel value  $u_i$  in (50) is obtained as a weighted average of some data samples  $f_j$ . Let us consider a single data pixel  $f_j$ , and analyse what is the weight by which this pixel contributes to the weighted result. For a single value of the dummy variable  $p$ , the sum

$$\sum_q G_\sigma(q) |u_{i+p+q} - f_{j+p+q}|^2$$

evaluates the weighted  $L_2$  distance between an image patch around pixel  $u_{i+p}$  on one hand, and an image patch around pixel  $f_{j+p}$  on the other hand (where the size of the patches is given by the weighting function  $G_\sigma$ ). In the notation used earlier in this paper, this patch distance is denoted  $d(u(\mathcal{P}_{i+p}), f(\mathcal{P}_{j+p}))$ . Note that the compared patches are offset with respect to the estimation and data positions  $i$  and  $j$ , respectively, by a common shift  $p$ .

Coming back to equation (50), after evaluating the patch distance, the nonlinearity  $\Psi'$  is applied next. We remark that this nonlinearity can be related to robust statistical estimation; its role is to downweight outliers, and convert patch distance to (robust) patch similarity. Then, the resulting patch similarities are summed over variable  $p$  in a second patch neighbourhood, again defined by the

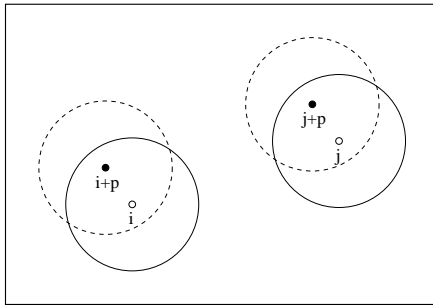


Figure 2: The weight computation in (50) between two pixels  $i$  and  $j$  involves the patch comparison – using patches  $G_\sigma$  (dashed lines) – between every pair of corresponding neighbours  $i+p$  and  $j+p$  within a neighbourhood  $G_\sigma$  (solid lines).

weighing function  $G_\sigma$ . Note that the inner and outer weighing functions are identical, which originates in the functional  $E_{GD}$  of (44) and the derivatives with respect to  $u_i$  which duplicated the inner weight also out of the nonlinearity (see Appendix A.1).

Summarising it in words, the equation (50) has the following meaning: For pixels  $u_i$  and  $f_j$ , calculate the patch distances of all patches at positions  $i+p$  and  $j+p$  taken with the offset  $p$  around  $u_i$  and  $f_j$ , respectively. Then, average these patch distances (transformed first by the nonlinearity  $\Psi'$ ) using the outer weighting  $G_\sigma$ . Thus, the pixel  $f_j$  will contribute to the result  $u_i$  with a high weight not only if the patches around  $u_i$  and  $f_j$  are similar, but also if the neighbouring patches  $u_{i+p}$  and  $f_{j+p}$  resemble each other.

### 5.3 GNDS Filter Family

In the previous section we discussed the roles of the inner (patch) weighting  $G_\sigma$  and the outer (similarity integration) weighting. Derived from the energy functional, these two weightings are identical. In the fixed point iteration though, these two weighting functions have a different role, and it is instructive to analyse what changes if they are decoupled.

In the following, we keep the parameter  $r_\sigma$  for the radius of the Gaussian  $G_\sigma$  of the inner pixel weighing for patch similarity calculation. The outer integration scale will use a different weighting function  $G_\rho$  of radius  $r_\rho$ , and the pixel averaging equation becomes

$$u_i = \frac{1}{M_{i,j}} \sum_j \sum_p G_\rho(p) \cdot \Psi' \left( \sum_q G_\sigma(q) |u_{i+p+q} - f_{j+p+q}|^2 \right) \cdot w_{i,j} \cdot f_j \quad (51)$$

where  $M_{i,j}$  is the corresponding normalisation factor.

Let us now study what is the effect of varying the parameters  $\sigma$  and  $\rho$  which determine the size of the inner and outer weighting windows, respectively.

First, let  $\rho \rightarrow 0$ , leading to the following outer weighting:

$$G_0(p) = \begin{cases} 1 & \text{if } p = 0; \\ 0 & \text{if } p \neq 0. \end{cases} \quad (52)$$

The equation (51) then simplifies to

$$u_i = \frac{1}{M_{i,j}} \sum_j \Psi' \left( \sum_q G_\sigma(q) |u_{i+q} - f_{j+q}|^2 \right) \cdot w_{i,j} \cdot f_j \quad (53)$$

which is exactly the non-iterative NL-means filter introduced by Buades, Coll and Morel in [14, 15]! NL-means weights the contribution of the pixel  $f_j$  using a single patch distance comparing patches around  $u_i$  and  $f_j$ , and omits any additional integration of these patch similarities using the outer summation.

Second, let  $\sigma \rightarrow 0$ . This leads to

$$u_i = \frac{1}{M_{i,j}} \sum_j \sum_p G_\rho(p) \cdot \Psi' (|u_{i+p} - f_{j+p}|^2) \cdot w_{i,j} \cdot f_j. \quad (54)$$

Comparing (54) with (53), we observe that these two equations have a highly similar structure, with a single difference: The position where the nonlinearity  $\Psi'$  is applied. For NL-means (53), we first sum the differences of individual pixels, thus evaluating the weighted  $L_2$  similarity, and then apply the robust weighting  $\Psi'$ . In the other case of equation (54), we apply the nonlinearity  $\Psi'$  to individual pixel differences, and then integrate the result over the window  $G_\rho$ . Even in this case, the weight of pixel  $f_j$  is influenced by the whole patches around  $u_i$  and  $f_j$ . The difference lies in the way the patch similarity is evaluated. Due to the structural resemblance of the filters (53) and (54) to isotropic and anisotropic penalisation [83] we call (54) *anisotropic NL-means*.

As a third example, let both  $\sigma \rightarrow 0$  and  $\rho \rightarrow 0$ . Then, the generalised NDS scheme (51) simplifies to the classical NDS scheme (11) which is based on simple pixel differences instead of patch distances.

Table 2: Examples of filtering methods belonging to the GNDS family (51) with varying inner scale  $\sigma$  and outer scale  $\rho$ .

Patch size	Integration scale	Method
$\sigma > 0$	$\rho = \sigma$	Generalised NDS (51)
$\sigma > 0$	$\rho \rightarrow 0$	NL-means (53), [15]
$\sigma \rightarrow 0$	$\rho > 0$	Anisotropic NL-means (54)
$\sigma \rightarrow 0$	$\rho \rightarrow 0$	Classical NDS (3), [56]

By varying the inner and outer scales  $\sigma$  and  $\rho$  in (51), we obtain a full family of highly nonlinear and robust filters. A single member of this family with  $\rho = \sigma$  can be derived from an energy functional. The well known NL-means method belongs to this family. It represents the case when the outer scale vanishes. Some of these special cases are summarised in Table 2.

Practically, the inner and outer scales both act in the direction that by increasing them, we increase the area used to evaluate image similarity: Higher values lead to a more thorough (and costly) patch comparison. Consider the special situation when this combined scale is kept constant, the amount of integration just shifts between the inner and outer scales. Such a setting forms a family of filters with approximately the same spatial extent of operations. What changes is the position at which the nonlinearity  $\Psi'$  enters the chain. The NL-means (53) and the summation of robust pixel similarities (54) represent the two extremes of this family. This is illustrated in Fig. 3. The practical effect of varying the inner and outer scales will be demonstrated later in this paper.

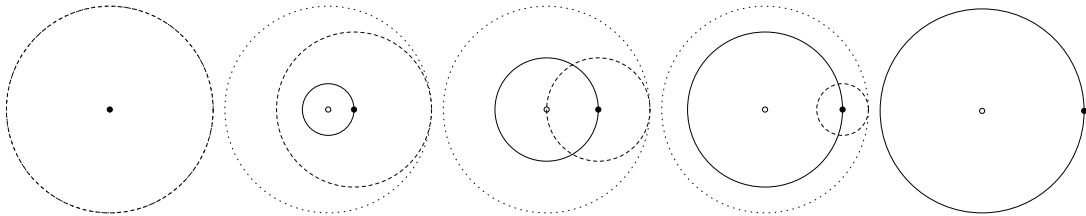


Figure 3: The tonal weight in (51) is computed over an area determined by the integration neighbourhood (solid lines) – Gaussian  $G_\rho$  of radius  $r_\rho$  – and the patch size (dashed lines) – Gaussian  $G_\sigma$  of radius  $r_\sigma$ . From left to right, different configurations where the total area described by a Gaussian of radius  $r_\rho + r_\sigma$  (dotted line) is kept constant. The first and the last configurations correspond to the weighting scheme of the filters (53) and (54), respectively.

#### 5.4 Alternative Formulations of the NL-Means Filter

Using the Whittaker-Tikhonov penaliser  $\Psi(s^2) = s^2$  we obtain  $\Psi'(s^2) := \partial_{s^2}\Psi(s^2) = 1$  and both filters (53) and (54) become equivalent to

$$u_i = \frac{1}{M_{i,j}} \sum_j w_{i,j} f_j. \quad (55)$$

In our setting the spatial function  $w$  acts uniquely as a search window, i.e. it delimits the spatial extent where the pixels  $j$ , neighbours of  $i$ , are taken from. However, in various works [37, 9, 20] it is argued that (55) can be regarded as

the NL-means filter by redefining the weights

$$\tilde{w}_{i,j} := w_{i,j} \cdot \exp\left(-\frac{|d(f(\mathcal{P}_i), f(\mathcal{P}_j))|^2}{h}\right), \quad (56)$$

with  $h > 0$  as a filter parameter. Note that the additional weighting term is constant as it depends on the input image  $f$ . This indicates that the filter (55) could be directly derived from the data term (1) of the NDS functional employing  $\Psi(s^2) = s^2$  and  $\tilde{w}$ . Analogously, a filter that averages over the evolving image  $u$  can be obtained from the smoothness term (2). Similar ideas have been considered in [36, 5, 13].

In [45, 21, 63] energy functionals with weights depending on the unknown solution  $u$  via  $d(u(\mathcal{P}_i), u(\mathcal{P}_j))$  have been considered. However, all these methods assume constant weights in the computation of the optimality conditions  $\nabla E(u) = 0$ . The variational filter proposed by Brox *et al.* in [13] also considers nonlocal weights depending on  $u$ . Although they do not assume constant weights in the derivation of the Euler-Lagrange equations, these become very complex and computationally expensive. To a certain extent, the mentioned filters could be obtained from the original NDS framework (3) by extending the definition of the weights  $w$  as in (56) (see also a related discussion in Section 4). In the proposed GNDS framework we keep regarding the weights  $w$  only as (nonlocal) spatial functions. Instead, we generalise the constraints being penalised in the energy functional. That is, in (44) we have replaced the *single pixel* similarity constraints of (3) by *patch similarity* constraints using the weighted  $L_2$  distance between patches, obtaining a new family of neighbourhood filters. The use of other similarity measures is discussed in Section 5.7. It is important to mention that we do take into account the dependency of the distance measures on the solution  $u$  when deriving the optimality conditions. As a result, the classical and also some iterative versions of the NL-means filter are obtained as special cases of the proposed filter family without needing a redefinition of the spatial weights  $w$ .

## 5.5 (Non-) Iterative and Steady-State Solutions

In Section 5.3 we explored the full family of filters that can be obtained from the proposed GNDS model by varying the inner and outer scales in the patch similarity computation. This entails the immediate extension of the filters presented in Section 3 to work with image patches rather than with single pixel differences. Let us consider the fixed point (49) that iteratively minimises the energy functional (44). For  $0 \leq \alpha < 1$  this process will converge to a stationary state due to the data term dependency on the input image. Note that for  $\alpha = 0$  we obtain a *generalised nonlocal  $M$ -smoothing* process. In this case we can think, for instance, of a novel NL-means filter with a steady-state solution. For  $\alpha = 1$  we have a *generalised nonlocal Bilateral filter* that needs to be stopped after certain number of



iterations before the image gets completely smoothed away. These issues will be further discussed later on in the experimental section.

## 5.6 Extension to Multichannel Images

The extension of the GNDS model to multichannel images is straightforward. Let  $\mathbf{f}, \mathbf{u} : \Omega \rightarrow \mathbb{R}^d$  be the noisy image and the unknown noise-free image, respectively, both with  $d$  channels. To obtain the multichannel counterpart of the scalar GNDS model (44) we just need to redefine the patch distance (43) as

$$|d(\mathbf{u}(\mathcal{P}_i), \mathbf{f}(\mathcal{P}_j))|^2 = \sum_p G_\sigma(p) \|\mathbf{u}_{i+p} - \mathbf{f}_{j+p}\|_2^2, \quad (57)$$

where  $\|\cdot\|_2$  is the Euclidean norm. Computing the optimality conditions  $\nabla E(\mathbf{u}) = 0$  we obtain a fixed point for every channel  $u^m$  ( $m = 1, \dots, d$ ), cf. (49):

$$u_i^m = \frac{(1 - \alpha) \sum_{j \in J} g_{i,j}^{GD} w_{i,j}^{D,r} f_j^m + \alpha \sum_{j \in J} g_{i,j}^{GS} w_{i,j}^{S,r} u_j^m}{(1 - \alpha) \sum_{j \in J} g_{i,j}^{GD} w_{i,j}^{D,r} + \alpha \sum_{j \in J} g_{i,j}^{GS} w_{i,j}^{S,r}}. \quad (58)$$

Note that all channels are coupled via the tonal or photometric weights

$$g_{i,j}^{GD} := G_\sigma * \Psi'_D \left( |d_D(\mathbf{u}(\mathcal{P}_{i-}), \mathbf{f}(\mathcal{P}_{j-}))|^2 \right) (0), \quad (59)$$

$$g_{i,j}^{GS} := 2 G_\sigma * \Psi'_S \left( |d_S(\mathbf{u}(\mathcal{P}_{i-}), \mathbf{u}(\mathcal{P}_{j-}))|^2 \right) (0), \quad (60)$$

which avoids the formation of discontinuities at different locations for the different image channels.

## 5.7 Extension to Other Distance Measures

The proposed energy functional (44) is very general in the sense that one could choose any suitable distance measures  $d_D, d_S$  to impose similarity of particular image characteristics. Once the distances have been chosen, the optimality conditions  $\nabla E(u) = 0$  need to be derived in order to prescribe the corresponding energy minimiser, for instance, via a fixed point or a gradient descent scheme. In the proposed GNDS model we have used the weighted  $L_2$  norm (43) to measure similarity between image patches. However, one can employ different distance measures as well. For example, Kervrann and Boulanger [45] use

$$|d(\mathbf{u}(\mathcal{P}_i), \mathbf{u}(\mathcal{P}_j))|^2 = \text{vec}(\mathbf{u}(\mathcal{P}_i) - \mathbf{u}(\mathcal{P}_j))^\top V_{ij}^{-1} \text{vec}(\mathbf{u}(\mathcal{P}_i) - \mathbf{u}(\mathcal{P}_j)), \quad (61)$$

where  $V_{ij}$  is a diagonal matrix whose entries are averaged local variances of the image patches. Similarly, Goossens *et al.* [39] replace  $V_{ij}$  by a local estimation of the noise covariance matrix to filter images corrupted by correlated noise.

Another example where the selection of the patch distance is driven by the noise type corrupting the image data is due to Coupé *et al.* [25]. Based on the Bayesian non-local means filter [46] and on the Speckle noise model introduced in [54], the authors propose a non-local filter for ultrasound images that uses the so-called Pearson distance for computing patch similarity:

$$|d(u(\mathcal{P}_i), u(\mathcal{P}_j))|^2 = \sum_p \frac{(u_{i+p} - u_{j+p})^2}{u_{j+p}}. \quad (62)$$

These and other measures of similarity can be utilised in the proposed functional (44) with accordingly derived minimisation algorithms. In addition, as classically done in variational methods, the choice of the data similarity constraint can be driven by the statistical properties of the type of noise present, whereas the smoothness term must reflect desirable properties of the solution. Therefore, the issue of selecting appropriate patch distances for the data and smoothness terms of the proposed GNDS model is still open. We will explore these issues in a future work.

## 6 Experiments

### 6.1 Evaluating the NDS Model

The NDS framework was originally proposed by Mrázek *et al.* in [56]. In that paper, preliminary experiments showed the smoothing properties of the model under different parameterisations. In particular, the behaviour of the data and smoothness constraints over neighbourhoods with varying size using robust and non-robust penalisers was analysed. In [29] several minimisation strategies for the NDS functional were compared, and in [64] the NDS model was juxtaposed with several well known filters from the literature. For the sake of completeness, we show some of the key results presented in those papers. We focus on two main issues: (i) We show that the NDS model is able to outperform a wide range of classical filters, and (ii) we study the relations among its smoothing parameters. As motivated from a statistical point of view [64], two well suited models for filtering signals degraded by Gaussian and salt-and-pepper noise are

$$E(u) = (1 - \alpha) \sum_{i \in J, j \in \mathcal{B}_D(i)} |u_i - f_j|^2 + \alpha \sum_{i \in J, j \in \mathcal{B}_S(i)} |u_i - u_j| \quad (63)$$

and

$$E(u) = (1 - \alpha) \sum_{i \in J, j \in \mathcal{B}_D(i)} |u_i - f_j| + \alpha \sum_{i \in J, j \in \mathcal{B}_S(i)} |u_i - u_j|^2, \quad (64)$$

respectively, where  $\mathcal{B}_\circ$  is the disk-shaped hard window function used as spatial kernel with radius  $r$ . We apply these models to reconstruct the noisy signals

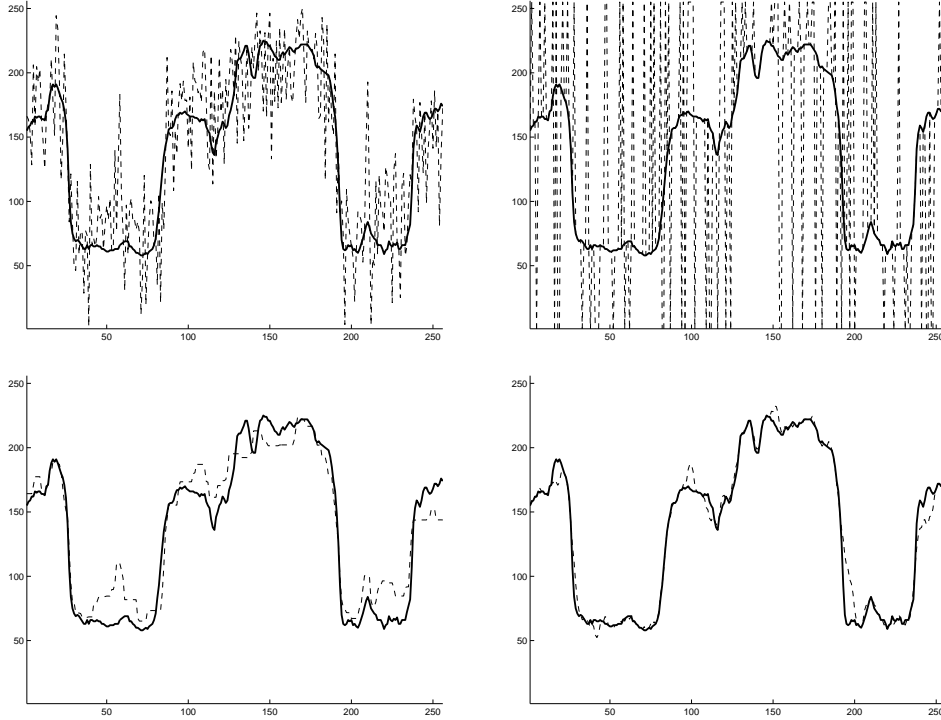


Figure 4: Signal denoising with the NDS functional. Original signals in solid lines, noisy and denoised signals in dashed lines. **Left Column:** Noisy signal perturbed by zero-mean Gaussian noise with  $\sigma_n = 40$ ,  $\ell^1 = 27.30$  (*top*) and its denoised version  $\ell^1 = 13.83$  (*bottom*). **Right Column:** Noisy signal perturbed by 40% of salt-and-pepper noise,  $\ell^1 = 48.04$  (*top*) and its denoised version  $\ell^1 = 4.61$  (*bottom*).

depicted in Fig. 4 (*top*). All parameters were optimised and the best five parameterisation for each model are shown in Table 3. We also report on the performance of the mean and median filters as representatives of M-smoothers (Section 3.1), and classical regularisation filtering (Section 3.3) with four different penalisers. Without exceptions, our designed models outperform all the well known filters obtained as particular cases of the NDS framework.

As it is noticeable in Table 3 there exists a trade-off between the parameter  $\alpha$  and the radii of the spatial kernels. For example, it is possible to achieve similar filtering results either by decreasing  $\alpha$  or by increasing  $r_S$ . On the one hand decreasing  $\alpha$  reduces the influence of the smoothness term, but on the other, increasing  $r_S$  considers contributions to the smoothness term from a larger neighbourhood. To illustrate this effect let us consider the original image shown in Fig. 5(a)(*top left*) and its degraded version with Gaussian noise of Fig. 5(a)(*top right*) that we restore employing the model (63). The radius  $r_D$  of the spatial kernel in the data term was fixed to 1. Fig. 5(b) displays the filtering results for a range of values  $\alpha$  and  $r_S$ . Fig. 5(a)(*bottom*) shows examples where similar restoration quality is

Table 3: Quantitative comparison of different filters. **Left:** Denoising results of the signal perturbed with Gaussian noise shown in Fig. 4 (*top left*). **Right:** Denoising results of the signal perturbed with salt-and-pepper noise shown in Fig. 4 (*top right*). The best results are written in bold letters and plotted in Fig. 4 (*bottom*).

Filter	$r_D$	$r_S$	$\alpha$	$\ell^1$	Filter	$r_D$	$r_S$	$\alpha$	$\ell^1$
	<b>2</b>	<b>2</b>	<b>0.93</b>	<b>13.83</b>		<b>0</b>	<b>1</b>	<b>0.10</b>	<b>4.61</b>
model (63)	2	3	0.87	13.83	model (64)	3	1	0.19	4.67
for Gaussian	2	4	0.82	13.85	for impulse	3	2	0.06	4.80
noise	2	5	0.78	13.93	noise	4	1	0.21	4.90
	3	2	0.90	14.00		2	1	0.24	4.91
mean	4	-	0.00	14.93	mean	6	-	0.00	23.95
median	4	-	0.00	14.90	median	6	-	0.00	6.98
Tikhonov	0	1	0.67	14.57	Tikhonov	0	1	0.90	23.22
TV	0	1	0.99	15.62	TV	0	1	0.99	35.04
Perona-Malik	0	1	0.70	14.47	Perona-Malik	0	1	0.90	23.21
Charbonnier	0	1	0.69	14.53	Charbonnier	0	1	0.90	23.21

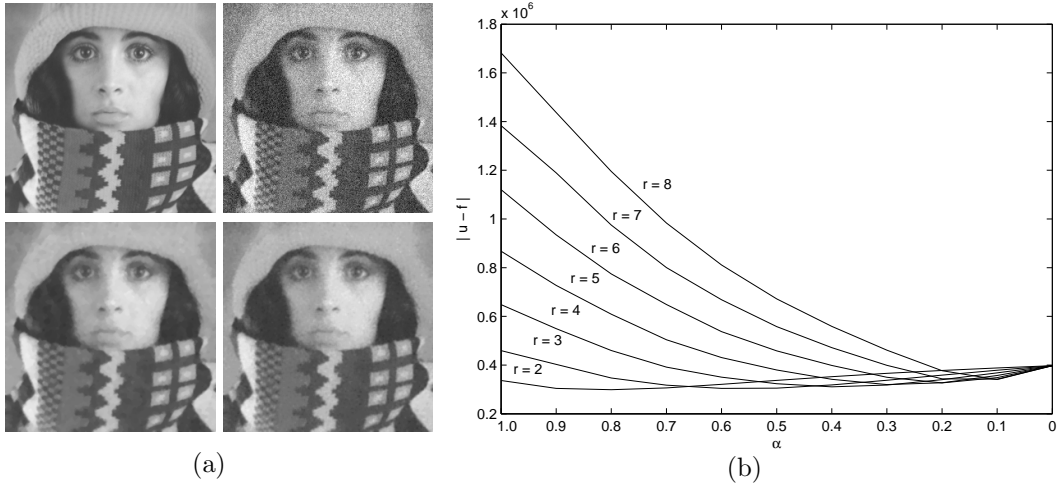


Figure 5: Denoising properties of the functional (63). (a) Original image (*top left*); noisy image perturbed with zero-mean Gaussian with  $\sigma = 20$ ,  $\ell^1 = 16.02$  (*top right*); restored image with  $\alpha = 0.8$ ,  $r_S = 2$ ,  $\ell^1 = 4.88$  (*bottom left*); restored image with  $\alpha = 0.2$ ,  $r_S = 6$ ,  $\ell^1 = 5.18$  (*bottom right*). (b)  $\ell^1$  distance between the original and the denoised image for different  $r$  values of  $\alpha$  and  $r_S$ .

achieved under different parameterisation. Moreover, slightly better results are attained for  $\alpha$  large and  $r_S$  small, which implies less operations and more efficiency. Additional experiments in [64] showed the proportional relation between the weight  $\alpha$  and the radius  $r_D$  of the spatial window in the data term.

Although the NDS framework allows nonlocal processing by extending the support of the spatial windows  $w_D$  and  $w_S$ , note in Table 3 that for the best denoising results the radii  $r_D$  and  $r_S$  do not take very large values. As mentioned before, the effective utilisation of larger neighbourhoods is hindered by the limited capability of single tonal differences to express local image structure and geometry. In the next section we show how the proposed *generalised* NDS overcomes this problem by employing more powerful ways of measuring tonal similarity.

## 6.2 Evaluating the Generalised NDS Model

### 6.2.1 Comparison of similarity measures

The filter (51) induces a novel similarity measure between two pixels  $u_i, u_j$  that can be considered as an *extended* patch similarity measure

$$S_{ext}(u_i, u_j) := \sum_p G_\rho(p) \cdot \Psi' \left( \sum_q G_\sigma(q) \cdot |u_{i+p+q} - u_{j+p+q}|^2 \right). \quad (65)$$

Choosing  $\rho \rightarrow 0$  one obtains an *isotropic* similarity measure

$$S_{iso}(u_i, u_j) := \Psi' \left( \sum_q G_\sigma(q) \cdot |u_{i+q} - u_{j+q}|^2 \right), \quad (66)$$

while with  $\sigma \rightarrow 0$ , equation (65) becomes an *anisotropic* similarity measure

$$S_{ani}(u_i, u_j) := \sum_p G_\rho(p) \cdot \Psi' (|u_{i+p} - u_{j+p}|^2). \quad (67)$$

Considering the penaliser of Leclerc [50] and Perona/Malik [61]

$$\Psi(s^2) = 2\lambda^2 \left( 1 - \exp \left( -\frac{s^2}{2\lambda^2} \right) \right) \quad (68)$$

with filter parameter  $\lambda$ , (66) corresponds exactly to the similarity measure used by Buades *et al.* [15] in their NL-means filter. A couple of recent works have proposed the use of other robust penalisers as well [39, 62]. We test these three measures on the noisy images displayed in Fig. 6. For each one of the 16 textures we select 30 random pixels and compute their similarity to all other pixels in the image. For every chosen pixel we take its best 20 matches (pixels with the largest similarity) and check whether they belong to the same texture or not. Table 4 shows the average number of matches within the same texture and the overall performance of each similarity measure. In the case of Gaussian noise we used the Leclerc penaliser and for salt-and-pepper noise the regularised  $L_1$  norm  $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$ . The radii of the Gaussians were set to  $r_\rho = r_\sigma = 4$ .

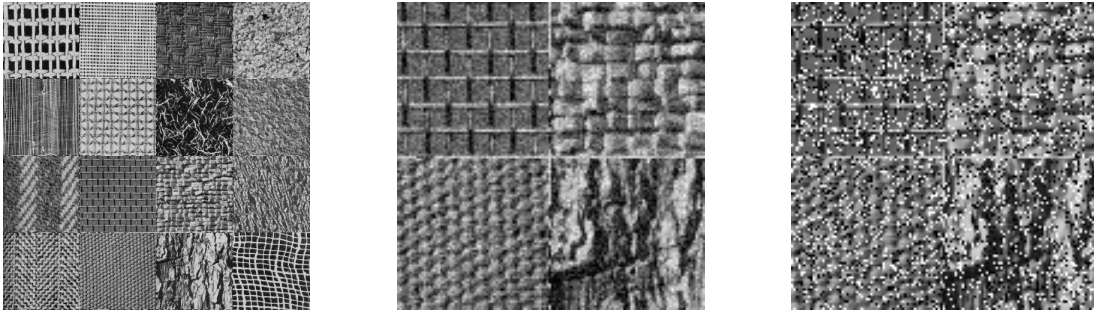


Figure 6: **Left:** Original image with 16 textures, each one identified by its coordinates in matrix notation  $T(x, y)$ ,  $x, y = \{1, 2, 3, 4\}$ . **Middle:** Added zero-mean Gaussian noise with  $\sigma = 20$ . **Right:** Added 20% of salt-and-pepper noise.

Table 4: Quantitative comparison of the three similarity measures  $S_{iso}$ ,  $S_{ani}$  and  $S_{ext}$  induced from the GNDS filter (51).  $S_{ext}$  is more suitable for images degraded with Gaussian noise, while  $S_{ani}$  is more robust under salt-and-pepper noise.

	Gaussian noise ( $\sigma = 20$ )			salt-and-pepper noise (20%)		
	$S_{iso}$	$S_{ani}$	$S_{ext}$	$S_{iso}$	$S_{ani}$	$S_{ext}$
T(1,1)	3	4	13	1	3	2
T(1,2)	19	19	19	19	19	19
T(1,3)	20	16	20	10	19	13
T(1,4)	10	8	12	11	10	8
T(2,1)	20	18	20	20	20	18
T(2,2)	14	5	19	9	18	12
T(2,3)	20	13	20	1	11	2
T(2,4)	16	6	15	4	11	6
T(3,1)	12	5	20	4	16	7
T(3,2)	18	17	19	11	20	14
T(3,3)	14	7	20	3	12	3
T(3,4)	12	13	18	1	13	3
T(4,1)	10	8	14	8	10	8
T(4,2)	20	14	20	4	20	5
T(4,3)	9	7	11	1	5	1
T(4,4)	17	10	16	1	18	6
matches	234	170	276	108	225	127
%	73.1	53.1	<b>86.2</b>	33.7	<b>70.3</b>	39.7

The results show that the extended similarity measure is more robust and perform best under Gaussian degradation. This is due to the fact that, via the outer Gaussian weighting, the selection of similar pixels relies more strongly on the underlying image structures. On the other hand, it performs poorly under impulse noise. In this case the best choice is the anisotropic similarity measure, which acts as a noise detector at every pixel location. The same holds for higher levels of noise. Smoothing experiments will be presented in the following sections.

### 6.2.2 Comparison of several patch-based methods

We now evaluate the smoothing capabilities of the proposed GNDS model on the set of test images *Barbara*, *House*, *Lena*, *Peppers*, *Boats* from Portilla *et al.* [65] which already contain Gaussian noise. The proposed GNDS filter is run iteratively via a gradient descent scheme. In all experiments we use the Leclerc penaliser (68) with fixed contrast parameter  $\lambda$  for successive iterations of the filter, a search window of size  $21 \times 21$  and circular patches of radius  $r_\sigma = 5$ . The radius  $r_\rho$  of the outer patch weighting was chosen between 0 and 2 pixels. Table 5 juxtaposes several patch-based filters proposed in the literature. We employ the peak signal-to-noise ratio (PSNR) as criterion for quality measure:

$$\text{PSNR (dB)} = 10 \log_{10} \left( \frac{255^2}{\frac{1}{|J|} \sum_{i \in J} (o_i - u_i)^2} \right), \quad (69)$$

where  $o$  denotes the original noise free image and  $u$  the estimated denoised version. The shown results for [15, 4, 37, 36] were taken from [13]. From those most competitive methods related to the proposed GNDS filter: Brox *et al.* [12, 13] run an iterative NL-means algorithm that uses the noisy image for averaging and updates the weights from the estimated solution  $u$  of the previous iteration. A similar strategy is due to Kervrann *et al.* [44, 45] who additionally adapt the size of the averaging neighbourhood at each pixel location to better capture local geometries. Azzabou *et al.* [5] developed a variational filter structurally similar to [38] and [13] that also adapts the spatial extent of the local neighbourhoods. Although the proposed GNDS filter does not utilise such powerful and sophisticated adaptive strategies, it also allows for a robust and more coherent selection of the similar pixels within the search window considered for averaging. This is achieved by making use of the extended patch similarity measure defined in (65). Note that in some cases the GNDS filter outperforms the more elaborated methods, though it is still below the state-of-the-art results provided by Dabov *et al.* [27].

It is worth mentioning that the nonlocal smoothness term of GNDS model (44) reaches higher PSNRs than the nonlocal data term, which is more pronounced for higher levels of noise. Interestingly, the combined use of both terms leads to slightly better results than the smoothness term alone. We also run experiments

Table 5: Denoising results of several patch-based filters on standard test images degraded with additive zero-mean Gaussian noise with standard deviation  $\{20, 50\}$ . The parameters  $(\lambda, \tau, \text{iterations})$  are displayed for the GNDS-D (data term of (44)) and the GNDS-S (smoothness term of (44)) filters. GNDS, i.e. the combination of both terms, yields better results than applying each of them independently, outperforming most of the listed methods.

	Gaussian noise ( $\sigma = 20$ )				
Filter \ PSNR (dB)	<i>Barbara</i>	<i>House</i>	<i>Lena</i>	<i>Peppers</i>	<i>Boats</i>
	22.18	22.11	22.13	22.19	22.17
Buades <i>et al.</i> [15]	30.31	32.49	31.78	29.62	29.34
Awate <i>et al.</i> [4]	30.14	32.59	31.79	29.75	29.54
Gilboa <i>et al.</i> [37]	30.20	32.55	31.95	30.28	29.89
Gilboa <i>et al.</i> [36]	29.43	32.17	31.39	30.04	29.53
Brox <i>et al.</i> [13]	30.33	32.74	32.08	30.04	29.69
Azzabou <i>et al.</i> [5]	30.46	32.34	32.12	30.67	29.94
Kervrann <i>et al.</i> [45]	30.37	32.90	32.64	30.59	30.12
Dabov <i>et al.</i> [27]	31.78	33.77	33.05	31.29	30.88
GNDS-D	30.62 (14,1.0,1)	32.66 (15,1.0,1)	31.98 (14,1.0,1)	30.21 (13,1.0,1)	29.78 (12,1.0,1)
GNDS-S	30.62 (14,1.0,1)	32.75 (11,0.8,2)	32.03 (10,0.8,2)	30.21 (13,1.0,1)	29.78 (12,1.0,1)
GNDS	30.64	32.78	32.05	30.22	29.80
	Gaussian noise ( $\sigma = 50$ )				
Filter \ PSNR (dB)	<i>Barbara</i>	<i>House</i>	<i>Lena</i>	<i>Peppers</i>	<i>Boats</i>
	14.76	14.56	14.62	14.68	14.59
Kervrann <i>et al.</i> [45]	24.09	28.67	28.38	25.29	25.93
Dabov <i>et al.</i> [27]	27.17	29.37	28.86	26.41	26.64
GNDS-D	25.40 (23,1.0,2)	27.66 (24,1.0,2)	27.30 (24,1.0,2)	25.25 (24,1.0,2)	25.16 (23,1.0,2)
GNDS-S	25.75 (20,0.9,2)	28.38 (19,1.0,2)	27.77 (19,1.0,2)	25.64 (19,1.0,2)	25.58 (20,0.9,2)
GNDS	25.78	28.40	27.81	25.67	25.60

considering models such as (i) local data terms  $\sum_{i \in J} \Psi(|u_i - f_i|^2)$  with a non-local smoothness term, and (ii) a nonlocal data term combined with semilocal



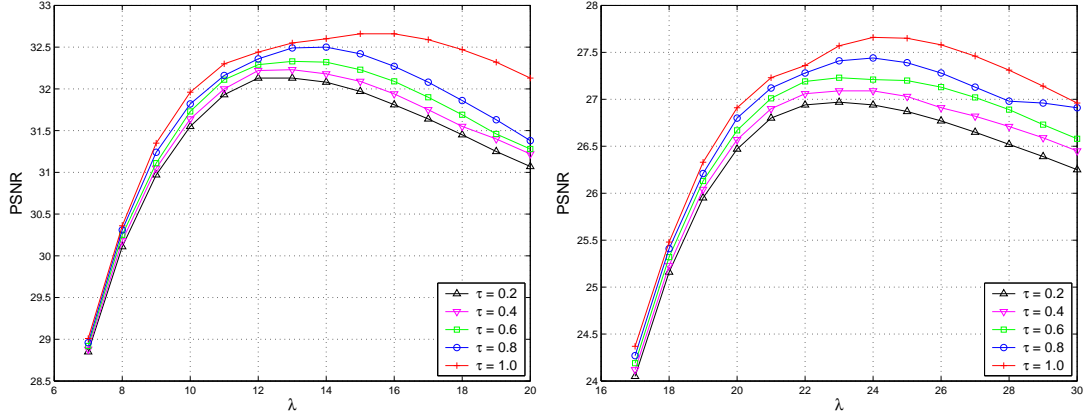


Figure 7: Performance of the GNDS-D filter (see Table 5) on the test image *House* degraded by Gaussian noise with standard deviation 20 (*Left*) and 50 (*Right*). The PSNR curves are displayed as functions of the time step size  $\tau$  and the filter parameter  $\lambda$ . The plots show that there exists an optimally global  $\lambda^*$  for a chosen time step size. Considering all the experiments carried out, the best results were obtained with  $\tau$  in the range  $[0.8, 1.0]$ .

smoothness terms  $\sum_{i \in J, j \in \mathcal{N}(i)} \Psi(|u_i - u_j|^2)$ , where the set  $\mathcal{N}(i)$  contains the 4 direct neighbours of pixel  $i$ . However, both models led to poorer results. This is in concordance with the findings in [37, 36], where the proposed variational filters perform better when a nonlocal regulariser is used and the data fidelity term is disregarded.

As was mentioned above, the GNDS filter was implemented using a steepest descent algorithm. We run the iterative scheme for different time-step size  $\tau = 0.1, 0.2, \dots, 1.0$ . Fig. 7 shows the performance of the GNDS-D filter applied to the noisy test image *House* as a function of the time step  $\tau$  and the filter parameter  $\lambda$ . Similar curves are obtained with the GNDS-S filter. As noted from Table 5 the best denoising results are attained with  $\tau$  in the range  $[0.8, 1.0]$ , in which case the number of iterations needed to reach the highest PSNR ranges between 1 and 3.

Fig. 8 shows a visual comparison of the proposed GNDS filter with the two most competitive methods [27, 45]. The *absolute method noise* (AMN)  $|o - u|$  ( $\times 5$ ) between the noise free images  $o$  and the restored versions  $u$  are shown in Fig. 9. All three approaches provide very good results, while the method of Dabov *et al.* [27] gives the highest PSNRs. Although our GNDS approach does not outperform these two methods in terms of PSNR, our results look much more pleasant and natural than those from Kervrann *et al.* [45]. That filter tends to over-enhance edges, creating staircasing artifacts that make the images look less natural, which can be observed in Fig. 10. Another visible effect of the method of Kervrann *et al.* is noticeable in the AMN images of Fig. 9. The black areas reveal that many

edges remain untouched in the filtering process, i.e. no noise is removed at those locations. Our GNDS results do not show any visible artifacts and almost no loss of structures is perceived in the method noise images. These findings suggest that PSNR is not a fully reliable measure for denoising capability and perceptual quality altogether. Alternative ways of assessing these criteria are necessary, but this goes beyond the scope of our paper.

Fig. 11 demonstrates the application of our GNDS filter to denoising colour images. The noisy *Boy* images were created adding zero-mean Gaussian noise in every  $\{R,G,B\}$  channel independently. As it was indicated in Subsection 5.6 we apply the filter (58) on every image channel using the so-called *channel coupling* technique in order to avoid the formation of false colours and the dislocation of edges. That is, the same tonal weights (59)–(60) are used in all channels. The accurate localisation and restoration of edges can be observed in the zoomed images of Fig. 12. This is especially visible in the transition between the *Boy*'s cheek and the red collar. Our filter is able to restore gentle facial features and to preserve small details such as the pullover's zip.

Finally, Fig. 13 shows that our GNDS model can successfully remove impulse noise as well. Compared to a semi-local median filter (over a  $3 \times 3$  window), our approach performs much better. We have used the regularised  $L_1$  penaliser in both the data term and the smoothness term with  $r_\rho = r_\sigma = 2$ .

## 7 Conclusions

We have introduced a general nonlocal discrete variational framework for image smoothing. It arises as a generalisation of the *Nonlocal Data and Smoothness* (NDS) filtering approach of Mrázek *et al.* [56]. Although the NDS model allows nonlocal interactions between pixels, these are effective only semi-locally. This is caused by that fact that its model constraints just penalise single pixel differences that cannot propagate reliable information about the local geometry too far away from a chosen pixel. Therefore, we propose the *Generalised NDS* (GNDS) model with data and smoothness terms penalising general dissimilarity measures defined on image patches. They allow us to incorporate structured pixel information from truly nonlocal neighbourhoods in the smoothing process. We showed that by using the weighted  $L_2$  norm as distance measure the energy minimiser results in a very robust and versatile neighbourhood filter that can be adjusted to restore vector-valued images corrupted by Gaussian and impulse noise. With respect to restoration quality our GNDS approach can outperform other related patch-based methods and compares fairly well to the more advanced ones [27, 45].

Our discrete variational framework includes as special cases patch-based generalisations of M-smoothers and bilateral filtering, as well as the NL-means filter of Buades *et al.* [15]. Other related approaches due to Kervrann *et al.* [45], Bogleux *et al.* [10] and Gilboa *et al.* [36, 37] could also be derived from our energy model by

employing a different similarity measure and/or by redefining the spatial weight functions that we use as search windows.

In this work we have mainly exploited the use of the weighted  $L_2$  norm to compute patch distances. However, there is a rich opportunity for future work concerning alternative similarity measures better suited for different types of noise contamination as well as for other applications such as deblurring, inpainting, super-resolution and segmentation.

## A Appendix

### A.1 GNDS Filter in Fixed-Point Form

The minimiser  $u$  of (44) necessarily satisfies

$$\frac{\partial E_G}{\partial u_i} = (1 - \alpha) \frac{\partial E_{GD}}{\partial u_i} + \alpha \frac{\partial E_{GS}}{\partial u_i} = 0 \quad \text{for all } i \in J. \quad (70)$$

Using the distance function  $d_{D;i,j}^2 := |d_D(u(\mathcal{P}_i), f(\mathcal{P}_j))|^2$  as in (43) we have for the data term:

$$\begin{aligned} \frac{\partial E_{GD}}{\partial u_k} &= \frac{\partial}{\partial u_k} \sum_{i,j \in J} \Psi_D(d_{D;i,j}^2) w_D(|x_i - x_j|^2) \\ &= \sum_{i,j \in J} \Psi'_D(d_{D;i,j}^2) \frac{\partial}{\partial u_k} (d_{D;i,j}^2) w_D(|x_i - x_j|^2) \\ &= \sum_{i,j \in J} \Psi'_D(d_{D;i,j}^2) \frac{\partial}{\partial u_k} \left( \sum_p G_\sigma(p) (u_{i+p} - f_{j+p})^2 \right) w_D(|x_i - x_j|^2) \\ &\stackrel{(i=k-p)}{=} 2 \sum_{j \in J} \sum_p \Psi'_D(d_{D;k-p,j}^2) G_\sigma(p) (u_k - f_{j+p}) w_D(|x_{k-p} - x_j|^2) \\ &\stackrel{(j=l-p)}{=} 2 \sum_{l \in J} \sum_p G_\sigma(p) \Psi'_D(d_{D;k-p,l-p}^2) (u_k - f_l) w_D(|x_{k-p} - x_{l-p}|^2) \\ \frac{\partial E_{GD}}{\partial u_k} &= 2 \sum_{j \in J} G_\sigma * \Psi'_D(d_{D;k-\cdot,j-\cdot}^2)(0) (u_k - f_j) w_D(|x_k - x_j|^2). \quad (71) \end{aligned}$$

Similarly, with the distance function  $d_{S;i,j}^2 := |d_S(u(\mathcal{P}_i), u(\mathcal{P}_j))|^2$  as in (43) we

have for the smoothness term:

$$\begin{aligned}
\frac{\partial E_{GS}}{\partial u_k} &= \sum_{i,j \in J} \Psi'_S(d_{S;i,j}^2) \frac{\partial}{\partial u_k} (d_{S;i,j}^2) w_S(|x_i - x_j|^2) \\
&= \sum_{i,j \in J} \Psi'_S(d_{S;i,j}^2) \frac{\partial}{\partial u_k} \left( \sum_p G_\sigma(p)(u_{i+p} - u_{j+p})^2 \right) w_S(|x_i - x_j|^2) \\
&= 2 \sum_{j \in J} \sum_p \Psi'_S(d_{S;k-p,j}^2) G_\sigma(p)(u_k - u_{j+p}) w_S(|x_{k-p} - x_j|^2) \\
&\quad + 2 \sum_{i \in J} \sum_p \Psi'_S(d_{S;i,k-p}^2) G_\sigma(p)(u_{i+p} - u_k)(-1) w_S(|x_i - x_{k-p}|^2) \\
\frac{\partial E_{GS}}{\partial u_k} &= 4 \sum_{j \in J} G_\sigma * \Psi'_S(d_{S;k-\cdot,j-\cdot}^2)(0)(u_k - u_j) w_S(|x_k - x_j|^2). \tag{72}
\end{aligned}$$

Plugging the partial derivatives (71)–(72) into (70) and employing the abbreviations (23)–(24) and (47)–(48) one obtains the fixed point form (49).

## Acknowledgements

The first author gratefully acknowledges partial funding by the *Deutscher Akademischer Austauschdienst* (DAAD), grant no. A/05/21715.

## References

- [1] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Springer, second edition, 2006.
- [2] V. Aurich and J. Weule. Non-linear Gaussian filters performing edge preserving diffusion. In G. Sagerer, S. Posch, and F. Kummert, editors, *Mustererkennung*, Informatik Aktuell, pages 538–545. Springer, 1995.
- [3] S. P. Awate and R. T Whitaker. Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. In *Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 44–51, 2005.
- [4] S. P. Awate and R. T Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):364–376, March 2006.

- [5] N. Azzabou, N. Paragios, F. Guichard, and F. Cao. Variable bandwidth image denoising using image-based noise models. In *Proc. of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007.
- [6] D. Barash and D. Comaniciu. A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift. *Image and Video Computing*, 22(1):73–81, 2004.
- [7] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proc. IEEE*, 76:869–889, 1988.
- [8] R. C. Bilcu and M. Vehvilainen. Fast nonlocal means for image denoising. In R. A. Martin, J. M. DiCarlo, and N. Samped, editors, *Digital Photography III*, volume 6502 of *Proc. of SPIE*, page 65020R, 2007.
- [9] S. Bougleux, A. Elmoataz, and M. Melkemi. Discrete regularization on weighted graphs for image and mesh filtering. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 128–139. Springer, Berlin, 2007.
- [10] S. Bougleux, A. Elmoataz, and M. Melkemi. Local and nonlocal discrete regularization on weighted graphs for image and mesh processing. *International Journal of Computer Vision*, 2009. In press.
- [11] L. E. J. Brouwer. Über Abbildungen von Mannigfaltigkeiten. *Mathematische Annalen*, 71(1):97–115, March 1911.
- [12] T. Brox and D. Cremers. Iterated nonlocal means for texture restoration. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 13–24. Springer, Berlin, 2007.
- [13] T. Brox, O. Kleinschmidt, and D. Cremers. Efficient nonlocal means for denoising of textural patterns. *IEEE Transactions on Image Processing*, 17(7):1083–1092, July 2008.
- [14] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60–65, 2005.
- [15] A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modeling and Simulation*, 4(2):490–530, 2005.

- [16] A. Buades, B. Coll, and J.-M. Morel. Neighborhood filters and PDEs. *Numerische Mathematik*, 105(1):1–34, 2006.
- [17] T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Transactions on Image Processing*, 10(2):231–241, 2001.
- [18] T. F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, wavelet, and stochastic methods*. SIAM, Philadelphia, 2005.
- [19] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, February 1997.
- [20] P. Chatterjee and P. Milanfar. A generalization of non-local means via kernel regression. In C. A. Bouman, E. L. Miller, and I. Pollak, editors, *Computational Imaging VI*, volume 6814 of *Proc. of SPIE*, page 68140P, 2008.
- [21] Y. Chen, J. Ma, Q. Feng, L. Luo, P. Shi, and W. Chen. Nonlocal prior Bayesian tomographic reconstruction. *Journal of Mathematical Imaging and Vision*, 30(2):133–146, 2008.
- [22] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, August 1995.
- [23] C. K. Chu, I. K. Glad, F. Godtlielsen, and J. S. Marron. Edge-preserving smoothers for image processing. *Journal of the American Statistical Association*, 93(442):526–541, June 1998.
- [24] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.
- [25] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot. Bayesian non local means-based speckle filtering. In *Proc. of the 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1291–1294, 2008.
- [26] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Transactions on Medical Imaging*, 27(4):425–441, April 2008.
- [27] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, August 2007.

- [28] J. Darbon, A. Cunha, T. F. Chan, S. Osher, and G. J. Jensen. Fast nonlocal filtering applied to electron cryomicroscopy. In *Proc. of the 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1331–1334, 2008.
- [29] S. Didas, P. Mrázek, and J. Weickert. Energy-based image simplification with nonlocal data and smoothness terms. In A. Iske and J. Levesley, editors, *Algorithms for Approximation*, pages 51–60. Springer, 2007.
- [30] S. Didas and J. Weickert. From adaptive averaging to accelerated nonlinear diffusion filtering. In K. Franke, K.-R. Müller, B. Nicolay, and R. Schäfer, editors, *Pattern Recognition*, volume 4174 of *Lecture Notes in Computer Science*, pages 101–110. Springer, Berlin, September 2006.
- [31] S. Didas and J. Weickert. Integrodifferential equations for continuous multi-scale wavelet shrinkage. *Inverse Problems and Imaging*, 1(1):47–62, February 2007.
- [32] M. Elad. On the origin of the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing*, 11(10):1141–1151, October 2002.
- [33] A. Elmoataz, O. Lézoray, and S. Bougleux. Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing. *IEEE Transactions on Image Processing*, 17(7):1047–1060, July 2008.
- [34] A. Elmoataz, O. Lézoray, S. Bougleux, and V.-T. Ta. Unifying local and nonlocal processing with partial difference operators on weighted graphs. In A. Foi and A. Gotchev, editors, *Proc. International Workshop on Local and Non-Local Approximation in Image Processing*, pages 11–26, 2008.
- [35] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- [36] G. Gilboa, J. Darbon, S. Osher, and T. F. Chan. Nonlocal convex functionals for image regularization. Technical Report CAM-06-57, Department of Mathematics, University of California at Los Angeles, CA, U.S.A., 2006.
- [37] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Multiscale Modeling and Simulation*, 6(2):595–630, 2007.
- [38] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *SIAM Multiscale Modeling and Simulation*, 7(3):1005–1028, 2008.

- [39] B. Goossens, H. Luong, A. Pizurica, and W. Philips. An improved non-local denoising algorithm. In A. Foi and A. Gotchev, editors, *Proc. International Workshop on Local and Non-Local Approximation in Image Processing*, pages 143–156, 2008.
- [40] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics*. Probability and Mathematical Statistics. Wiley & Sons, New York, 1986.
- [41] M. Hein, J.-Y. Audibert, and U. von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8:1325–1370, 2007.
- [42] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics*, 6:813–827, 1977.
- [43] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.
- [44] C. Kervrann and J. Boulanger. Optimal spatial adaptation for patch-based image denoising. *IEEE Transactions on Image Processing*, 15(10):2866–2878, October 2006.
- [45] C. Kervrann and J. Boulanger. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *International Journal of Computer Vision*, 79(1):45–69, August 2008.
- [46] C. Kervrann, J. Boulanger, and P. Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 520–532. Springer, Berlin, 2007.
- [47] S. Kindermann, S. Osher, and P. W. Jones. Deblurring and denoising of images by nonlocal functionals. *SIAM Multiscale Modeling and Simulation*, 4(4):1091–1115, 2005.
- [48] O. Kleinschmidt, T. Brox, and D. Cremers. Nonlocal texture filtering with efficient tree structures and invariant patch similarity measures. In A. Foi and A. Gotchev, editors, *Proc. International Workshop on Local and Non-Local Approximation in Image Processing*, pages 103–113, 2008.
- [49] J. J. Koenderink and A. L. Van Doorn. The structure of locally orderless images. *International Journal of Computer Vision*, 31(2/3):159–168, 1999.
- [50] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3(1):73–102, 1989.

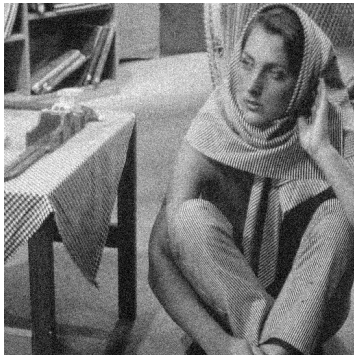


- [51] J.-S. Lee. Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24:255–269, 1983.
- [52] Y.-L. Liu, J. Wang, X. Chen, and Y.-W. Guo. A robust and fast non-local means algorithm for image denoising. *Journal of Computer Science and Technology*, 23(2):270–279, March 2008.
- [53] Y. Lou, P. Favaro, S. Soatto, and A. Bertozzi. Nonlocal similarity image filtering. In P. Foggia, C. Sansone, and M. Vento, editors, *Image Analysis and Processing - ICIAP 2009*, volume 5716 of *Lecture Notes in Computer Science*, pages 62–71. Springer, Berlin, 2009.
- [54] T. Loupas, W. N. McDicken, and P. L. Allan. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *IEEE Transactions on Circuits and Systems*, 36(1):129–135, January 1989.
- [55] M. Mahmoudi and G. Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE Signal Processing Letters*, 12(12):839–842, December 2005.
- [56] P. Mrázek, J. Weickert, and A. Bruhn. On robust estimation and smoothing with spatial and tonal kernels. In R. Klette, R. Kozera, L. Noakes, and J. Weickert, editors, *Geometric Properties for Incomplete Data*, volume 31 of *Computational Imaging and Vision*, pages 335–352. Springer, Dordrecht, 2006.
- [57] D. Mumford. The Bayesian rationale for energy functionals. In B. M. ter Haar Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, volume 1 of *Computational Imaging and Vision*, pages 141–153. Kluwer, Dordrecht, 1994.
- [58] M. Nikolova. Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *SIAM Multiscale Modeling and Simulation*, 4(3):960–991, 2005.
- [59] N. Nordström. Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection. *Image and Vision Computing*, 8:318–327, 1990.
- [60] J. Orchard, M. Ebrahimi, and A. Wong. Efficient non-local-means denoising using the SVD. In *Proc. of the 15th IEEE International Conference on Image Processing*, pages 1732–1735. IEEE Computer Society Press, 2008.
- [61] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

- [62] J. D. Peter, V. K. Govindan, and A. T. Mathew. Robust estimation approach for NL-means filter. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, K.C. Yu, T-M. Rhyne, and L. Monroe, editors, *Advances in Visual Computing*, volume 5359 of *Lecture Notes in Computer Science*, pages 571–580. Springer, Berlin, 2008.
- [63] G. Peyré, S. Bogleux, and L. Cohen. Non-local regularization of inverse problems. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part III*, volume 5304 of *Lecture Notes in Computer Science*, pages 57–68. Springer, Berlin, October 2008.
- [64] L. Pizarro, S. Didas, F. Bauer, and J. Weickert. Evaluating a general class of filters for image denoising. In B. K. Ersbøll and K. S. Pedersen, editors, *Image Analysis*, volume 4522 of *Lecture Notes in Computer Science*, pages 601–610. Springer, Berlin, 2007.
- [65] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using Gaussian scale mixtures in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, November 2003.
- [66] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [67] P. Saint-Marc, J.-S. Chen, and G. Medioni. Adaptive smoothing: A general tool for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):514–529, June 1991.
- [68] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12:43–63, 2000.
- [69] C. Schnörr. Unique reconstruction of piecewise smooth images by minimizing strictly convex non-quadratic functionals. *Journal of Mathematical Imaging and Vision*, 4:189–198, 1994.
- [70] S. M. Smith and J. M. Brady. SUSAN – A new approach to low level image processing. *International Journal of Computer Vision*, 23(1):43–78, 1997.
- [71] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs. *SIAM Journal on Numerical Analysis*, 42(2):686–713, 2004.
- [72] R. L. Stevenson, B. E. Schmitz, and E. J. Delp. Discontinuity preserving regularization of inverse visual problems. *IEEE Transactions on Systems, Man and Cybernetics*, 24:455–469, 1994.

- [73] V.-T. Ta, O. Lézoray, A. Elmoataz, and S. Schüpp. Graph-based tools for microscopic cellular image segmentation. *Pattern Recognition*, 42(6):1113–1125, 2009.
- [74] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-posed Problems*. W. H. Winston, Washington, DC, 1977.
- [75] C. Tomasi and R. Manduchi. Bilateral filtering for gray and colour images. In *Proc. of the 1998 IEEE International Conference on Computer Vision*, pages 839–846, Bombay, India, January 1998. Narosa Publishing House.
- [76] J. van de Weijer and R. van den Boomgaard. Local mode filtering. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 428–433, Kauai, HI, December 2001. IEEE Computer Society Press.
- [77] R. van den Boomgaard and J. van de Weijer. On the equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks. In *Proc. 16th International Conference on Pattern Recognition*, volume 3, pages 927–930, Quebec City, Canada, August 2002.
- [78] B. van Ginneken and B. M. ter Haar Romeny. Applications of locally orderless images. *Journal of Visual Communication and Image Representation*, 11:196–208, 2000.
- [79] E. R. Vrscay. A simple model for affine self-similarity of images and its applications. In A. Foi and A. Gotchev, editors, *Proc. International Workshop on Local and Non-Local Approximation in Image Processing*, pages 61–76, 2008.
- [80] J. Weickert. Anisotropic diffusion filters for image processing based quality control. In A. Fasano and M. Primicerio, editors, *Proc. Seventh European Conf. on Mathematics in Industry*, pages 355–362, Stuttgart, 1994. Teubner.
- [81] J. Weickert. A review of nonlinear diffusion filtering. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 3–28. Springer, Berlin, 1997.
- [82] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.
- [83] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, 2001.

- [84] G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer, second edition, 2003.
- [85] G. Winkler, V. Aurich, K. Hahn, and A. Martin. Noise reduction in images: Some recent edge-preserving methods. *Pattern Recognition and Image Analysis*, 9(4):749–766, 1999.
- [86] L. P. Yaroslavsky. *Digital Picture Processing*. Springer, New York, 1985.
- [87] E. Zeidler. *Nonlinear Functional Analysis and Applications I: Fixed-Point Theorems*. Springer, New York, 1986.
- [88] D. Zhou and C. Burges. High-order regularization on graphs. In *International Workshop on Mining and Learning with Graphs*, 2008.
- [89] D. Zhou and B. Schölkopf. A regularization framework for learning from graph data. In *Proc. ICML Workshop on Statistical Relational Learning and Its Connections to Other Fields*, pages 132–137, 2004.
- [90] D. Zhou and B. Schölkopf. Regularization on discrete spaces. In W. G. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 361–368. Springer, Berlin, September 2005.
- [91] S. Zimmer, S. Didas, and J. Weickert. A rotationally invariant block matching strategy improving image denoising with non-local means. In A. Foi and A. Gotchev, editors, *Proc. International Workshop on Local and Non-Local Approximation in Image Processing*, pages 135–142, 2008.



(a) *Barbara*, PSNR = 22.18



(b) *Boats*, PSNR = 22.17



(c) *Lena*, PSNR = 22.13



(d) PSNR = 31.78, [27]



(e) PSNR = 30.88, [27]



(f) PSNR = 33.05, [27]



(g) PSNR = 30.37, [45]



(h) PSNR = 30.12, [45]



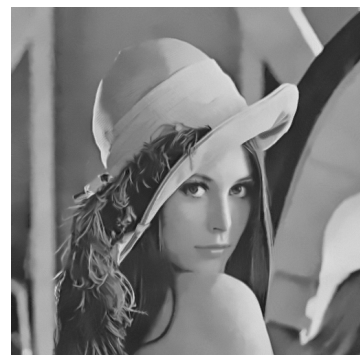
(i) PSNR = 32.64, [45]



(j) PSNR = 30.64, GNDS

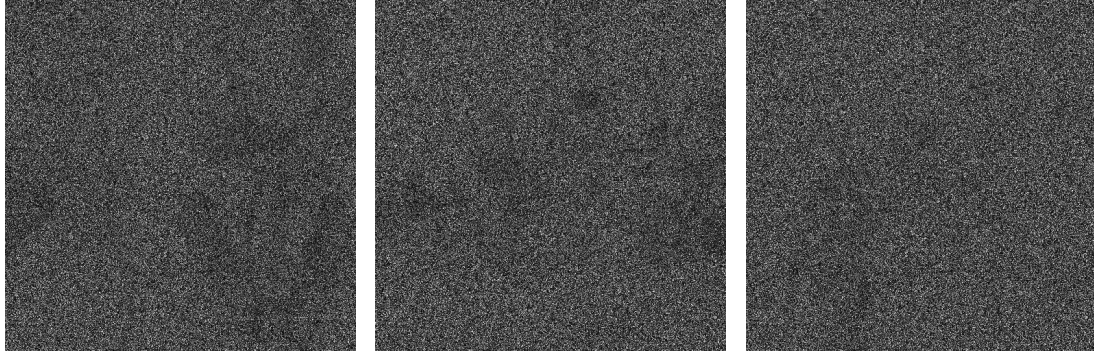


(k) PSNR = 29.80, GNDS



(l) PSNR = 32.05, GNDS

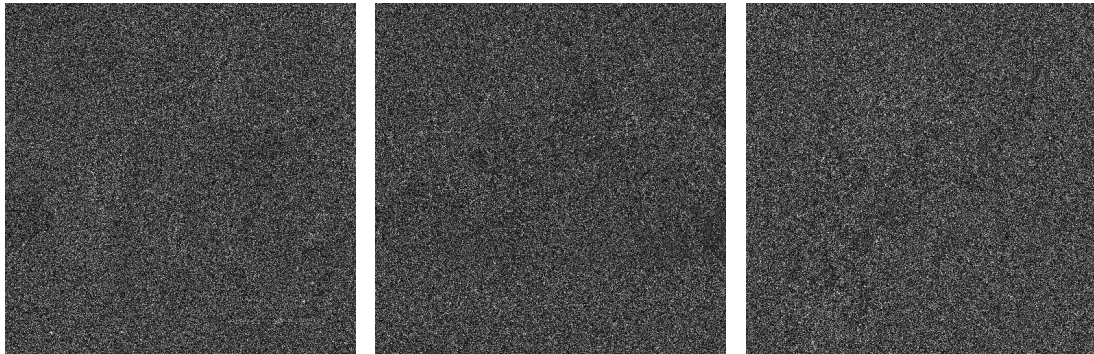
Figure 8: Comparison to state-of-the-art methods. **Top Row:** Test images degraded by Gaussian noise with standard deviation 20. **2nd Row:** Restored images by Dabov *et al.* [27]. **3rd Row:** Restored images by Kervrann *et al.* [45]. **Bottom Row:** Restored images by the proposed GNDS model (44).



(a) AMN for the restored *Barbara*, *Boats* and *Lena* by Dabov *et al.* [27].



(b) AMN for the restored *Barbara*, *Boats* and *Lena* by Kervrann *et al.* [45].



(c) AMN for the restored *Barbara*, *Boats* and *Lena* by the proposed GNDS model (44).

Figure 9: *Absolute method noise* (AMN)  $|o - u| (\times 5)$  for the smoothing results shown in Fig. 8 obtained by (a) Dabov *et al.* [27], (b) Kervrann *et al.* [45], and (c) the proposed GNDS filter.



(a) Noisy test images degraded by Gaussian noise with standard deviation 20.



(b) Restored images by Dabov *et al.* [27].



(c) Restored images by Kervrann *et al.* [45].



(d) Restored images by the proposed GNDS model (44).

Figure 10: Zoomed restored images from Fig. 8. All methods provide pleasant visual results, although the method by Kervrann *et al.* [45] also shows some staircasing artifacts that make the images look less natural.





Figure 11: GNDS filtering on colour images. **Top Row:** From left to right: Original noise free image and noisy versions degraded by Gaussian noise with standard deviation  $\{20, 40\}$ . **Bottom Row:** From left to right: Slightly smoothed original image and the corresponding restored images by the proposed GNDS filter.



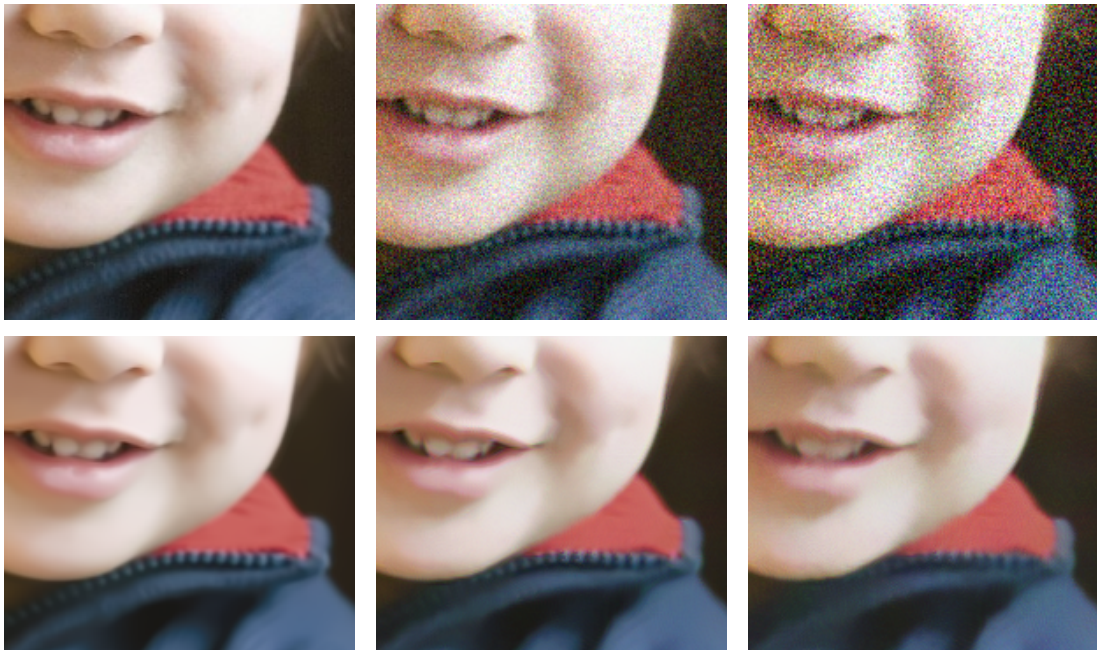


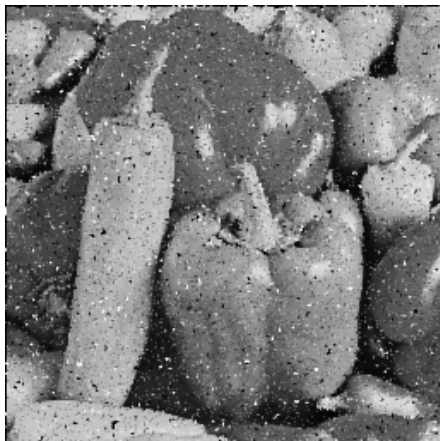
Figure 12: Zoomed images from Fig. 11. Gentle facial features are well restored, and the edges are well localised thanks to the channel coupling.



(a) Original *Peppers* image



(b) PSNR = 12.27



(c) PSNR = 20.17



(d) PSNR = 25.14

Figure 13: Filtering impulse noise. **Top Left:** Noise free image. **Top Right:** Image degraded with 20% of salt-and-pepper noise. **Bottom Left:** Restored image with a median filter. **Bottom Right:** Restored image with our GNDS model.