Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising

Aleksandra Pižurica, *Member, IEEE* and Wilfried Philips, *Member, IEEE*

***Abstract*— We develop three novel wavelet domain denoising methods for *subband-adaptive*, *spatially-adaptive* and *multivalued* image denoising. The core of our approach is the estimation of the probability that a given coefficient contains a significant noise-free component, which we call "signal of interest". In this respect we analyze cases where the probability of signal presence is (i) fixed per subband, (ii) conditioned on a local spatial context and (iii) conditioned on information from multiple image bands. All the probabilities are estimated assuming a generalized Laplacian prior for noise-free subband data and additive white Gaussian noise. The results demonstrate that the new subband-adaptive shrinkage function outperforms Bayesian thresholding approaches in terms of mean squared error. The spatially adaptive version of the proposed method yields better results than the existing spatially adaptive ones of similar and higher complexity. The performance on color and on multispectral images is superior with respect to recent multiband wavelet thresholding.**

***Index Terms*— Image denoising, wavelets, Bayesian estimation, generalized likelihood ratio, color and multispectral images.**

I. Introduction

In image denoising, where a trade-off between noise suppression and the preservation of actual image discontinuities must be made, solutions are sought which can "detect" important image details and accordingly adapt the degree of noise smoothing. In the wavelet transform domain [1]–[4], noise reduction results from *shrinking* the noisy coefficient magnitudes: ideally, the wavelet coefficients that contain primarily noise should be reduced to negligible values while the ones containing a "significant" noise-free component should be reduced less [5]. A common shrinkage approach is thresholding [5]–[8], where the coefficients with magnitudes below a certain threshold are treated as "non significant" and are set to zero, while the remaining, "significant" ones are kept unmodified (hard-thresholding) or reduced in magnitude (soft-thresholding).

Shrinkage estimators can also result from a Bayesian approach [9]–[37], which imposes a prior distribution on noise-free data. Common priors for noise-free subband data include (generalized) Laplacian [2], [9], [16], [20], alpha-stable model [18], [19], double stochastic (Gaussian scale mixture) models [28]–[30] and mixtures of two distributions [10]–[15] where one distribution models the statistics of "significant" coefficients and the other one models the statistics of "insignificant" data. Combined with these marginal priors, Hidden Markov Tree (HMT) [25]–[27] and Markov Random Field (MRF) [33]–[36] models are often employed to incorporate inter- and intra-scale dependencies.

Regardless of the particular prior, Bayesian wavelet domain denoising methods have been developed along the following two main lines. The first class of methods optimizes the threshold selection for hard- or soft-thresholding [9]–[12]. The second class of methods derives shrinkage functions by minimizing a Bayesian risk, typically under a quadratic cost function (minimum mean squared error - MMSE estimation [14]–[18]) or under a delta cost function (maximum a posteriori - MAP estimation [19], [20]). The above listed methods are *subband adaptive*: they are optimized with respect to the marginal subband statistics. The use of joint and bivariate statistics of wavelet coefficients is addressed in [21], [22], respectively. In practice, *spatially adaptive* Bayesian estimators are effective, where a given parameter of the marginal prior is refined with respect to the local spatial context [23]–[32].

In this paper we develop three novel Bayesian methods for *subband-adaptive*, *spatially-adaptive* and *multivalued* image denoising. The core of our approach is estimating the probability that a given coefficient contains a significant noise-free component, which we call "*signal of interest*". In this respect we analyze cases where the involved probabilities are (i) fixed per subband, (ii) conditioned on a local spatial context and (iii) conditioned on information from multiple image bands in case of multivalued images. For actual denoising, we adopt a simple shrinkage rule that was also used in [33]–[37], where empirical wavelet coefficients are multiplied with the probability of containing a significant noise-free component. The heuristics behind this rule is to define a "softer" and yet simple alternative to the classical thresholding functions while satisfying the two main requirements: the shrinkage factor is always between zero and one and the coefficients that are more likely to represent the noise are heavier shrunk. A mathematical motivation for this shrinkage rule in the form of a posterior expected action can be found in [35]. Here, the proposed approach for estimating the required probabilities is essentially different: previous methods were relying on preliminary coefficient classifications yielding binary masks that were combined with MRF priors [33]–[36] or used for empirical density estimation followed by fitting of log-likelihood ratios [37]. In contrast to this, our new approach

A. Pižurica and W. Philips are with the Department for Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium. E-mail: Aleksandra.Pizurica@telin.UGent.be, philips@telin.UGent.be, Tel: +32 9 264 34 12, Fax: +32 9 264 42. 95.

removes the need for preliminary coefficient classifications and derives all the required probabilities analytically starting from the generalized Laplacian marginal prior. Significant advantages of this new approach are that it does not depend on any preliminary edge detection (classification) steps, it is simpler to implement and faster while it yields better results than the more complex ones based on MRFs. Moreover, we extend the new method for multiband images as well.

The main novelties and contributions of this paper are: (1) A novel *subband-adaptive* shrinkage function, which shrinks each coefficient according to the probability that it presents a signal of interest. Experimental results indicate that for natural images this estimator outperforms, in terms of MSE, any classical soft-thresholding rule with a uniform threshold per subband. (2) We develop a *spatially adaptive* version of the proposed method. The results demonstrate that the new method outperforms spatially adaptive thresholding with context modelling as well as MMSE approaches that employ much more complex HMTs and related methods based on MRFs. (3) We extend the proposed method for *multivalued* data. The results on color and on multispectral images demonstrate a significant improvement with respect to recent multiband wavelet thresholding approaches.

The paper is organized as follows. In Section II, we develop a new subband adaptive shrinkage function for natural images. In Section III, we extend it first to a spatially adaptive method and in Section IV to denoising multivalued images. Section V concludes the paper.

II. Subband adaptive Bayesian wavelet shrinkage

We assume the input image is contaminated with signal-independent additive white Gaussian noise of zero mean and variance σ^2. An orthogonal wavelet transformation [1]–[4] of the noisy input yields an equivalent additive white noise model in the transform domain. In each wavelet subband at a given scale and orientation we have

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

where β_i are noise-free wavelet coefficients, ϵ_i are independent identically distributed (i.i.d.) normal random variables $\epsilon_i \sim N(0, \sigma^2)$, which are statistically independent from β_i and n is the number of coefficients in a subband. For compactness, we suppressed here the indices that denote the scale and the orientation and we denoted the spatial position with a single index, like in a raster scanning. In the remainder of this Section we suppress the spatial index too because the same shrinkage rule will be applied to all the coefficients in a given subband.

Our approach is aimed for priors that are sharply peaked at zero and heavy-tailed like Laplacian, generalized Laplacian and alpha-stable distributions. The generalized Laplacian (also called generalized Gaussian) prior for the noise-free subband data [2], [9], [16], [20] is

$$f(\beta) = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})} \exp(-\lambda|\beta|^{\nu}) \tag{2}$$

where $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$ is the Gamma function, $\lambda > 0$ is the *scale* parameter and ν is the *shape* parameter, which is typically $\nu \in [0, 1]$ for natural images. The variance and the kurtosis of the noise-free histogram are [16]: $\sigma_\beta^2 = \Gamma(\frac{3}{\nu})/\Big(\lambda^2\Gamma(\frac{1}{\nu})\Big)$ and $\kappa_\beta = \Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})/\Gamma^2(\frac{3}{\nu})$, respectively.

Let us define a "*signal of interest*" as a noise-free coefficient component that exceeds a specific threshold T and formulate the following two hypotheses: H_0: "the signal of interest is absent" and H_1: "the signal of interest is present" (in a given coefficient) precisely as:

$$H_0 : |\beta| \leq T \quad \text{and} \quad H_1 : |\beta| > T \tag{3}$$

Let $P(H_1|y)$ denote the conditional probability that a wavelet coefficient contains a signal of interest, given its observed value. The Bayes' rule yields $P(H_1|y) = \mu\eta/(1 + \mu\eta)$ where $\mu = P(H_1)/P(H_0)$ and $\eta = f(y|H_1)/f(y|H_0)$ and the product $\mu\eta$ is called the generalized likelihood ratio [38]. We now consider a simple shrinkage rule

$$\hat{\beta} = P(H_1|y)y = \frac{\mu\eta}{1 + \mu\eta}y \tag{4}$$

that we call *ProbShrink* hereafter[1]. Its calculation and the resulting shape for the assumed prior are illustrated in Fig. 1 that we will now explain in somewhat more detail. Fig. 1(a) shows an example of the conditional densities of *noise-free* coefficients $f(\beta|H_0)$ and $f(\beta|H_1)$. Note that the conditional density of noise free coefficients given H_0 (signal of interest is absent) is in other words the probability density function of *insignificant* noise free coefficients and that it is proportional to $f(\beta)$ for $|\beta| \leq T$ and equal to zero otherwise. Analogously, the conditional density of β given H_1 is the density of *significant* noise-free coefficients and is proportional to $f(\beta)$ for $|\beta| > T$ and equal to zero otherwise. For the assumed additive noise model where the noise coefficients and the noise-free coefficients are respectively realizations of two stochastic processes that are statistically independent, the conditional densities of the *noisy* coefficients, $f(y|H_0)$ and $f(y|H_1)$ result from the following convolutions

$$\begin{aligned} f(y|H_0) &= \int_{-\infty}^{\infty} \phi(y - \beta; \sigma) f(\beta|H_0) d\beta \\ f(y|H_1) &= \int_{-\infty}^{\infty} \phi(y - \beta; \sigma) f(\beta|H_1) d\beta \end{aligned} \tag{5}$$

where $\phi(y; \sigma)$ is the zero mean Gaussian density with the standard deviation σ. Fig. 1(b) illustrates these conditional densities of the noisy coefficients. Fig. 1(c) shows the resulting *ProbShrink* rule (4), where $P(H_1)$ (i.e., the prior ratio μ) is left as a free parameter. One can see that the smallest coefficients are heavily shrunk towards zero while the largest ones tend to remain unchanged. Between these extremes, there is a smooth transition which depends on the global subband statistics expressed through $P(H_1)$. Next we address the specification of this parameter in a given subband.

[1]For the assumed mixture prior, this shrinkage rule can be interpreted as an approximation of the MMSE estimate $E(\beta|y) = P(H_1|y)E(\beta|y, H_1) + P(H_0|y)E(\beta|y, H_0)$, which follows from simplifying $E(\beta|y, H_1) \cong y$ and $E(\beta|y, H_0) \cong 0$.

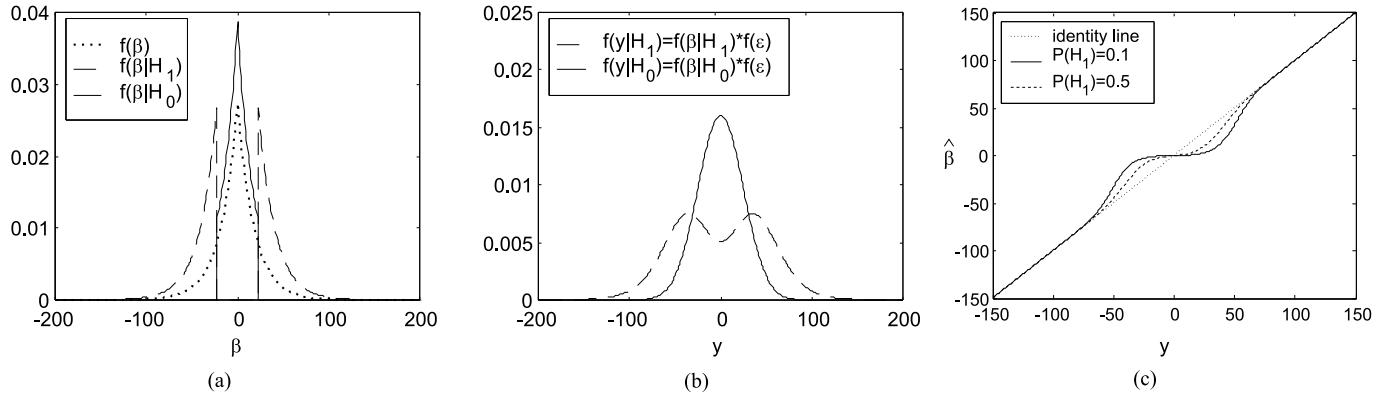

Fig. 1. (a) An illustration of the probability density functions of noise-free coefficients: $f(\beta)$ (dotted), $f(\beta|H_0)$ (solid) and $f(\beta|H_1)$ (dashed). (b) The resulting conditional densities of noisy coefficients $f(y|H_0)$ (solid) and $f(y|H_1)$ (dashed). (c) *ProbShrink* rule $\hat{\beta} = P(H_1|y)y$, where $P(H_1)$ is a parameter.

A. Adapting the prior probabilities to the subband statistics

The first novelty of the proposed subband adaptive shrinkage method is the way we estimate the prior probability of signal presence $P(H_1)$. In related approaches it has been usually assumed $P(H_1) = P(H_0) = 0.5$ (e.g., [17], [33]–[36]) or $P(H_1)$ was estimated empirically as a given fraction of the observed *noisy* coefficients [14], [37]. Here we derive the probability $P(H_1)$ from the prior model for the noise-free coefficients in a given subband. In particular, we note that for the hypotheses model (3) the probability of signal presence amounts to the area under the tails of $f(\beta)$ for $|\beta| > T$ and thus we estimate $P(H_1)$ as

$$P(H_1) = 1 - \int_{-T}^{T} f(\beta) d\beta \tag{6}$$

Next we develop this expression for the generalized Laplacian prior and analyze the performance of the resulting *ProbShrink* rule (4).

B. ProbShrink rule for the generalized Laplacian prior

Under the assumed prior (2), the conditional densities of noise-free coefficients are

$$f(\beta|H_0) = \begin{cases} B_0 \exp(-\lambda|\beta|^\nu), & \text{if } |\beta| \le T \\ 0, & \text{if } |\beta| > T \end{cases} \tag{7}$$

and

$$f(\beta|H_1) = \begin{cases} 0, & \text{if } |\beta| \le T \\ B_1 \exp(-\lambda|\beta|^\nu), & \text{if } |\beta| > T \end{cases} \tag{8}$$

with the normalization constants (see Appendix):

$$B_0 = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)}$$
$$B_1 = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Big[1 - \Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)\Big]} \tag{9}$$

where $\Gamma_{inc}(x,a) = \frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}dt$ is the *incomplete gamma function*. From (6) we have that (see also Appendix): $P(H_1) = 1 - \Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)$ and thus

$$\mu = \frac{P(H_1)}{P(H_0)} = \frac{1 - \Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)}{\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)} \tag{10}$$

For the Laplacian prior ($\nu = 1$) the above expression reduces to $\mu = P(H_1)/P(H_0) = \exp(-\lambda T)/[1 - \exp(-\lambda T)]$. Together with the likelihood ratio $\eta = f(y|H_1)/f(y|H_0)$, which is calculated using (5), this completes the specification of the subband adaptive estimator (4).

C. Estimation of the prior parameters

In the proposed method, the parameters λ and ν of the generalized Laplacian prior for noise-free data are estimated from the noisy histogram in each subband, like in [9], [16]. In particular, the variance σ_y^2 and the fourth moment $m_{4,y}$ of the generalized Laplacian signal corrupted by additive white Gaussian noise with standard deviation σ are [16]

$$\sigma_y^2 = \sigma^2 + \frac{\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})}, \quad m_{4,y} = 3\sigma^4 + \frac{6\sigma^2\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})} + \frac{\Gamma(\frac{5}{\nu})}{\lambda^4\Gamma(\frac{1}{\nu})} \tag{11}$$

From the above equations, we find

$$\kappa_y = \frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})} = \frac{m_{4,y} + 3\sigma^4 - 6\sigma^2\sigma_y^2}{(\sigma_y^2 - \sigma^2)^2} \tag{12}$$

and

$$\lambda = \Big((\sigma_y^2 - \sigma^2)\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}\Big)^{-\frac{1}{2}} \tag{13}$$

The expression $\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})/\Gamma^2(\frac{3}{\nu})$ in the left hand side of (12) is a monotonic decreasing function of ν. We solve the parameter ν numerically from this equation. Using the estimated value of the shape parameter ν, the scale parameter λ follows directly from (13).

In this paper we use orthogonal wavelets of Daubechies [1]. In particular, all the results in this Section were produced using a five-level orthogonal transform with the least asymmetrical wavelet *symmlet* with eight vanishing moments. In our denoising experiments with artificial noise we assume that the noise standard deviation σ is known (as it is usually for reporting

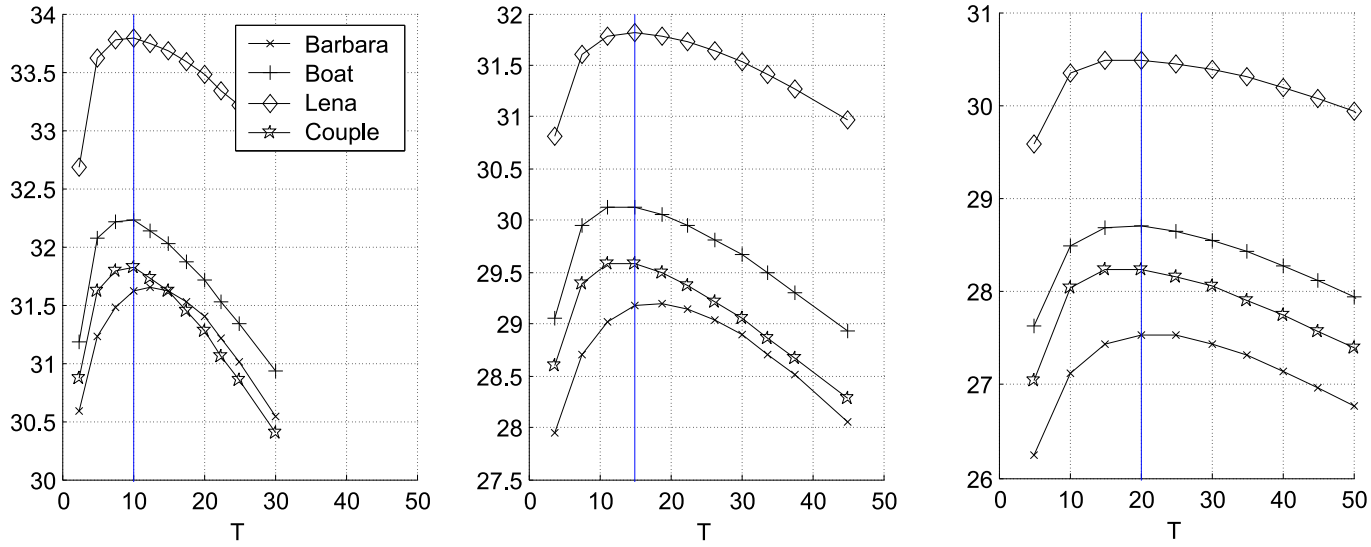

Fig. 2. Resulting PSNR[dB] values for the subband adaptive *ProbShrink* estimator as a function of the threshold T. From left to right the noise standard deviation is $\sigma = 10$, $\sigma = 15$ and $\sigma = 20$.

the results in case of artificially added noise). In practice the noise standard deviation is usually not known in advance, but its reliable estimate can be obtained as the median absolute deviation of the coefficients in the highest frequency subband divided by 0.6745 [6].

D. Experimental performance evaluation

The threshold T that specifies the notion of the signal of interest is the only parameter of the proposed shrinkage rule which is not estimated directly from the observed images. Ideally, we wish to define this threshold such to minimize the resulting mean squared error. Mathematical derivations seem untractable for the assumed prior. Therefore, like in [35], [36] we define the significant signal component based on the oracle thresholding and we also use the simulations to validate that this choice indeed optimizes the mean squared error performance of our overall method. The ideal oracle thresholding minimizes MSE by zeroing all the coefficients with the *signal component* below σ [4]. In other words, an oracle would tell us that the *noise-free* signal component above σ is the signal of interest, i.e., $T = \sigma$. While this choice minimizes the MSE of an idealized hard-thresholding, we need to verify its performance for the proposed estimator through simulations. Our experiments on different natural images confirm that the resulting PSNR of the proposed estimator peaks in the vicinity of $T = \sigma$ (see Fig. 2).

Table I shows the peak signal to noise ratio[2] (PSNR) performance of the proposed *ProbShrink* rule with $T = \sigma$ in comparison to two reference methods: *BayesShrink* of [9] and adaptive Bayesian wavelet shrinkage (*ABWS*) of [14]. *BayesShrink* is Bayesian soft-thresholding with the threshold σ^2/σ_β, which was for natural images (i.e., for the generalized Laplacian prior) shown to be optimal in terms of mean squared error. The *ABWS* method of [14] is a MMSE estimator under a different prior for noise free data, which is a mixture of two Gaussians. We implemented all the methods in the orthogonal wavelet representation with five decomposition levels and using the least asymmetrical wavelet of [1] (often called *"symmlet"*) with eight vanishing moments (*sym8*). For the *ABWS* method we used the default choices of the hyperparameters from [14]. The results demonstrate that *ProbShrink* rule outperforms both *BayesShrink* and *ABWS* on all tested images. The improvement with respect to *ABWS* is on most images greater than 1dB. This significant difference results mainly from using different priors: the generalized Laplacian

[2]PSNR is defined as $\mathrm{PSNR} = 10\log_{10}(255^2/\mathrm{MSE})$, where MSE is the mean squared error.

TABLE I

PSNR[dB] RESULTS OF THE PROPOSED SUBBAND ADAPTIVE $ProbShrink$ AND TWO OTHER SUBBAND ADAPTIVE BAYESIAN METHODS FOR $sym8$ WAVELET.

	Standard deviation of noise			
Estimator	10	15	20	25
	BARBARA			
noisy image	28.12	24.59	22.09	20.17
ABWS [14]	30.71	28.27	26.47	25.10
BayesShrink [9]	31.24	28.86	27.32	26.20
ProbShrink	**31.62**	**29.17**	**27.54**	**26.32**
	BOAT			
noisy image	28.15	24.62	22.10	20.17
ABWS [14]	31.01	29.04	27.66	26.68
BayesShrink [9]	32.01	29.98	28.55	27.54
ProbShrink	**32.23**	**30.13**	**28.70**	**27.69**
	COUPLE			
noisy image	28.15	24.60	22.11	20.18
ABWS [14]	30.48	28.36	27.15	26.27
BayesShrink [9]	31.70	29.46	28.08	27.09
ProbShrink	**31.82**	**29.60**	**28.24**	**27.24**
	LENA			
noisy image	28.13	24.60	22.12	20.16
ABWS [14]	32.84	30.97	29.66	28.74
BayesShrink [9]	33.47	31.53	30.26	29.30
ProbShrink	**33.80**	**31.82**	**30.49**	**29.51**

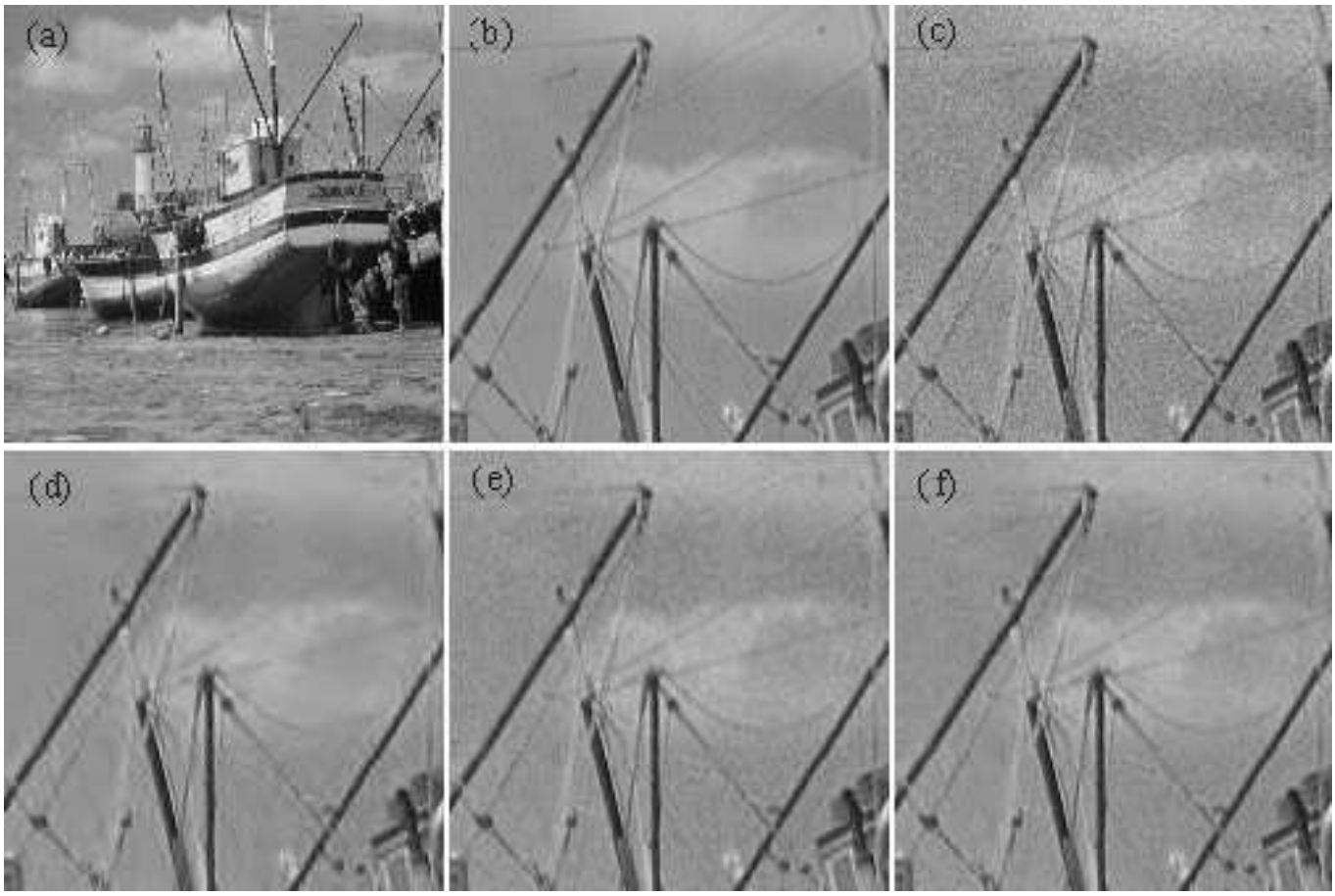

Fig. 3. Visual performance of the subband adaptive approach. (a) *Boat* image. (b) noise-free part (c) noisy part, $\sigma = 10$ and denoised versions using (d) ABWS [14], (e) *BayesShrink* [9] and (f) the proposed subband adaptive *ProbShrink*.

that we use models the image wavelet coefficients much better than the mixture of two Gaussians. Compared to *BayesShrink* under the same, generalized Laplacian prior *ProbShrink* yields a slight improvement on all images. Since *BayesShrink* is soft-thresholding with the MSE optimum threshold, we can deduce that *ProbShrink* (at least on the tested images) outperforms soft thresholding with any threshold that is constant per subband. We believe that this is an important argument in favor of the new shrinkage rule, especially because it is of similar complexity as Bayesian thresholding.

Visual quality of the three estimators from Table I is illustrated in Fig. 3. It can be seen that *ABWS* blurs image details more than the other two methods. *ProbShrink* suppresses noise slightly better than *BayesShrink* while preserving the image details equally well.

III. Spatially adaptive Bayesian shrinkage

The shrinkage approach analyzed so far was subband-adaptive: if two noisy coefficients from the same subband were of equal magnitudes than they were shrunk by the same amount no matter their spatial position and no matter their local surrounding. Now we adapt the estimator to the local spatial context in the image using a *local spatial activity indicator* (LSAI) z_l for each spatial position l as follows:

$$\hat{\beta}_l = P(H_1|y_l, z_l)y_l = \frac{\eta_l \xi_l \mu}{1 + \eta_l \xi_l \mu} y_l \tag{14}$$

where

$$\eta_l = \frac{f(y_l|H_1)}{f(y_l|H_0)}, \quad \xi_l = \frac{f(z_l|H_1)}{f(z_l|H_0)} \quad \text{and} \quad \mu = \frac{P(H_1)}{P(H_0)} \tag{15}$$

The characteristic parts of the method are illustrated in Fig. 4, where the generalized likelihood ratio denotes the product $\eta_l \xi_l \mu$. The figure summarizes the whole method: from the noisy subband, i.e., from the noisy coefficient histogram we estimate the parameters of the noise-free prior $f(\beta)$. Using $f(\beta)$ we estimate $P(H_1)$ (by integrating the tails) and we find the conditional densities of the noisy coefficients using (5). The conditional densities of LSAI are calculated using the conditional densities of the noisy coefficients. Fig. 4 illustrates also a heuristic explanation of the proposed method: each coefficient is shrunk according to how probable it is that it presents useful information, based on its value (via η_l), based on a measurement from the local surrounding (via ξ_l) and based on the global statistical properties of the coefficients in a given subband (via μ).

Note that even though the general form of the estimator (14) is the same as in our previous work [37], there is a great difference between the new, here proposed approach and that

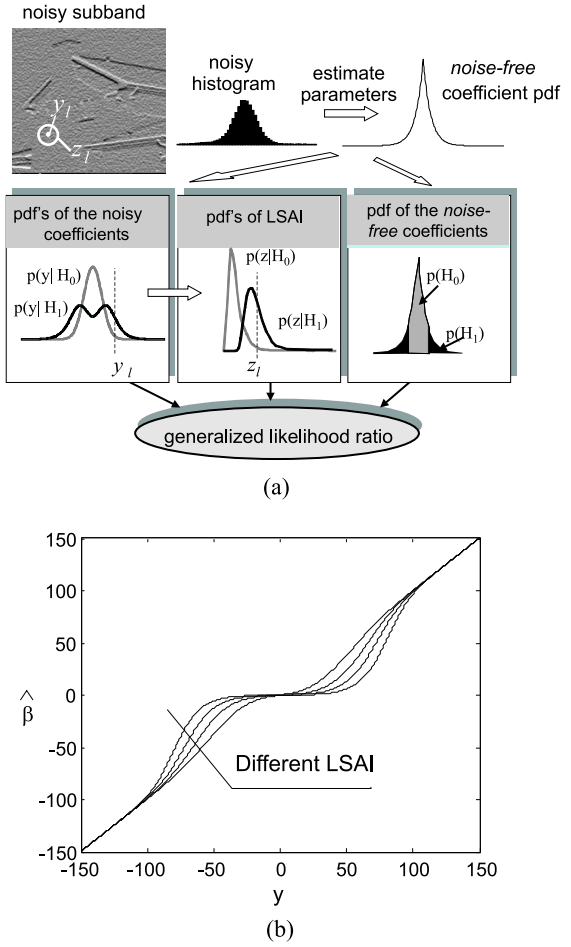

Fig. 4. (a) An illustration of the proposed denoising method, where pdf denotes the probability density function and where LSAI denotes the local spatial activity indicator. (b) The resulting shrinkage rule is a family of characteristics, which correspond to different values of LSAI.

of [37]. In the first place, the approach of [37] assumes no particular prior on the noise-free data and no particular noise distribution. Instead, it uses interscale products for heuristic preliminary coefficient classification (mask determination) and involves an empirical density estimation using detected masks and piece-wise linear fitting of the empirical log-likelihood ratios. None of these steps is required in the new method: we do not need any binary masks here, nor empirical density estimation or fitting procedures. All the required probabilities and probability density functions are now expressed analytically, starting from the generalized Laplacian prior.

We define LSAI as the locally averaged magnitude of the coefficients in a relatively small square window $\delta(l)$ of a fixed size N, within the same subband:

$$z_l = \frac{1}{N} \sum_{k \in \delta(l)} \omega_k \qquad (16)$$

where ω_l denotes the coefficient magnitude $\omega_l = |y_l|$. For practical reasons, we simplify the statistical characterization of z_l considerably assuming that all the coefficients within the small window are equally distributed and *conditionally* independent[3] (given H_0 or H_1). Under these assumptions, $f(Nz_l|H_{0,1})$ equals N convolutions of $f(\omega_l|H_{0,1})$ with itself, where the densities of coefficient magnitudes are $f(\omega_l|H_{0,1}) = 2f(y_l|H_{0,1})$ for $\omega_l \geq 0$ and $f(\omega_l|H_{0,1}) = 0$ for $\omega_l < 0$.

For simplicity of implementation, we experimented with square windows only. For all the tested images (the sizes of which were up to 512x512) the window size 3x3 yielded maximum PSNR. Experimental results in Fig. 5 show that the resulting spatially adaptive estimator (14) always yields an improved PSNR as compared to the subband adaptive version (4). For some images the PSNR gains are up to 0.9dB. Even though these PSNR gains may seem marginal, the difference can be seen visually. The visual improvement mainly consists of better suppressing noise in uniform areas as can be seen from Fig. 6.

A. Results in the orthogonal wavelet representation

We compared the performance of the proposed method with several recent denoising methods that use the orthogonal wavelet transform: spatially adaptive bivariate shrinkage [23], the locally adaptive Wiener method of [28], the Hidden Markov Tree (HMT) approach of [25] and a two-dimensional extension of the block thresholding method from [8]. For the sake of fair comparison, we implemented all the methods using the same wavelet decomposition with five decomposition levels and using *symmlet* with eight vanishing moments. We implemented all the reference methods with the parameter values that yield the best PSNR performance. In particular, for [23], [28] we used the parameter values that were reported by the authors to yield the best performance. For the HMT method [25], we used the publicly available code of the Rice University. Here we report the results we obtained with the full HMT training for each image, which yields the best PSNR performance of this method. In our two-dimensional extension of the block thresholding of [8] we optimize the constant of this method to yield the maximum resulting PSNR.[4]

Fig. 7 shows the results of the five tested methods. The results demonstrate that the proposed method yields similar or better results than the existing methods. On all the tested images the new method outperforms a much more complex approach based on HMTs.

B. Results in the redundant wavelet representation

The denoising performance of any wavelet shrinkage method improves by using a redundant instead of the orthogonal wavelet representation (see the discussions, e.g., in [24], [26], [30], [32]). In this Section, we present the

[3]Such assumptions were earlier used, e.g., in [28], [31]

[4]The block thresholding method "NeighBlock" from [8] applies the shrinkage factor $(1 - \lambda\sigma^2/s^2)_+$ where λ is a constant, s^2 is the average of the squared coefficient values within a local window and the window size is related to the number of the data samples. The authors have considered one-dimensional signals only for which the optimal value of the parameter was $\lambda = 4.50524$. A direct extension to two dimensions (by relating the window size to the number of the coefficients in the corresponding direction and by keeping the constant λ unchanged) results in a poor performance. We achieved the best performance of this method by setting $\lambda = 1$ which is also motivated by the Proposition 1 from [7].

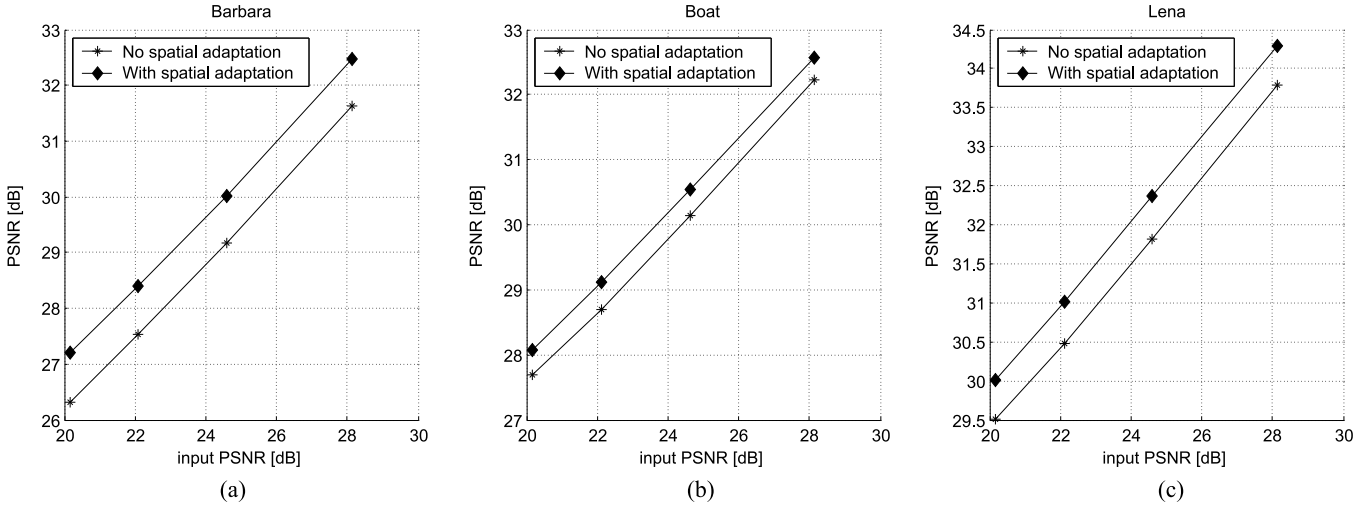

Fig. 5. Performance of the proposed *ProbShrink* method without spatial adaptation (subband adaptive shrinkage) and with spatial adaptation on three test images: (a) *Barbara*, (b) *Boat* and (c) *Lena*.

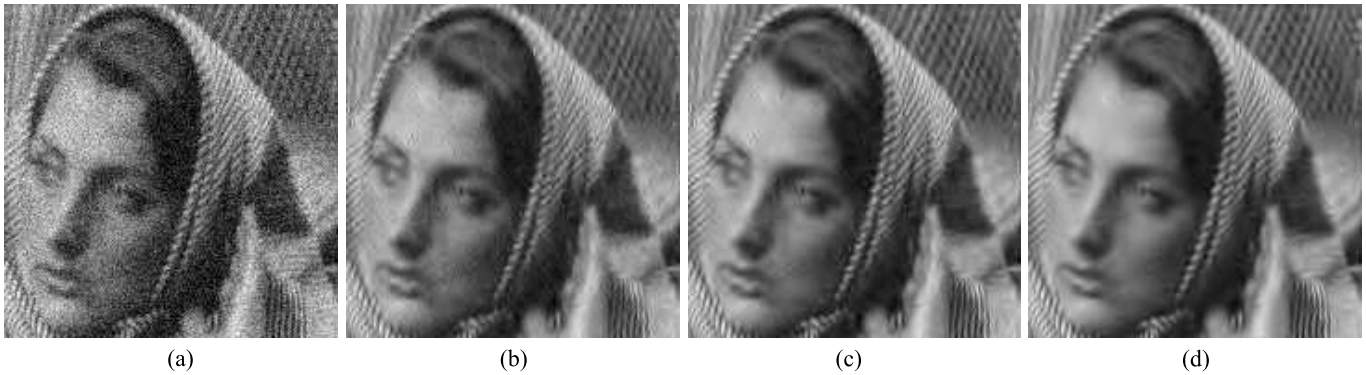

Fig. 6. Visual performance of different versions of the proposed *ProbShrink* method. (a) Noisy *Barbara* image, $\sigma = 20$, PSNR=22.09dB. (b) Subband adaptive shrinkage in the orthogonal transform, PSNR=27.54dB. (c) Spatially adaptive shrinkage in the orthogonal transform PSNR=28.4dB. (d) Spatially adaptive shrinkage in the non-decimated transform PSNR=29.53dB.

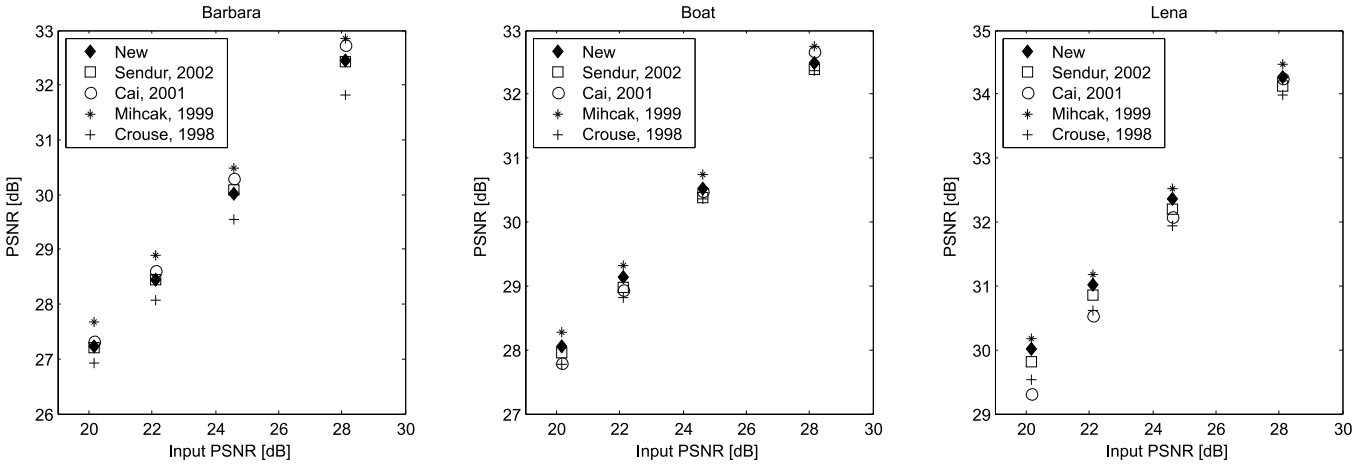

Fig. 7. Performance of the proposed spatially adaptive shrinkage and several recent methods using orthogonal wavelet transform: [25] (Crouse, 1998), [28] (Mihçak, 1999), [8] (Cai, 2001) and [23] (Şendur, 2002).

implementation and the performance of the proposed method in a redundant wavelet representation. In particular, we use the *non-decimated* wavelet transform implemented with the algorithm *à trous* [4]. We optimized the window size of our method implemented in the non-decimated wavelet representation experimentally. For simplicity, we considered the

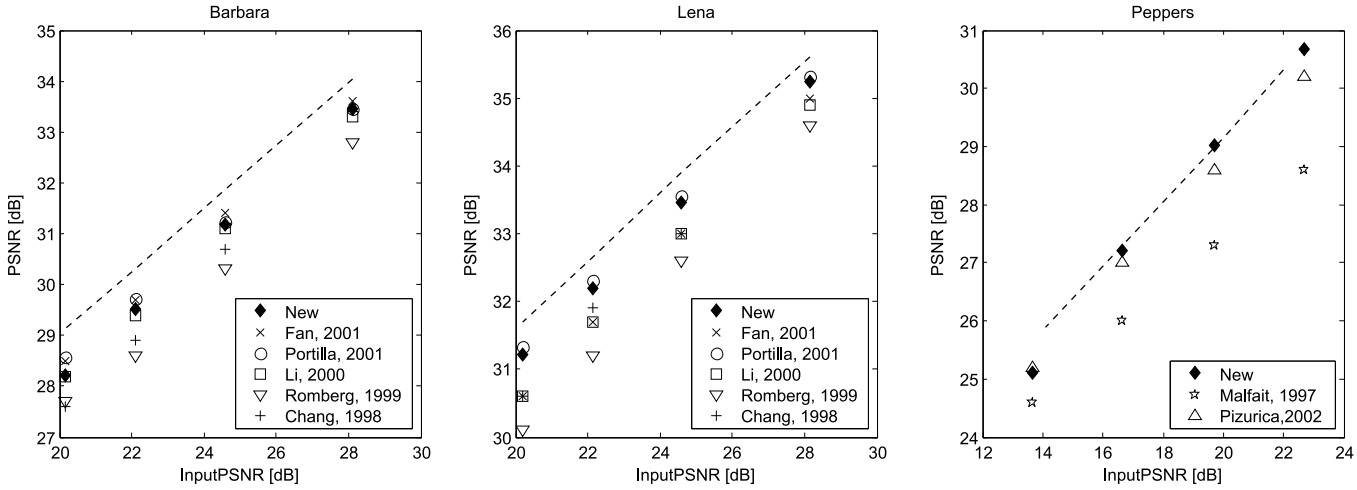

Fig. 8. Results of several recent methods, which use redundant wavelet representation with three orientations: [33] (Malfait, 1997), [24] (Chang, 1998), [26] (Romberg, 1999), [32] (Li, 2000), [27] (Fan, 2001), [30] (Portilla, 2001) and [36] (Pizurica, 2002). Dashed lines show the best available results, obtained with 8-orientation steerable pyramid in [31].

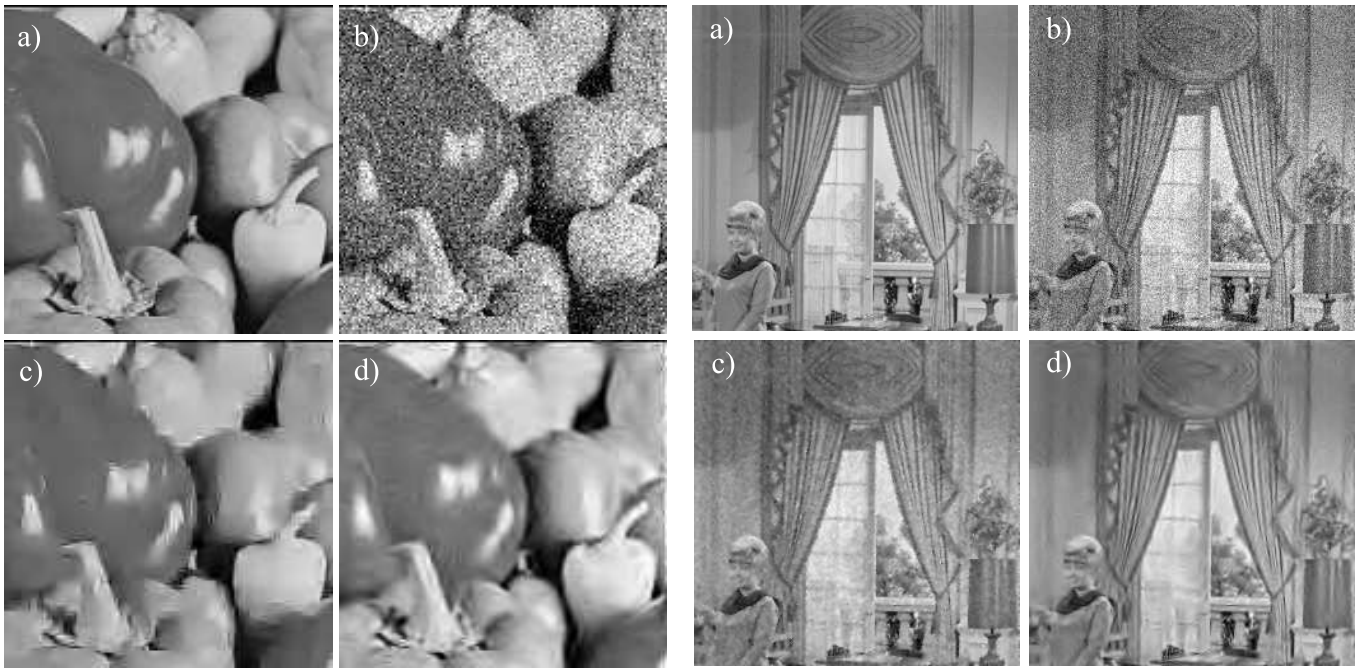

Fig. 9. **Left**: (a) noise-free part of the *Peppers* image, (b) noisy image, $\sigma = 37.5$, (c) the result of the MRF-based method of [36] and (d) the new method. **Right**: noise-free part of the *Couple* image, (b) noisy image, $\sigma = 20$, (c) the result of the HMT-based method of [25] and (d) the new method.

square window shape only. For all the tested images with sizes up to 512x512 the window size 7x7 yielded the maximum PSNR. We experimented with two types of orthogonal wavelets: Daubechies' wavelets and symmlets [1], [4]. Among these, *symmlet* with eight vanishing moments yielded the best results for most of the tested images. For some images (*Lena* and *Peppers*) slightly better results were obtained using the Daubechies' wavelet with two vanishing moments. Table II lists the resulting PSNR values in comparison with the results that we obtained using the orthogonal wavelet representation. The sizes of all images in this table are 512x512, except for *Peppers*, which is 256x256. One can see that on different images the use of the non-decimated representation brings an improvement in the range of 0.8 to 1.5dB. The visual quality also improves significantly as it can be seen by comparing Fig. 6(c) and Fig. 6(d).

In Fig. 8, the results are plotted in comparison with seven recently published methods, which also use overcomplete wavelet transforms with three orientations per scale: spatially adaptive thresholding of [24], locally adaptive Wiener filtering of [32], MMSE estimation with a Gaussian scale mixture prior of [30], MMSE estimators with two different HMT models [26], [27] and two wavelet shrinkage methods based on MRF priors [33], [36]. Dashed lines in these diagrams show the best published results so far that were obtained with an 8-orientation redundant steerable pyramid in [31]. The results

TABLE II

PSNR[dB] RESULTS OF THE PROPOSED SPATIALLY ADAPTIVE METHOD ($ProbShrink-SP$) USING THE ORTHOGONAL ($-ort$) AN THE REDUNDANT ($-red$) WAVELET REPRESENTATIONS.

	Standard deviation of noise				
Estimator	10	15	20	25	30
	BARBARA				
noisy image	28.12	24.59	22.09	20.17	18.57
ProbShrink-SP-ort	32.45	30.02	28.44	27.22	26.26
ProbShrink-SP-red	33.46	31.19	29.53	28.23	27.17
	BOAT				
noisy image	28.15	24.62	22.10	20.17	18.58
ProbShrink-SP-ort	32.50	30.52	29.14	28.05	27.23
ProbShrink-SP-red	33.25	31.32	29.93	28.89	28.04
	COUPLE				
noisy image	28.15	24.60	22.11	20.18	18.58
ProbShrink-SP-ort	32.20	30.03	28.63	27.56	26.76
ProbShrink-SP-red	32.94	30.81	29.41	28.38	27.52
	LENA				
noisy image	28.13	24.60	22.12	20.16	18.60
ProbShrink-SP-ort	34.26	32.36	31.01	30.01	29.16
ProbShrink-SP-red	35.24	33.46	32.20	31.21	30.33
	PEPPERS				
noisy image	28.10	24.61	22.07	20.19	18.59
ProbShrink-SP-ort	32.69	30.49	28.88	27.62	26.76
ProbShrink-SP-red	33.90	31.81	30.30	29.23	28.35

are plotted for three test images: 512x512 *Lena* and *Barbara* and 256x256 *Peppers*. The results of the reference methods were copied from the corresponding publications making sure that the same versions of the test images are used. The *Peppers* is the same as the one in [36], which was also used in [31]. We verified that the standard images *Lena* and *Barbara* are the same as in [30], [31] and there the authors confirm that the same versions were used in [24], [32]. The comparisons from [27] indicate that the same image versions are used for the results of [26], [27] as well.

The results in Fig. 8 demonstrate that in comparison with the approaches of similar complexity [24], [32], the new method yields a significant improvement and that it also outperforms more complex methods that are based on HMT priors [26] and on MRF priors [33], [36]. Visual improvement is illustrated in Fig. 9. The new method is for image sizes 512x512 approximately three times as fast as our MRF-based method from [36].

IV. EXTENSIONS TO MULTIVALUED DATA

The proposed denoising approach leads to efficient low-complexity noise filters for *multivalued* data like color images, multispectral and hyperspectral data or multimodal magnetic resonance images. In all these cases different image bands are correlated: an image discontinuity from one band is likely to occur in at least some of the remaining bands. The simplest approach to extend our method for multivalued images is to include the interband correlation in the definition of the local spatial activity indicator.

Let $\omega^b_{l,s}$ denote the noisy coefficient magnitude in the image band b, wavelet subband s and spatial position l. A possible

TABLE III

PSNR[dB] RESULTS FOR COLOR 512X512 IMAGES USING RGB REPRESENTATION. $ProbShrink-MB$ DENOTES THE PROPOSED MULTIBAND DENOISING AND $ProbShrink-SP$ DENOTES COMPONENTWISE DENOISING USING THE SPATIALLY ADAPTIVE SINGLE BAND METHOD FROM SECTION III.

	Standard deviation of noise			
Method	10	15	20	25
	BABOON			
MBT [39]	28.50	26.78	25.53	24.56
ProbShrink-SP	29.80	27.34	25.79	24.60
ProbShrink-MB	**30.17**	**27.83**	**26.38**	**25.27**
	LENA			
MBT [39]	33.84	32.29	31.14	30.15
ProbShrink-SP	34.19	32.46	31.22	30.29
ProbShrink-MB	**34.60**	**33.03**	**31.92**	**31.04**
	PEPPERS			
MBT [39]	31.19	30.22	29.45	28.77
ProbShrink-SP	33.20	31.65	30.61	29.75
ProbShrink-MB	**33.44**	**32.05**	**31.12**	**30.35**

multiband extension of the LSAI from (16) is:

$$z^b_{l,s} = \frac{1}{NB} \sum_{i=1}^{B} \sum_{k \in \delta(l)} \omega^i_{k,s} \tag{17}$$

where B is the number of image bands. With this definition of the LSAI the probability of signal presence is conditioned on the spatial context as well as on information from other image bands. Based on experiments with standard color and with high-resolution multispectral *Landsat* images, we found that best results are obtained when the neighborhood $\delta(l)$ is reduced to a single pixel, i.e., when LSAI includes only the coefficients at the same spatial position from different image bands: $z^b_{l,s} = \frac{1}{B}\sum_{i=1}^{B} \omega^i_{l,s}$. Its conditional densities are estimated by convolving the corresponding densities of the coefficient magnitudes. We take into account a high correlation between image bands by assuming that the wavelet coefficients at the same positions in different image bands are distributed either according to the presence of a signal, i.e., as $f(y|H_1)$, or according to the absence of a signal, i.e., as $f(y|H_0)$. For the sake of tractability we also assume that the coefficients are *conditionally* independent given that either H_0 or H_1 is true.[5]

Note that in (17) the spectral and the spatial window are constrained to have the same shape and size. An alternative is to define the square spatial window and to limit the spectral window to one pixel from each of the remaining spectral channels. In our experiments such a definition of LSAI did not improve PSNR significantly on any of the tested color and multispectral images and it has even resulted in a slight performance loss compared to the case when LSAI includes only the coefficients at the same spatial position from different image bands. Nevertheless, for some applications it might be

[5]Note that the *conditional* independence does not imply that the corresponding coefficients are statistically independent. We take into account the statistical dependence between these coefficients by imposing that they all follow either $f(y|H_1)$ or $f(y|H_0)$ meaning that they all contain either a large or a small noise-free component.

Fig. 10. Parts of the noise-free, noisy (σ=25, PSNR=20.17dB) and the denoised (PSNR=30.73dB) color image. Three-band (RGB) representation is used.

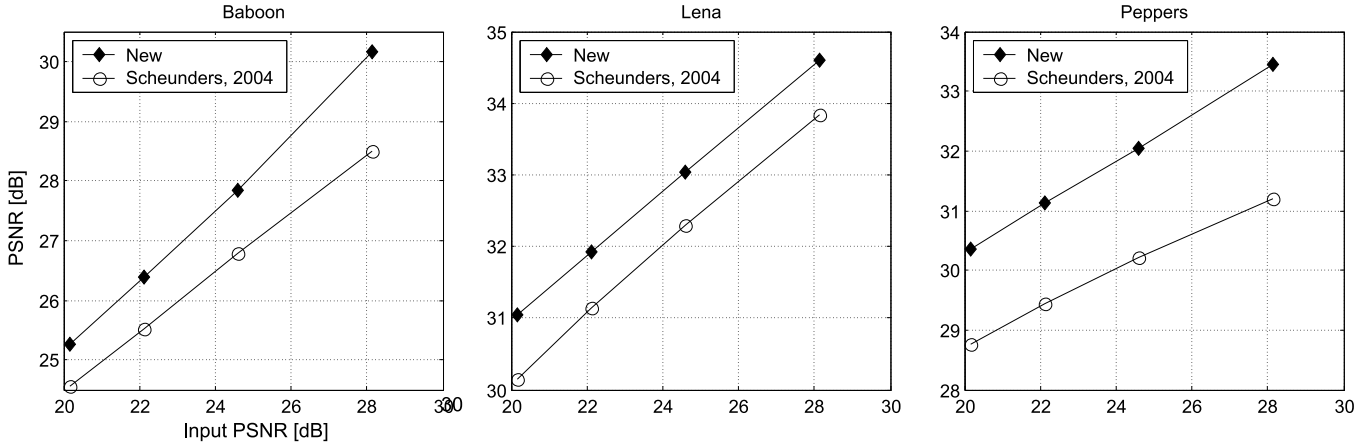

Fig. 11. PSNR results on color images in comparison with the method of [39] (Scheunders, 2004) using the RGB (three-band) representation.

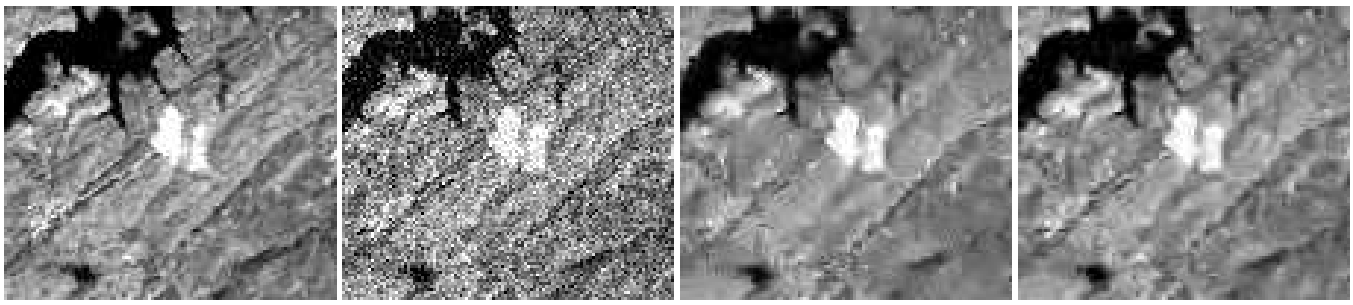

Fig. 12. Left to right: parts of the 4th noise-free band of the 7-band *Landsat* image from [39], noisy image ($\sigma = 35, PSNR = 17.24dB$) result of [39] (PSNR=21.16dB) and the result of the proposed method (PSNR=22.42dB).

interesting to define different spatial and spectral windows in order to optimally incorporate the spatial and the spectral information.

A. Experimental results for color images

Let $\mathbf{u} = [\mathbf{u}^1 ... \mathbf{u}^B]$ denote the reference noise-free multi-band image, where each image band is presented as a one-dimensional vector $\mathbf{u}^b = [u_1^b ... u_L^b]$ obtained according to raster scanning. Also, let $\mathbf{g} = \mathbf{u} + \mathbf{n}$ denote the noisy multi-band image, where $\mathbf{n}$ is added white Gaussian noise, and let $\hat{\mathbf{u}}$ denote the estimated noise-free image (i.e., the image denoised by wavelet shrinkage). As a quantitative performance measure we use the peak signal to noise ratio defined as $PSNR = 10log_{10}(255^2/MSE)$ where the mean squared error MSE is averaged over all the spectral channels. The input and the output PSNR are thus calculated respectively as

$$PSNR_{input} = 10log_{10}\frac{255^2}{\frac{1}{LB}\sum_{l=1}^{L}\sum_{b=1}^{B}(g_l^b - u_l^b)^2}$$

$$PSNR_{res} = 10log_{10}\frac{255^2}{\frac{1}{LB}\sum_{l=1}^{L}\sum_{b=1}^{B}(\hat{u}_l^b - u_l^b)^2}$$

For color images, visual results are illustrated in Fig. 10. The PSNR results for three representative color images are tabulated in Table III. The reference methods in this Table are the componentwise denoising using the single-band spatially adaptive method from Section III and the recent multiband wavelet thresholding (MBT) of [39]. We implemented all three methods with the same wavelet transform. The results show that the proposed multiband denoising improves PSNR on all the tested images as compared to the componentwise denoising. The results also demonstrate a superior performance in comparison with the multiband thresholding of [39]. For the ease of comparison we also present the PSNR plots of these two methods in Fig. 11. One can see that on some images our method achieves an improvement of more than 2dB.

B. Application to multispectral images

We also experimented with multispectral satellite images. In the satellite systems it may be desirable to perform denoising before the image compression step in order to improve the compression efficiency. Also deconvolution of satellite images is often useful and can be performed as an inverse filtering operation followed by denoising [40]. There are several noise sources in optical satellite images (photonic noise, electronic noise, quantization errors...) and the additive Gaussian noise model is a realistic approximation [39], [41], [42].

We used *Landsat* images that consist of seven bands. Fig. 12 shows the results for one band from the *Landsat* image from [39]. Our method yields in this case an improvement of 1.3dB over *MBT* [39]. The new method achieves an improvement on all the other bands as well. The gain is different for different image bands and for different noise standard deviations σ. For example, for $\sigma = 35$ the total gain over *MBT* summed over all the bands is approximately 1dB and for $\sigma = 15$ this total gain is around 3.5dB. We also verified that the proposed multiband *ProbShrink* better denoises any multispectral band than the componentwise denoising using the method of Section III. The total gain of the multiband method with respect to the single-band version is approximately 2.5dB irrespective of the input noise level.

The effectiveness of the proposed multiband denoising method results from adapting the wavelet shrinkage to both the interband correlations and to the local statistics in each image band. In other words, the estimated probabilities of signal presence are different in each image band even though they are dependant on information from other bands as well as on the measurements from the given band.

C. Implementation details

Matlab implementations of the proposed method for greyscale and for color images are available at the web-site http://telin.UGent.be/ ∼sanja.

V. Conclusion

We developed a new wavelet domain denoising method based on the probability that a given coefficient represents a significant noise-free component, which we call "signal of interest". First we developed a novel subband-adaptive wavelet shrinkage function. Experiments on natural images yielded a better MSE performance than Bayesian soft-thresholding with the MSE-optimum threshold. The proposed spatially adaptive denoising method yields superior results as compared to some much more complex recent approaches based on HMT and MRF models. These results motivate strongly a further development of the presented concept. Also, improvements are expected by implementing the proposed method with a transform of a better orientation selectivity, like complex wavelets [43], steerable pyramids [31] or curvelets [44].

We also demonstrated that the proposed method can be easily extended to deal with multivalued images simply by defining the local spatial activity indicator as a function of the coefficients from multiple image bands. Our initial experiments on color and on multispectral Landsat images already showed a significant improvement over multiband wavelet thresholding.

Acknowledgement

We thank Paul Scheunders for generously providing the source code for multiband wavelet thresholding of [39] and for the fruitful discussions on multiband image denoising. We also thank to Javier Portilla for providing test images for the performance comparison with [30], [31].

VI. Appendix

For the generalized Laplacian prior $f(\beta)$ from (1), we have

$$\int_{-T}^{T} f(\beta)d\beta = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\int_{-T}^{T}\exp(-|\lambda\beta|^{\nu})d\beta = \frac{\lambda\nu}{\Gamma(\frac{1}{\nu})}\int_{0}^{T}\exp(-(\lambda\beta)^{\nu})d\beta$$

By introducing the change of variables $t = (\lambda\beta)^{\nu}$ it follows that $d\beta = \frac{1}{\lambda\nu}\, t^{\frac{1}{\nu}-1}\, dt$ and thus

$$\int_{-T}^{T} f(\beta)d\beta = \frac{1}{\Gamma\Big(\frac{1}{\nu}\Big)}\int_{0}^{(\lambda T)^{\nu}} t^{\frac{1}{\nu}-1}e^{-t}dt = \Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)$$

where $\Gamma_{inc}(x,a) = \frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}dt$ is the *incomplete gamma function*. From (6), we have

$$\mu = \frac{P(H_1)}{P(H_0)} = \frac{1-\int_{-T}^{T} f(\beta)d\beta}{\int_{-T}^{T} f(\beta)d\beta} = \frac{1-\Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)}{\Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)}$$

as it was given in (10). For the conditional densities $f(\beta|H_0)$ and $f(\beta|H_1)$ of noise-free coefficients from (7) and (8), the normalization constants B_0 and B_1 are

$$B_0 = \Big(\int_{-T}^{T}\exp(-|\lambda\beta|^{\nu})d\beta\Big)^{-1} = \Big(\frac{2\Gamma(\frac{1}{\nu})}{\lambda\nu}\int_{-T}^{T} f(\beta)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)}$$

and

$$B_1 = \Big(2\int_T^\infty \exp(-|\lambda\beta|^\nu)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\Big(2\int_T^\infty f(\beta)d\beta\Big)^{-1}$$
$$= \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\Big(1-\int_{-T}^{T} f(\beta)d\beta\Big)^{-1} \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Big[1-\Gamma_{inc}\Big((\lambda T)^\nu,\frac{1}{\nu}\Big)\Big]}$$

REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.

[2] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. and Machine Intel.*, vol. 11, no. 7, pp. 674–693, 1989.

[3] A. Cohen and J. Kovačević, "Wavelets: the mathematical background," *Proc. IEEE*, vol. 84, no. 4, pp. 514–522, Apr. 1996.

[4] S. Mallat, *A wavelet tour of signal processing*. Academic Press, London, 1998.

[5] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.

[6] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[7] T. Cai, "Adaptive wavelet estimation: a block thresholding and oracle inequality approach," *The Annals of Statistics*, vol. 27, pp. 898–924, 1999.

[8] T. Cai and B. Silverman, "Incorporating information on neighboring coefficients into wavelet estimation," *Sankhya*, vol. 63, pp. 127–148, 2001.

[9] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Proc.*, vol. 9, no. 9, pp. 1532–1546, Sept. 2000.

[10] B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *J. of the American Statistical Association*, vol. 93, pp. 173–179, 1998.

[11] ——, "Wavelet-based nonparametric Bayes methods," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, ser. Lecture Notes in Statistics, D. D. Dey, P. Müller, and D. Sinha, Eds., vol. 133. Springer Verlag, New York, 1998, pp. 133–155.

[12] F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a Bayesian approach," *J. of the Royal Statist. Society B*, vol. 60, pp. 725–749, 1998.

[13] D. Leporini, J.-C. Pesquet, and H. Krim, "Best basis representation with prior statistical models," in *Lecture Notes in Statistics*, P. Müller and B. Vidakovic, Eds. Springer Verlag, 1999, pp. 155–172.

[14] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. of the Amer. Statist. Assoc*, vol. 92, pp. 1413–1421, 1997.

[15] M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika*, vol. 85, no. 2, pp. 391–401, 1998.

[16] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, Lausanne, Switzerland, 1996.

[17] M. Hansen and B. Yu, "Wavelet thresholding via MDL for natural images," *IEEE Trans. Inform. Theory*, vol. 46, no. 5, pp. 1778–1788, Aug. 2000.

[18] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Medical Imaging*, vol. 20, no. 8, pp. 772–783, Aug. 2001.

[19] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosc. and Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, Aug. 2003.

[20] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inform. Theory*, vol. 45, pp. 909–919, Apr. 1999.

[21] E. P. Simoncelli, "Modeling the joint statistics of image in the wavelet domain," in *Proc. SPIE Conf. on Wavelet Applications in Signal and Image Processing VII*, Denver, CO.

[22] L. Şendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet -based denoising exploiting interscale dependency," *IEEE Trans. Signal Proc.*, vol. 50, no. 11, pp. 2744 –2756, Nov. 2002.

[23] ——, "Bivariate shrinkage with local variance estimation," *IEEE Signal Proc. Letters*, vol. 9, no. 12, pp. 438–441, Dec. 2002.

[24] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," in *Proc. IEEE Internat. Conf. on Image Proc.*, Chicago, IL, Oct. 1998.

[25] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 46, no. 4, pp. 886–902, 1998.

[26] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree structured image modeling using wavelet-domain hidden Markov models," in *Proc. SPIE Technical Conf. on Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, Denver, CO, July 1999.

[27] G. Fan and X. G. Xia, "Image denoising using local contextual hidden Markov model in the wavelet domain," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 125–128, May 2001.

[28] 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.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.

[29] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising using a local Gaussian scale mixture model in the wavelet domain," in *Proc. SPIE, 45th Annual Meeting*, San Diego, July 2000.

[30] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain," in *Proc. IEEE Internat. Conf. on Image Proc.*, Thessaloniki, Greece, Oct. 2001.

[31] ——, "Image denoising using Gaussian scale mixtures in the wavelet domain," *IEEE Trans. Image Proc.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[32] X. Li and M. Orchard, "Spatially adaptive denoising under overcomplete expansion," in *Proc. IEEE Internat. Conf. on Image Proc.*, Vancouver, Canada, Sept. 2000.

[33] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Trans. Image processing*, vol. 6, no. 4, pp. 549–565, Apr. 1997.

[34] M. Jansen and A. Bultheel, "Geometrical priors for noisefree wavelet coefficients in image denoising," in *Bayesian inference in wavelet based models*, ser. Lecture Notes in Statistics, P. Müller and B. Vidakovic, Eds., vol. 141. Springer Verlag, 1999, pp. 223–242.

[35] ——, "Empirical Bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Stat. Assoc.*, vol. 96, no. 454, pp. 629–639, 2001.

[36] 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*, vol. 11, no. 5, pp. 545–557, May 2002.

[37] ——, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Medical Imaging*, vol. 22, no. 3, pp. 323–331, Mar. 2003.

[38] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inform. Theory*, vol. 14, no. 3, pp. 434–443, May 1968.

[39] P. Scheunders, "Wavelet thresholding of multivalued images," *IEEE Trans. Image Proc.*, vol. 13, no. 4, pp. 475–483, Apr. 2004.

[40] J. Kalifa and S. Mallat, "Minimax restoration and deconvolution," in *Bayesian Inference in Wavelet-Based Models*, ser. Lecture Notes in Statistics, P. Müller and B. Vidakovic, Eds. Springer Verlag, 1999, pp. 115–138,.

[41] B. Corner, R. Narajanan, and S. Reichenbach, "Noise estimation in remote sensing imagery using data masking," *Internat. J. Remote Sensing*, vol. 24, no. 4, pp. 689–702, 2003.

[42] M. Abrams and S. Cain, "Sampling, radiometry and image reconstruction for polar and geostationary meteorological remote sensing systems," in *Proc. SPIE, Image Reconstruction from Incomplete Data II*, P. Bone, M. Fiddy, and R. Milane, Eds., vol. 4792, pp. 207–215.

[43] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.

[44] J. Starck, E. Candes, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Proc.*, vol. 11, no. 6, pp. 670 –684, June 2002.

Aleksandra Pižurica was born in Novi Sad, Yugoslavia in 1969. In 1994, she received the Diploma degree in electrical engineering from the University of Novi Sad (Serbia and Montenegro), in 1997 the M.Sc. diploma in telecommunications from the University of Belgrade (Serbia and Montenegro) and in 2002 the Ph.D. degree in applied sciences from the Ghent University (Belgium). Currently, she is a postdoctoral researcher at the Department of Telecommunications and Information Processing of the Ghent University.

Her research interests include image restoration, multiresolution representations, Markov Random Field models, signal detection and estimation and optical communication systems.

Wilfried Philips was born in Aalst, Belgium on October 19, 1966. In 1989, he received the Diploma degree in electrical engineering and in 1993 the Ph.D. degree in applied sciences, both from the Ghent University, Belgium.

Since November 1997 he is a lecturer at the Department of Telecommunications and Information Processing of the Ghent University.

His main research interests are image restoration, image analysis, and lossless and lossy data compression of images and video.