Adaptive-Size Block Transforms
for Signal-Dependent Noise Removal

Alessandro Foi[a], Radu Bilcu[b], Vladimir Katkovnik[a], and Karen Egiazarian[a]

[a]Institute of Signal Processing, Tampere University of Technology,
P.O. Box 553, Tampere 33101, FINLAND - firstname.lastname@tut.fi
[b]Multimedia Technologies Laboratory, Nokia Research Center, Tampere, FINLAND - firstname.lastname@nokia.com

ABSTRACT

We present a new transform-based method for adaptive denoising. It is assumed that the observations are given by a broad class of models with a signal-dependent variance. Denoising is performed by coefficient shrinkage in local block-transform domain. The intersection of confidence intervals (ICI) rule is used in order to determine the spatially-adaptive size of the block transforms. It enables both a simpler modelling of the noise in the transform domain and a sparser decomposition of the signal. Consequently, coefficient shrinkage is very effective and the reconstructed estimate's quality is high. Experiments with simulated as well as with real data demonstrate the advanced performance of the proposed algorithm.

1. INTRODUCTION AND MOTIVATION

In many applications the observed signal is corrupted by a signal-dependent noise. The most widely encountered models are Poisson, film-grain, multiplicative and speckle noise. Their common feature is that the variance of the noise is directly related to the true-signal's intensity. Starting with the classical filters by Lee and Kuan, a number of adaptive approaches to this sort of observations have been developed and proposed (e.g. [5], [11], [12], [13], [9], [8], [2]). The predominantly signal-dependent nature of the noise of modern digital imaging sensors makes such algorithms of fundamental importance even for consumer applications.

In this paper we propose a new transform-based denoising technique for a broad class of signal-dependent noise observations. It is based on filtering in local block-transform domain, where the size of the block is adaptively selected. The use of an adaptive-size block allows a simpler and more direct noise modelling in the transform domain, which enables the correct use of standard coefficient shrinkage techniques (e.g. hard-thresholding, Wiener filtering). At the same time, an adaptive-size block improves the sparsity of the signal representation in the transform domain, further facilitating the denoising. It results in a simple yet very effective denoising algorithm.

As a particular case, we consider the two-dimensional image denoising problem. However, our approach is rather general and can be utilized for data of arbitrary dimensions.

The presented method can be considered as an improvement and generalization of the technique [10] (developed for standard additive Gaussian noise) to the following very general signal-dependent noise observation model.

1.1 Signal-dependent noise model

We consider observations $z(x)$, $x \in X \subset \mathbb{Z}^2$, with the expectations $E\{z(x)\} = y(x) \geq 0$, where the errors (noise) $\eta(x) = z(x) - y(x)$ are independent and the variance of these observations is modeled as

$$\sigma_z^2(x) = var\{z(x)\} = var\{\eta(x)\} = \rho(y(x)), \quad (1)$$

ρ being a given positive function of y called the *variance function*. For example, $\rho(y) = y$, $\rho(y) = y^2$, and $\rho(y) = (Ky^\alpha)^2$ for the Poisson, gamma, and film-grain observation models, respectively.

The problem is to reconstruct the true image y from the noisy observations z.

1.2 Filtering in transform-domain

Orthonormal transforms are commonly used in conjuction with a shrinkage operator (e.g. hard-thresholding or Wiener filtering) which attenuates the noisy transform coefficients depending on their magnitude and standard deviation. When considering the conventional additive white Gaussian noise model, $\rho(y) \equiv \sigma^2$, there are no particular complications because, due to the orthonormality of the transform, each coefficient has variance σ^2.

On the contrary, when $\rho(y)$ is not constant, the variance of each coefficient needs to be estimated separately, significantly increasing the complexity of the overall approach. On the other hand, performing coefficient shrinkage imposing a constant variance for the noisy coefficients can lead to a consistent degradation of the restoration accuracy.

2. SPATIALLY-ADAPTIVE BLOCK-SIZE

The typically encountered variance functions ρ, such as those mentioned in Section 1.1, are smooth functions of their argument y. As a consequence, on a local neighborhood where the signal is nearly constant, so is its variance. It means that with a good approximation, the variance can be assumed as constant if the transform's support is restricted to a region of signal uniformity.

In our approach we rely on the LPA-ICI technique [7] to identify adaptive neighborhoods where the signal can be

assumed to be locally constant. Such neighborhoods are found by comparing pointwise estimates of the signal calculated over an enlarging family of nested neighborhoods. For a fixed point x, the ICI (intersection of confidence intervals) criterion [6, 4] is implemented in the following way. Given a set of pointwise LPA (local polynomial approximation) estimates $\{\hat{y}_{B_j}^{\text{LPA}}(x)\}_{j=1}^J$ calculated on a nested sequence $\{B_1 \subset \cdots \subset B_J\}$ of neighborhoods of x, we determine a sequence of confidence intervals

$$\mathcal{D}_j = \left[\hat{y}_{B_j}^{\text{LPA}}(x) - \Gamma \sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)}, \ \hat{y}_{B_j}^{\text{LPA}}(x) + \Gamma \sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)}\right],$$

where $\Gamma > 0$ is a threshold parameter and $\sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)}$ is the standard-deviation of $\hat{y}_{B_j}^{\text{LPA}}(x)$. The ICI rule can be stated as follows: *Consider the intersection of confidence intervals $\mathcal{I}_j = \bigcap_{i=1}^j \mathcal{D}_i$ and let j^+ be the largest of the indexes j for which $\mathcal{I}_j$ is non-empty, $\mathcal{I}_{j^+} \neq \varnothing$ and $\mathcal{I}_{j^++1} = \varnothing$. Then the adaptive neighborhood B_x^+ of x is defined as $B_x^+ = B_{j^+}$.*

In particular, we exploit the zero-order local polynomial model, using uniform weighting over nested square blocks $\{B_j\}_{j=1}^J$. It means that the estimate $\hat{y}_{B_j}^{\text{LPA}}(x)$ and its standard deviation $\sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)}$ are calculated as

$$\hat{y}_{B_j}^{\text{LPA}}(x) = \frac{1}{|B_j|}\sum_{v \in B_j} z(v) \quad \text{and} \quad \sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)} = \frac{1}{|B_j|}\sqrt{\sum_{v \in B_j} \sigma_z^2(v)},$$

where $|B_j|$ denotes the size (cardinality) of B_j. In this way, the LPA estimate $\hat{y}_{B_j}^{\text{LPA}}(x)$ is simply the mean of the signal over B_j, and – up to a factor $\sqrt{|B_j|}$ – it coincides with the DC coefficient of a DCT or DFT transform or with the coarsest approximation coefficient of a full wavelet decomposition (with the transforms supported on B_j).

Roughly speaking, the adaptive B_x^+ selected by the ICI is the largest neighborhood of x which is compatible with the data contained within all smaller neighborhoods. In practice, this means that significant details of the image are not oversmoothed by the LPA estimate. At the same time, because of the zero-order (local constant) polynomial model, it means that y is nearly constant within B_x^+. It enables the hypothesis of constant variance within the adaptive B_x^+. This significantly simplifies the shrinkage, as the constant variance on B_x^+ does not require the calculation of the variance of the individual transform coefficients. Moreover, because of the absence of singularities within B_x^+ the data enjoy a sparse representation in the transform domain, significantly improving the effectiveness of shrinkage.

Further simplification can be achieved by approximating $\sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)}^2 = \frac{1}{|B_j|^2}\sum_{v \in B_j} \sigma_z^2(v)$ as $\frac{1}{|B_j|}\rho\left(|\hat{y}_{B_j}^{\text{LPA}}(x)|\right)$. Observe that for the Poissonian noise model (where the variance depends linearly on the signal), this approximation becomes accurate.

Let us remark that all the formulas and considerations in the present and in the previous section assume accurate knowledge of the variance σ_z^2 of the observations. This is of course an idealistic assumption, and in all cases where the variance function ρ is invertible (e.g. for the speckle,

film-grain, or Poissonian case), exact knowledge of the variance is equivalent to knowing the signal intensity itself.

It is common practice to estimate the variance through the variance function on the noisy data as $\hat{\sigma}_z^2 = \rho(|z|)$. In most situations this proves to be a good-enough approximation. However, when the signal-to-noise ratio is particularly low, such direct relation can be rather erroneous and recursive variance update procedures can be exploited in order to iteratively refine the estimate of the variance (e.g. similarly to [2]).

3. ALGORITHM

Our denoising approach is based on sliding window transform-domain filtering with adaptive size for the transform's support. The LPA-ICI is used to define the size of the support. Any orthonormal transform which provides a coefficient corresponding to the mean value of the data over its support (i.e. a DC coefficient) can be directly used in the following algorithm. Let us observe that the near totality (including the DCT, the DFT/FFT, and wavelets) of transforms used for image processing applications satisfy this requirement.

The overall algorithm comprises two major parts. The first part (steps I, II, and III) is based on hard-thresholding. The second part (steps IV and V) is based on empirical Wiener filtering, using the hard-thresholding estimate obtained after step III as a reference estimate. Although the second part is optional, it can be very beneficial to the estimates quality.

(I) LPA-ICI ADAPTIVE BLOCK-SIZES

For a fixed j, all the estimates $\hat{y}_{B_j}^{\text{LPA}}(x) \ \forall x \in X$ are simultaneously calculated as the convolution $\hat{y}_{B_j}^{\text{LPA}} = z \circledast 1_{B_j}$, where 1_{B_j} is a uniform kernel equal to h_j^{-2} on a square support of size $h_j \times h_j$. The variance of $\hat{y}_{B_j}^{\text{LPA}}(x)$ is estimated as $\sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)}^2 = \frac{1}{|B_j|}\rho\left(|\hat{y}_{B_j}^{\text{LPA}}(x)|\right)$.

This procedure is repeated for a set $H = \{h_1 < \cdots < h_J\}$ of increasing h_j and with nested supports of the kernels 1_{B_j}, $j = 1, \ldots, J$. Using established terminology, such h_j's are called *scales*, and H is thus the *set of scales*.

Given the estimates $\{\hat{y}_{B_j}^{\text{LPA}}(x)\}_{j=1}^J$ and their variances $\{\sigma_{\hat{y}_{B_j}^{\text{LPA}}(x)}^2\}_{j=1}^J$, the ICI selects for every $x \in X$ the pointwise adaptive-size block B_x^+ (of size $h_{j^+} \times h_{j^+}$), as described in Section 2.

(II) TRANSFORM-DOMAIN HARD-THRESHOLDING

For every $x \in X$, the transform is calculated on the adaptive-size block B_x^+. We denote the corresponding transform coefficients as $\varphi_{z,B_x^+}(i)$, $i = 1, \ldots, |B_x^+|$, with $\varphi_{z,B_x^+}(1)$ being the DC coefficient.

The variance of z on the adaptive block, denoted as $\sigma_{z(B_x^+)}^2$, is assumed constant and equal to $\rho\left(|\varphi_{z,B_x^+}(1)|/\sqrt{|B_x^+|}\right)$. Thus, the standard-deviation of

each transform coefficient is simply $\sigma_{z(B_x^+)}$ and hard-thresholding is performed in the form

$$\hat{\varphi}_{y,B_x^+}(i) = \begin{cases} 0 & \text{if } |\varphi_{z,B_x^+}(i)| < \tau\sigma_{z(B_x^+)} \text{ and } i > 1 \\ \varphi_{z,B_x^+}(i) & \text{if } |\varphi_{z,B_x^+}(i)| \geq \tau\sigma_{z(B_x^+)} \text{ or } i = 1 \end{cases},$$

where $\tau = \gamma\sqrt{2\ln|B_x^+|+1}$ is a size-dependent threshold parameter which is essentially the so-called "universal" threshold. The factor γ is a fixed constant, invariant with respect to the block-size. Only AC coefficients are thresholded, hence the DC is always preserved.

Inverse transformation of the hard-thresholded coefficients $\hat{\varphi}_{y,B_x^+}$ yields a local estimate of the signal $\hat{y}_{B_x^+}^{\text{TRA}}$: $B_x^+ \to \mathbb{R}$ defined on the adaptive-size block B_x^+.

In what follows, we make use of the total variance of this local estimate. Because of the orthonormality of the transform, such average variance can be calculated as the sum of the variances of all non-thresholded coefficients. Therefore $\sigma_{\hat{y}_{B_x^+}^{\text{TRA}}}^2 = \sigma_{z(B_x^+)}^2 N_x^\varphi$, where N_x^φ stands for the number of non-zero coefficients after thresholding (i.e. the number of basis functions used for reconstruction).

(III) AVERAGING WITH ADAPTIVE WEIGHTS

For each $x \in X$, a local estimate $\hat{y}_{B_x^+}^{\text{TRA}}$ supported on an adaptive-size block B_x^+ has been obtained from the previous steps. It should be observed that usually blocks corresponding to adjacent pixels overlap. This overlap corresponds to the overcompleteness of our approach and requires that, in order to obtain an estimate $\hat{y} : X \to \mathbb{R}$ for the whole image, some averaging or fusing of all the many local estimates is performed. How this averaging is done can have a dramatic impact on the quality of the final estimate. We use the following convex combination with adaptive weights w_x which depend on the variance as well as on the size of the reconstructed local estimates:

$$\hat{y} = \frac{\sum_{x \in X} w_x \hat{y}_{B_x^+}^{\text{TRA}}}{\sum_{x \in X} w_x \chi_{B_x^+}}, \quad w_x = \frac{1}{\sigma_{\hat{y}_{B_x^+}^{\text{TRA}}}^2 |B_x^+|}, \quad (2)$$

where $\chi_{B_x^+}$ is the characteristic function of B_x^+. In (2) we implicitly assume that the local estimates $\hat{y}_{B_x^+}^{\text{TRA}}$ are zero-padded outside their domain B_x^+. The weights w_x have this form because the total variance $\sigma_{\hat{y}_{B_x^+}^{\text{TRA}}}^2$ is obviously an upper bound for the pointwise residual-noise variance of the local estimate $\hat{y}_{B_x^+}^{\text{TRA}}$ (such pointwise variance is not necessarily uniform over B_x^+), while the extra factor $|B_x^+|$ addresses the correlation that exists between overlapping blocks (the number of overlapping blocks is loosely proportional to their size). Qualitatively speaking, these weights favour estimates which correspond to sparser representations (fewer coefficients survived thresholding, and thus lower variance) and at the same time avoid that estimates with a small support (thus representing image details) are oversmoothed by other overlapping estimates which have a large support (which usually are strongly correlated among themselves and outnumber estimates of a smaller support).

The estimate $\hat{y} : X \to \mathbb{R}$ obtained from (2) can already be considered as a satisfactory estimate of the unknown signal y. Alternatevely, it can be used as a reference estimate for the second part of the algorithm, providing a further improvement in the restoration performance.

(IV) TRANSFORM-DOMAIN WIENER-FILTERING

For every $x \in X$, the transform is calculated on the adaptive-size block B_x^+ for the noisy data z as well as for the reference estimate $\hat{y}$. We denote the corresponding transform coefficients as φ_{z,B_x^+} and $\varphi_{\hat{y},B_x^+}$, respectively.

The Wiener filter in transform-domain can be given as

$$\hat{\varphi}_{y,B_x^+}^{\text{W}}(i) = \frac{\varphi_{\hat{y},B_x^+}^2(i)}{\varphi_{\hat{y},B_x^+}^2(i) + \sigma_{z(B_x^+)}^2} \varphi_{z,B_x^+}(i), \quad i = 1, \ldots, |B_x^+|,$$

where $\sigma_{z(B_x^+)}^2$ is the variance of the observations within B_x^+ (assumed as constant), which is calculated as $\sigma_{z(B_x^+)}^2 = \rho\left(|\varphi_{\hat{y},B_x^+}(1)|/\sqrt{|B_x^+|}\right)$, $\varphi_{\hat{y},B_x^+}(1)$ being the DC coefficient of the transform of $\hat{y}$ on B_x^+.

The local Wiener estimate $\hat{y}_{B_x^+}^{\text{TRA-W}} : B_x^+ \to \mathbb{R}$ is obtained by inverse transformation of the attenuated coefficients $\hat{\varphi}_{y,B_x^+}^{\text{W}}$. The variance of this local estimate is computed as

$$\sigma_{\hat{y}_{B_x^+}^{\text{TRA-W}}}^2 = \sigma_{z(B_x^+)}^2 \sum_{i=1}^{|B_x^+|} \left(\frac{\varphi_{\hat{y},B_x^+}^2(i)}{\varphi_{\hat{y},B_x^+}^2(i) + \sigma_{z(B_x^+)}^2}\right)^2.$$

(V) AVERAGING WITH ADAPTIVE WEIGHTS

Averaging of the local estimates $\hat{y}_{B_x^+}^{\text{TRA-W}}$ is done analogously to step (III), using adaptive weights $w_x = \left(\sigma_{\hat{y}_{B_x^+}^{\text{TRA-W}}}^2 |B_x^+|\right)^{-1}$, yielding the final estimate $\hat{y}^{\text{W}} : X \to \mathbb{R}$.

The above algorithm can be implemented in recursive mode, progressively improving the initial estimate of the variance. However, in our experiments we found no concrete advantage in doing so. Therefore, for the experiments presented in the next section, we followed the above five steps only once, exactly as described here, and the results are given for the final estimate $\hat{y}^{\text{W}}$ from step (V).

4. EXPERIMENTAL RESULTS

We show experimental results for three common types of signal-dependent noise: the "scaled" *Poisson* noise, $z \sim \mathcal{P}(\chi y)/\chi$, $\chi \in \mathbb{R}^+$, the *film-grain* noise, $z = y + Ky^\alpha\eta$, $K, \alpha \in \mathbb{R}^+$ and $\eta \sim \mathcal{N}(0,1)$, and the "multiple-look" *speckle* noise, $z = L^{-1}\sum_{i=1}^L y\epsilon_i$, $\epsilon_i \sim \mathcal{E}(\beta)$, $\beta \in \mathbb{R}^+$. The calligraphic letters $\mathcal{P}$, $\mathcal{N}$, and $\mathcal{E}$ denote, respectively, the Poisson, Gaussian, and exponential distributions. For the above observation models, the variance functions $\rho(y) = \sigma_z^2$ are $\rho(y) = y/\chi$, $\rho(y) = K^2 y^{2\alpha}$, and $\rho(y) = y^2\beta/L$, respectively. For all these experiments, the DCT transform is chosen because of its better energy compaction properties for stationary signals. In our implementation we use $\Gamma = 1.2$, $H = [4, 6, 8, 12, 16]$ and $\gamma = 0.85$. The true signal y is assumed to have range [0,255].
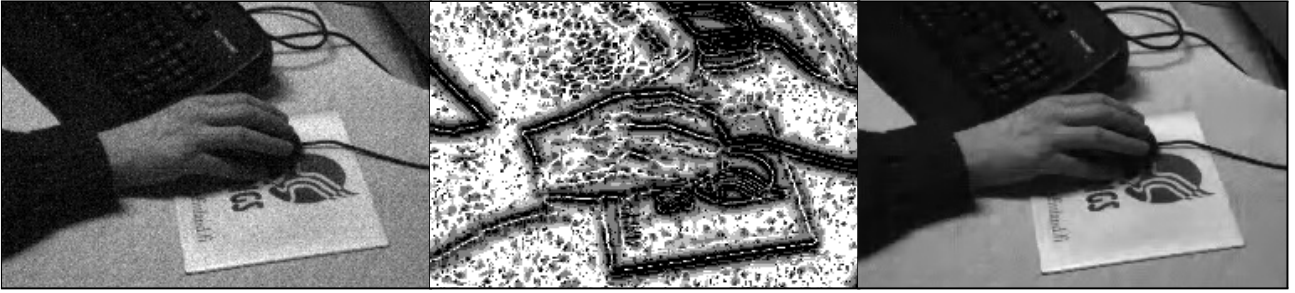
Figure 1: Raw-data from CMOS sensor, adaptive block-sizes (black h_{j+}=4, white h_{j+}=16), and denoised estimate $\hat{y}^{\mathrm{w}}$.

	noise type	noisy	[5]	[11]	[2, 3]	AS B-DCT
Peppers	*Poisson*	1243	160	145	120	**104**
	Film-grain	1351	169	150	123	**109**
	Speckle	4442	372	378	286	**225**
Aerial	*Poisson*	758	252	179	183	**149**
	Film-grain	829	267	188	185	**154**
	Speckle	1698	387	318	330	**257**

Table 1: *MSE* values for different models, and methods.

	χ	noisy	[12]	[9]	[2, 3]	AS B-DCT
Lena	*30/255*	1054	168	143	73	**61**
	60/255	525	117	96	50	**42**
	90/255	349	93	75	40	**34**
	120/255	262	81	63	34	**30**
Cameraman	*30/255*	1015	199	154	136	**107**
	60/255	504	140	97	89	**70**
	90/255	336	113	74	70	**55**
	120/255	254	97	61	60	**46**

Table 2: *MSE* comparison against algorithms for Poissonian noise.

Firstly, in Table 1 we give a comparison for the simulations presented in [11] for the above noise models with parameters χ=0.1, K=3.3, α=0.5, L=4, and β=1. In the table our results are compared against the adaptive-neighborhood filter [11], the noise-updating repeated Wiener filter [5] (as quoted in [11]), and the recursive anisotropic LPA-ICI technique [2, 3].

Next, in Table 2 we compare our results for removal of Poissonian noise ($\chi = \frac{30}{255}, \frac{60}{255}, \frac{90}{255}, \frac{120}{255}$) against those obtained by two other transform-based methods [12],[9] recently developed specifically for this type of noise and by the recursive anisotropic LPA-ICI technique.

The results in the tables show that the proposed *adaptive-size block-DCT* (AS B-DCT) algorithm outperforms all other methods, for all considered noise models and noise levels.

Finally, we show a result of denoising real data acquired by the CMOS sensor of a Nokia cameraphone. The statistical characteristics of the sensor's raw-data have been studied, and were found to follow very accurately the observation model (1). The corresponding variance function $\rho(y)$ has been estimated and used in the algorithm. In low-light conditions, or for short exposure-times, the signal-to-noise ratio can be dramatically low. Figure 1 shows the raw-data captured with an exposure time of 4ms, the adaptive block-sizes selected by the LPA-ICI (step I), and the denoised estimate $\hat{y}^{\mathrm{w}}$ obtained using the proposed adaptive-size block DCT method. In the figure it is possible to observe how the adaptive block-sizes reveal the image features such as edges and details. Despite the moderately high level of the noise in the raw-data, the quality of the denoised estimate is rather high: smooth areas are faithfully restored and finer details are accurately preserved.

REFERENCES

[1] Fan, J., and I. Gijbels, *Local polynomial modelling and its application*. London, Chapman and Hall, 1996.

[2] Foi, A., R. Bilcu, V. Katkovnik, and K. Egiazarian, "Anisotropic local approximations for pointwise adaptive signal-dependent noise removal", *Proc. of XIII European Signal Proc. Conf., EUSIPCO 2005*, Antalya, September 2005.

[3] Foi, A., D. Paliy, V. Katkovnik, and K. Egiazarian, "Anisotropic nonparametric image restoration demobox" (MATLAB software), *LASIP (Local Approximations in Signal and Image Processing) Project*, http://www.cs.tut.fi/~lasip/ , 2005.

[4] Goldenshluger, A., and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression", *Math. Meth. Statistics*, vol. 6, pp. 135-170, 1997.

[5] Jiang, S.S., and A.A. Sawchuk, "Noise updating repeated Wiener filter and other adaptive noise smoothing filters using local image statistics", *Appl. Opt.*, vol. 25, pp. 2326-2337, 1986.

[6] Katkovnik, V., "A new method for varying adaptive bandwidth selection", *IEEE Trans. Signal Proc.,* vol. 47, no. 9, pp. 2567-2571, 1999.

[7] Katkovnik, V., K. Egiazarian, and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule", *J. Math. Imaging and Vision*, vol. 16, no. 3, pp. 223-235, 2002.

[8] Katkovnik, V., A. Foi, K. Egiazarian, and J. Astola, "Anisotropic local likelihood approximations", *Proc. of Electronic Imaging 2005*, 5672-19, 2005.

[9] Lu, H., Y. Kim, and J.M.M. Anderson, "Improved Poisson intensity estimation: denoisng application using poisson data," *IEEE Trans. Image Processing,* vol. 13, no. 8, pp. 1128-1135, 2004.

[10] Öktem, H., V. Katkovnik, K. Egiazarian, and J. Astola, "Local adaptive transform based image de-noising with varying window size", *Proc. IEEE Int. Conference on Image Process., ICIP 2001,* Thessaloniki, Greece, pp. 273-276, October 2001.

[11] Rangarayyan, R.M., M. Ciuc, and F. Faghih, "Adaptive-neighborhood filtering of images corrupted by signal-dependent noise", *Appl. Opt.*, vol. 37, pp. 4477-4487, 1998.

[12] Timmermann, K.E., and R. Nowak, "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging", *IEEE Trans. Inf. Theory,* vol. 45, no. 3, pp. 846-862, 1999.

[13] Willett, R.M., and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging", *IEEE Trans. Medical Imaging,* vol. 22, no. 3, pp. 332-350, 2003.