

Evaluating a General Class of Filters for Image Denoising

Luis Pizarro, Stephan Didas, Frank Bauer, and Joachim Weickert

Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Building E1 1, Saarland University, 66041 Saarbrücken, Germany
{pizarro,didas,bauer,weickert}@mia.uni-saarland.de
<http://www.mia.uni-saarland.de>

Abstract. Recently, an energy-based unified framework for image denoising was proposed by Mrázek et al. [10], from which existing nonlinear filters such as M-smoothers, bilateral filtering, diffusion filtering and regularisation approaches, are obtained as special cases. Such a model offers several degrees of freedom (DOF) for tuning a desired filter. In this paper, we explore the generality of this filtering framework in combining nonlocal tonal and spatial kernels. We show that Bayesian analysis provides suitable foundations for restricting the parametric space in a noise-dependent way. We also point out the relations among the distinct DOF in order to guide the selection of a combined model, which itself leads to hybrid filters with better performance than the previously mentioned special cases. Moreover, we show that the existing trade-off between the parameters controlling similarity and smoothness leads to similar results under different settings.

1 Introduction

Many denoising techniques have been proposed in literature, many of them are application-dependent. However, there are only few strategies that combine different approaches and allow further generalizations. One of them is the energy-based approach proposed by Mrázek et al. [10], which combines the well-known M-smoothers [5] with bilateral filtering [17]. By extending the spatial influence that neighbouring pixels have on a central pixel, they end up with a general nonlocal filtering framework that rewards similarity to the input image and penalises large deviations from smoothness. Many other existing nonlinear filters are obtained as special cases of this framework. In this paper we explore the relations among the different degrees of freedom available for tuning a specific filter. Particularly, we focus on degradation processes governed by Gaussian and impulse perturbations. Moreover, in Bayesian analysis we find suitable foundations to instantiate probabilistically this model. Finally, we point out that the use of combined (nonlocal) *data* and (nonlocal) *smoothness* constraints leads to better denoising results than considering those filters obtained as special cases of this general approach. Our paper is organised as follows: Section 2 presents

the *Nonlocal Data and Smoothness* (NDS) filtering framework. Section 3 provides brief insides into Bayesian analysis that will help us in tuning the model parameters in Section 4. Finally, some conclusions are outlined in Section 5.

2 The NDS Framework

Let $f \in \mathbb{R}^n$ be a given degraded version of the unknown 1-D or 2-D image $u \in \mathbb{R}^n$, and let i, j be pixel indices running along the set $\Omega = \{1, \dots, n\}$. In [10] a unifying variational approach to restore the original image u was proposed. It minimises the energy function

$$E(u) = \alpha \sum_{i,j \in \Omega} \Psi_D(|u_i - f_j|^2) w_D(|x_i - x_j|^2) + (1 - \alpha) \sum_{i,j \in \Omega} \Psi_S(|u_i - u_j|^2) w_S(|x_i - x_j|^2). \tag{1}$$

This energy-based approach comprises a linear combination of two constraints, i. e. $E(u) = \alpha E_D(u) + (1 - \alpha) E_S(u)$, where the constant $\alpha \in [0, 1]$ determines the relative importance of both assumptions. The *data term* $E_D(u)$ rewards similarity to the input image f , while the *smoothness term* $E_S(u)$ penalises deviations from (piecewise) homogeneity in the restored u . On one hand, the kernels $\Psi(\cdot)$ are increasing functions that penalise large *tonal distances* s^2 ; that is, the distance between two pixel grey values s_i and s_j . On the other hand, the kernels $w(\cdot)$ are nonnegative and (possibly) nonlocal windows that ponder the influence of distant pixels; the Euclidean distance between two pixel locations x_i and x_j is called *spatial distance* x^2 . See Table 1 and Table 2 for a non-exhaustive list of kernels $\Psi(\cdot)$ and $w(\cdot)$ proposed in literature.

Table 1. Popular choices for tonal weights Ψ

$\Psi(s^2)$		$\Psi'(s^2)$	known in the context of
s^2		1	Tikhonov regularisation [16]
$2(\sqrt{s^2 + \epsilon^2} - \epsilon)$		$(s^2 + \epsilon^2)^{-\frac{1}{2}}$	regularised total variation (TV) [15]
$2\lambda^2 \left(\sqrt{1 + \frac{s^2}{\lambda^2}} - 1 \right)$		$\left(1 + \frac{s^2}{\lambda^2}\right)^{-\frac{1}{2}}$	nonlinear regularisation, Charbonnier <i>et al.</i> [4]
$\lambda^2 \log \left(1 + \frac{s^2}{\lambda^2}\right)$		$\left(1 + \frac{s^2}{\lambda^2}\right)^{-1}$	nonlinear diffusion, Perona-Malik 1 [14]
$2\lambda^2 \left(1 - \exp\left(-\frac{s^2}{2\lambda^2}\right)\right)$		$\exp\left(-\frac{s^2}{2\lambda^2}\right)$	nonlinear diffusion, Perona-Malik 2 [14]
$\min(s^2, \lambda^2)$		$\begin{cases} 1 & s < \lambda \\ 0 & \text{else} \end{cases}$	segmentation, Mumford and Shah [11]

Table 2. Possible choices for spatial weights w

$w(s^2)$		known in the context of
$\begin{cases} 1 & s < \theta \\ 0 & \text{else} \end{cases}$		hard window locally orderless images, Koenderink and van Doorn [8]
$\exp\left(-\frac{s^2}{\theta^2}\right)$		soft window Chu <i>et al.</i> [5]

Table 3. Nonlinear filters derived as special cases of the NDS framework (1)

	Filter	Model
(a)	histogram operations	$\sum_{i,j} \Psi(u_i - f_j ^2)$
(b)	M-smoothing	$\sum_{i,j} \Psi_D(u_i - f_j ^2) w_D(x_i - x_j ^2)$
(c)	bilateral filtering	$\sum_{i,j} \Psi_S(u_i - u_j ^2) w_S(x_i - x_j ^2)$
(d)	Bayesian/regularisation/diffusion	$\int (u - f ^2 + \alpha \Psi_S(\nabla u ^2)) dx$

The NDS function (1) presented above exhibits wide generality in choosing the parameters Ψ_D, Ψ_S, w_D, w_S and α to derive a desired filter. In particular, Table 3 shows some well-known nonlinear filters obtained as special cases by tuning the different degrees of freedom.

To minimise the expression (1) we make use of the Jacobi method. See [6] for a comparison study among different minimisation methods for the NDS functional.

3 Statistical Background

The most common degradation model is given by $f = u + \eta$, where u is the true image, η represents a zero-mean additive noise with standard deviation σ , and f is the recorded image. In the following we consider these quantities as realisations of the random variables U, η , and F , denoting as p_U, p_η , and p_F their probability density function (pdf), respectively. In Bayesian analysis [18], the *maximum a posteriori* (MAP) estimator

$$\begin{aligned} \hat{u}_{\text{MAP}} &= \arg \max_u \log p_{U|F}(u|f) \\ &= \arg \min_u (-\log p_{F|U}(f|u) - \log p_U(u)) \end{aligned} \tag{2}$$

yields the most likely image u given f . The conditional distribution $p_{F|U}$, also called *likelihood*, models the degradation process of U to F and is therefore considered as the noise distribution, i. e. $p_{F|U}(f|u) = p_\eta(\eta) = \prod p_\eta(f_i - u_i)$, assuming that the noise is independent and identically distributed (i.i.d.). We will focus on Gaussian and impulse noise. These types of noise are well modelled by the *Gaussian* and *Laplacian* distributions, being respectively their pdf's

$$p_{\eta_G} = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i \in \Omega} |\eta_i|^2 \right), \tag{3}$$

$$p_{\eta_I} = \left(\frac{1}{2\sigma} \right)^n \exp \left(-\frac{1}{\sigma} \sum_{i \in \Omega} |\eta_i| \right). \tag{4}$$

Plugging these noise models into the MAP estimator (2), and observing the structural resemblance to our NDS function (1), suggests that the penalisers for the data term can be instantiated as $\Psi_D(s^2) = s^2$ for Gaussian noise, and as $\Psi_D(s^2) = |s|$ for impulse noise, see Table 1.

The previous noise distributions are special cases of a more general probabilistic law: the *generalized Gaussian* distribution [7], with parameters mean μ , variance σ^2 , and $\nu > 0$ (Gaussian case $\nu = 2$, Laplacian case $\nu = 1$); and pdf

$$p_Z(z) = \frac{\nu \Gamma(3/\nu)^{1/2}}{2\sigma \Gamma(1/\nu)^{3/2}} \exp \left(-\frac{|z - \mu|^\nu}{\sigma^\nu} \left(\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)} \right)^{\nu/2} \right), \tag{5}$$

where $\Gamma(\cdot)$ is the Euler Gamma function. This distribution has been also utilized for modelling probabilistic *prior* knowledge about the signal u to recover. In (2), this information is represented in terms of the *prior* distribution p_U of the grey values of U . Besag [2] proposed the Laplacian law as model for p_U , which was later extended by Bouman & Sauer with their *generalized Gaussian Markov Random Field* [3] based on the distribution (5) for $\nu \in [1, 2]$. Since choosing a particular model for the prior distribution is essentially equivalent to specify the penaliser Ψ_S for the smoothness term in our NDS framework, we can instantiate such tonal kernel as $\Psi_S(s^2) = |s|^\nu$. This function is nonconvex for $0 < \nu < 1$, what may give rise to local minima in (1). However, nonconvex penalisers can allow almost exact restoration quality [9,12,13].

In summary, the Bayesian framework provides a founded basis for choosing appropriate tonal kernels $\Psi(\cdot)$ for the data and smoothness terms in (1). Studying other types of noise and the properties of the signal to recover, will lead to different criteria for selecting the penalisers.

4 Tuning the Model Parameters

4.1 Linear Combination of Kernels

The problem of determining α for the image simplification approach described in the Section 2 is crucial to obtain an optimal combination of similarity and smoothness. We intend to justify the use of such framework for $\alpha \notin \{0, 1\}$, i. e. for a wider spectrum of filters than those special cases outlined in Table 3.

A function $\varphi : [0, 1] \rightarrow \mathbb{R}$ is called *unimodal* on $[0, 1]$ if it contains a single minimum in that interval. Then, we obtain an estimate of the true image u as

$$\hat{u} = \arg \min_{\alpha} \varphi(\alpha), \tag{6}$$

assuming that we deal with the unimodal function $\varphi(\alpha) := \|u - u_\alpha\|_1$, where u_α is the solution image for a specific value of α from (1). Exploiting the empirical unimodality¹ of φ on $[0, 1]$, we employ the *Fibonacci Search* method to find an optimal value for α that solves (6). This line-search strategy ensures fast convergence. For *multimodal* functions it is better to utilize the *Simulated Annealing* technique as minimization strategy, which guarantees finding a global minimum in finite time. See [1] for the implementation details.

Another important ingredient to achieve an appropriate mixture of similarity and smoothness is the determination of the support of the spatial kernels $w(\cdot)$ in both terms of (1). Moreover, we want to find out if there is a certain interrelation between both supports. In the following, we use as spatial kernels the disk-shaped hard window function of Table 2, denoted by $\mathcal{B}(\cdot)$ with supporting radius θ , i. e. with diametrical support $(2\theta + 1)$.

Let us first consider the case of Gaussian noise. As suggested by the statistical framework, we select the penaliser $\Psi_D(s^2) = s^2$ for the data term, and the function $\Psi_S(s^2) = |s|^\nu$ for the smoothness term, focusing on $\nu \in \{1, 2\}$. Thus, our designed nonlocal filter for Gaussian noise reads

$$E(u) = \alpha \sum_{i \in \Omega, j \in \mathcal{B}(i)} |u_i - f_j|^2 + (1 - \alpha) \sum_{i \in \Omega, j \in \mathcal{B}(i)} |u_i - u_j|^\nu. \quad (7)$$

Let us apply this model to reconstruct the Gaussian noise signal depicted in *top left* of Fig. 1. All model parameters were optimised and the best five results for both values of ν are shown in the first two sections of the left-hand side of Table 4. In the third section we present the performance of those filters from Table 3(d), also optimising their parameters. Mean and median stand as representatives of M-smoothers, and four instantiations of Ψ_S for regularisation were included. Without exceptions, our designed model outperforms all the well known filters obtained as particular cases of the unifying NDS filtering framework.

Now, if we have data contaminated by impulse noise, for instance salt-and-pepper noise, we just need to modify from our previous model the penaliser in the data term by $\Psi_D(s^2) = |s|$, as it is proposed in the Bayesian framework. Our new model reads

$$E(u) = \alpha \sum_{i \in \Omega, j \in \mathcal{B}(i)} |u_i - f_j| + (1 - \alpha) \sum_{i \in \Omega, j \in \mathcal{B}(i)} |u_i - u_j|^\nu. \quad (8)$$

We use this filter for denoising the salt-and-pepper noise signal plotted on *top right* of Fig. 1. We conduct the same comparative analysis as before, and conclude again that our model beats all the other particular filters.

Exploiting the NDS image denoising framework for $\alpha \notin \{0, 1\}$ and nonlocal window functions leads to hybrid methods for image simplification. It is important to mention that we have deliberately chosen penalisers that do not require any gradient threshold parameter. This keeps our model simple and efficient.

¹ Even though we can guarantee neither the continuous dependence of φ with respect to α nor a unique solution, we have observed this behavior in most of our experiments.

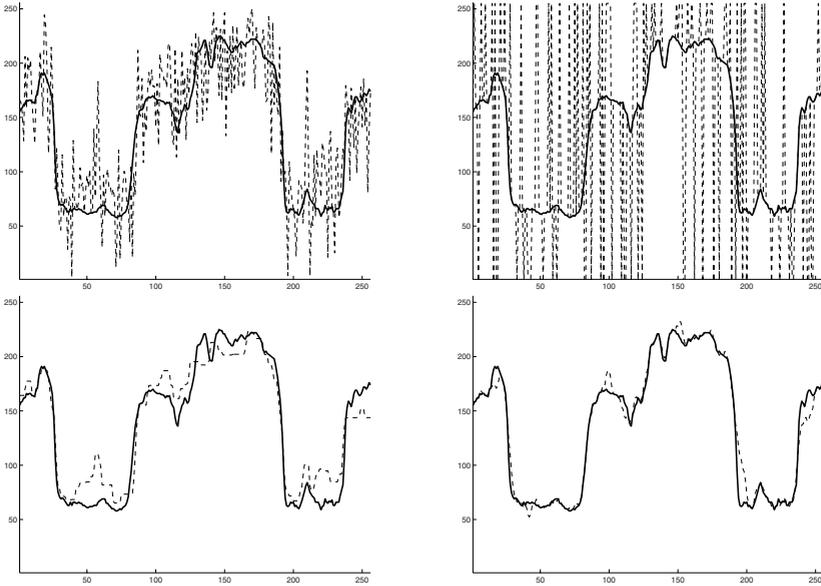


Fig. 1. Examples of signal denoising using the NDS function. Original signal in solid line. *Top left:* Noisy signal perturbed by zero-mean Gaussian noise with $\sigma = 40$ in dashed line, $\ell^1 = 27.30$. *Bottom left:* Recovered version of the Gaussian noise signal in dashed line, $\ell^1 = 13.83$. *Top right:* Noisy signal perturbed by 40% of a zero-mean salt-and-pepper noise in dashed line, $\ell^1 = 48.04$. *Bottom right:* Recovered version of the salt-and-pepper noise signal in dashed line, $\ell^1 = 4.61$.

4.2 Smoothing Effects

Trade-off Between α and the Radius of w_D . If we consider a functional which only consists of a data term, we notice that increasing the support of the spatial window leads to smoothing. On the other hand, if we leave the spatial window of the data term small and add a smoothness term, this has visually almost the same effect. In this experiment we want to quantify the difference more accurately and search for α corresponding to a certain support. To this end we consider the two following functions. The first function

$$E_D(u) = \sum_{i,j \in \Omega} (u_i - f_j)^2 w_D(|x_i - x_j|^2) \tag{9}$$

consists only of a data term, but allows for a larger window given by the dischaped hard window function w_D with supporting radius θ_D . The second function has a local data term and a smoothness term which only takes the direct neighbours $\mathcal{N}(i)$ of pixel i into consideration

$$E_C(u) = \alpha \sum_{i \in \Omega} (u_i - f_i)^2 + (1 - \alpha) \sum_{i \in \Omega, j \in \mathcal{N}(i)} (u_i - u_j)^2 \tag{10}$$

Table 4. Numerical comparison of different filters. *Left:* Denoising results of the Gaussian noise signal of Fig. 1. *Right:* Denoising results of the salt-and-pepper noise signal of Fig. 1. The best results are written in bold letters and plotted in Fig. 1.

Filter	θ_D	θ_S	α	ℓ^1	Filter	θ_D	θ_S	α	ℓ^1
model (7), $\nu = 2$	3	1	0.223	14.18	model (8), $\nu = 2$	0	1	0.903	4.61
	3	2	0.572	14.24		3	1	0.810	4.67
	2	1	0.208	14.26		3	2	0.936	4.80
	2	2	0.542	14.30		4	1	0.793	4.90
	3	3	0.793	14.32		2	1	0.757	4.91
model (7), $\nu = 1$	2	2	0.072	13.83	model (8), $\nu = 1$	3	8	0.895	5.20
	2	3	0.133	13.83		3	9	0.910	5.27
	2	4	0.178	13.85		3	10	0.921	5.38
	2	5	0.223	13.93		4	9	0.918	5.59
	3	2	0.098	14.00		3	7	0.892	5.51
mean	4	-	1.000	14.93	mean	6	-	1.000	23.95
median	4	-	1.000	14.90	median	6	-	1.000	6.98
Tikhonov	0	1	0.329	14.57	Tikhonov	0	1	0.096	23.22
TV	0	1	0.001	15.62	TV	0	1	0.001	35.04
Perona-Malik 1	0	1	0.298	14.47	Perona-Malik 1	0	1	0.095	23.21
Charbonnier	0	1	0.314	14.53	Charbonnier	0	1	0.096	23.21

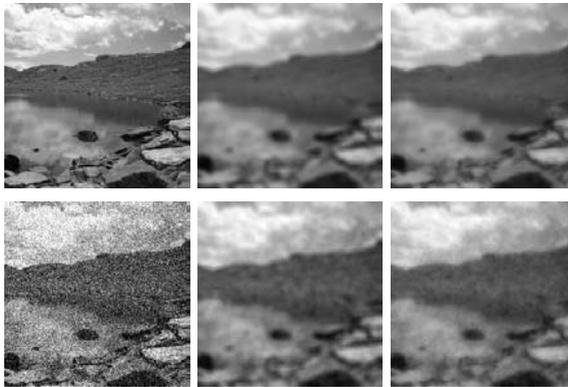


Fig. 2. Example of the trade-off between different parameters. *Top left:* Original image (256 × 256 pixels). *Top middle:* Smoothed version with E_D , radius $\theta_D = 5$. *Top right:* Smoothing with E_C , $\alpha = 0.503$. *Bottom left:* Image with additive Gaussian noise, standard deviation $\sigma = 50$. *Bottom middle:* Denoising with E_D , radius $\theta_D = 5$. *Bottom right:* Denoising with E_C , $\alpha = 0.424$.

Here, only changing the value α is used to steer the amount of smoothness. Fig. 2 shows two examples of the trade-off between these parameters. We see that in both cases with and without noise it is possible to obtain very similar results with both functionals. We are interested in knowing how far away the results

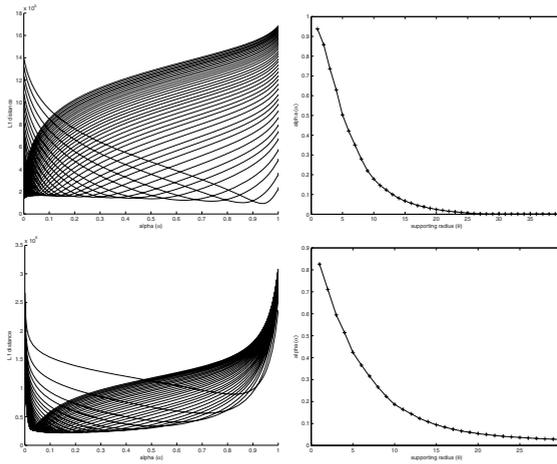


Fig. 3. Trade-off between large kernels and small α . *Top left:* ℓ^1 -error between E_D and E_C depending on α , for noise-free image. *Top right:* Optimal value of α depending on the radial support θ_D . *Bottom:* Same plots for the noisy input image.

obtained with the functions E_D and E_C are from each other, and how they approach each other by means of tuning θ_D and α , respectively. In Fig. 3, we display some measurements to quantify the trade-off between these parameters. In the left column, each curve stands for a certain radius size, and there is one value of α that minimizes the ℓ^1 distance between their estimates. The minimum distance achieved for every pair (θ_D, α) is displayed in the right column.

Trade-off Between α and the Radius of w_S . Similarly to the previous experiment, we want now to quantify the trade-off between decreasing the value of α and increasing the support of the spatial window in the smoothness term; both procedures lead to smoothing. Let us assume a level $\sigma = 20$ of Gaussian noise. From our experiments in Section 4.1, we know that the following function produces satisfactory results under Gaussian perturbations:

$$E(u) = \alpha \sum_{i \in \Omega, j \in \mathcal{B}(i)} |u_i - f_j|^2 + (1 - \alpha) \sum_{i \in \Omega, j \in \mathcal{B}(i)} |u_i - u_j|. \quad (11)$$

Keeping fixed $\theta_D = 1$, Fig. 4 shows the ℓ^1 distance between the original and the restored images for α ranging in $[0, 1]$ and different radial support θ_S . One can see that slightly better results are attained when both α and θ_S are small. Fig. 5 shows an example where similar restoration quality is achieved under different parameterization.

These experiments show that it is possible to interchange certain smoothing approaches by some others. This is important when one is searching for fast and efficient denoising algorithms. By using the NDS framework one does not have necessarily to give up quality for speed of computation.

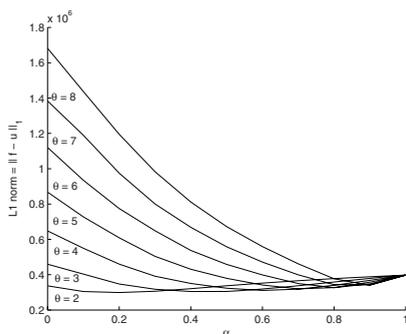


Fig. 4. Denoising properties of the functional (11). The plot outlines the ℓ^1 distance between the original and the denoised images for different values of the smoothing parameters α and θ_S . Original image in Fig. 5 *left*, and noisy image in Fig. 5 *middle left*.



Fig. 5. Denoising properties of the functional (11). *Left*: original image. *Middle left*: disturbed image with Gaussian noise $\sigma = 20$, $\ell^1 = 16.02$. *Middle right*: restored image with $\alpha = 0.2$, $\theta_S = 2$, $\ell^1 = 4.88$. *Right*: restored image with $\alpha = 0.8$, $\theta_S = 6$, $\ell^1 = 5.18$.

5 Conclusions

We have shown the capabilities of the NDS framework as unifying filtering approach. We saw that excellent results can be obtained when terms that reward fidelity to the observations and penalise smoothness in the solution are non-locally combined. By tuning its different degrees of freedom it is possible to design hybrid filters that outperform the performance of classical filters. The NDS functional possesses such a versatility that it is even possible to attain very similar results by tuning the parameters with different criteria and directions, what is particularly useful in looking for alternative ways to solve a denoising problem.

Acknowledgements. We gratefully acknowledge partly funding by *Deutscher Akademischer Austauschdienst (DAAD)* and by the priority programme SPP1114 of the *Deutsche Forschungsgemeinschaft (DFG)*, project WE 2602/2-3.

References

1. Bard, J.: Practical Bilevel Optimization. Algorithms and Applications. Springer, Heidelberg (1999)
2. Besag, J.: Toward Bayesian image analysis. *Journal of Applied Statistics* 16(3), 395–407 (1989)
3. Bouman, C., Sauer, K.: A generalized gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing* 2(3), 296–310 (1993)
4. Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proc. IEEE International Conference on Image Processing (ICIP-94, Austin, November. 13-16, 1994), vol.2, pp. 168–172 (1994)
5. Chu, C.K., Glad, I.K., Godtlielsen, F., Marron, J.S.: Edge-preserving smoothers for image processing. *Journal of the American Statistical Association* 93(442), 526–541 (1998)
6. Didas, S., Mrazek, P., Weickert, J.: Energy-based image simplification with non-local data and smoothness terms. In: Iske, A., Levesley, J. (eds.) *Algorithms for Approximation*, pp. 51–60. Springer, Heidelberg (2006)
7. Herman, G.T., Hurwitz, H., Lent, A., Lung, H-P.: On the Bayesian Approach to Image Reconstruction. *Information and Control* 42, 60–71 (1979)
8. Koenderink, J.J., Van Doorn, A.L.: The structure of locally orderless images. *International Journal of Computer Vision* 31(2/3), 159–168 (1999)
9. Künsch, H.R.: Robust priors for smoothing and image restoration. *Annals of the Institute of Statistical Mathematics* 46, 1–19 (1994)
10. Mrázek, P., Weickert, J., Bruhn, A.: On robust estimation and smoothing with spatial and tonal kernels. In: Klette, R., Kozera, R., Noakes, L., Weickert, J. (eds.) *Geometric Properties for Incomplete Data*, November 2005. *Computational Imaging and Vision*, vol. 31, Springer, Heidelberg (2005)
11. Mumford, D., Shah, J.: Optimal approximation of piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics* 42, 577–685 (1989)
12. Nikolova, M.: Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM Journal on Numerical Analysis* 40(3), 965–994 (2002)
13. Nikolova, M.: Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *Multiscale Modeling & Simulation* 4(3), 960–991 (2005)
14. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 629–639 (1990)
15. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D*. 60, 259–268 (1992)
16. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady* 4(2), 1035–1038 (1963)
17. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and colour images. In: Proc. of the 1998 IEEE International Conference on Computer Vision, Bombay, India, January 1998. pp. 839–846, Narosa Publishing House (1998)
18. Winkler, G.: *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Springer, Heidelberg (2006)