Noise Reduction in Video Sequences Using Wavelet-Domain and Temporal Filtering

Aleksandra Pižurica, Vladimir Zlokolica and Wilfried Philips

Ghent University, Dept. Telecommunications and Information Processing,
Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

ABSTRACT

We develop a sequential wavelet domain and temporal filtering scheme, with jointly optimized parameters, which results in high-quality video denoising over a large range of noise levels. In this scheme, spatial filtering is performed by a spatially adaptive Bayesian wavelet shrinkage in a redundant wavelet representation. In the next filtering stage, a motion detector controls selective, recursive averaging of pixel intensities over time. The results demonstrate that the proposed filter outperforms recent single-resolution representatives as well as some recent motion-compensated wavelet based video filters.

We also analyze important practical issues for possible industrial applications. In particular, we investigate the performance degradations that result from making the wavelet domain filtering part less complex, by removing the redundancy of the representation and/or by replacing a sophisticated spatially adaptive shrinkage method by soft-thresholding.

Keywords: Video denoising, wavelets, Bayesian estimation, motion adaptation, recursive filtering.

1. INTRODUCTION

Numerous applications, including multimedia services, teleconferencing, surveillance, object tracking, medical and astronomical imaging..., utilize digital video. Usually, noise reduction can significantly improve visual quality of a video. Also, noise reduction is often crucial for the effectiveness of the subsequent processing tasks, like video coding. Among various noise sources, the more prevalent ones are the noise introduced by the camera, shot noise (caused by the electronic hardware) and the channel (thermal) noise [1]. Most noise sources are modelled well by additive white Gaussian noise model, which is also treated in this paper.

Noise filters for video, which make use of both spatial and time correlations among pixel intensities are in general called *spatio-temporal* or three-dimensional (3-D) filters. Numerous existing approaches range from lower complexity solutions like 3-D weighted mean [2], 3-D rational [3] and 3-D order-statistic algorithms [4, 5] to sophisticated Bayesian methods based on 3-D Markov models [6, 7]. Approaches that attempt at taking full advantage of the time-redundancy incorporate motion detection and compensation [8] or at least use motion detection and some special operations in case of detected motion [3, 9]. An effective way to minimize the number of frames to be stored is *recursive* filtering [9–11]. A common approach is also to apply a 2-D spatial filter and 1-D temporal filter separately [11] and most often *sequentially* [9, 10]. The motivation is: after spatial denoising, detection and estimation of motion are facilitated and thus the temporal filtering is more effective. A systematic overview of 3-D filters is in [1].

Above listed video filters are single-resolution solutions. In contrast to still image denoising literature, relatively few publications have addressed so far *multiresolution* video denoising. Roosmalen *et al* [12] proposed video denoising by thresholding the coefficients of a specific 3D multiresolution representation, which combines 2-D steerable pyramid [13] decomposition (of the spatial content) and a 1-D wavelet decomposition (in time). Related to this, Selesnick and Li [14] investigated wavelet thresholding in a non-separable 3-D dual-tree complex wavelet representation [15]. A motion adaptive filtering scheme, which uses 2-D wavelet transform of several successive video frames is proposed in [16]. We proposed recently [17] a filtering scheme where 2-D wavelet denoising is followed by selective, recursive temporal filtering.

Send correspondence to Aleksandra.Pizurica@telin.ugent.be.

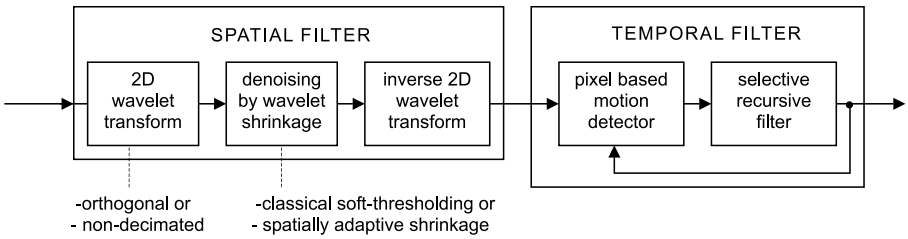

Figure 1. The analyzed denoising scheme.

The main motivation for the sequential wavelet domain and temporal filtering approach is the following. Since the pioneering thresholding works [18, 19], many sophisticated, spatially adaptive wavelet shrinkage techniques were developed and led to impressive results in still image denoising (see e.g., [20–24]). The extension of such approaches to 3-D in the wavelet domain (being undoubtedly an interesting research topic) imposes also a significant increase in complexity, which seems less attractive for (near) real-time applications. An alternative solution that we proposed in [17], where 2-D wavelet denoising is followed by a motion adaptive temporal filter, brings practically negligent increase in complexity over 2-D wavelet filtering, while it improves significantly the resulting visual quality. Moreover, as compared to other existing wavelet domain filters for video [12, 14, 16], the scheme of [17] minimizes the frame storage.

In this paper, we further develop and improve our initial filtering scheme from [17] and we perform an extensive experimental evaluation for video sequences corrupted by different noise levels. The main improvement consists in optimizing the temporal filtering part, the parameters of which were earlier chosen ad-hoc. We also analyze two important practical issues, which are of particular interest for industrial applications. Firstly, there is a question of choosing the wavelet domain filtering method. Many different approaches exist. The question is how important is it to use a sophisticated estimation method (like in [17]) instead of a simple thresholding (given the fact that the imperfections of the wavelet filter are partly compensated by the temporal filter)? Secondly, our initial filtering scheme was previously tested with the non-decimated wavelet transform only. Real-time applications would preferably use the orthogonal transform. It is interesting to investigate how significant are the performance degradations in this case. Well known comparisons between redundant and orthogonal denoising schemes for still images do not hold here, because the temporal filter reduces the denoising artifacts of the 2-D wavelet filter.

The main novelties and contributions of this paper can be summarized as follows. We develop a sequential wavelet domain and temporal filtering scheme, with jointly optimized parameters, which results in high-quality video denoising over a large range of tested noise levels. Moreover, we analyze the proposed sequential filtering scheme with two types of wavelet based denoising

- denoising by soft-thresholding with the uniform threshold
- spatially adaptive Bayesian wavelet shrinkage

and we analyze the use of

- orthogonal and
- non-decimated

wavelet transform. The video denoising scheme is depicted in Fig. 1. We test the method on four different video sequences corrupted by different amounts of additive white Gausian noise. The results demonstrate that the proposed filter outperforms not only recent single-resolution representatives [3, 5] but it also compares favorably with some recent motion-compensated wavelet based video filters, like [16].

The paper is organized as follows. Section 2 reviews the wavelet transform and explains two considered wavelet based denoising approaches. Section 3 addresses temporal filtering part and the joint parameter optimization. The results are presented and discussed in Section 4. Section 5 concudes the paper.

2. 2-D WAVELET DOMAIN NOISE FILTERING

Here we review briefly the wavelet decomposition and the denoising approaches that are considered in our method. For a comprehensive treatment of wavelets see, e.g., [25–27].

2.1. The Discrete Wavelet Transform (DWT)

From the signal processing point of view, the discrete wavelet transform is a *filter bank* algorithm iterated on the low-pass output [26]. A filter bank is a pair of lowpass and highpass filters followed by downsampling by two. The lowpass filtering produces an *approximation* of the signal, which is expressed by the *scaling* coefficients, while the highpass filtering reveals the *details* (i.e., the differences between two successive approximations) that are expressed by the *wavelet* coefficients. At the reconstruction, the scaling and the wavelet coefficients are first up-sampled (by introducing a zero between each two samples) and then filtered with a lowpass and a highpass filter, respectively, followed by summation of the filtered outputs. If the wavelet transform is *orthogonal*, the reconstruction highpass and the lowpass filter coefficients are simply the mirrored versions of their counterparts at the decomposition stage. The conventional *separable* two-dimensional (2D) DWT follows from applying the filter bank algorithm successively to the rows and to the columns of an image.

The above described DWT is critically sampled (non-redundant). It is well known that noise suppression improves in a *redundant* representation. In this respect, two approaches are common: (1) *Cycle spinning* [28]: apply the orthogonal DWT to several cyclically shifted image versions and average over unshifted denoising results and (2) Denoising in a *non-decimated* (*stationary*) wavelet representation, which is computed with the algorithm *à trous* [26].

Here we consider the following options: non-decimated or orthogonal wavelet transform. In both cases the experiments were done with the Daubechies least asymmetrical wavelet (symmlet) with eight vanishing moments [25].

2.2. On marginal statistics of image wavelet coefficients

For natural noise-free images, the wavelet coefficient histograms in each subband are typically long-tailed and sharply peaked at zero. A common marginal prior is generalized Laplacian (also called generalized Gaussian) density [26]

$$p(y) = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})} \exp(-\lambda|y|^{\nu}), \quad \lambda, \nu > 0, \tag{1}$$

where $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$ is the Gamma function. For natural images, the shape parameter ν is typically $\nu \in [0, 1]$. The variance and the courtosis of a generalized Laplacian signal are [29]

$$\sigma_y^2 = \frac{\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})}, \quad \kappa_y = \frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})}. \tag{2}$$

In case of additive white Gaussian noise, the model parameters ν and λ are estimated from the noisy coefficient histogram using the following equations [29, 30]

$$\frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})} = \frac{m_4 + 3\sigma^4 - 6\sigma^2\sigma_w^2}{(\sigma_w^2 - \sigma^2)^2}, \qquad \lambda = \Big((\sigma_w^2 - \sigma^2)\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}\Big)^{-1/2}, \tag{3}$$

where σ_w^2 and $m_{4,w}$ denote the variance and the fourth moment of the noisy histogram, respectively. A special case in the family (1) with $\nu = 1$, called Laplacian density, is often used due to analytical tractability. The scale parameter is then estimated as

$$\lambda = [0.5(\sigma_w^2 - \sigma^2)]^{-1/2}. \tag{4}$$

In denoising experiments, this simplification usually does not produce a significant degradation in performance.

2.3. Denoising by wavelet shrinkage

We assume that each input video frame $\mathbf{f} = [f_1, ..., f_n]$ is contaminated with additive white Gaussian noise of zero mean and variance σ^2. Due to linearity of the wavelet transform, the noise remains additive in the transform domain as well

$$w_i = y_i + \epsilon_i, \quad i = 1, ..., n \tag{5}$$

where y_i are unknown noise-free wavelet coefficients and ϵ_i are noise contributions. If the wavelet transform is orthogonal, ϵ_i are independent identically distributed (i.i.d.) normal random variables $\epsilon_i \sim N(0, \sigma^2)$. In case where the input standard deviation σ is not known, one usually estimates it as the median absolute deviation of the highest-frequency subband coefficients divided by 0.6745 [18].

Regardless of the type of the employed discrete wavelet transform (e.g., critically sampled or non-decimated), noise reduction is commonly done by *wavelet shrinkage*: the magnitude of each coefficient is reduced by a given amount depending on the noise level and depending on how likely it is that a given coefficient represents an actual discontinuity. A common shrinkage approach is *thresholding* [18, 19], which sets the wavelet coefficients with "small" magnitudes to zero while keeping ("hard-thresholding") or shrinking in magnitude ("soft-thresholding") the remaining ones. Thresholding with a uniform per subband threshold is attractive due to its simplicity. Of course, the performance is limited and the denoising quality is often not satisfactory. Many spatially adaptive wavelet shrinkage methods (e.g., [20–24]) have been developed over recent years.

In this paper we consider two approaches:

- *Soft-thresholding with a uniform threshold* in each subband - as a representative of this approach we use the threshold selection of [31] which is, for natural images, optimal in terms of mean squared error.
- *Spatially adaptive wavelet shrinkage* - we use here our recently proposed approach from [17, 32].

Both of the here considered denoising methods are Bayesian - they make use of prior distributions of noise-free wavelet coefficients. The two methods are reviewed briefly below.

2.3.1. A Bayesian thresholding approach

The classical soft-thresholding [18, 19] estimates the noise-free coefficient value as

$$\hat{y}_i = \begin{cases} \text{sgn}(w_i)(|w_i| - T), & \text{if } |w_i| > T; \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Chang *et al* [31] found experimentally that under the prior (1), the threshold value that minimizes the minimum mean squared error is

$$T = \sigma_y / \sigma^2. \tag{7}$$

For the classical soft-thresholding with a uniform threshold per subband, the above threshold selection yields for natural images the best results both in terms of mean squared error and visually (for details, see [31]).

2.3.2. A Bayesian spatially adaptive shrinkage approach

For spatially adaptive shrinkage, we use here our recent method from [17, 32], which shrinks each wavelet coefficient according to probability that it represents a "signal of interest". We define this signal of interest as a significant noise-free coefficient component, the magnitude of which is above a certain threshold. The estimate how probable it is that a coefficient represents a "signal of interest" we rely on three sources:

- the coefficient value w_l,
- a *local spatial activity indicator* (LSAI) defined as the locally averaged coefficient magnitude in a small window $\delta(l)$: $z_l = \sum_{k \in \delta(l)} |w_k|$ and
- the global statistical distribution of the coefficients in a given subband.

If we define two hypotheses H_1 - "signal of interest is present: $|y| > T$ " and H_0 - "signal of interest is absent: $|y| \leq T$", our shrinkage estimator from [17] is

$$\hat{y}_l = \frac{\rho \xi_l \eta_l}{1 + \rho \xi_l \eta_l} w_l, \tag{8}$$

where

$$\rho = \frac{P(H_1)}{P(H_0)}, \quad \xi_l = \frac{p(w_l|H_1)}{p(w_l|H_0)} \quad \text{and} \quad \eta_l = \frac{p(z_l|H_1)}{p(z_l|H_0)}. \tag{9}$$

and where $p(w_l|H_0)$ and $p(w_l|H_1)$ denote the conditional probability density functions of the noisy coefficients given the absence and given the presence of a signal of interst. Similarly, $p(z_l|H_0)$ and $p(z_l|H_1)$ denote the corresponding conditional probability density functions of the local spatial activity indicator.

We estimate the required ratios ρ, ξ_l and η_l directly from the observed image coefficients. In particular, for the prior (1), with $\nu = 1$, the conditional densities of *noise-free* coefficients are

$$p(y|H_0) = \begin{cases} A_0 \exp(-\lambda|y|^\nu) & \text{if } y \leq T, \\ 0 & \text{if } y > T, \end{cases} \tag{10}$$

$$p(y|H_1) = \begin{cases} 0 & \text{if } y \leq T, \\ A_1 \exp(-\lambda|y|^\nu) & \text{if } y > T. \end{cases} \tag{11}$$

where it is straightforward to show that $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T} - 1)$ and $A_1 = (\lambda/2)e^{\lambda T}$. Since the additive white noise model $w = y + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$ is assumed, the densities of noisy coefficients $p(w|H_0)$ and $p(w|H_1)$ result from convolving the normal density $N(0, \sigma^2)$ with $p(y|H_0)$ and with $p(y|H_1)$, respectively. One can carry out the above described procedure numerically. We derived analytical expressions in [32]. Further on, we simplify the statistical characterization of z_l considerably assuming that all the coefficients within the small window are equally distributed and conditionally independent. With these simplifications, and if we denote the coefficient magnitude by $m_l = |w_l|$, we have that $f(Nz_l|H_1)$ is given by N convolutions of $f(m_l|H_1)$ with itself and $f(Nz_l|H_0)$ is given by N convolutions of $f(m_l|H_0)$ with itself, where $p(m_l|H_{0,1}) = 2p(w_l|H_{0,1})$, $m_l > 0$.

Finally, we estimate the prior ratio $\rho = P(H_1)/P(H_0)$ as follows. Starting from $P(H_1) = \int_{-\infty}^{-T} p(y)dy + \int_T^{\infty} p(y)dy$, for the prior (1) we derive [30, 32]

$$\rho = \frac{P(H_1)}{P(H_0)} = \frac{1 - \Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)}{\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)}, \tag{12}$$

where $\Gamma_{inc}(x, a) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$ is the *incomplete gamma function*. For the Laplacian prior, with $\nu = 1$, the previous expression reduces to

$$\rho = \frac{P(H_1)}{P(H_0)} = \frac{\exp(-\lambda T)}{1 - \exp(-\lambda T)}. \tag{13}$$

An extensive analysis of the proposed method is in [30,32], where we show that the optimal value of the threshold in terms of the mean squared error is $T = \sigma$, and we also show that on standard test images the proposed method yields results that are currently among the best reported.

Our method is implemented with the generalized Laplacian prior. The Laplacian prior ($\nu = 1$), yields on still images minor performance degradation. Usually, the peak signal to noise ratio (PSNR) drops for 0.1 to 0.3dB..

3. TEMPORAL FILTERING

In video denoising literature it is well known that spatial denoising alone produces annoying artifacts and unsatisfactory video quality [1]. This holds even in case where a sophisticated wavelet domain denoising method is used for spatial filtering [17]. The main reason is that the residual noise and denoising artifacts differ from frame to frame causing in this way an unpleasant "flickering" effect. In the proposed video denoising scheme (Fig. 1), a temporal filter suppresses the residual noise and artifacts produced by the 2-D wavelet domain filter.

3.1. A selective recursive temporal filter

In the sequential spatio-temporal denoising scheme from Fig. 1, motion detection and temporal filtering are performed over spatially denoised frames. The realization of a motion adaptive temporal filter here benefits from the use of a high-quality spatial denoising. After 2-D wavelet denoising, the inter-frame differences due to remaining noise (and due to artifacts and degradations) are relatively small as compared to the actual inter-frame differences produced by motion and by changes of the video content in time. Due to this, the motion detector in the analyzed scheme can be quite simple, yet efficient.

We use a pixel-based motion detector and we switch off the recursive filtering at those positions (in space and time) where motion is detected. A formal description follows.

Let $\mathbf{f}^k$ denote the k-th frame of a *noise-free* video sequence and $\mathbf{d}^k = \mathbf{f}^k + \mathbf{n}^k$ the corresponding noisy frame, where $\mathbf{n}^k$ is the noise field. Further on, let

$$\hat{\mathbf{f}}^{2D,k} = [\hat{f}_1^{2D,k}, ..., \hat{f}_L^{2D,k}] \tag{14}$$

denote the k-th *2-D denoised frame*. Define the *motion field* $\mathbf{m}^k = [m_1^k, ..., m_L^k]$ of the k-th frame with respect to the previous frame as follows

- $m_l^k = 0$ if there is *no* (significant) motion from the frame $k-1$ to the frame k at the spatial position l, meaning that $f_l^k \approx f_l^{k-1}$.
- $m_l^k = 1$ if there *is* motion from the frame $k-1$ to the frame k at the spatial position l, meaning that f_l^k and f_l^{k-1} differ significantly.

We estimate this motion field from the denoised frames as

$$\hat{m}_l^k = \begin{cases} 0, & \text{if } |\hat{f}_l^{2D,k} - \hat{f}_l^{3D,k-1}| < T, \\ 1, & \text{otherwise,} \end{cases} \tag{15}$$

where T is the motion threshold. At the spatial positions where no motion was detected ($m_l^k = 0$), we apply recursive time averaging $\hat{f}_l^{3D,k} = \alpha \hat{f}_l^{2D,k} + (1-\alpha)\hat{f}_l^{3D,k-1}$, with $0 \leq \alpha \leq 1$. At positions where motion is detected, the temporal filtering is switched off and the spatially filtered value is kept: $\hat{f}_l^{3D,k} = \hat{f}_l^{2D,k}$. Thus the 3-D filtered pixel intensity is:

$$\hat{f}_l^{3D,k} = (1 - m_l^k)[\alpha \hat{f}_l^{2D,k} + (1-\alpha)\hat{f}_l^{3D,k-1}] \quad + \quad m_l^k \hat{f}_l^{2D,k}. \tag{16}$$

Note that this recursive filter accumulates and averages the pixel intensities at a given position from *all the previous* frames if the motion was not present at that position. The detection of a motion resets the filter.

3.2. Experimental parameter optimization

The above described motion-adaptive recursive temporal filter involves two parameters: the motion threshold T and the weighting factor α for the recursive filtering. To keep the method simple, here we shall use fixed parameter values. We optimize jointly (α, T) to provide a robust and good denoising performance for a range of noise levels and for different types of sequences.

As the optimization criterion, we choose here to maximize the mean PSNR gain that the complete filter achieves over 2-D spatial filtering alone. According to this criterion, we optimize the parameters experimentally. We used four representative sequences (*flower*, *miss America*, *salesman* and *tennis*) and each of these was corrupted with three different noise levels (standard deviations $\sigma = 10$, 15 and 20). For each of these 12 noisy sequences we ran simulations with 2-D filtering only and with the spatio-temporal filtering for different (α, T) pairs. The range of α values was 0 to 1 with the step 0.1 and the range of T values was 5 to 25 with the step 5. For each pair (α, T) and for each of the 12 sequences we calculate the mean PSNR gain of the temporal filter with respect to spatial filtering alone. Several examples illustrating the mean PSNR gain as a function of α and T are shown in Fig. 2. By averaging (α, T) values which yielded the maximum PSNR gain on different sequences, we got the following values: $\alpha = 0.6$ and $T = 23$. We used these values for all the subsequent results in this paper.

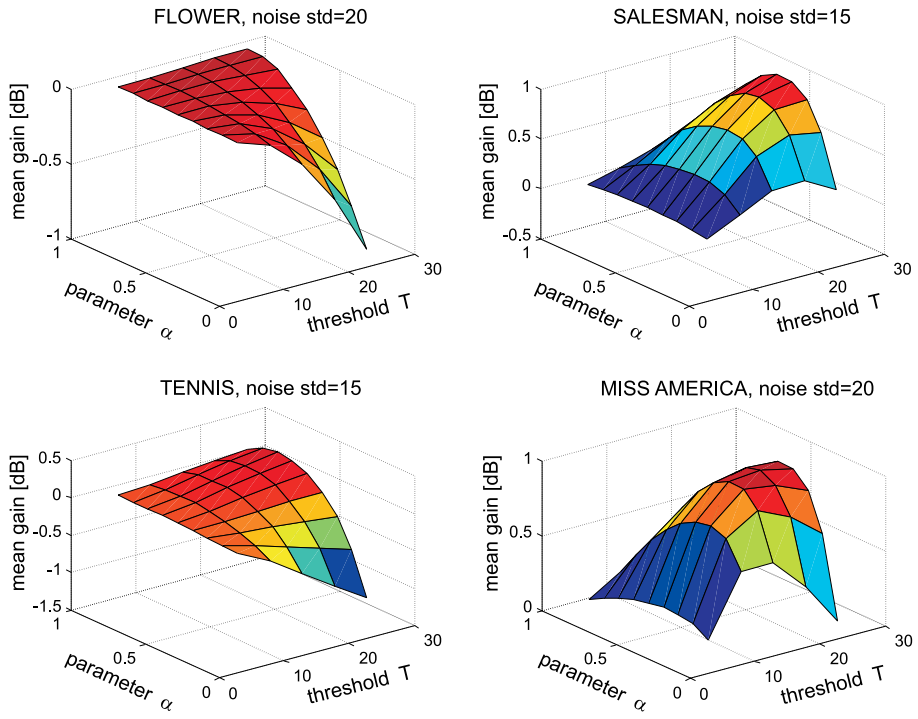

Figure 2. Examples showing mean PSNR gain of the temporal filtering part, as a function of the parameters α and T.

4. RESULTS AND DISCUSSION

4.1. Implementation with a non-decimated wavelet transform

Here we present the results of the proposed method implemented with a non-decimated wavelet transform with 4 decomposition levels. The wavelet used is *symmlet* [25, 26] with eight vanishing moments. We abbreviate the proposed sequential wavelet domain and temporal filtering as *seqwt*. Its implementation with the spatially adaptive wavelet shrinkage (Section 2.3.2) is abbreviated by *seqwt-sas* and the implementation with the simple wavelet thresholding (Section 2.3.1) is abbreviated by *seqwt-thr*.

We used four test sequences *Miss America*, *salesman*, *tennis* and *flower*, each of which was corrupted by additive white Gaussian noise with standard deviations $\sigma = 10$, 15 and 20. Fig. 3 illustrates resulting PSNR of the proposed method in comparison with three recent filters: 3-D rational filter (*3D rational*) [3] , 3-D K-Nearest Neighbors filter (*3D knn*) [5] and the Motion Compensated Wavelet based filter (*mcw*) of [16]. The results demonstrate a significant improvement of the new filter over all the other tested ones. In terms of *visual quality*, the proposed filter also outperforms the reference filters from [3, 5, 16] on all test sequences. This is illustrated with three examples in Fig. 4, Fig. 5 and Fig. 6. The advantage of the new filter is even more evident when viewing video.

Fig. 3 compares also the performance of two versions of the analyzed sequential scheme: (1) with spatially adaptive wavelet shrinkage (*seqwt-sas*) and (2) with soft-thresholding (*seqwt-thr*). The performance degradations that result from replacing spatially adaptive wavelet shrinkage by simple wavelet thresholding depend on the sequence content. On some sequences, the PSNR decrease is less than 0.3dB (see *flower* in Fig. 3(a)). On other sequences, the PSNR drops even more than 1.5dB (e.g., last 50 frames of *tennis* in Fig. 3(b)). Our experiments showed that the corresponding degradations of visual quality are significant.

4.2. Implementation with orthogonal transform

Now we apply 2-D wavelet denoising in the orthogonal wavelet representation instead of using the non-decimated one and we investigate the resulting performance degradations of the proposed 3-D filter. In case of 2-D wavelet

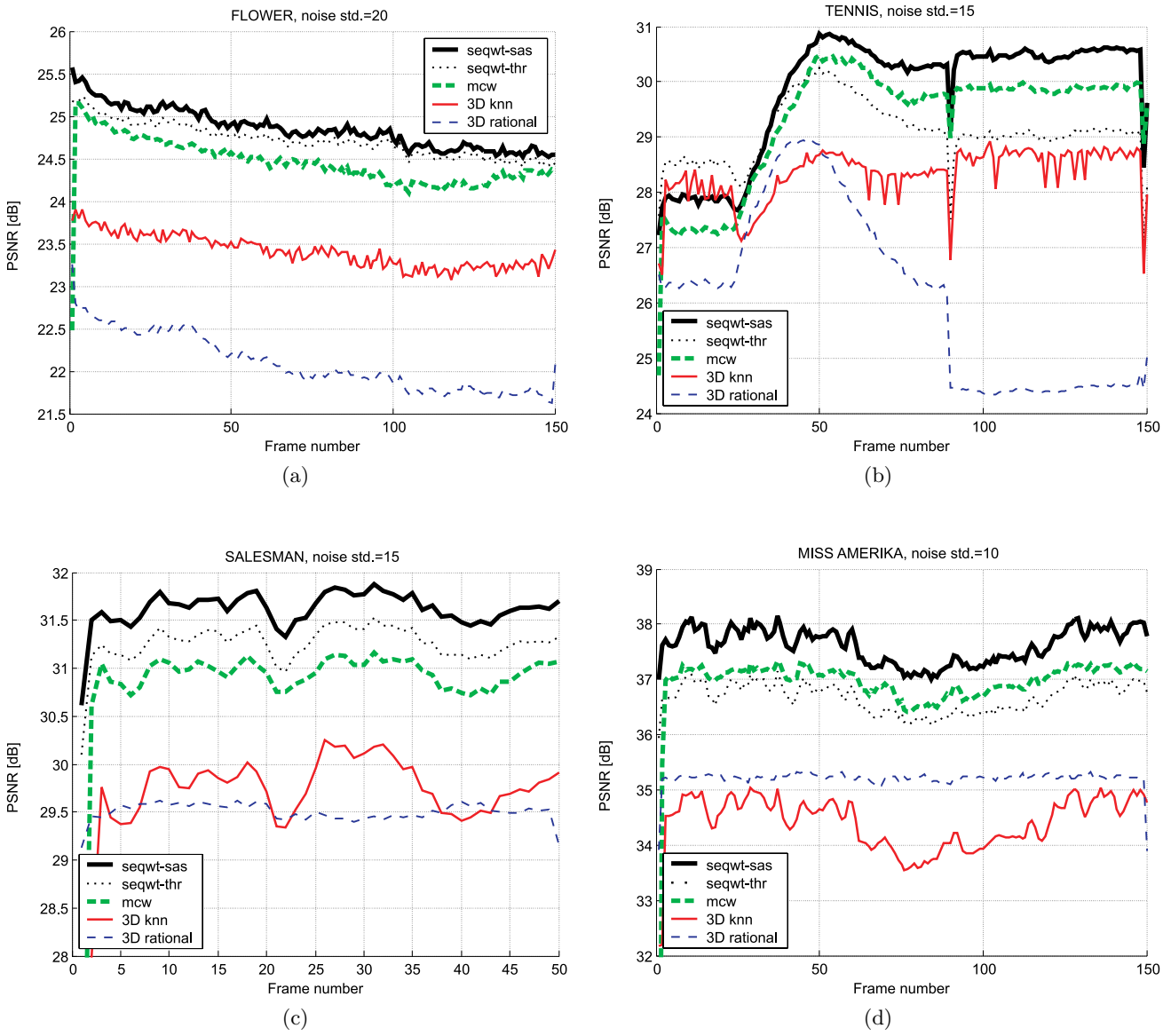

Figure 3. Quantitative performance of the proposed sequential wavelet domain and temporal filtering in comparison with other recent video filters. *seqwt-sas*: the proposed method with spatially adaptive wavelet shrinkage; *seqwt-thr*: a simplification of the proposed method using wavelet thresholding; *mcw* [16]; *3D knn* [5] and *3D rational* [3]. Sequences (a) *flower*, $\sigma = 20$, (b) *tennis*, $\sigma = 15$, (c) *salesman*, $\sigma = 15$ and (d) *miss America*, $\sigma = 10$.

domain filtering alone the performance degradations that result from using an orthogonal transform instead of a non-decimated one are usually ∼1dB (see e.g., [22,30]). In the proposed 3-D denoising scheme the corresponding degradations are smaller because the temporal filtering part reduces the artifacts of the wavelet domain filter.

In order to use the classical orthogonal wavelet transform, we cropped the frame sizes in each direction to the closest size that is a power of two. We used spatially adaptive shrinkage from Section 2.3.2. PSNR values for three test sequences are shown in Fig. 7. The results demonstrate that degradations imposed by using the orthogonal wavelet transform range from 0.25dB to 0.8dB on different test sequences. Visual differences in case of *tennis* and *flower* sequences are difficult to notice. For the *salesman* sequence the differences are clearly visible only when viewing the video sequence. Our experiments also showed that visual quality degrades less by

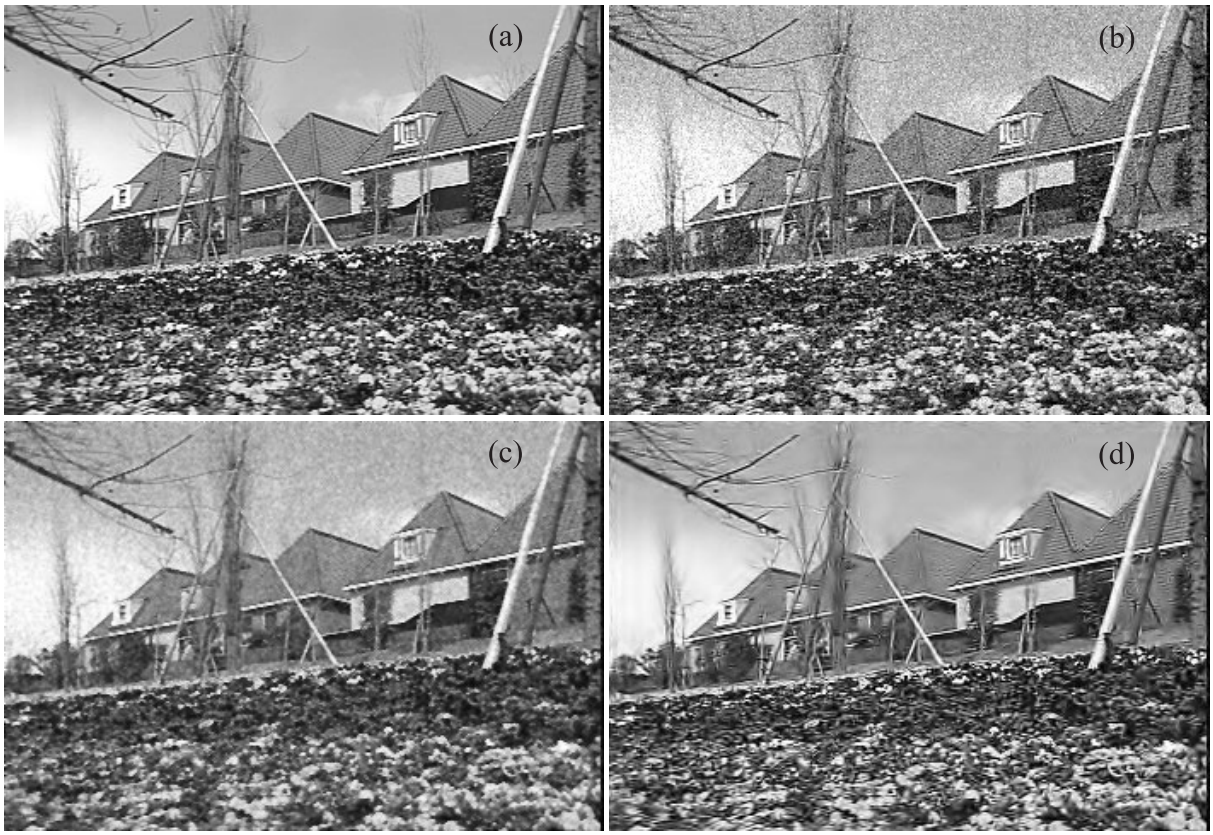

Figure 4. (a) A noise-free frame of the *flower* sequence. (b) Noisy frame, $\sigma = 20$. (c) The result of the 3-D rational filter of [3]. (d) The result of the proposed *seqwt-sas* filter.

removing the redundancy of the wavelet representation than in the case where spatially adaptive shrinkage is replaced by thresholding in a redundant representation.

5. CONCLUSION

We developed a sequential spatio-temporal scheme for video denoising, where 2-D wavelet denoising is followed by selective recursive temporal filtering. To achieve a high-quality video denoising, we used a non-decimated wavelet transform and a spatially adaptive wavelet shrinkage method. The parameters of the temporal filter were optimized to maximize the mean PSNR gain for different test sequences. The proposed scheme outperforms tested single-resolution and wavelet domain spatio-temporal filters both in terms of PSNR and visually.

We also investigated performance degradations that result from making the wavelet domain filtering part less complex, by removing the redundancy of the representation and by replacing spatially adaptive shrinkage method by soft-thresholding. The latter simplification degrades the performance significantly on some test sequences. We conclude that the use of a sophisticated spatially adaptive wavelet denosing is essential in the analyzed scheme.

Further improvements are expected from using a more sophisticated recursive temporal filter. One possibility in this respect is Kalman filtering, like in [11]. Further research will be done towards integrating the recursive temporal filter in the wavelet domain.

REFERENCES

1. J. Brailean, R. Kleihorst, S. Efstratiadis, A. Katsaggelos, and A. Lagendijk, "Noise reduction filters for dynamic image sequences: A review," *Proceedings of the IEEE* **83**(9), pp. 1272–1292, 1995.

Figure 5. (a) A noise-free frame of the *Miss America* sequence. (b) Noisy frame, $\sigma = 15$. (c) The result of the spatio-temporal filter from [5]. (d) The result of the proposed *seqwt-sas* filter.

2. M. Özkan, A. Erdem, and A. Sezan, M.I. Tekalp, "Efficient multiframe wiener restoration of blurred and noisy image sequences," *IEEE Trans. Image Proc* **1**, pp. 453–476, 1992.
3. F. Cocchia, S. Carrato, and G. Ramponi, "Design and real-time implementation of a 3-D rational filter for edge preserving smoothing," *IEEE Trans. Consumer Electron.* **43**, pp. 1291–1300, Nov. 1997.
4. G. Arce, "Multistage order statistic filters for image sequence processing," *IEEE Trans. Acoust., Speech, Signal Process.* **39**, pp. 1147–1163, May 1991.
5. V. Zlokolica and W. Philips, "Motion and detail adaptive denoising of video," in *IS&T/SPIE Symposium on Electronic Imaging*, (San Jose, California, USA, Jan. 2004. (accepted)).
6. L. Hong and D. Brzakovic, "Bayesian restoration of image sequences using 3-d markov random fields," in *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, pp. 1413–1416, (Glasgow, UK, May 1989.).
7. J. Brailean and A. Katsaggelos, "Simultaneous displacement estimation and restoration of noisy-blurred image sequences," *IEEE Trans. Image Process.* **4**, pp. 1236–1251, Sept. 1995.
8. G. de Haan, "Ic for motion-compensated de-interlacing,noise,reduction and picture rate conversion," *IEEE Trans. on Cosumers Electron.* **45**, pp. 617–623, Aug. 1999.
9. K. Jostschulte, A. Amer, M. Schu, and H. Schroder, "Perception adaptive temporal tv-noise reduction using contour preserving prefilter techniques," *IEEE Trans. Consumer Electron.* **44**(3), pp. 1091–1096, 1998.

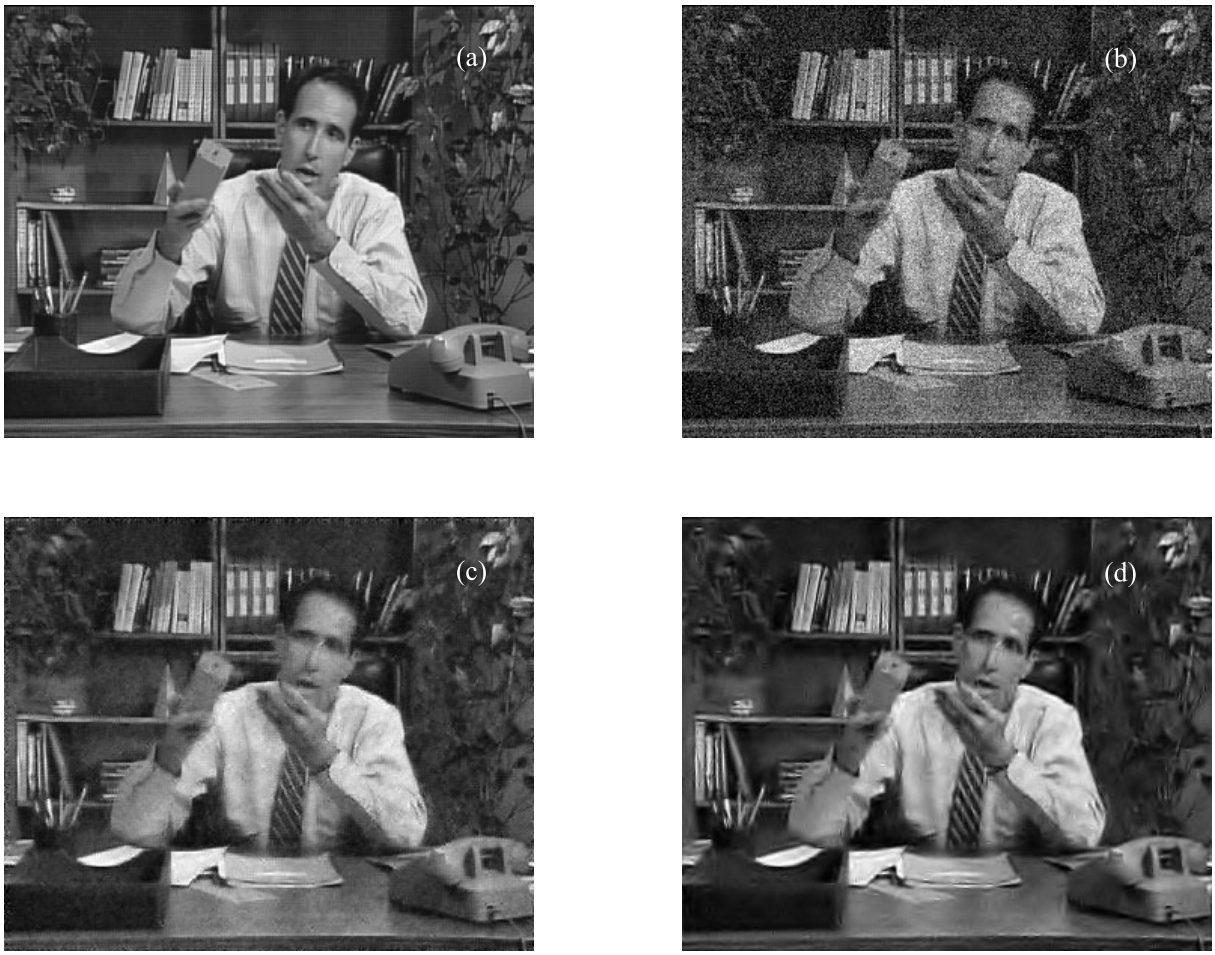

Figure 6. (a) A noise-free frame of the *salesman* sequence. (b) Noisy frame, $\sigma = 20$. (c) The result of the motion compensated wavelet filter from [16]. (d) The result of the proposed *seqwt-sas* filter.

10. A. Katsaggelos, J. Driessen, S. Efstratiadis, and R. Lagendijk, "Temporal motion compensated noise filtering of image sequences," in *Proc. SPIE Vis. Comm. and Image Process.*, pp. 61–70, (Boston, MA, Nov. 1989.).
11. D. Dugad and N. Ahuja, "Video denoising by combining kalman and wiener estimates," in *Proc. IEEE Internat. Conf. on Image Proc.*, pp. 152–156, (Cobe, Japan, 1999.).
12. P. Roosmalen, S. Westen, R. Lagendijk, and J. Biemond, "Noise reduction for image sequences using an oriented pyramid thresholding technique," in *Proc. IEEE Conf. on Image Process.*, pp. 375–378, (Lausanne, Switzerland, Sep. 1996).
13. E. P. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inform. Theory* **38**, pp. 587–607, 1992.
14. I. Selesnick and K. Li, "Video denoising using 2d and 3d dual-tree complex wavelet transforms," in *Proc. SPIE Wavelet Applications in Signal and Image Processing*, **5207**, (San Diego, Aug 4-8, 2003).
15. N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis* **10**, pp. 234–253, May 2001.
16. V. Zlokolica, A. Pižurica, and W. Philips, "Video denoising using multiple class averaging with multiresolution," in *International Workshop VLBV03*, (Madrid, Spain, Sep. 2003).

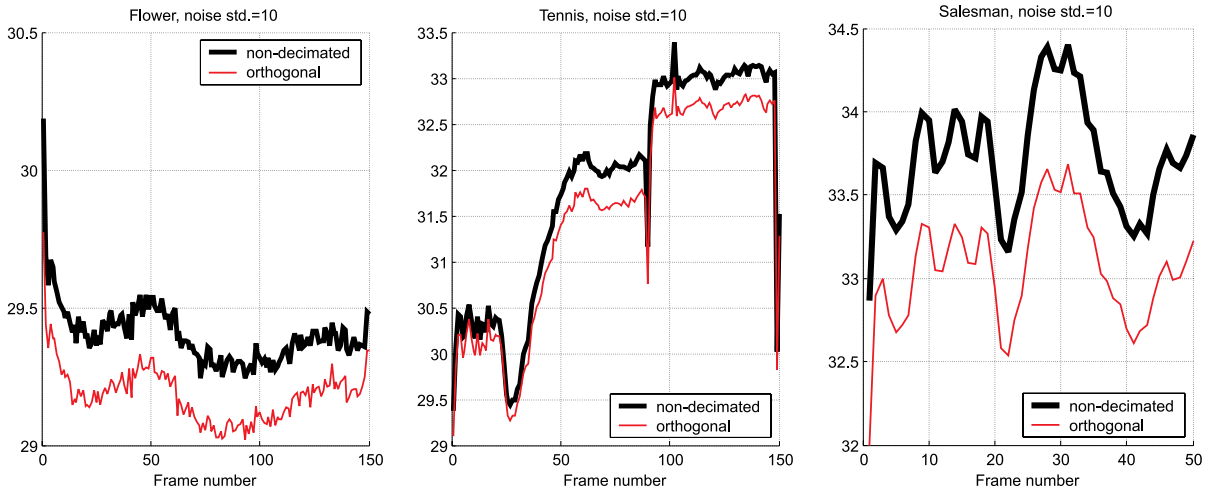

Figure 7. Performance of the proposed *seqwt-sas* filter with orthogonal and with non-decimated wavelet transform on parts of the *flower*, *tennis* and *salesman* sequences corrupted by additive white Gausian noise of standard deviation $\sigma = 10$. (Frames of each sequance were cropped to sizes that are powers of two).

17. A. Pižurica, V. Zlokolica, and W. Philips, "Combined wavelet domain and temporal video denoising," in *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance*, pp. 334–341, (Miami, FL,USA, July 21-22, 2003.).
18. D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika* **8**, pp. 425–455, 1994.
19. D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory* **41**, pp. 613–627, May 1995.
20. M. Malfait and D. Roose, "Wavelet-based image denoising using a markov random field a priori model," *IEEE Trans. Image processing* **6**(4), pp. 549–565, 1997.
21. M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Proc. Lett.* **6**, pp. 300–303, Dec. 1999.
22. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Proc.* **9**, pp. 1522–1531, Sept. 2000.
23. A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for wavelet based bayesian image denoising," *IEEE Trans. Image Proc* **11**, pp. 545–557, May 2002.
24. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using gaussian scale mixtures in the wavelet domain," *IEEE Trans. Image Proc. (accepted)* .
25. I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.
26. S. Mallat, *A wavelet tour of signal processing*, Academic Press, London, 1998.
27. M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, 1995.
28. R. R. Coifman and D. L. Donoho, "Translation-invariant denoising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., pp. 125–150, (Springer Verlag, New York), 1995.
29. E. P. Simoncelli and E. H. Adelson, "Noise removal via bayesian wavelet coring," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, (Lausanne, Switzerland, 1996).
30. A. Pižurica, *Image Denoising Using Wavelets and Spatial Context Modeling*. PhD thesis, Ghent University, Belgium, 2002.
31. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Proc.* **9**, pp. 1532–1546, Sept. 2000.
32. A. Pižurica and W. Philips, "Adaptive probabilistic wavelet shrinkage for image denoising." Technical Report, June 2003, (submitted to *IEEE Trans. Image Proc.*).