

Context-based Denoising of Images Using Iterative Wavelet Thresholding

Detlev Marpe^{*a}, Hans L. Cycon^{**b}, Gunther Zander^{**b}, and Kai-Uwe Barthel^{**b}
^aHeinrich-Hertz-Institute (HHI); ^bUniversity of Applied Sciences (FHTW Berlin)

ABSTRACT

In this paper, we propose a spatially adaptive wavelet thresholding method using a context model that has been inspired by our prior work on image coding. The proposed context model relies on an estimation of the weighted variance in a local window of scale and space. Appropriately chosen weights are used to model the predominant correlations for a reliable statistical estimation. By iterating the context-based thresholding operation, a more accurate reconstruction can be achieved. Experimental results show that our proposed method yields significantly improved visual quality as well as lower mean squared error compared to the best recently published results in the denoising literature.

Keywords: Wavelet thresholding, context modeling, image denoising, image restoration

1. INTRODUCTION

Electronically captured signals always suffer from perturbations in the sense that there is a superposition of noise and the signal in question. One classical task of signal processing is to discriminate between noise and signal, and to remove the unwanted noise from the signal. Especially for the case of additive white Gaussian noise a number of techniques using wavelet-based thresholding have been proposed.^{2, 4-6, 10} The idea of *wavelet thresholding* rests on the assumption that the signal magnitudes dominate the magnitudes of the noise in a wavelet representation, so that wavelet coefficients can be set to zero if their magnitudes are less than a pre-determined threshold. Donoho and Johnstone⁵ proposed *hard-* and *soft-thresholding* methods for denoising, where the former leaves the magnitudes of coefficients unchanged if they are larger than a given threshold, while the latter just shrinks them to zero by the threshold value. They proved that the performance of these thresholding methods is close to that of an ideal coefficient selection method where the coefficients of the underlying signal are known in advance.

However, the major problem with both methods and most of its variants is the choice of a suitable threshold value. The definition of a global, *i.e.*, coefficient independent threshold given by Donoho and Johnstone⁵ depends on the noise power and the signal size. It was derived by proving an asymptotically optimal upper bound on the approximation error in the limit of an arbitrary large signal size. In practice, however, one deals with signals of finite length, where the applicability of such a theoretical result is rather questionable. In addition, most signals show a spatially non-uniform energy distribution, which motivates the choice of a non-uniform threshold. Since a given noisy signal may consist of some parts where the magnitudes of the signal are below the globally defined threshold and other parts where the noise magnitudes exceed that given threshold, methods relying on a globally defined threshold cut off parts of the signal, on the one hand, and leave some noise untouched, on the other hand. This observation led to the idea of a spatially adaptive threshold choice depending on the relationship of local energy (variance) of the observed signal and the noise variance. Chang et al.^{1, 2} were the first to propose this kind of spatially adaptive wavelet thresholding for image denoising. Their method of selecting a spatially adaptive threshold is based on a context model, which involves neighboring coefficients of the wavelet decomposition for the estimation of the local variance.

* marpe@hhi.de; phone +49 30 31002-619; fax +49 30 3927200; <http://bs.hhi.de>; Image Processing Department, Heinrich-Hertz-Institute (HHI), Einsteinufer 37, 10587 Berlin, Germany; ** [hcycon,zander,barthel]@fhtw-berlin.de; phone +49 30 54699-363; fax +49 30 54699-329; <http://www.fhtw-berlin.de>; University of Applied Sciences (FHTW Berlin), Allee der Kosmonauten 20-22, 10315 Berlin, Germany

In this paper, we propose an adaptive method for denoising of images called *context-based iterative denoising (CBID)*, which follows the ideas of Chang et al., but introduces two distinct new features. First, we developed a more elaborate context model, which was inspired by our prior work on context-based image and video coding.^{8,9} This context model allows us to modify the threshold directly on a coefficient level without performing a sorting pass on the context related variance estimates. As a measure for local activity, the context of a given wavelet coefficient is determined by the weighted variance in a local window containing neighbouring coefficients as well as corresponding parent coefficients in the next higher decomposition level of the multiresolution decomposition, where the chosen weights depend on the decomposition level and the orientation of the given subband.

The second main idea of our proposed method is to iterate the context-based thresholding process on the denoised wavelet representation. By varying the context weights in each iteration step appropriately, we are able to improve the visual quality substantially. In addition, we used an overcomplete, *i.e.*, non-subsampled wavelet representation as proposed by Coifman and Donoho,⁴ which is known to lead to more satisfactory results than thresholding in a critically subsampled representation.

2. PROPOSED ALGORITHM

2.1. Problem Formulation, Notations and Basic Solutions

Suppose that a given image $\mathbf{x} = \{x[m, n] | m, n = 1 \dots, N\}$ has been corrupted by additive noise such that we have observations

$$y[m, n] = x[m, n] + z[m, n], \quad m, n = 1, \dots, N, \quad (1)$$

where $\{z[m, n] | m, n = 1 \dots, N\}$ are assumed to be *independent and identically distributed (iid)* as zero-mean Gaussian with variance σ_n^2 . Then the goal of denoising the observed image $\mathbf{y} = \{y[m, n]\}$ is to estimate an image $\hat{\mathbf{x}} = \{\hat{x}[m, n]\}$ as ‘close’ as possible to the original image \mathbf{x} in the sense of minimizing the *mean squared error (MSE)* between $\hat{\mathbf{x}}$ and \mathbf{x} given by $MSE(\mathbf{x}, \hat{\mathbf{x}}) = 1/N^2 \sum_{m,n=1}^N (x[m, n] - \hat{x}[m, n])^2$. However, we should keep in mind that minimizing the MSE is not the ultimate criterion since in most of the practical interesting cases we want to obtain a denoised image, which is *visually the most accurate reconstruction*.

Donoho and Johnstone⁵ proposed a very simple solution to the denoising problem consisting of three steps:

1. Transform the observed image \mathbf{y} into the wavelet domain, *i.e.*, $\mathbf{Y} = \mathcal{W}\mathbf{y}$, where \mathcal{W} denotes the 2-D critically sampled dyadic *discrete wavelet transform (DWT)*.
2. Apply a thresholding operator to the resulting wavelet domain representation (except the lowpass part), *i.e.*, apply either the so-called *hard-thresholding* operator T_τ^{hard} given by

$$T_\tau^{\text{hard}}(Y[m, n]) = \begin{cases} Y[m, n], & |Y[m, n]| > \tau \\ 0, & |Y[m, n]| \leq \tau \end{cases}$$

or the *soft-thresholding* operator T_τ^{soft} given by

$$T_\tau^{\text{soft}}(Y[m, n]) = \begin{cases} \text{sgn}(Y[m, n])(|Y[m, n]| - \tau), & |Y[m, n]| > \tau \\ 0, & |Y[m, n]| \leq \tau \end{cases}, \quad (2)$$

where τ denotes the pre-selected threshold.

3. Transform the thresholded representation back to the original domain, *i.e.*, $\hat{\mathbf{x}} = \mathcal{W}^{-1}(T_\tau \mathbf{Y})$, which yields an estimate of \mathbf{x} .

This approach has been studied extensively both with respect to theoretical and practical aspects (see Mallat⁷ and references therein), and it is the basis of a number of extensions and modifications, which have been proposed so far. In the next section, we will sketch of one of the most successful ideas developed in this context.

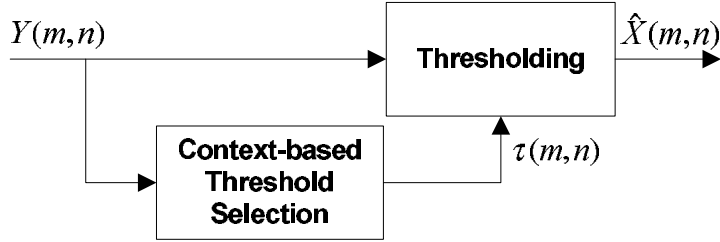


Figure 1: One stage of the proposed adaptive thresholding scheme.

2.2. Wavelet Thresholding in Overcomplete Representations

When applying wavelet thresholding algorithms to image denoising problems, we are faced with the problem of visually unpleasant artifacts such as pseudo-Gibbs phenomena or speckled noise fragments. To suppress these undesirable artifacts, Coifman and Donoho⁴ proposed to extend the thresholding method to *non-subsampled wavelet representations (NS-DWT)*. Let's consider, for example, a one level decomposition using the NS-DWT. Then, the resulting coefficients can be separated into four sets of uncorrelated coefficients, *i.e.*, $\mathbf{Y}^{(0,0)} = \{Y[2m, 2n]\}$, $\mathbf{Y}^{(1,0)} = \{Y[2m+1, 2n]\}$, $\mathbf{Y}^{(0,1)} = \{Y[2m, 2n+1]\}$ and $\mathbf{Y}^{(1,1)} = \{Y[2m+1, 2n+1]\}$. Now, the main idea behind using an overcomplete representation is that ringing artifacts caused by mis-alignments between basis functions and image features are considerably reduced, if we build the average of the reconstructions related to the four thresholded 'copies' of the observed image \mathbf{y} , where each single representation can be interpreted as a shifted version of \mathbf{y} . More specifically, the denoised estimate $\hat{\mathbf{x}}$ using a one level NS-DWT is formed by $\hat{\mathbf{x}} = 1/4 \sum_{j,k=0}^1 \mathcal{W}^{-1}(T_\tau \mathbf{Y}^{(j,k)})$. Obviously, this method can be generalized to a dyadic non-subsampled decomposition of arbitrary depth by simply applying the averaging procedure to each reconstruction step of all lowpass branches in the dyadic NS decomposition.*

Due to its effectiveness both in terms of objective and subjective quality, we decided to build our proposed algorithm on the non-subsampled DWT, although the conceptual ideas behind our approach are not confined to any specific wavelet representation.

2.3. Context-based Threshold Selection

The most critical issue in wavelet thresholding is the choice of an appropriate threshold. In the original work of Donoho et al.,^{4,5} a universal threshold $\tau(\sigma_n, N) = \sigma_n \sqrt{2 \log(N^2)}$ has been derived, which depends on the image size (N^2) and the noise standard deviation σ_n . However, following the strategy of minimizing the MSE, a more flexible choice is often preferred:

$$\tau(\sigma_n, \mathcal{C}) = \lambda(\mathcal{C}) \sigma_n, \quad (3)$$

where the constant λ may be optimized with respect to the underlying image class \mathcal{C} . We adopt the rather general choice of Eq. (3) as a basis of our context-based threshold selection method.

Fig. 1 illustrates the concept of our proposed adaptive algorithm. Rather than applying a pre-selected uniform threshold, we propose a *context-based threshold selection* scheme for a coefficient-dependent choice of the threshold. Our model relies on an estimation of the local weighted variance $\sigma_w[m, n]^2$ of each wavelet coefficient $Y^{(l,o)}[m, n]$ at level l and orientation $o \in \{V, H, D\}$ using a window \mathcal{N} , which covers a 5×5 neighborhood of $Y^{(l,o)}[m, n]$ and a 3×3 neighborhood of the corresponding parent coefficient $Y^{(l+1,o)}[\lfloor m/2 \rfloor, \lfloor n/2 \rfloor]$ as shown in Fig. 2 (a). This model was motivated by our prior work on image coding,^{8,9} where similar shaped neighborhoods were successfully used for context modeling of wavelet coefficients. However, in contrast to image coding applications, where we are restricted to a causal neighborhood or a rather small collection of different context states, the denoising framework offers much more freedom in the design of appropriate context models. Thus, we are able not only to adapt the shape of the window \mathcal{N} , *i.e.*, the locality of reference, but also the relevance of each local neighbor to the statistics of a given subband by weighting its energy appropriately. Both aspects can

*Note that the coefficients of the NS-DWT can be organized using a full balanced quadtree, where each node corresponds to one lowpass branch.

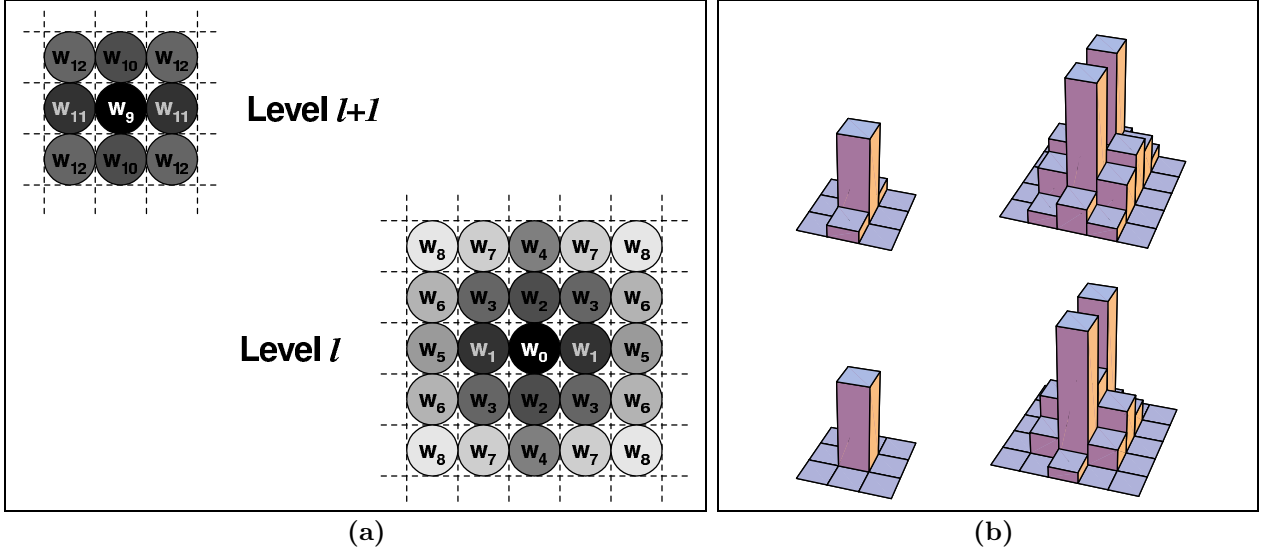


Figure 2. (a) Footprint of the local window \mathcal{N} covering two levels (scales) of the dyadic decomposition. Weight w_0 corresponds to the position of the current coefficient and w_9 is related to the position of the corresponding parent coefficient. (b) Visualization of the weights chosen for bands with vertical orientation: (top row) level $l = 2$; (bottom row) level $l = 3$.

be integrated into one model of weighted variances using a fixed window with a sufficiently large size as shown in Fig. 2 (a). According to this design, we define the *estimated weighted variance* of a coefficient $Y^{(l,o)}[m, n]$ with respect to the window \mathcal{N} and a corresponding set of weights $\mathbf{w} = \{w_{j,k}^{(l,o)} \mid j, k \in \mathcal{N}\}$ by

$$\sigma_w^{(l,o)}[m, n]^2 = \frac{\sum_{j,k \in \mathcal{N}} w_{j,k}^{(l,o)} Y^{(l,o)}[j, k]^2}{\sum_{j,k \in \mathcal{N}} w_{j,k}^{(l,o)}}. \quad (4)$$

Having an estimation of the underlying weighted variance field via context modeling, we are ready to build a coefficient-dependent threshold by weighting the uniform threshold of Eq. (3) with the quotient of noise variance and estimated local variance:

$$\tau^{(l,o)}[m, n] = \lambda(\mathcal{C}) \sigma_n \frac{\sigma_n^2}{\sigma_w^{(l,o)}[m, n]^2}. \quad (5)$$

This definition of an adaptive threshold $\tau^{(l,o)}[m, n]^\dagger$ has a rather intuitive explanation: In smooth image regions, where the noise variance dominates the observed local variance, a large threshold is chosen according to Eq. (5). Thus, most of the noise components, which are also highly visible in flat areas, are removed. In textured or edge-dominated regions, however, where the observed local variance is expected to be much higher than the noise power (which in turn is not as visible here due to masking effects of the human visual system), the corresponding thresholds are small guaranteeing that most of the signal components are retained.

2.4. Iterated Thresholding

The method we are proposing is an iterative process. Since the variance estimation in the first pass of the thresholding operation is based on the observed noisy coefficients $Y[m, n]$, a considerable amount of unreliable estimates leads to visually annoying noise specks in the reconstruction. Thus, we propose to add some additional steps of thresholding, each relying on a local variance estimation that uses already denoised coefficients of the preceding step. Suppose, we have given the denoised coefficients of the $(i-1)$ -th iteration step by $\hat{X}^{(i-1)}[m, n]$,

[†]For mathematical convenience, we omit the superscript (l, o) indicating the subband in the following.

then the local variance of the i -th iteration step is given by

$$\sigma_w^{(i)}[m, n]^2 = \frac{\sum_{j, k \in \mathcal{N}} w_{j, k}^{(i)} \hat{X}^{(i-1)}[j, k]^2}{\sum_{j, k \in \mathcal{N}} w_{j, k}^{(i)}}$$

with a suitable set of weights $\mathbf{w}^{(i)} = \{w_{j, k}^{(i)} | j, k \in \mathcal{N}\}$. Correspondingly, the adaptive thresholds of the i -th iteration step are defined by

$$\tau^{(i)}[m, n] = \lambda^{(i)}(\mathcal{C}) \sigma_n \frac{\sigma_n^2}{\sigma_w^{(i)}[m, n]^2}. \quad (6)$$

Note that the additional computational complexity of performing subsequent steps of thresholding is negligible, since no additional transform or inverse transform steps are involved.

2.5. Choice of Parameters

Our model involves multiple sets of weights $\mathbf{w}^{(i, l, o)}$ depending on the iteration step i and the subband characterized by its level l and orientation o . The MSE-optimal choice of weights for a given class \mathcal{C} of images would be obtained by selecting those weights which yield weighted variance estimates according to Eq. (4) such that the thresholding process using the corresponding context-based thresholds minimizes the $MSE(\hat{X}^{(i)}, X)$. Unfortunately, this is an numerically intractable optimization problem.

However, it is intuitively clear that the local weighted variance should somehow reflect the correlation structure of wavelet coefficients typically observed for natural images. In general, magnitudes of wavelet coefficients show correlations which decay exponentially with the distance. Moreover, in a separable 2-D wavelet decomposition, the decay depends strongly on the orientation o of the given band, *i.e.*, along the direction of highpass filtering the correlation typically falls off more rapidly than in lowpass direction. In addition, the correlation also depends on the level l of decomposition, such that on higher levels one observes a much stronger decay than on lower levels. By putting these observations together, we finally arrived at a model of weights, which is illustrated in Fig. 2.

The symmetric structure of the weights within the context template shown in Fig. 2 (a) reflects the main dependencies in horizontal and vertical direction due to the separable filtering process. It also simplifies the model by reducing the number of free parameters from 34 to 13. Fig. 2 (b) visualizes two typical instances of weights chosen for level $l = 2$ (top) and $l = 3$ (bottom), both for the vertical (lowpass filtered, $o = V$) band and the first thresholding pass ($i = 0$). Two design criteria are noteworthy. First, it can be observed that in both cases the weight w_2 corresponding to the vertical neighbors of the current coefficient is the most dominant one followed by the weight w_9 of the corresponding parent coefficient in the next upper level, hence capturing the most significant correlation patterns of both intra- and interband type. Second, the impact of the current coefficient to process is suppressed by choosing the corresponding weight w_0 to be much lower than that of the immediate vertical neighbors. This helps to further discriminate between correlated signal coefficients and isolated noise coefficients.

For all subsequent thresholding iterations, a different design resulted in a unique set of weights $\mathbf{w}^{(i > 0, l)}$ independent of the subband orientation. Since a thresholded band is sparsely populated, a less confident variance estimation is obtained by using only directly neighbored coefficients ($w_1 = w_2 = w_3 \neq 0$). Thus, the influence of the parent coefficient (and its neighbors) in the variance estimation has been significantly increased in that case, at least for the lowest levels where the parent-child dependency is most helpful to distinguish between signal components and coefficients related to noise specks in flat regions. Note, that the main task of the iterative process is to eliminate these remaining noise fragments.

The universal constant $\lambda^{(0)}(\mathcal{C})$ used in definition (6) for the initial thresholding has been numerically optimized for a certain class \mathcal{C} of natural images of size 512×512 pixels. For all remaining thresholding iterations ($i > 0$), we have chosen $\lambda^{(i)}(\mathcal{C}) = 1/2\lambda^{(i-1)}(\mathcal{C})$. Halving the proportional factor of the local variance estimation from one iteration step to the next corresponds to the fact that the estimates become more and more unreliable.

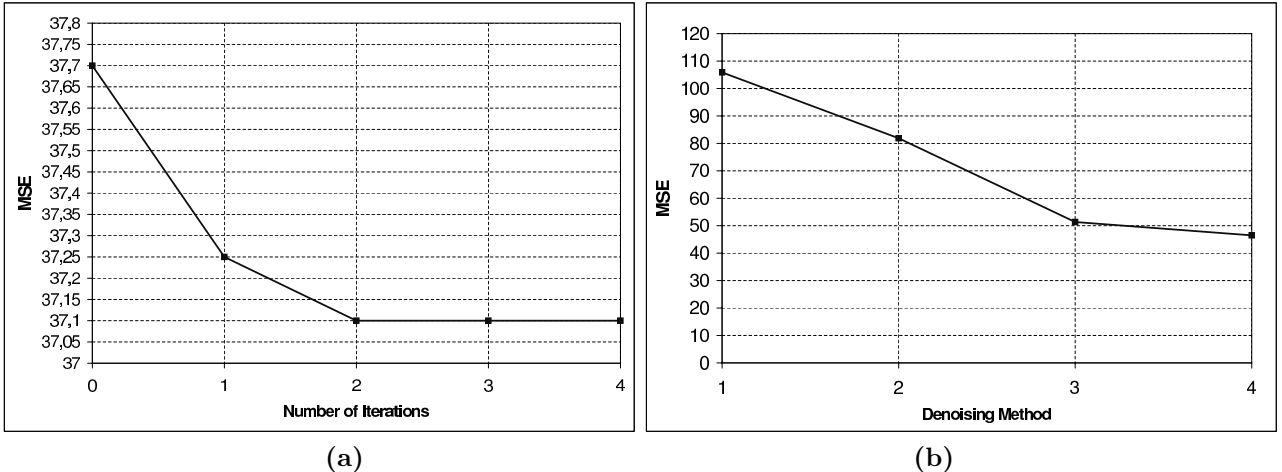


Figure 3. Denoising of the *Lena* image. (a) Noise $\sigma_n = 20$: MSE vs. number of iterations used for our proposed algorithm. (b) Noise $\sigma_n = 25$: MSE vs. different denoising methods; *Method 1*: Soft-Thresholding (ST) using DWT; *Method 2*: ST using NS-DWT; *Method 3*: Simplified CBID using NS-DWT; *Method 4*: Proposed CBID using NS-DWT.

3. EXPERIMENTAL RESULTS

Simulations have been carried out for a set of various test images. In this presentation, we confine ourselves to results obtained for *Lena* and *Barbara*, both of size 512×512 pels. For all experiments, a four level NS-DWT based on the biorthogonal 9/7-tap spline filter³ has been used. For the actual thresholding we employed soft-thresholding as specified in Eq. (2). The estimation $\hat{\sigma}_n^2$ of the ‘real’ noise variance σ_n^2 has been performed by using the robust median estimator in the highest subband (highpass filtered in both directions, $l = 0$, $o = D$) of each of the four different ‘branches’ of the whole quadtree of subbands as proposed by Donoho et al.^{4,5}:

$$\hat{\sigma}_n = \frac{1}{0.6745} \text{Median}(|Y^{(0,D)}(j,k)|)_{0 \leq j,k < \frac{N}{2}}.$$

In our first experiment, we tried to evaluate the impact of the iteration. The graph in Fig. 3 (a) shows the reduction in MSE measured for up to 4 iterations using the *Lena* image with noise $\sigma_n = 20$. It can be observed that the iteration process converges very quickly after 2–3 iterations resulting in an objective gain of up to nearly 2% MSE reduction. However, the gain in subjective quality is much higher due the elimination of visually annoying specks in flat areas.

Fig. 3 (b) shows the MSE reduction for different denoising methods using *Lena* corrupted by noise $\sigma_n = 25$. *Method 1* relates to the basic soft-thresholding method using a critically downsampled DWT and a global uniform threshold, as originally proposed by Donoho.⁵ *Method 2* corresponds to the extension of *Method 1* described in section 2.2. As can be seen from the graph, by using the NS-DWT instead of the DWT, a MSE reduction of approx. 25% can be achieved by this so-called ‘second-generation denoising method’. *Third-generation denoising methods* are characterized by using a spatially adaptive threshold and *Method 3* is actually a very simple representative of this kind of methods. It uses in addition to *Method 2* a flat, *i.e.*, uniformly weighted 3×3 context for a context-based threshold selection in an one-pass thresholding operation. Most remarkable is the fact, that this relatively simple extension yields an additional MSE reduction of more than 35%. *Method 4* in Fig. 3 (b) corresponds to our proposed method *CBID* using 3 iterations. Compared to its simple counterpart of *Method 3*, additional 10% MSE reduction can be achieved, such that in comparison to the first-generation method the performance benefit amounts to more than 50%.

Table 1 finally compares our proposed method (CBID) against the recently published adaptive algorithm of Chang et al.² For both test images at various noise contamination levels CBID outperforms the reference method by 3.5–8.5% in terms of MSE performance. For a comparison of subjective quality, Fig. 4 shows a magnified part of the *Lena* image for $\sigma_n = 22.5$. Although both methods suffer slightly from ringing artifacts, the reconstruction obtained by CBID is subjectively more pleasing due to less blotchiness.

Table 1. MSE vs. noise standard deviation σ_n for two standard test images (512×512 pels) obtained by our proposed method (CBID) and the method presented in Chang et al.²

σ_n	<i>Lena</i>		<i>Barbara</i>	
	Chang ²	CBID	Chang ²	CBID
10.0	w/o	19.2	w/o	29.1
12.5	24.9	23.8	39.5	38.5
15.0	29.9	28.1	50.4	48.3
17.5	35.2	32.7	60.7	58.7
20.0	40.2	37.1	73.2	69.6
22.5	45.2	42.2	85.3	80.5
25.0	50.8	46.5	96.2	91.6

4. CONCLUSIONS

We have introduced a relatively simple context-based model for adaptive threshold selection within a wavelet thresholding framework. Estimations of local weighted variance with appropriately chosen weights are used to adapt the threshold to the local statistics of the underlying image. An iterative application of our method helps to considerably improve the subjective reconstruction quality. Simulation results have shown that despite its simplicity our proposed algorithm yields significantly better results both in terms of visual quality and mean squared error than those recently reported for other spatially adaptive methods.

REFERENCES

1. S. G. Chang, B. Yu and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising", *Proc. IEEE Int. Conf. on Image Processing*, Oct. 1998.
2. S. G. Chang, B. Yu and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising", *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, Sept. 2000.
3. A. Cohen, I. Daubechies, J.-C. Feauveau, "Biorthogonal Bases of Compactly Supported Wavelets", *Comm. on Pure and Appl. Math.* **45**, pp. 485–560, 1992.
4. R. R. Coifman and D. L. Donoho, "Translation-invariant denoising" in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim (Eds.), Berlin, Germany, Springer-Verlag, 1995.
5. D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation Via Wavelet Shrinkage", *Biometrika*, vol. 81, pp. 425–455, 1994.
6. X. Li and M. T. Orchard: "Spatially Adaptive Image Denoising under Overcomplete Expansions", *Proc. IEEE Int. Conf. on Image Processing*, Vancouver, 2000.
7. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.
8. D. Marpe and H. L. Cycon, "Very Low Bit-Rate Video Coding Using Wavelet-Based Techniques", *IEEE Trans. on Circ. and Sys. for Video Techn.*, vol. 9, no. 1, pp. 85–94, Feb. 1999.
9. D. Marpe, G. Blättermann, J. Ricke and P. Maaß: "A Two-Layered Wavelet-Based Algorithm for Efficient Lossless and Lossy Image Compression", *IEEE Transactions on Circ. and Sys. for Video Techn.*, vol. 10, no. 7, pp. 1094–1102, Oct. 2000.
10. M. K. Mihçak, I. Kozintsev, and K. Ramchandran, "Spatially Adaptive Statistical Modeling of Wavelet Image Coefficients and Its Application to Denoising," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 6, Mar. 1999, pp. 3253–3256.



Figure 4. Sample reconstruction of the *Lena* image (512×512 pels): (top left) original, (top right) corrupted by noise of standard deviation $\sigma = 22.5$, (bottom left) denoised by Chang et al.,² (bottom right) denoised by CBID.