SURE-Based Non-Local Means

Dimitri Van De Ville, *Member, IEEE*, and Michel Kocher

***Abstract*—Non-local means (NLM) provides a powerful framework for denoising. However, there are a few parameters of the algorithm—most notably, the width of the smoothing kernel—that are data-dependent and difficult to tune. Here, we propose to use Stein's unbiased risk estimate (SURE) to monitor the mean square error (MSE) of the NLM algorithm for restoration of an image corrupted by additive white Gaussian noise. The SURE principle allows to assess the MSE without knowledge of the noise-free signal. We derive an explicit analytical expression for SURE in the setting of NLM that can be incorporated in the implementation at low computational cost. Finally, we present experimental results that confirm the optimality of the proposed parameter selection.**

***Index Terms*—Denoising, non-local means, Stein's unbiased risk estimate.**

I. Introduction

NON-LOCAL means (NLM) is a recent denoising method [1] that has received a lot of attention from the signal processing community. While standard linear filtering relies on local spatial correlation, the non-local principle exploits the fact that similar neighborhoods can occur anywhere in the image and can contribute for denoising.

The standard NLM algorithm is computationally expensive. The initial approach by Buades *et al.* [1] proposed to limit the search region within which similar neighborhoods are looked for. Numerous methods were proposed to accelerate the NLM approach such as a preselection of the contributing neighborhoods based on average value and gradient [2], average and variance [3] or higher-order statistical moments [4], cluster tree arrangement [5], and singular value decomposition [6]. Also the computation of the distance measure between different neighborhoods can be optimized using the fast Fourier transform [7] or a moving average filter [8]. Variations of the NLM algorithm have also been proposed to improve the denoising performance; e.g., adaptive neighborhoods [9], iterative application [5], combination with kernel regression [10] and spectral analysis [11], and other similarity measures based on principal component analysis [12] or rotation invariance [13]. The most evolved version of the NLM framework is probably BM-3D [14], which further processes the selected neighborhoods and gives high quality results.

A remaining question of the NLM methodology is how to optimally set the parameters of the algorithm; i.e., the width of the smoothing kernel that serves to determine the weights, the size of the neighborhood, and the search region. In this paper, we propose to use the principle of Stein's unbiased risk estimate (SURE) [15] as an estimator for the mean squared error (MSE) from the noisy image only. SURE-based selection of optimal parameters has received a renewed interest for wavelet-based methods [16], [17]. Also, SURE can be used in a Monte-Carlo fashion when its expression is analytically not tractable [18] (e.g., for optimizing the nonquadratic regularization functional in total-variation), and it can be extended for non-Gaussian distributions [19].

In Section II, we briefly review the NLM algorithm and the SURE principle. Next, in Section III, we show that it is possible to obtain the analytical divergence of the NLM algorithm with respect to the measurements, which serves as the key ingredient of SURE. In Section IV, we present the experimental results, which are discussed in Section V.

Manuscript received April 16, 2009; revised June 23, 2009. First published July 14, 2009; current version published August 26, 2009. This work was supported by in part by the Swiss National Science Foundation (PP00P2-123438, first author) and in part by the Centre for Biomedical Imaging (CIBM). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Vicky Zhao.

D. Van De Ville is with the Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland and the Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

M. Kocher is with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne,, Switzerland.

Digital Object Identifier 10.1109/LSP.2009.2027669

II. Preliminaries

A. *The Non-Local Means Algorithm*

We consider the observation model

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^N$ stands for the vector representation of the noise-free image, $\mathbf{n}$ is the zero-mean white Gaussian noise of variance $\sigma^2\mathbf{I}$, and $\mathbf{y}$ is the observed noisy data. We denote the grayscale value of the individual pixel at position $\mathbf{l} \in \mathcal{I}$ as $y_\mathbf{l}$, where we use vector indexing to better reflect the spatial dependencies of the image. The pixel-based NLM algorithm [1] is a spatially adaptive filter that maps the measured data $\mathbf{y}$ into $\hat{\mathbf{x}}$ as follows:

$$\hat{x}_\mathbf{l} = \frac{\sum_{\mathbf{k}\in\mathcal{S}_\mathbf{l}} w_{\mathbf{k},\mathbf{l}} y_\mathbf{k}}{\sum_{\mathbf{k}\in\mathcal{S}_\mathbf{l}} w_{\mathbf{k},\mathbf{l}}} \tag{2}$$

where $\mathcal{S}_\mathbf{l}$ is the search region around $\mathbf{l}$ and $w_{\mathbf{k},\mathbf{l}}$ are the weights that compare the neighborhoods around pixels $\mathbf{k}$ and $\mathbf{l}$, respectively. The weights are defined as

$$w_{\mathbf{k},\mathbf{l}} = \exp\left(-\frac{\sum_{\mathbf{b}\in\mathcal{B}}(y_{\mathbf{k}+\mathbf{b}} - y_{\mathbf{l}+\mathbf{b}})^2}{B\,2\lambda^2}\right) \tag{3}$$

where $\mathcal{B}$ defines the neighborhood and B is its total size; e.g., $\mathcal{B} = [-3,3]\times[-3,3]$ and $B = 49$ for a 7×7 neighborhood.

B. *Mean Squared Error and Stein's Unbiased Risk Estimate*

The MSE of the denoised image with respect to its noise-free version is

$$\mathrm{MSE}(\hat{\mathbf{x}}) = \frac{1}{N}\|\mathbf{x}-\hat{\mathbf{x}}\|^2 = \frac{1}{N}\sum_{\mathbf{l}\in\mathcal{I}}(x_{\mathbf{l}}-\hat{x}_{\mathbf{l}})^2 \tag{4}$$

where $\|\cdot\|^2$ is the Euclidean norm. SURE provides a means for unbiased estimation of the true MSE. It is specified by the following analytical expression [15]–[18]:

$$\mathrm{SURE} = \frac{1}{N}\|\mathbf{y}-\hat{\mathbf{x}}\|^2 - \sigma^2 + 2\sigma^2\frac{\mathrm{div}_{\mathbf{y}}\{\hat{\mathbf{x}}\}}{N} \tag{5}$$

where $\mathrm{div}_{\mathbf{y}}\{\hat{\mathbf{x}}\}$ is the divergence of the NLM algorithm with respect to the measurements

$$\mathrm{div}_{\mathbf{y}}\{\hat{\mathbf{x}}\} = \sum_{\mathbf{l}\in\mathcal{I}}\frac{\partial \hat{x}_{\mathbf{l}}}{\partial y_{\mathbf{l}}} \tag{6}$$

which needs to be well defined in the weak sense. The derivation of SURE relies on the additive white Gaussian noise hypothesis and assumes the knowledge of the noise variance σ^2. In practice, σ^2 can be easily estimated from the measured data.

III. SURE-Based Non-Local Means

The divergence term in (5) plays a crucial role in the expression of SURE. Fortunately, for the NLM algorithm, we can explicitly obtain the divergence from (2) as shown in the following Proposition.

Proposition 1 (Divergence of NLM): The individual terms of the divergence $\mathrm{div}_{\mathbf{y}}\{\hat{\mathbf{x}}\}$ of the NLM algorithm are given by

$$\frac{\partial \hat{x}_{\mathbf{l}}}{\partial y_{\mathbf{l}}} = \frac{\widehat{x_{\mathbf{l}}^2}-\hat{x}_{\mathbf{l}}^2}{B\lambda^2} + \frac{1}{W_{\mathbf{l}}} + \frac{1}{W_{\mathbf{l}}B\lambda^2}\sum_{\mathbf{b}\in\mathcal{B}} w_{\mathbf{l}-\mathbf{b},\mathbf{l}}(y_{\mathbf{l}}-y_{\mathbf{l}+\mathbf{b}})(\hat{x}_{\mathbf{l}}-y_{\mathbf{l}-\mathbf{b}}) \tag{7}$$

with $W_{\mathbf{l}} = \sum_{\mathbf{k}\in\mathcal{S}_{\mathbf{l}}} w_{\mathbf{k},\mathbf{l}}$. Moreover, $\widehat{x_{\mathbf{l}}^2}$ is the result of the NLM algorithm that uses the squared input values

$$\widehat{x_{\mathbf{l}}^2} = \frac{\sum_{\mathbf{k}\in\mathcal{S}_{\mathbf{l}}} w_{\mathbf{k},\mathbf{l}}y_{\mathbf{k}}^2}{\sum_{\mathbf{k}\in\mathcal{S}_{\mathbf{l}}} w_{\mathbf{k},\mathbf{l}}}. \tag{8}$$

Notice that the weights $w_{\mathbf{k},\mathbf{l}}$ in this filter are identical to the ones from the regular algorithm; i.e., the same as in (2).

We see that the divergence term in (7) has a variance-like contribution (the first part) and a cross-correlation-like contribution (the last part). Plugging the divergence into the SURE expression leads to our proposed estimator.

Proposition 2 (SURE for NLM): The SURE for the NLM algorithm can be expressed as

$$\begin{aligned}\mathrm{SURE} &= \frac{1}{N}\|\mathbf{y}-\hat{\mathbf{x}}\|^2 - \sigma^2 \\ &+ \frac{2\sigma^2}{NB\lambda^2}\left(\sum_{\mathbf{l}\in\mathcal{I}}\widehat{x_{\mathbf{l}}^2} - \|\hat{\mathbf{x}}\|^2\right) + \frac{2\sigma^2}{N}\sum_{\mathbf{l}\in\mathcal{I}}\frac{1}{W_{\mathbf{l}}} \\ &+ \frac{2\sigma^2}{NB\lambda^2}\sum_{\mathbf{l}\in\mathcal{I}}\frac{1}{W_{\mathbf{l}}}\sum_{\mathbf{b}\in\mathcal{B}} w_{\mathbf{l}-\mathbf{b},\mathbf{l}}(y_{\mathbf{l}}-y_{\mathbf{l}+\mathbf{b}})(\hat{x}_{\mathbf{l}}-y_{\mathbf{l}-\mathbf{b}}).\end{aligned} \tag{9}$$

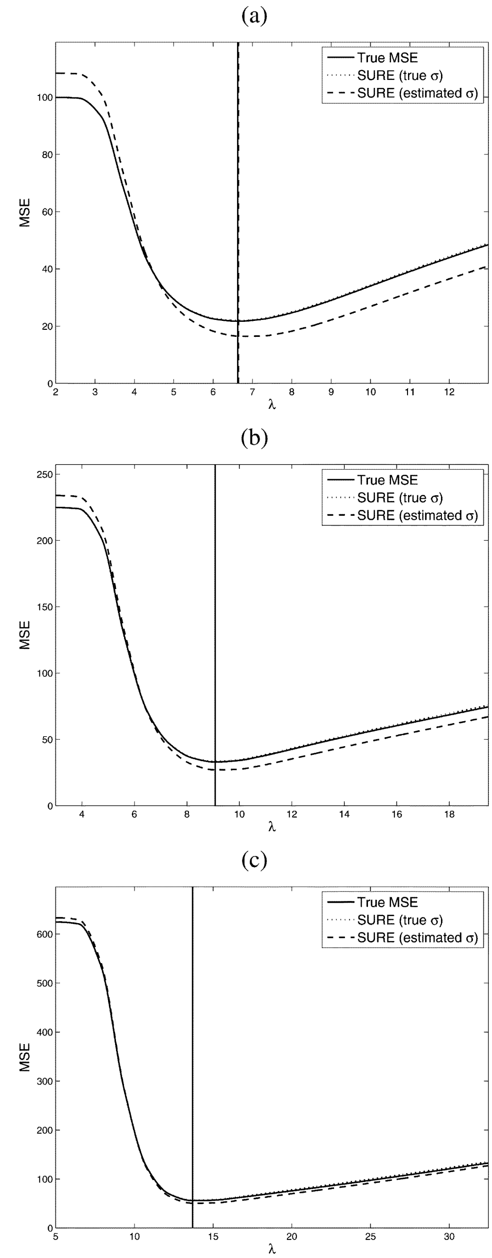

Fig. 1. Performance measures (true MSE, SURE for exact σ, SURE for estimated σ) as a function of the smoothing parameter λ. The optimal setting of λ for each measure is indicated by a vertical line in the same drawing style. Lena test image, neighborhood 7×7. (a) $\sigma = 10$. (b) $\sigma = 15$. (c) $\sigma = 25$.

The computation of the divergence term can be readily incorporated within the core of the NLM algorithm. Specifically, implementing (9) requires two additional memory arrays to store $\widehat{y_{\mathbf{l}}^2}$ and $W_{\mathbf{l}}$, and its computational complexity takes only $O(B\cdot N)$ operations, which is negligible compared to $O(B\cdot N\cdot S)$ of the NLM algorithm, where S is the total size of the search region.

IV. Experimental Results

We implemented the proposed approach in Matlab (R2008b) using C for the core calculations. For all experiments, we used a search region $\mathcal{S} = [-20,20]\times[-20,20]$. The computation time on an Apple MacBook Pro 2.53 GHz Intel Core 2 Duo

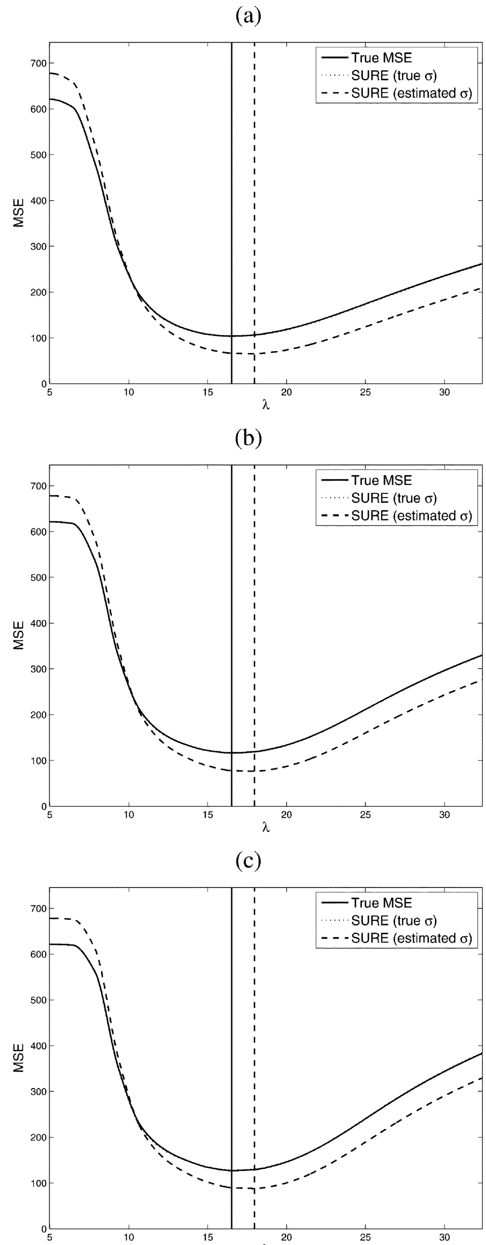

Fig. 2. Performance measures (true MSE, SURE for exact σ, SURE for estimated σ) as a function of the smoothing parameter λ. The optimal setting of λ for each measure is indicated by a vertical line in the same drawing style. Cameraman test image, $\sigma = 25$. (a) Neighborhood 5×5. (b) 7×7. (c) 9×9.

(only a single core is used) was 132 s for a 256×256 image and 7×7 neighborhood. The additional cost of SURE was completely negligible; indeed, $S = 1681$ for the chosen search region $\mathcal{S}$.

We evaluated SURE using both the true noise variance σ^2 and the estimated one; i.e., we used the well-known wavelet estimator where we filtered the image by the diagonal highpass filter of the Haar wavelet transform and computed a fraction $(1/.6745)$ of the median of absolute deviation. In Fig. 1, we plot true MSE and SURE for true σ and estimated σ, as a function of λ, for various noise levels. In Fig. 2, we show the same measures for various neighborhood sizes. Finally, in Table I, we report the peak signal-to-noise ratios (PSNR) for various test images and neighborhood sizes; we used $\mathrm{PSNR} = -10\log_{10}(\mathrm{MSE}/255^2)$.

V. Discussion and Outlook

The results in Fig. 1 and Table I confirm that the performance of the NLM algorithm depends on λ and that the MSE can be monitored using SURE. The optimal λ is data-dependent and noise-level dependent; i.e., the ratio of optimal λ versus true σ is changing for different σ. When SURE is based on the estimated σ, the result is also satisfactory—although σ is known be slightly overestimated for images with high-frequency content (e.g., barbara). Fixing $\lambda = .7\sigma$ seems to be good compromise for our test set, but it would be slightly suboptimal and also suffer from the bias in estimating σ. In practice, one useful approach is to combine our SURE with an optimization strategy to find the optimal λ in a few iterations; e.g., golden section search does fine due to the convex behavior of MSE/SURE.

Fig. 2 and Table I illustrate that the influence of the neighborhood size is minor and that the smallest size performs best. However, this result might depend on the image content and the noise level and thus monitoring with SURE is useful.

The surprising aspect of our work is that it is possible to obtain an explicit analytical form for the divergence term of NLM, which is rather exceptional for nonlinear algorithms. Moreover, the additional computational cost to calculate the SURE is extremely low. We hope that this work will inspire further improvements of the NLM algorithm and its variations.

Appendix A
Derivation of the Divergence Term

To obtain the divergence term, we introduce $W_\mathbf{l} = \sum_{\mathbf{k}\in\mathcal{S}_\mathbf{l}} w_{\mathbf{k},\mathbf{l}}$ and we derive $\hat{x}_\mathbf{l}$ with respect to $y_\mathbf{l}$, which results in

$$\frac{\partial \hat{x}_\mathbf{l}}{\partial y_\mathbf{l}} = \frac{W_\mathbf{l}\frac{\partial}{\partial y_\mathbf{l}}(W_\mathbf{l}\hat{x}_\mathbf{l}) - (W_\mathbf{l}\hat{x}_\mathbf{l})\frac{\partial W_\mathbf{l}}{\partial y_\mathbf{l}}}{W_\mathbf{l}^2} = \frac{1}{W_\mathbf{l}}\left(\sum_{\mathbf{k}\in\mathcal{S}}\frac{\partial w_{\mathbf{k},\mathbf{l}}}{\partial y_\mathbf{l}}y_\mathbf{k} + \underbrace{w_{\mathbf{l},\mathbf{l}}}_{=1} - \hat{x}_\mathbf{l}\sum_{\mathbf{k}\in\mathcal{S}}\frac{\partial w_{\mathbf{k},\mathbf{l}}}{\partial y_\mathbf{l}}\right). \tag{10}$$

Further on, deriving the weights gives

$$\frac{\partial w_{\mathbf{k},\mathbf{l}}}{\partial y_\mathbf{l}} = \begin{cases} w_{\mathbf{k},\mathbf{l}}\left(\frac{y_\mathbf{k}-y_\mathbf{l}}{B\lambda^2} + \frac{y_{2\mathbf{l}-\mathbf{k}}-y_\mathbf{l}}{B\lambda^2}\right), & \mathbf{l}-\mathbf{k}\in\mathcal{B} \\ w_{\mathbf{k},\mathbf{l}}\left(\frac{y_\mathbf{k}-y_\mathbf{l}}{B\lambda^2}\right), & \mathbf{l}-\mathbf{k}\notin\mathcal{B} \end{cases}$$

which allows us to further manipulate (10) into

$$\frac{1}{W_\mathbf{l}}\left(\sum_{\mathbf{k}\in\mathcal{S}} w_{\mathbf{k},\mathbf{l}}\frac{(y_\mathbf{k}-y_\mathbf{l})}{B\lambda^2}y_\mathbf{k} + \sum_{\mathbf{b}\in\mathcal{B}} w_{\mathbf{l}-\mathbf{b},\mathbf{l}}\frac{(y_{\mathbf{l}+\mathbf{b}}-y_\mathbf{l})}{B\lambda^2}y_{\mathbf{l}-\mathbf{b}} + 1 - \hat{x}_\mathbf{l}\sum_{\mathbf{k}\in\mathcal{S}} w_{\mathbf{k},\mathbf{l}}\frac{(y_\mathbf{k}-y_\mathbf{l})}{B\lambda^2}\right.$$

TABLE I
PSNR RESULTS (dB). FOR THE SURE-BASED NLM, THE OPTIMAL λ IS INDICATED BETWEEN PARENTHESES. ALL EXAMPLES USE THE SEARCH WINDOW $\mathcal{S} = [-20, 20] \times [-20, 20]$

	$\mathcal{B}$	cameraman (256×256)			barbara (512×512)		
		$\sigma = 10$	15	25	10	15	25
Fixed ($\lambda = .7\sigma$)	5×5	32.6 (7.00)	30.3 (10.50)	27.9 (17.50)	33.1 (7.00)	31.0 (10.50)	27.6 (17.50)
	7×7	32.2 (7.00)	29.7 (10.50)	27.4 (17.50)	32.8 (7.00)	30.7 (10.50)	27.4 (17.50)
	9×9	31.9 (7.00)	29.3 (10.50)	27.1 (17.50)	32.4 (7.00)	30.4 (10.50)	27.1 (17.50)
SURE (true σ)	5×5	32.8 (6.41)	30.3 (9.85)	28.0 (16.64)	33.2 (7.38)	31.0 (10.04)	28.1 (14.91)
	7×7	32.3 (6.46)	29.7 (9.35)	27.5 (16.80)	32.9 (7.49)	30.8 (10.14)	28.1 (14.80)
	9×9	32.0 (6.50)	29.3 (9.22)	27.1 (16.91)	32.6 (7.57)	30.5 (10.24)	27.9 (14.91)
SURE (est. σ)	5×5	32.5 (7.38)	30.1 (11.66)	27.9 (17.46)	32.7 (8.83)	30.7 (11.26)	28.0 (15.96)
	7×7	32.1 (7.18)	29.6 (11.79)	27.4 (17.51)	32.4 (8.90)	30.5 (11.30)	28.0 (15.69)
	9×9	31.9 (7.16)	29.2 (12.16)	27.1 (17.51)	32.1 (8.98)	30.2 (11.36)	27.7 (15.69)

		lena (512×512)			montage (256×256)		
		$\sigma = 10$	15	25	10	15	25
Fixed ($\lambda = .7\sigma$)	5×5	35.1 (7.00)	33.0 (10.50)	30.5 (17.50)	35.1 (7.00)	32.6 (10.50)	29.2 (17.50)
	7×7	34.7 (7.00)	32.6 (10.50)	30.0 (17.50)	34.5 (7.00)	31.9 (10.50)	28.6 (17.50)
	9×9	34.4 (7.00)	32.2 (10.50)	29.5 (17.50)	34.0 (7.00)	31.4 (10.50)	28.0 (17.50)
SURE (true σ)	5×5	35.1 (6.58)	33.2 (9.31)	30.7 (14.97)	35.1 (7.42)	32.6 (10.41)	29.3 (16.09)
	7×7	34.8 (6.60)	33.0 (9.18)	30.7 (14.09)	34.6 (7.51)	31.9 (10.31)	28.7 (15.76)
	9×9	34.5 (6.69)	32.7 (9.21)	30.4 (13.92)	34.1 (7.64)	31.4 (10.34)	28.2 (15.65)
SURE (est. σ)	5×5	35.1 (6.91)	33.2 (9.58)	30.7 (15.19)	35.1 (7.93)	32.6 (10.90)	29.3 (16.58)
	7×7	34.7 (6.91)	32.9 (9.38)	30.7 (14.20)	34.5 (7.91)	31.9 (10.74)	28.7 (16.14)
	9×9	34.4 (6.98)	32.7 (9.38)	30.4 (13.98)	34.1 (8.02)	31.4 (10.74)	28.2 (16.03)

$$-\hat{x}_{\mathbf{l}} \sum_{\mathbf{b}\in\mathcal{B}} w_{\mathbf{l}-\mathbf{b},\mathbf{l}} \frac{(y_{\mathbf{l}+\mathbf{b}} - y_{\mathbf{l}})}{B\lambda^2}\Bigg)$$
$$= \frac{1}{B\lambda^2} \underbrace{\frac{\sum_{\mathbf{k}\in\mathcal{S}} w_{\mathbf{k},\mathbf{l}} y_{\mathbf{k}}^2}{W_{\mathbf{l}}}}_{\widehat{x_{\mathbf{l}}^2}} - \frac{y_{\mathbf{l}}\hat{x}_{\mathbf{l}}}{B\lambda^2}$$
$$+ \frac{1}{W_{\mathbf{l}}} - \frac{\hat{x}_{\mathbf{l}}^2}{B\lambda^2} + \frac{y_{\mathbf{l}}\hat{x}_{\mathbf{l}}}{B\lambda^2}$$
$$+ \frac{1}{W_{\mathbf{l}} B\lambda^2}\sum_{\mathbf{b}\in\mathcal{B}} w_{\mathbf{l}-\mathbf{b},\mathbf{l}} (y_{\mathbf{l}+\mathbf{b}}-y_{\mathbf{l}})(y_{\mathbf{l}-\mathbf{b}}-\hat{x}_{\mathbf{l}}).$$

ACKNOWLEDGMENT

The authors thank Dr. S. Ramani for the interesting discussions and Dr. G. Peyré for making available to the community his Matlab/C implementation of the NLM algorithm. The modified source code is available on request.

REFERENCES

[1] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *SIAM Interdisc. J.: Multiscale Model. and Simul.*, vol. 4, no. 2, pp. 290–530, 2005.

[2] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 839–842, Dec. 2005.

[3] 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 Trans. Med. Imag.*, vol. 27, no. 4, pp. 425–441, Apr. 2008.

[4] A. Dauwe, B. Goossens, H. Luong, and W. Philips, "A fast non-local image denoising algorithm," in *Proc. SPIE Electron. Imag.*, 2008, vol. 6812, pp. 681210–681210-8.

[5] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient nonlocal means for denoising of textural patterns," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1083–1092, Jul. 2008.

[6] J. Orchard, M. Ebrahimi, and A. Wong, "Efficient nonlocal-means denoising using the SVD," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, San Diego, CA, 2008.

[7] J. Wang, Y. Guo, Y. Ying, Y. Liu, and Q. Peng, "Fast non-local algorithm for image denoising," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2006, pp. 1429–1432.

[8] B. Goossens, H. Luong, A. Pizurica, and W. Philips, "An improved non-local denoising algorithm," in *Proc. Int. Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*, 2008, pp. 143–156.

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

[10] P. Chatterjee and P. Milanfar, "A generalization of non-local means via kernel regression," in *Proc. SPIE Electron. Imag.*, 2008, vol. 6814, pp. 68140P–68140P-9.

[11] G. Peyré, "Image processing with non-local spectral bases," *SIAM Multiscale Model. and Simul.*, vol. 7, no. 2, pp. 703–730, 2008.

[12] N. Azzabou, N. Paragios, and F. Guichard, "Image denoising based on adapted dictionary computation," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2007, vol. 3, pp. 109–112.

[13] S. Zimmer, S. Didas, and J. Weickert, "A rotationally invariant block matching strategy improving image denoising with non-local means," in *Proc. Int. Workshop on Local and Non-Local Approximation in Image Processing*, 2008, pp. 135–142.

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

[15] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, pp. 1135–1151, 1981.

[16] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1814–1830, Nov. 2005.

[17] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 593–606, Mar. 2007.

[18] S. Ramani, T. Blu, and M. Unser, "Monte-carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.

[19] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.