On Robust Estimation and Smoothing with Spatial and Tonal Kernels[*]

Pavel Mrázek,[1,2] Joachim Weickert,[1] and Andrés Bruhn[1]

[1] Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Building 27
Saarland University, 66123 Saarbrücken, Germany
`http://www.mia.uni-saarland.de`
`{mrazek,weickert,bruhn}@mia.uni-saarland.de`

[2] Pattern Recognition Department
Institute for Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 18208 Praha 8, Czech Republic
`http://www.utia.cas.cz/RO`

Abstract

This paper deals with establishing relations between a number of widely-used nonlinear filters for digital image processing. We cover robust statistical estimation with (local) M-estimators, local mode filtering in image or histogram space, bilateral filtering, nonlinear diffusion, and regularisation approaches. Although these methods originate in different mathematical theories, we show that their implementation reveals a highly similar structure. We demonstrate that all these methods can be cast into a unified framework of functional minimisation combining nonlocal data and nonlocal smoothness terms. This unification contributes to a better understanding of the individual methods, and it opens the way to new techniques combining the advantages of known filters.

Keywords: image analysis, M-estimators, mode filtering, nonlinear diffusion, bilateral filter, regularisation
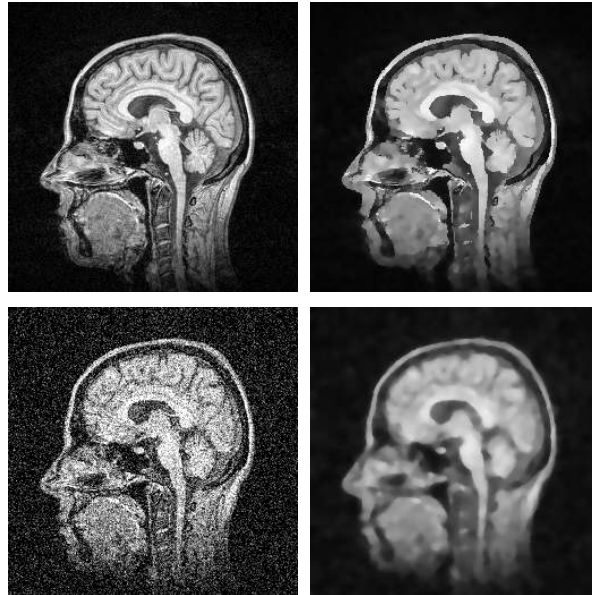

Figure 1: Examples of the tasks addressed by methods studied in this paper. Top: image simplification (input image on the left, processed with local mode filter on the right). Bottom: image denoising (noisy image on the left, filtered using TV flow on the right).

1 Introduction

Image smoothing for the task of denoising or simplification of the visual information is a well established and thoroughly studied topic. A large number of methods have been proposed and new ones continue to appear. However, it is still not easy to see the advantages of the various approaches, and the relations between different methods are only partly understood.

This paper is intended as a contribution in this direction: by studying several methods and their relations, we end up with a better understanding of each of them. We focus on M-estimators from robust statistics, median filters, mode filtering, bilateral filter, nonlinear diffusion filtering, and regularisation techniques. Although these methods seem very different at the first glance and originate in different mathematical theories, we will show that they can lead to highly similar discrete algorithms. From there, it is not far to the observation that all these methods can be cast in a single unified framework of discrete regularisation theory: they can all be derived from minimisation of a single energy functional with (possibly nonlocal) data and (possibly nonlocal) smoothness terms.

This unifying framework has several advantages. Firstly, it explicitly shows all the freedom in selecting the penaliser type, the parameters, and the balance between smoothness and data terms; seing that freedom, it is natural to require

that any reasonable smoothing methods motivates the choice of these parameters using some assumptions about the data to be reconstructed, and the noise present in the signal. Secondly, it makes explicit what assumptions are used to derive a given, previously known method from the general settings. Thirdly, after showing known methods as just special members of a whole family of discrete filters, the unifying framework allows to design novel methods tailored to the particular properties of the data and noise, and combine the advantages of known filters.

Related work Several recent papers have offered interesting connections between different filtering strategies. In [31] the relations between statistical methods (M-estimators) and iterative solvers are studied. The link between iterative mean shift algorithm, mode filters and clustering was analysed in [5]. The relations between mean, median and mode filters in the continuous settings were addressed by [11]. Mode filters and their connections to other approaches represented the main topic of [27, 28]. Finally, [9] studied the relations between bilateral filters, robust estimation and diffusion filters. Compared to these papers, our work covers a larger number of methods and it includes them all into a single, unified framework.

Organisation of the paper Our paper is organised as follows. Sections 2 to 7 represent a brief tour of several nonlinear filters for image processing, from robust statistical estimation and histogram operations to local M-smoothers, regularisation theory, diffusion filtering and bilateral filters. The methods differ in the use of information from local, global or windowed neighbourhood, and in computing the estimates relying either on the original data directly, or using a gradually smoothed image. Section 8 then proposes a unified framework which covers all of the presented methods by combining a nonlocal smoothness term and a nonlocal data term with tonal and spatial weight functions into a single functional. Section 9 shows some image filtering examples, and the paper is concluded with a summary in Section 10.

2   Statistical estimation

Let us assume there is an unknown (constant) signal u, and it is observed N-times. We obtain the noisy samples f_i, $i = 1, \ldots, N$ according to $f_i = u + n$ where n stands for the noise. If n is a zero-mean Gaussian random variable, one can estimate u by calculating the sample mean $\bar{u} = \frac{1}{N} \sum_{j=1}^{N} f_j$. The mean $\bar{u}$ is the maximum a posteriori (MAP) estimate of u, and minimises the l_2 error $E(u) = \sum_{j=1}^{N} (u - f_j)^2$.

Complications arise if the noise n is not normally distributed, e.g. if it has heavier tails (i.e. there are more outliers, or more distant outliers in the data). This can be caused either directly by the noise properties, or when not a single constant u, but e.g. two constants get mixed in the data (we have to estimate a constant value near a discontinuity in the signal). The classical statistical

solution is to use more robust error norms, and this leads to the theory of M-estimation [13, 12]. An M-estimate of a constant value u from noisy data f_j is found by minimising

$$E(u) = \sum_{j=1}^{N} \Psi\big(|u - f_j|^2\big) \tag{1}$$

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

Table 1: Examples of error norms for M-estimators.

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

The right column of Table 1 gives an overview of what element minimises the functional (1) with the given error penaliser Ψ. For the l_2 norm (a), the solution is the mean of the noisy samples. The l_1 norm (b), the formula 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 (1) approximates a mode (maximum) of the probability density underlying the noisy samples. 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 [5] 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.

3   Histogram operations

In image analysis, the data (grey values) f_i are measured at positions (pixels) x_i, and we want to find a solution vector $\mathbf{u} = (u_i)_{i=1,\ldots,N}$ where each output value u_i belongs to the position x_i. We construct the M-estimates u_i by minimising

$$E(\mathbf{u}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \Psi\big(|u_i - f_j|^2\big). \tag{2}$$

$E(\mathbf{u})$ can be minimised by gradient descent (converging towards a local minimum if Ψ is nonconvex), where each element u_i may be processed independently. Initialising by $u_i^0 = f_i$, the gradient descent becomes

$$
\begin{aligned}
u_i^{k+1} &= u_i^k - \tau \frac{\partial E}{\partial u_i} \\
&= u_i^k - \tau \sum_{j=1}^{N} \Psi'\big(|u_i^k - f_j|^2\big) \, 2 \, (u_i^k - f_j) \\
&= \Big(1 - 2\tau \sum_{j=1}^{N} \Psi'\big(|u_i^k - f_j|^2\big)\Big) u_i^k \\
&\quad + 2\tau \sum_{j=1}^{N} \Psi'\big(|u_i^k - f_j|^2\big) \, f_j
\end{aligned} \tag{3}
$$

Here τ is the step size. To speed up convergence, τ can be chosen adaptively to the data such that it is larger in plateaus and smaller in areas of large slope; see [5]. Setting

$$
\tau := \frac{1}{2 \sum_{j=1}^{N} \Psi'\big(|u_i^k - f_j|^2\big)}, \qquad g(s^2) := \Psi'(s^2), \tag{4}
$$

we can rewrite (3) into the iterative formula

$$
u_i^{k+1} = \frac{\sum_{j=1}^{N} g\big(|u_i^k - f_j|^2\big) \, f_j}{\sum_{j=1}^{N} g\big(|u_i^k - f_j|^2\big)}. \tag{5}
$$

Note that in this formulation, the spatial distance between solution u_i and the input samples f_j is not taken into consideration since the index j runs through all pixels. This procedure is equivalent to operations with the histogram of the input image. As an example, Fig. 2 (b) shows the steady state when iterating (5) with a weighting function g that was chosen in order to correspond to the penaliser Ψ from Table 1(c). This results in replacing each pixel by a local mode of the image histogram. We may thus regard it as an image adaptive quantisation strategy.

4 Local M-smoothers

While the previous method was a *global*, histogram-based technique, it is often desirable to estimate a grey value of a pixel from a *local* neighbourhood only. In the framework of M-estimation, this can be achieved by introducing a second weighting term, which depends on the spatial distance between the position of restored pixel u_i and the input sample f_j. For the *local M-smoothers*, the functional to minimise has the following structure [6, 31]:

$$
E(\mathbf{u}) = \sum_{i=1}^{N} \sum_{j \in \mathcal{B}(i)} \Psi\big(|u_i - f_j|^2\big) \, w\big(|x_i - x_j|^2\big) \tag{6}
$$

Figure 2: Local mode filtering. *(a) Left:* input image. *(b) Center:* iteratively smoothed using eq. (5) and the penaliser from Table 1(c) in a global spatial window. This approximates mode filtering of the histogram. *(c) Right:* processed with the local iterative filter from eq. (9). The tonal weight $g(s^2) = exp(-s^2/\lambda^2)$ was combined with the soft spatial window (8) with $\theta = 5$. This approximates a local mode filter.

where the spatial weights w represent e.g. a hard disk-shaped window around the current position x_i,

$$w(s^2) = \begin{cases} 1 & s^2 < \theta \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

or a soft window [6],

$$w(s^2) = e^{-s^2/\theta^2}. \tag{8}$$

The local window $\mathcal{B}(i)$ is introduced in (6) 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.

In the same way as in the previous section, a minimisation of (6) with adaptive time steps leads to the iterative formula

$$u_i^{k+1} = \frac{\sum_{j \in \mathcal{B}(i)} g(|u_i^k - f_j|^2)\, w(|x_i - x_j|^2)\, f_j}{\sum_{j \in \mathcal{B}(i)} g(|u_i^k - f_j|^2)\, w(|x_i - x_j|^2)} \tag{9}$$

where the function w is called *spatial weight*, in order to distinguish it from the *tonal weight* g. The iterative process is initialised with $u_i^0 := f_i$. Note that, both in (5) and in (9), we are only interested in the *steady state* for $k \to \infty$, not in the evolution towards this minimiser.

As stated in [31], the procedure (9) is called *W-estimator*, and represents one possibility to obtain a solution to the local M-estimation problem. It converges to a local minimum of (6) close to the input data. Depending on the penaliser Ψ, the iterations may lead e.g. to a local mode approximation as in Fig. 2 (c) [11, 27, 28, 7], or to an approximation of a windowed median filter or Gaussian smoothing.

5 Bayesian and regularisation frameworks

By taking the windowed M-estimator (6) and decreasing the spatial window size θ, we arrive at the weighting

$$w\big(|x_i - x_j|^2\big) = \begin{cases} 1 & \text{if } x_i = x_j, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

This leads to the functional

$$E_D(\mathbf{u}) = \sum_{i=1}^{N} \Psi\big(|u_i - f_i|^2\big). \tag{11}$$

For any reasonable penaliser Ψ, (11) is minimised by $u_i = f_i$.

It is clear that such a solution is not desired: a good estimate cannot be obtained by looking at a single noisy sample. However, estimation formulated using a local neighbourhood is a highly successful practice, but it has to be combined with some assumptions about the signal to be recovered. In the Bayesian terminology such an assumptions is called *prior information*, in the framework of regularisation theory it is named *smoothness term* or *regulariser*; see e.g. [2, 10, 15].

We construct a smoothness term to express our assumptions about the signal. For the sake of convenience, let us now focus on a continuous modeling where (11) is replaced by

$$E_D(u) = \int_{\Omega} \Psi\big(|u - f|^2\big)\, dx \tag{12}$$

where $\Omega \subset \mathbb{R}^m$ is the image domain, and $f, u : \Omega \to \mathbb{R}$ denote the original and filtered image, respectively. In the classical example of the Mumford–Shah functional [16] where the signal is assumed to be piecewise constant with step-like discontinuities, we have

$$E_S(u) = \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \beta\,|\Gamma|$$

where Γ is the set of discontinuities and $|\Gamma|$ denotes its length (one-dimensional Hausdorff measure). The smoothness of the image is measured by the squared gradient magnitude $|\nabla u|^2$. We see that deviations from the smoothness are not penalised at the discontinuities. The parameter β balances the image smoothness against the measure of the discontinuity set Γ.

It was shown e.g. in [31] that the explicit boundaries Γ can be expressed implicitly using a robustified prior: Let γ denote the discontinuity indicator function

$$\gamma(x) = \begin{cases} 1 & \text{on } \Gamma \text{ (edge)}, \\ 0 & \text{on } \Omega \setminus \Gamma \text{ (no edge)}, \end{cases}$$

and

$$\Psi_\beta(|\nabla u|^2) := \min\{|\nabla u|^2, \beta\} \tag{13}$$

the cup function from Table 1(d). Then it follows that

$$
\begin{aligned}
\min_u E_S(u) &= \min_{u,\Gamma} \left(\int_{\Omega \backslash \Gamma} |\nabla u|^2 dx + \beta\, |\Gamma| \right) \\
&= \min_{u,\gamma} \int_\Omega \Big((1-\gamma)\, |\nabla u|^2 + \beta\, \gamma \Big)\, dx \\
&= \min_u \int_\Omega \Big(\min_{\gamma \in \{0,1\}} (1-\gamma)\, |\nabla u|^2 + \beta\, \gamma \Big)\, dx \\
&= \min_u \int_\Omega \Psi_\beta \big(|\nabla u|^2 \big)\, dx.
\end{aligned}
\tag{14}
$$

For smoother penalisers Ψ, the discontinuity indicator γ may also attain intermediate values from the interval $[0,1]$ [19].

In the Bayesian / regularisation framework, the data and smoothness terms are combined into a single functional, thus balancing the measured data against the smoothness assumptions. The resulting functional has e.g. the form

$$
\begin{aligned}
E(u) &= E_D(u)\ +\ \alpha\, E_S(u) \\
&= \int_\Omega \Big(\Psi_D\big(|u-f|^2 \big)\ +\ \alpha\, \Psi_S(|\nabla u|^2) \Big)\, dx
\end{aligned}
\tag{15}
$$

with some regularisation parameter $\alpha > 0$. As an example, the continuous Mumford–Shah functional fits into this framework if we choose $\Psi_D(s^2) := s^2$ and $\Psi_S(s^2) := \min(s^2, \lambda^2)$. Discrete versions of such functionals are considered e.g. in [3, 17, 18].

6 Diffusion filtering

For the sake of completeness, it should be mentioned that equation (15) also covers nonlinear diffusion filters with and without a fidelity term: Choosing $\Psi_D(s^2) := s^2$, every mininiser of (15) has to satisfy necessarily the Euler–Lagrange equation [8]

$$
0 = \operatorname{div}\Big(g(|\nabla u|^2)\, \nabla u \Big) - \frac{u-f}{\alpha}
\tag{16}
$$

where $g(s^2) := \Psi_S'(s^2)$. Its solution can be regarded as the steady state of the diffusion–reaction process

$$
\frac{\partial u}{\partial t} = \operatorname{div}\Big(g(|\nabla u|^2)\, \nabla u \Big) - \frac{u-f}{\alpha}.
\tag{17}
$$

where the "time" t is a purely numerical parameter. Such "biased" diffusion processes with a fidelity term have been considered by [19] in the nonconvex and by [22, 25, 4] in the convex case. They yield the filtered image at *infinite* time $(t \to \infty)$.

Alternatively, in [21] it has been argued that by rewriting (16) as

$$\frac{u-f}{\alpha} = \text{div}\left(g(|\nabla u|^2)\,\nabla u\right) \tag{18}$$

it becomes evident that this process can be regarded as an implicit time discretisation of the diffusion process

$$\frac{\partial u}{\partial t} = \text{div}\left(g(|\nabla u|^2)\,\nabla u\right), \tag{19}$$
$$u(t=0) = f \tag{20}$$

with a single time step of size α. This is a classical, "unbiased" nonlinear diffusion filter as is considered e.g. in [20, 29]. Note that, in contrast to (16), such a filter gives the desired result at *finite* diffusion time $t = \alpha$.

7 Bilateral filtering

Since digital images are sampled on a quadratic pixel grid, it becomes necessary to consider discrete variants of the continuous functional (15). For the data term, this has already been discussed. Therefore let us now focus on the smoothness term $E_S(u) = \Psi_S(|\nabla u|^2)$.

One possibility is to estimate the image gradient magnitude as a sum of squared differences from a pixel to its neighbours. The discrete smoothness penaliser is then expressed in the following way:

$$E_S(\mathbf{u}) = \sum_{i=1}^{N} \Psi_S\left(\sum_{j\in\mathcal{N}(i)} |u_i - u_j|^2\right) \tag{21}$$

where $\mathcal{N}(i)$ stands for the set of 4-neighbours of a pixel i. Just by exchanging the order of summation and penalisation in the last term, we can express the assumption of image smoothness in a slightly different way:

$$E_S(\mathbf{u}) = \sum_{i=1}^{N} \sum_{j\in\mathcal{N}(i)} \Psi_S\left(|u_i - u_j|^2\right). \tag{22}$$

This change of operation ordering leads to an *anisotropic* smoothness measure; see also [30].

Let us now increase the size of the neighbourhood from which the expression (22) is estimated. Then the smoothness term becomes

$$E_S(\mathbf{u}) = \sum_{i=1}^{N} \sum_{j\in\mathcal{B}(i)} \Psi\left(|u_i - u_j|^2\right) w\left(|x_i - x_j|^2\right) \tag{23}$$

where $\mathcal{B}(i)$ is the larger neighbourhood set, and the summation is additionally weighted by a function w of the spatial distance between pixels.

The functional (23) can be minimised by an iterative procedure

$$u_i^{k+1} = u_i^k - \tau \, \frac{\partial E}{\partial u_i}, \qquad \tau := \frac{1}{2 \, \sum_{j \in \mathcal{B}(i)} \Psi'\big(|u_i^k - u_j^k|^2\big)}.$$

Setting $g(s^2) := \Psi'(s^2)$ leads to the weighted averaging scheme

$$u_i^{k+1} = \frac{\sum_{j \in \mathcal{B}(i)} g\big(|u_i^k - u_j^k|^2\big) \, w\big(|x_i - x_j|^2\big) \, u_j^k}{\sum_{j \in \mathcal{B}(i)} g\big(|u_i^k - u_j^k|^2\big) \, w\big(|x_i - x_j|^2\big)} \qquad (24)$$

Equation (24) is exactly the *bilateral filter* [26]; see also [23] for related ideas. While bilateral filtering was originally proposed as a heuristic algorithm, we derived it here as an iterative solver to minimise the anisotropic smoothness term (23) which is evaluated in a nonlocal window.

One should observe the large amount of structural similarities between the local M-smoother (9) and the bilateral filter (24). However, there is one significant difference: Local M-smoothing uses the *initial* image in the averaging procedure and searches for the steady state, while bilateral filtering uses the *evolving* image and has to stop after a certain number of iterations in order to avoid obtaining a flat image.

Moreover, it should be noted that an alternative functional to justify the bilateral filter was proposed in [9]. Differently to the one presented here, the functional of Elad contains a windowed smoothness term combined with a local data term. Also, bilateral filter can be viewed as an approximation to the Beltrami flow [1, 24].

8 Unifying framework

Figure 3 presents an overview of the energy functionals minimised by all the methods discussed so far. Starting from statistical *M-estimation* at the top, we went counterclockwise down the left branch via histogram operations, introduced spatial window weighting w into the functional, and derived *local M-estimators*. Reducing the window size and combining data with a smoothness assumption, we arrived at *regularisation methods* fitting into the *Bayesian framework* at the bottom of Fig. 3. Concentrating on the smoothness term only, we estimated the gradient magnitude $|\nabla u|$ using discrete samples, extended the size of the estimation window, formed an anisotropic smoothness measure, and derived the *bilateral filter* on the right. The circle can be closed to histogram-based global methods by extending the spatial window size.

We observe that the methods classify into two main branches: the data-based on the left, and the smoothness-based on the right of Fig. 3. The data-based methods correspond to *statistical estimation* from noisy data, while the right branch leads to the known methods of *image smoothing*. Unfortunately, these labels are sometimes confused and a method optimising the smoothness term is claimed to represent a robust statistical estimator. We have seen that the
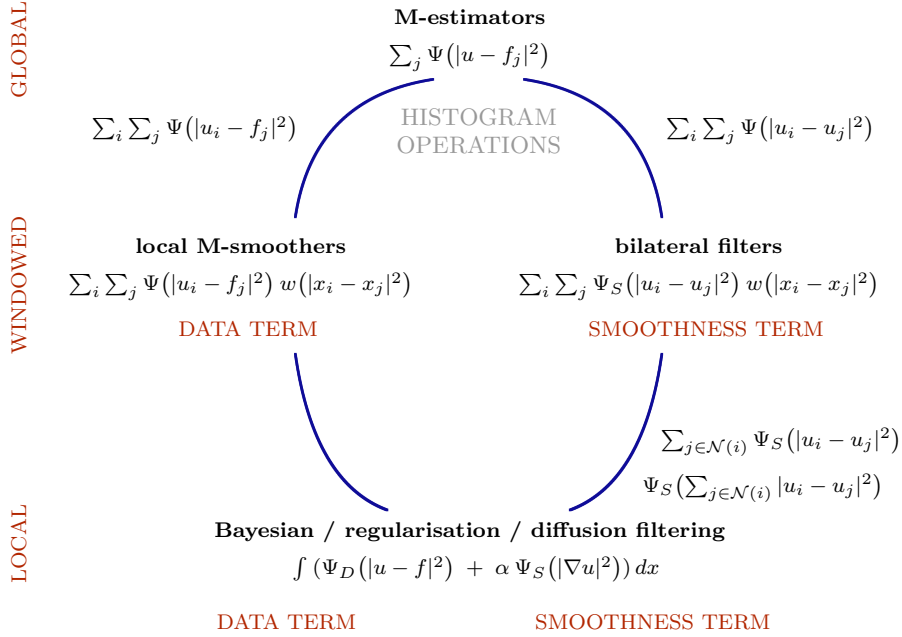
Figure 3: Overview of the methods studied in this paper and their corresponding energy functionals.

methods have a highly similar structure, but believe that the terms *estimation* and *smoothing* should not be used as equivalent.

The spatial extent of each filter is controlled by the weight function w, from global methods at the top ($w = 1$) to local approaches at the bottom. Let us focus on the methods in the middle, where the weight w specifies a finite window. We said that the regularisation methods contain a (local) data term and a (local) smoothness term. If we understand the local M-estimators as a data term (expressed using a nonlocal window for each pixel) and the bilateral filter as a smoothness term (again using a finite window), it is natural to combine them into a single, unified functional:

$$E(\mathbf{u}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma \, \Psi_D\big(|u_i - f_j|^2\big) \, w_D\big(|x_i - x_j|^2\big)$$
$$+ (1 - \sigma) \, \Psi_S\big(|u_i - u_j|^2\big) \, w_S\big(|x_i - x_j|^2\big) \tag{25}$$

Depending on the choice of the penalising functions Ψ_D, Ψ_S and on the extent of the spatial weighting functions w_D and w_S, the single formulation (25) covers all the filters discussed so far. They are summarised in Table 2.

The unification of methods into a single framework has several advantages. It contributes to the understanding of each method as it makes explicit what

Table 2: Filtering methods structured according to the used penaliser (from Table 1 (a), (b), (c)), extent of the spatial weight w, and correspondence to the data or smoothness terms.

penaliser ‖	windowed data	local smoothness	windowed smoothness
(a)	mean	linear diffusion	
(b)	median	TV diffusion	
(c)	mode approximation	nonlinear diffusion	bilateral filter

parameters $\Psi_D, w_D, \Psi_S, w_S, \sigma$ are needed to derive a given filter. We can see all the freedom that this class of methods offers: four weighting functions (and their parameters) plus the parameter $\sigma \in [0,1]$ balancing the data against the smoothness term. Obviously, to obtain a reasonable and well performing filter, the choice of parameters should be motivated by some arguments about the data and noise properties. Last but not least, we have seen that known filters represent just several special cases in the framework of functional (25). New methods can be designed to combine the advantages of known filters. This remains the topic for further research.

Some more questions are left open, though. For example, what is the meaning of the smoothness term calculated from a larger window as in (23)? A single pixel may then be directly connected to quite distant pixels, which leads to large-scale smoothing effects of a single filter iteration, but the local topology (e.g. the classical notion of connected regions) is lost, similarly to the locally orderless images [14].

9   Experiments

In this section we present several examples in order to demonstrate the effect of individual filter components on the final results. We stress that the pictures are intended to visualise the main effects. They are not intended to claim that one method performs better than the other: optimised results in terms of image simplification or denoising can be obtained by each of them by tuning the parameters.

The first image simplification example is shown in Fig. 5. The penaliser $\Psi(s^2) = 1 - \exp(-s^2/\lambda^2)$ was employed, and the soft spatial window (8) had varied size ($\theta = 1$, $\theta = 3$, $\theta = 10$, cropped circularly into windows of sizes 3×3, 7×7, and 21×21, respectively). We observe that the image filtered via the data term minimisation (i.e. local M-estimator, top row of Fig. 5) becomes smoother as the window size increases. In this case the steady state is depicted. In the bottom row (smoothness term, bilateral filter) the effect of different window sizes is fairly small if the same number of iterations is used. We observe that already a small window applied iteratively leads to global effects.

Figures 6 and 7 demonstrate the influence of the penaliser type on the re-

Figure 4: Input images for the filtering examples. Noise-free on the left (used in Figures 5 and 6), noisy on the right (Gaussian noise, SNR=4; used in Fig. 7).

DATA TERM

SMOOTHNESS

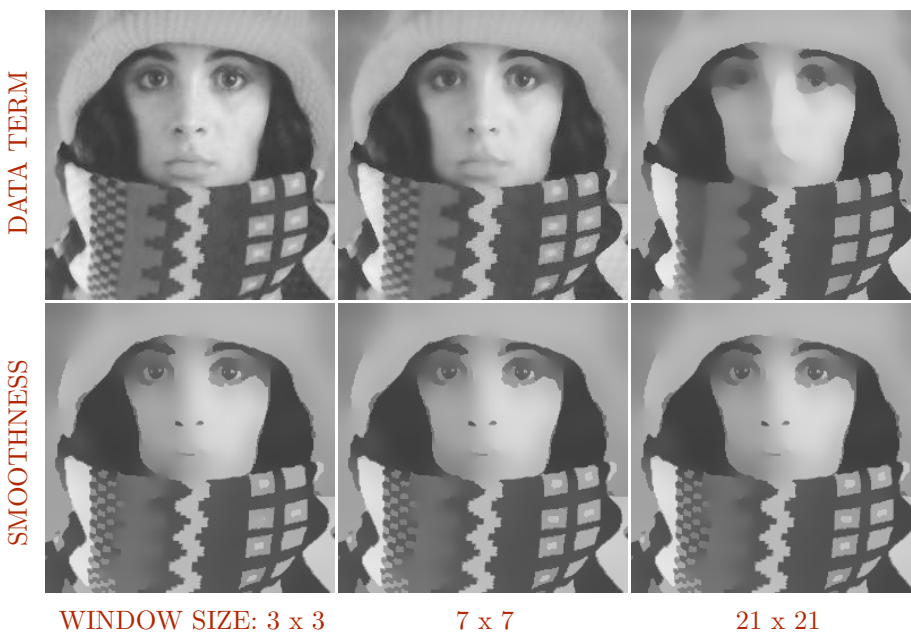WINDOW SIZE: 3 x 3   7 x 7   21 x 21

Figure 5: Filtering using the penaliser $\Psi(s^2) = 1 - e^{-s^2/\lambda^2}$ (Table 1d) with varied size of the spatial neighbourhood. Top: local M-smoothers (data term, steady state of iterating (9)). Bottom: bilateral filtering (smoothness term, 200 iterations based on (24)).

sult. All images were created using a 7×7 soft spatial window. The l_2 penaliser blurs the image most and removes noise very well, while the local mode approximations on the right of Figures 6 and 7 perform better at preserving the discontinuities, but the result is also more sensitive to noise. The l_1 penalisation in the center column can represent a good compromise between contrast preservation and noise removal, depending on the particular task and data properties.
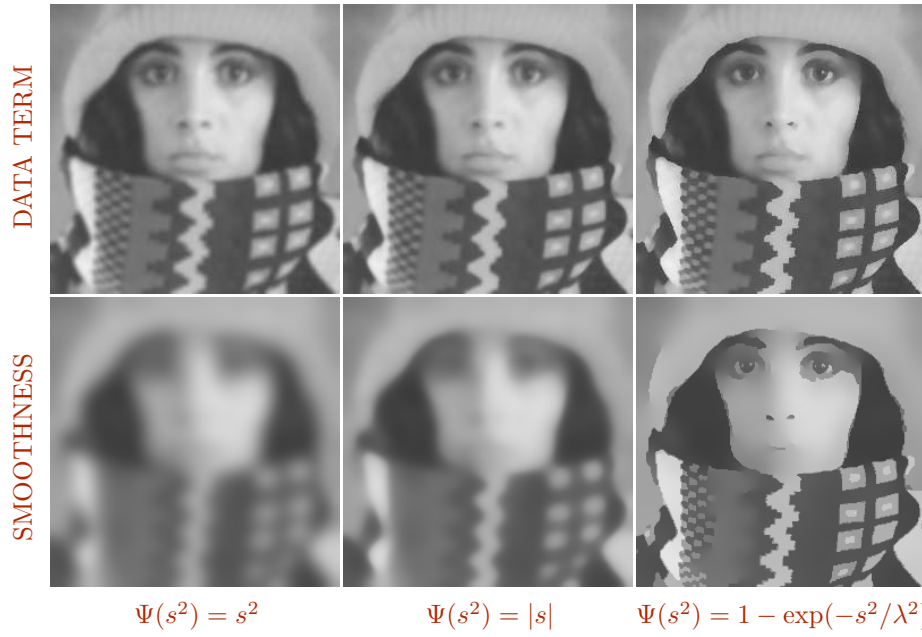
13

Figure 6: Effect of the penaliser type on the filtering result (with soft spatial weighting (8), $\theta = 3$).
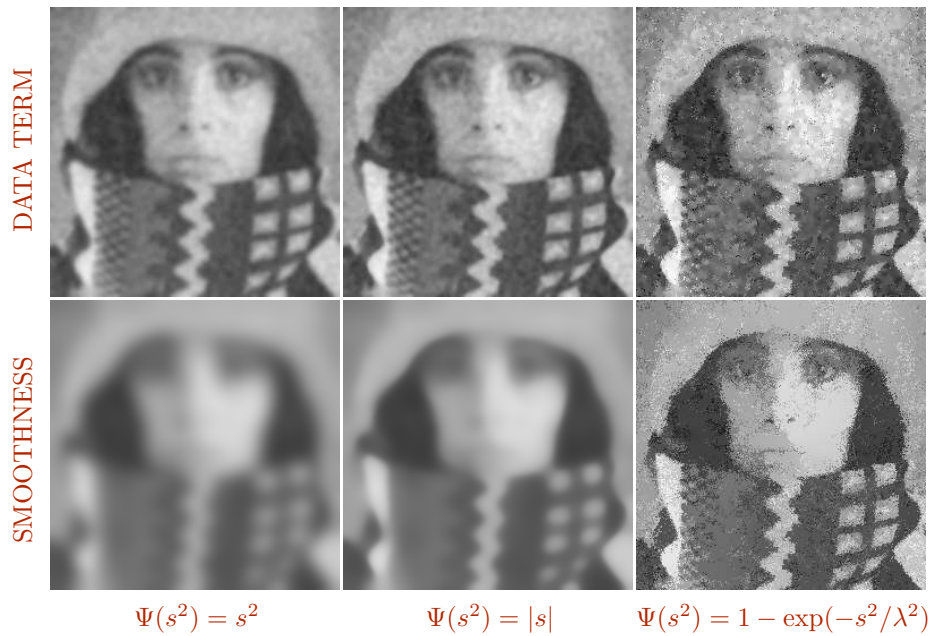

Figure 7: Effect of the penaliser type on the filtering result, starting from the noisy image in Fig 4 right (filtering in a soft window (8), $\theta = 3$).

10   Conclusion

In this paper we focused on the relations between nonlinear filters for digital image processing. We covered statistical M-estimation, mean and median filtering, mode approximation, regularisation and nonlinear diffusion approaches, and bilateral filtering. We have shown that all these methods can be cast into the unified framework of functional minimisation where the functional consists of a (possibly nonlocal) data and (possibly nonlocal) smoothness term. The mutual influence of image pixels is controlled by weighting functions depending on the spatial and tonal distances.

The unified formulation brings new insight and clarifies the relations between different methods. It makes explicit what assumptions are needed to derive known methods (often proposed ad hoc) from the general framework. Then, novel methods can be designed to combine the advantages of known filters and suit the particular data properties.

References

[1] D. Barash. A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):844–847, June 2002.

[2] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, Aug. 1988.

[3] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.

[4] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.

[5] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, Aug. 1995.

[6] C. K. Chu, I. Glad, F. Godtliebsen, and J. S. Marron. Edge-preserving smoothers for image processing. *Journal of the American Statistical Association*, 93(442):526–556, 1998.

[7] 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.

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

[9] M. Elad. On the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing*, 11(10):1141–1151, Oct. 2002.

[10] 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.

[11] L. D. Griffin. Mean, median and mode filtering of images. *Proceedings of the Royal Society of London, Series A*, 456(2004):2995–3004, 2000.

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

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

[14] J. J. Koenderink and A. J. van Doorn. The structure of locally orderless images. *International Journal of Computer Vision*, 31(2/3):159–168, Apr. 1999.

[15] 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.

[16] D. Mumford and J. Shah. Optimal approximation of piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–685, 1989.

[17] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2):633–658, 2000.

[18] M. Nikolova. Image restoration by minimizing objective functions with nonsmooth data-fidelity terms. In *Proc. First IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pages 11–18, Vancouver, Canada, July 2001. IEEE Computer Society Press.

[19] N. Nordström. Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection. *Image and Vision Computing*, 8:318–327, 1990.

[20] 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.

[21] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63, Feb. 2000.

[22] 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.

[23] S. M. Smith and J. M. Brady. SUSAN: A new approach to low-level image processing. *International Journal of Computer Vision*, 23(1):45–78, May 1997.

[24] A. Spira, R. Kimmel, and N. Sochen. Efficient Beltrami flow using a short time kernel. In L. Griffin and M. Lillholm, editors, *Scale Space Methods in Computer Vision*, volume 2695 of *Lecture Notes in Computer Science*, pages 511–522. Springer, Berlin, 2003.

[25] 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.

[26] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Sixth International Conference on Computer Vision*, pages 839–846, Bombay, India, Jan. 1998. Narosa Publishing House.

[27] 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, Dec. 2001. IEEE Computer Society Press.

[28] 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, Aug. 2002.

[29] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

[30] 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, Dec. 2001.

[31] 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.