Nonlocal Total Variation for Image Denoising

Haijuan Hu, Jacques Froment

Nonlocal Total Variation for Image Denoising

Haijuan Hu
UMR 6205, Laboratoire de Mathématiques
de Bretagne Atlantique, Université de Bretagne-Sud
Campus de Tohannic, 56017 Vannes
Université Européenne de Bretagne, France
Email: haijuan.hu@univ-ubs.fr

Jacques Froment
UMR 6205, Laboratoire de Mathématiques
de Bretagne Atlantique, Université de Bretagne-Sud
Campus de Tohannic, 56017 Vannes
Université Européenne de Bretagne, France
Email: Jacques.Froment@univ-ubs.fr

***Abstract*—A nonlocal total variation (NLTV) scheme for image debluring has already been proposed in the literature. The goal of the present article is to study this scheme in the context of image denoising. We establish that its performance is comparable to non-local means and better than the classical total variation denoising approach. However, we show that the nonlocal total variation scheme is essentially a neighborhood filter and therefore a local one. In order to obtain a truly nonlocal scheme and so as to use redundancy in the whole image, we propose a new energy functional that includes a Fourier term. We call this new scheme *spatial-frequency domain nonlocal total variation* (SFNLTV). Experiments show that SFNLTV outperforms in most cases non-local means and NLTV algorithms, both in term of Euclidean criteria (PSNR) and visually.**

***Index Terms*—Gaussian noise, denoising, total variation, non-local variation, variational models, non-local means, Fourier transform.**

I. Introduction

In this work, we consider recovering images degraded by additive Gaussian noise. There is a large literature for removing such noise. See for example the approaches based on wavelet transformations [1], [2], the total variational (ROF) model [3], neighborhood filters (such as Yaroslavsky filter [4] and bilateral filter [5], [6]), non-local means [7] and its variant [8] and combinations of them [9], [10], [11]. We focus on nonlocal operators [9], [12], [13], [14], which combine the idea of variational models and patch-based methods in non-local means (NL-means).

In this paper, we first study the nonlocal total variation (NLTV) model introduced in [13], [14]. We establish a relation between NLTV and neighborhood filters which shows that NLTV is essentially a local denoising scheme and is therefore unable to use redundancies in the whole image to further decrease the noise variance. We experimentaly find that in the case of image denoising, NLTV improves the classical ROF [3] and is as good as NL-means [7] for non-textured images. In order to get a truly nonlocal denoising scheme and to better recover oscillating patterns such as microtextures, we propose a new variational model by adding to NLTV a term based on the Fourier transform. We call this new scheme *spatial-frequency domain nonlocal total variation* (SFNLTV). Our SFNLTV improves NLTV both visually and in terms of PSNR. In addition, the choice of parameters is less sensitive to images and therefore, we use the same choice of parameters for all tested images.

II. Related Works

As usual, a digital image is denoted by a $M \times N$ matrix $u = \{u(i) : i \in I\}$, where $I = \{0, 1, ..., M-1\} \times \{0, 1, ..., N-1\}$, and $0 \leq u(i) \leq 255$. The additive Gaussian noise model is:

$$\mathcal{V}(i) = \mathcal{U}(i) + \eta(i),$$

where $\mathcal{U}$ and $\mathcal{V}$ are the original and noisy images respectively, and η is the Gaussian noise: $\eta(i)$ are independent and identically distributed Gaussian random variables with mean 0 and standard deviation $\sigma > 0$. A denoised image is denoted as $\bar{\mathcal{V}}$. For simplicity, we assume symmetric boundary conditions in this paper.

A. Total Variation Model

In [3], Rudin, Osher and Fatemi (ROF) introduced the total variation (TV) as a regularizing functional for image denoising. The denoised image $\bar{\mathcal{V}}$ is the minimizer of the following functional:

$$E(u) := \lambda \sum_{i \in I} \|\nabla u(i)\| + \frac{1}{2} \sum_{i \in I} (u(i) - \mathcal{V}(i))^2, \qquad (1)$$

where $\nabla u(i) = (u(i_1 + 1, i_2) - u(i_1, i_2), u(i_1, i_2 + 1) - u(i_1, i_2))$ is the gradient of u at point $i = (i_1, i_2)$, $\|\|$ is the Euclidean norm, and λ is the Lagrange multiplier. $\sum_{i \in I} \|\nabla u(i)\|$ is called the total variation of u.

B. Non-Local Means

For $i \in I$ and d an odd integer, let $\mathcal{N}_i(d) = \{j \in I : \|j - i\|_\infty \leq (d-1)/2\}$ be the window with center i and size $d \times d$, simply written as $\mathcal{N}_i$. Similarly, denote $U_i(D)$ the window with center i and size $D \times D$, simply written as U_i.

The denoised image by NL-means [7] is given by

$$\bar{\mathcal{V}}(i) = \frac{\sum_{j \in U_i(D)} w(i,j) \mathcal{V}(j)}{\sum_{j \in U_i(D)} w(i,j)}$$

with

$$w(i,j) = e^{-||\mathcal{V}(\mathcal{N}_i) - \mathcal{V}(\mathcal{N}_j)||_a^2 / (2\sigma_r^2)}, \qquad (2)$$

where $\sigma_r > 0$ is a control parameter,

$$||\mathcal{V}(\mathcal{N}_i) - \mathcal{V}(\mathcal{N}_j)||_a^2 = \frac{\sum_{k \in \mathcal{N}_i(d)} a(i,k) |\mathcal{V}(k) - \mathcal{V}(\mathcal{T}(k))|^2}{\sum_{k \in \mathcal{N}_i(d)} a(i,k)}, \qquad (3)$$

$\mathcal{T} = \mathcal{T}_{ij}$ being the translation mapping of $\mathcal{N}_i$ onto $\mathcal{N}_j$: $\mathcal{T}(k) = k - i + j, k \in \mathcal{N}_i$, and $a(i,k) = e^{-\|i-k\|^2/2a^2}$ ($a = (d-1)/4$ is a good choice).

C. Nonlocal Total Variation Model

In [13], [14] the authors considered a nonlocal total variation functional that replaces the total variation one in ROF (1):

$$E(u) := \lambda J(u) + \frac{1}{2}\sum_{i\in I}(u(i) - \mathcal{V}(i))^2, \tag{4}$$

where

$$J(u) = \sum_{i\in I}\sqrt{\sum_{j\in I}(u(i) - u(j))^2 w(i,j)} \tag{5}$$

is called the nonlocal total variation of u. The variational model (4) is denoted as NLTV. We consider $w(i,j)$ used in [14], where $w(i,j)$ is taken as the one used in NL-means (3) without normalization,

$$w(i,j) = \begin{cases} e^{-||\mathcal{V}(\mathcal{N}_i) - \mathcal{V}(\mathcal{N}_j)||_a^2/(2\sigma_r^2)} & \text{if} \quad j \in U_i(D) \\ 0 & \text{else} \end{cases}. \tag{6}$$

The weight $w(i,j)$ is used to estimate the similarity between two pixels i and j. When the two pixels i and j are similar, $w(i,j)$ is large, which makes the recovered values $\bar{\mathcal{V}}(i)$ and $\bar{\mathcal{V}}(j)$ close. Note that for $i = (i_1, i_2)$, if we use

$$w(i,j) = \begin{cases} 1 & \text{if} \quad j \in \{(i_1+1, i_2), (i_1, i_2+1)\} \\ 0 & \text{else} \end{cases},$$

then (4) reduces to (1). Thus NLTV is more general and more adaptive to image content than ROF.

For any fixed i, $|\nabla_w u(i)| := \sqrt{\sum_{j\in I}(u(i) - u(j))^2 w(i,j)}$ is a convex functional of u. Therefore $J(u)$ is convex. Since the fidelity term is strictly convex, so is the energy function $E(u)$. Thus we can use the gradient descent method to find the minimizer. Write

$$W(i,j) = \frac{w(i,j)}{|\nabla_w u(i)|} + \frac{w(j,i)}{|\nabla_w u(j)|}.$$

The gradient of $E(u)$ is $\nabla_u E(u) = \{\nabla_u E(u)(i)\}_{i\in I}$, with

$$\nabla_u E(u)(i) = \lambda\sum_{y\in I}(u(i) - u(j))W(i,j) + (u(i) - \mathcal{V}(i)). \tag{7}$$

Notice that by (6), the summation of $y \in I$ in (7) can be replaced by $y \in U_i^0 = U_i \backslash \{i\}$.

III. Relation with Neighborhood Filters

We consider the gradient descent algorithm

$$u^{k+1}(i) = u^k(i) - t_k \nabla_u E(u^k)(i), \quad k = 0, 1, 2, \cdots \tag{8}$$

with $u^0 = \mathcal{V}$, $t_k > 0$ such that $E(u^{k+1}) < E(u^k)$.

By (8) and (7), we have

$$u^{k+1}(i) = \sum_{j\in U_i} u^k(j) W^k(i,j) + t_k \mathcal{V}(i), \tag{9}$$

where

$$W^k(i,j) = \begin{cases} t_k \lambda W(i,j) & \text{if} \quad i \neq j \\ 1 - t_k - t_k\lambda\sum_{j\in U_i^0} W^k(i,j) & \text{if} \quad i = j. \end{cases} \tag{10}$$

Note that

$$\sum_{j\in U_i} W^k(i,j) + t_k = 1.$$

Let $U_i = U_i(D)$. We can easily obtain that $u^k(i)$ is the weighted average of its neighbors $\mathcal{V}(j), j \in U_i(Dk - k + 1)$. In our experiments, we find that after 15 iterations, the iterative process for all of our tested image approximately converges (See Section IV for the tested images). Thus and despite the name "nonlocal", NLTV should be considered as a local neighborhood filter.

IV. Simulations

We use the Peak Signal-to-Noise Ratio (PSNR) to measure the quality of restored images:

$$\text{PSNR } (\bar{\mathcal{V}}) = 10\log_{10}\frac{255^2 MN}{\sum_{i\in I}(\bar{\mathcal{V}}(i) - \mathcal{U}(i))^2}.$$

Recall that $\mathcal{U}$ and $\bar{\mathcal{V}}$ are the original image and the restored one respectively. The larger the value of PSNR, the better the restored image. In our experiments we mainly use (512×512) images, Lena, Barbara, Bridge, Boats and (256×256) images, Peppers, House, Cameraman. They are all available online[1]. In the experiments, we consider the case $\sigma = 20$.

A. Choices of Parameters for NLTV Model

We search for the optimal choice of parameters for each image in terms of PSNR value with $\sigma_r = 20$ fixed and other parameters varying in some ranges. We find that except the Bridge image, one can use the same choice of parameters $D = 3, d = 9, \lambda = 15$. The PSNR value of each image is close to the best one (that we have tested) with a difference less than 0.1. So the optimal parameters are not sensitive to general images. Since the Bridge image has many irregular fine details, the optimal parameters are a little different: $D = 3, d = 15, \lambda = 11$. Let us give some remarks about the choices of the parameters.

- Note that our choice for D is 3, the smallest square window. When D increases, restored images look smoother globally and many fine details disappear, while in the areas where there are great changes of gray values, there is more noise left. Though some regular textures are a little better recovered and the staircase artifact is less obvious for larger D, the global quality is deteriorated. See the denoised images with $D = 3$ and 19 in Fig.1. Thus the proposed choice $D = 21$ in [14] in the context of image deconvolution is not a good choice for image denoising. As our choice D is small, the denoising process is fast.

[1]Lena, Peppers, Boats and House:
http://decsai.ugr.es/∼javier/denoise/test_images/index.htm
Bridge: www.math.cuhk.edu.hk/∼rchan/paper/dcx/
Barbara, Cameraman: www.dcs.qmul.ac.uk/∼phao/CIP/Images/.

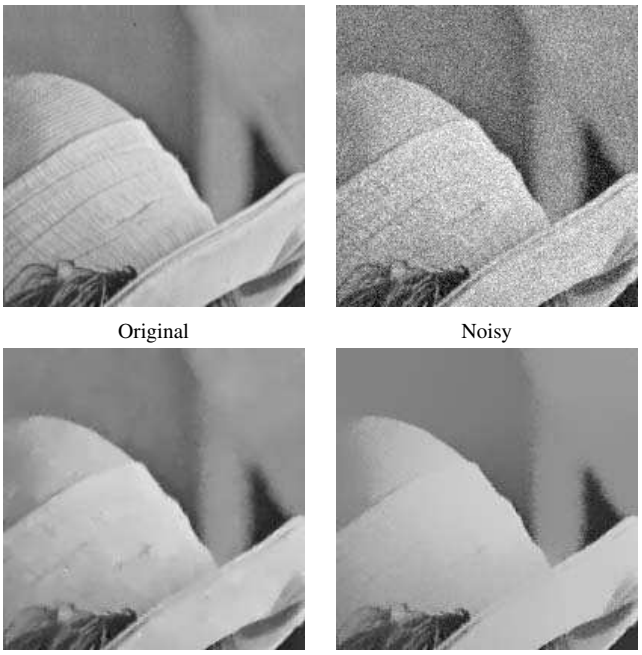

Fig. 1. Denoised image with NLTV for Lena (upper right part).

• The influence of d is less obvious than D. In fact, as the patch size d increases, all the nonzero weights $w(i,j)$ tend to be identical, thus NLTV is close to ROF, if $D = 3$.

• With other parameters fixed, the optimal choice of λ depends on the nonlocal variation (thus depends on the other parameters d, D, h) and the noise level.

B. Comparisons with ROF Model and NL-means

We now compare NLTV with ROF and NL-means. The PSNR values are shown in Table I and the denoised images are shown in Fig.2. For NL-means, we use $D = 5, d = 3, \sigma_r = 24$ for the Bridge image, and $D = 11, d = 7, \sigma_r = 18$ for other images. From the comparisons, we can see that:

• in terms of PSNR, NLTV is better than ROF for all the images, and is similar to NL-means except the Barbara image;

• NLTV is better than NL-means for isolated pixels (for which there is few similar patches). In fact, by (9) and (10), when i is an isolated point, there are no or few similar points in its neighbors, so $\sum_{j\in U_i^0} W(i,j)$ is small. Thus the recovered value is close to the noisy one. Compared to ROF, the denoised images with NLTV have less staircase artifacts, and have less noise left.

TABLE I

PSNR VALUES FOR DIFFERENT IMAGES WITH NLTV, NL-MEANS (NL), ROF, AND SFNLTV (SF). THE VALUES MARKED WITH * ARE OBTAINED BY OPTIMAL PARAMETERS FOR THE CORRESPONDING IMAGES. FOR ROF, WE USE THE OPTIMAL PARAMETER FOR EACH IMAGE. IN THE FIRST ROW, THE IMAGES ARE NAMED BY THEIR FIRST THREE OR FOUR LETTERS.

Image	Lena	Bar	Pep	Boat	Bri	Hou	Cam
NLTV	31.56	28.46	30.21	29.49	26.81*	31.74	29.45
NL	31.56	<u>29.68</u>	30.18	29.32	26.81*	31.92	29.35
ROF*	31.12	27.10	29.56	29.20	26.68	31.31	28.82
SF	<u>31.77</u>	29.19	<u>30.29</u>	<u>29.89</u>	<u>26.92</u>	<u>32.14</u>	<u>29.64</u>

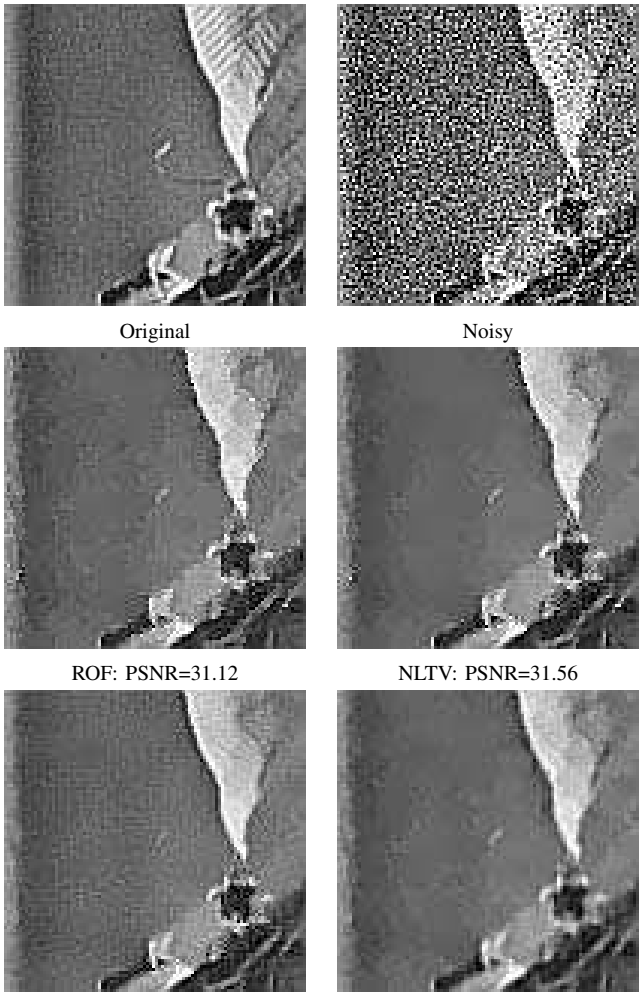

Fig. 2. Zoom in the left area of the Lena's hat. In order to enhance visual artifacts, the same sharpen filter has been applied to all displayed images.

V. IMPROVEMENTS OF NLTV MODEL

A. Spatial-Frequency Domain Nonlocal Total Variation Model

In order to better recover oscillating patterns in images such as microtextures, and to get a truly nonlocal denoising scheme, we propose to add in NLTV a term based on the Fourier transform. Denote $\hat{u}$ the Fourier transform of u. Let

$$J_1(u) = \sum_{i\in I} |\nabla_{w_1} u(i)|, \quad \text{and} \quad J_2(u) = \sum_{\omega\in I} |\nabla_{w_2} \hat{u}(\omega)|,$$

with

$$|\nabla_{w_1} u(i)| = \sqrt{\sum_{j\in I} (u(i) - u(j))^2 w_1(i,j)}, \tag{11}$$

and

$$|\nabla_{w_2} \hat{u}(\omega)| = \sqrt{\sum_{\xi\in I} |\hat{u}(\omega) - \hat{u}(\xi)|^2 w_2(\omega,\xi)}, \tag{12}$$

where $w_1(i,j)$ is taken as (6) in NLTV, and $w_2(\omega,\xi)$ is taken as (6) with $\hat{\mathcal{V}}$, d_f, D_f, σ_{rf} replacing $\mathcal{V}$, d, D, σ_r respectively.

We consider the following energy functional called *spatial-frequency domain nonlocal total variation* (SFNLTV) model:

$$E(u) := \lambda J_1(u) + \mu J_2(u) + \frac{1}{2}\sum_{i\in I}(u(i) - \mathcal{V}(i))^2. \quad (13)$$

Because Fourier transform is linear, the functional in (13) is also strictly convex. So we can also use the gradient descent method (9) to minimize it.

Similar to (7),

$$\nabla_u J_1(u)(i) = \sum_{j\in I}(u(i) - u(j))\left(\frac{w_1(i,j)}{|\nabla_{w_1}u(i)|} + \frac{w_1(j,i)}{|\nabla_{w_1}u(j)|}\right). \quad (14)$$

We will formally prove that $\nabla_u J_2(u)(i) = \Re f(i)$, with

$$\hat{f}(w) = \sum_{\xi\in I}(\hat{u}(\omega) - \hat{u}(\xi))\left(\frac{w_2(\omega,\xi)}{|\nabla_{w_2}\hat{u}(\omega)|} + \frac{w_2(\xi,\omega)}{|\nabla_{w_2}\hat{u}(\xi)|}\right) \quad (15)$$

in the following, where $\Re$ denotes the real part of a complex number. Firstly,

$$\begin{aligned} &\frac{d}{dt}J_2(u+tv)|_{t=0} \quad (16)\\ =\ &\sum_{\omega\in I}\sum_{\xi\in I}[(\hat{u}(\omega) - \hat{u}(\xi))(\bar{\hat{v}}(\omega) - \bar{\hat{v}}(\xi))\\ &+(\hat{v}(\omega) - \hat{v}(\xi))(\bar{\hat{u}}(\omega) - \bar{\hat{u}}(\xi))w_2(\omega,\xi)\frac{1}{|\nabla_{w_2}\hat{u}(\omega)|}]. \end{aligned}$$

By exchanging ω and ξ and changing the order of summation, we obtain

$$\begin{aligned} &\sum_{\omega\in I}\sum_{\xi\in I}(\hat{u}(\omega) - \hat{u}(\xi))(\bar{\hat{v}}(\omega) - \bar{\hat{v}}(\xi))w_2(\omega,\xi)\frac{1}{|\nabla_{w_2}\hat{u}(\omega)|}\\ =\ &\sum_{\omega\in I}\frac{1}{2}\hat{f}(\omega)\hat{v}(\omega). \quad (17) \end{aligned}$$

Then by (16) and (17), we get

$$\begin{aligned} &\frac{d}{dt}J_2(u+tv)|_{t=0}\\ =\ &\sum_{\omega\in I}\frac{1}{2}\hat{f}(\omega)\hat{v}(\omega) + \overline{\sum_{\omega\in I}\frac{1}{2}\hat{f}(\omega)\hat{v}(\omega)}\\ =\ &\sum_{i\in I}\frac{1}{2}f(i)v(i) + \sum_{i\in I}\frac{1}{2}\bar{f}(i)v(i)\\ =\ &\sum_{i\in I}\Re f(i)v(i). \end{aligned}$$

Thus, it follows that $\nabla_u J_2(u)(i) = \Re f(i)$.

B. Simulations and Conclusions

Note that we have now eight parameters λ, d, D, σ_r, μ, d_f, D_f, σ_{rf}. We use the same choices of parameters in $w_1(i,j)$ as $w(i,j)$ in NLTV for general natural images, i.e. $d = 9$, $D = 3$, $\sigma_r = 20$, and test the other parameters. We find that $\lambda = 11$, $\mu = 2$, $d_f = 9$, $D_f = 5$, $\sigma_{rf} = 16$ is a good choice for all our testing images. With this choice, SFNLTV is better than NLTV for all the images. Therefore, despite SFNLTV needs more parameters than NLTV, the choice of parameters for SFNLTV is less sensitive to images. We show PSNR values in Table I. SFNLTV is the best for almost all tested images. Carefully comparing the denoised images, we can see that SFNLTV is better than NLTV for fine details, but a little noisier in homogeneous regions (see Fig.2).

Let P denote the matrix composed of the basis of $\mathbb{R}^{MN}$ or $\mathbb{C}^{MN}$. Recall that the hard and soft thresholding methods [1] can be considered as the solutions [15] of

$$\bar{\mathcal{V}} = \arg\min_u\{\|u - \mathcal{V}\|^2 + T^2\|P^T u\|_0\}, \quad (18)$$

and

$$\bar{\mathcal{V}} = \arg\min_u\{\|u - \mathcal{V}\|^2 + 2T\|P^T u\|_1\}, \quad (19)$$

where T is the threshold. Therefore $J_2(u)$ is a variant of the l^0 or l^1 norm in the thresholding methods (18) and (19) if we use the Fourier basis as P. Thus minimizing in both spatial and frequency domain may improve the performance of energy minimization algorithms. We declare that such mixed approaches are very promising.

ACKNOWLEDGMENT

The authors would like to thank Xiaoqun ZHANG for the code of NLTV.

REFERENCES

[1] D. Donoho and J. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[2] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, pp. 1532–1546, 2000.

[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[4] L. P. Yaroslavsky, *Digital Picture Processing*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1985.

[5] S. Smith and J. Brady, "SUSAN—A new approach to low level image processing," *Int. J. Com. Vis.*, vol. 23, no. 1, pp. 45–78, 1997.

[6] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision*. IEEE, 1998, pp. 839–846.

[7] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[8] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, 2006.

[9] S. Kindermann, S. Osher, and P. W. Jones, "Deblurring and denoising of images by nonlocal functionals," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1091–1115, 2006.

[10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[12] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *Multiscale Model. Simul.*, vol. 6, no. 2, pp. 595–630, 2008.

[13] ——, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul*, vol. 7, no. 3, pp. 1005–1028, 2008.

[14] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via nonlocal operators," *J. Sci. Comput.*, vol. 42, no. 2, pp. 185–197, 2010.

[15] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, 2010.