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

Aleksandra Pižurica and Wilfried Philips

Abstract

We develop three novel wavelet domain denoising methods for *subband-adaptive*, *spatially-adaptive* and *multivalued* image denoising. The core of our approach is 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 generalized Laplacian prior for noise-free subband data and additive white Gaussian noise. The results demonstrate that the new subband-adaptive shrinkage function outperforms in terms of mean squared error Bayesian thresholding approaches. Spatially adaptive version of the proposed method yields better results than the existing spatially adaptive ones of similar and of higher complexity. The performance on color and on multispectral images is superior with respect to recent multiband wavelet thresholding.

Keywords: Image denoising, wavelets, generalized likelihood ratio, color, 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

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.rug.ac.be, philips@telin.rug.ac.be, Tel: +32 9 264 34 12, Fax: +32 9 264 42. 95. **Corresponding author:** Aleksandra Pižurica

the ones containing a "significant" noise-free component should be reduced less [5]. A common shrinkage approach is thresholding [5,6], 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 [7–32], which imposes a prior distribution on noise-free data. Common priors for noise-free subband data include (generalized) Laplacian [2,7,14,16], double stochastic (Gaussian scale mixture) models [23–25] and mixtures of two distributions [8–13] 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) [20–22] and Markov Random Field (MRF) [28–31] 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 [7–10]. 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 [12–15]) or under a delta cost function (maximum a posteriori - MAP estimation [16]). The above listed methods are *subband adaptive*: they are optimized with respect to the marginal subband statistics. The use of bivariate and joint statistics of wavelet coefficients is addressed in [17, 18], 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 [19–27].

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 [28–32], where empirical wavelet coefficients are multiplied with the probability of containing a significant noise-free component. Here, the proposed approach for estimating these probabilities is essentially different: previous methods were relying on preliminary coefficient classifications yielding binary masks that were combined with MRF priors [28–31] or used for empirical density estimation followed by fitting of log-likelihood ratios [32]. In contrast to this, our new approach 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 probability that it presents a signal of interest. We show 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 further on for denoising multivalued images. Section IV concludes the paper.

II. Subband adaptive Bayesian wavelet shrinkage

We assume the input image is contaminated with 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 each wavelet subband[1]

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

where β_i are noise-free wavelet coefficients, ϵ_i are independent identically distributed (i.i.d.) normal random variables $\epsilon_i \sim N(0, \sigma^2)$ and n is the number of coefficients in a subband. A widely used generalized Laplacian (also called generalized Gaussian) prior for the noise-free subband data [2, 7, 14, 16] 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 for natural images typically $\nu \in [0, 1]$. The variance σ_β^2 and the kurtosis κ_β of the noise-free histogram are $\sigma_\beta^2 = \frac{\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})}$ and $\kappa_\beta = \frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})}$, respectively [14].

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: "signal of interest is absent" and H_1: "signal of interest is present" (in a given coefficient) precisely as:

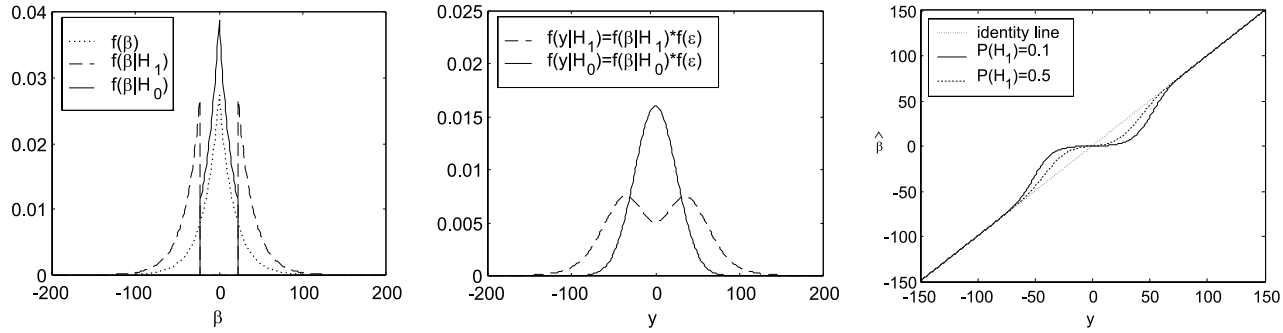
$$H_0 : |\beta| \leq T \quad \text{and} \quad H_1 : |\beta| > T. \tag{3}$$

We consider a simple estimator where each wavelet coefficient is multiplied with the probability that it contains a signal of interest, given its observed value

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

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 generalized likelihood ratio [33]. In the remainder, we call the above shrinkage rule *ProbShrink*. Fig. 1(a) shows an example of the conditional densities of *noise-free* coefficients $f(\beta|H_0)$ and $f(\beta|H_1)$ and

[1] As it is common in the related literature, for compactness we omit here the indices that denote the scale and the orientation and we denote the spatial position with a single index, like in a raster scanning.



(a) (b) (c)

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.

Fig. 1(b) illustrates the corresponding conditional densities of the noisy coefficients, $f(y|H_0)$ and $f(y|H_1)$ which result from the following convolutions

$$f(y|H_0) = \int_{-\infty}^{\infty} \phi(y-\beta;\sigma) f(\beta|H_0) d\beta \ , \quad f(y|H_1) = \int_{-\infty}^{\infty} \phi(y-\beta;\sigma) f(\beta|H_1) d\beta \qquad (5)$$

where $\phi(y;\sigma)$ is the zero mean Gaussian density with the standard deviation σ. 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. Next we address the specification of this parameter in a given subband.

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., [15, 28–31]) or $P(H_1)$ was estimated empirically as a given fraction of the observed *noisy* coefficients [12, 32]. Here we derive the probability $P(H_1)$ from the prior model for the noise-free coefficients in a given subband. In particular, for the model (3), we propose to estimate $P(H_1)$ as

$$P(H_1) = \int_{-\infty}^{\infty} f(\beta|H_1) d\beta = 1 - \int_{-T}^{T} f(\beta) d\beta. \qquad (6)$$

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

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

and

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

with the normalization constants (see Appendix):

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

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. Experimental performance evaluation

The threshold 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. It is reasonable to relate this threshold to the noise standard deviation choosing $T = \sigma$ like in related approaches [30, 31], where this choice was motivated by oracle thresholding [4]. Our experiments on different natural images confirm that this choice yields the best mean squared error performance of the proposed estimator (see Fig. 2).

Table I compares the peak signal to noise ratio[2] (PSNR) performance of the proposed *ProbShrink* rule with $T = \sigma$ and *BayesShrink* soft-thresholding of [7]. The latter is Bayesian

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

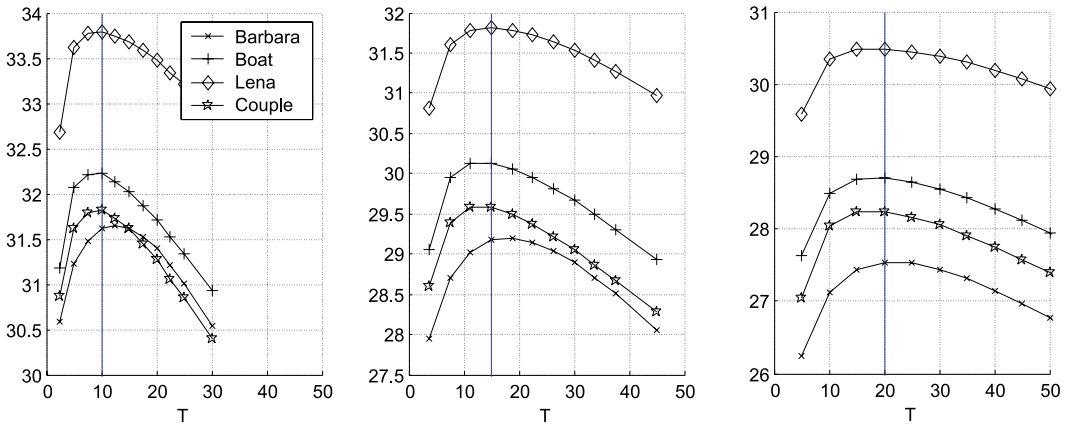


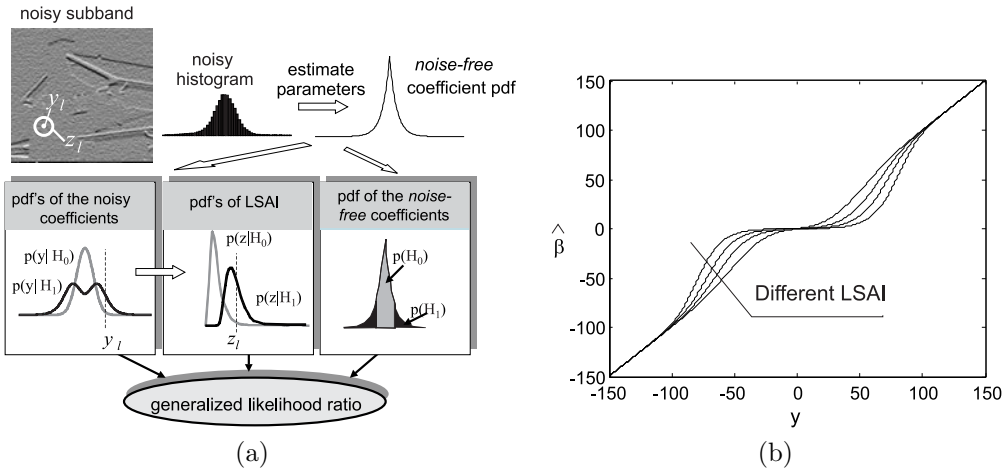
Fig. 2. Resulting PSNR[dB] values for the subband adaptive *ProbShrink* estimator as a function of the threshold T. From left to right the noise standard deviation is $\sigma = 10$, $\sigma = 15$ and $\sigma = 20$.

TABLE I

PSNR[dB] RESULTS OF THE PROPOSED *ProbShrink* RULE COMPARED TO MSE-OPTIMUM SOFT-THRESHOLDING *BayesShrink* UNDER THE GENERALIZED LAPLACIAN PRIOR FOR *sym*8 WAVELET.

	Standard deviation of noise			
Estimator	10	15	20	25
	BARBARA			
noisy image	28.12	24.59	22.09	20.17
BayesShrink	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
BayesShrink	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
BayesShrink	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
BayesShrink	33.47	31.53	30.26	29.30
ProbShrink	**33.80**	**31.82**	**30.49**	**29.51**

soft-thresholding with the threshold σ^2/σ_β, which is for natural images (i.e., for the generalized Laplacian prior) optimal in terms of mean squared error. The results demonstrate that *ProbShrink* rule outperforms *BayesShrink* on all tested 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,



(a) (b)

Fig. 3. (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.

especially because it is of similar complexity to Bayesian thresholding.

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:

$$\hat{\beta}_l = P(H_1|y_l, z_l)y_l = \frac{\eta_l \xi_l \mu}{1 + \eta_l \xi_l \mu} y_l, \tag{11}$$

where

$$\eta_l = \frac{f(y_l|H_1)}{f(y_l|H_0)}, \quad \xi_l = \frac{f(z_l|H_1)}{f(z_l|H_0)} \quad \text{and} \quad \mu = \frac{P(H_1)}{P(H_0)}. \tag{12}$$

The characteristic parts of the method are illustrated in Fig. 3, where the generalized likelihood ratio denotes the product $\eta_l \xi_l \mu$. The proposed method has a nice heuristic explanation: 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 (11) is the same as in our previous work [32], there is a great difference between the new, here proposed approach and that of [32]. In the first place, the approach of [32] assumes no particular prior on noise-free data and no particular noise distribution. Instead, it uses interscale products for heuristic preliminary coefficient classification (mask determination) and involves 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:

$$z_l = \frac{1}{N} \sum_{k \in \delta(l)} \omega_k, \tag{13}$$

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 $\omega_l < 0$. The resulting spatially adaptive estimator (11) yields a significant improvement with respect to the subband-adaptive estimator (4) as it is illustrated in Fig. 4. In all cases, the window size 3x3 was used, which was experimentally found optimal. Fig. 5 demonstrates that the visual improvement resulting from the spatial adaptation of the estimator is also evident.

A. Results in the orthogonal wavelet representation

Fig. 6 shows the results of the proposed method in comparison with several representative denoising methods using the orthogonal transform: the bivariate shrinkage of [17], the locally adaptive Wiener method of [23], the Hidden Markov Tree (HMT) approach of [20] and the

[3] Such assumptions were earlier used, e.g., in [23, 26]

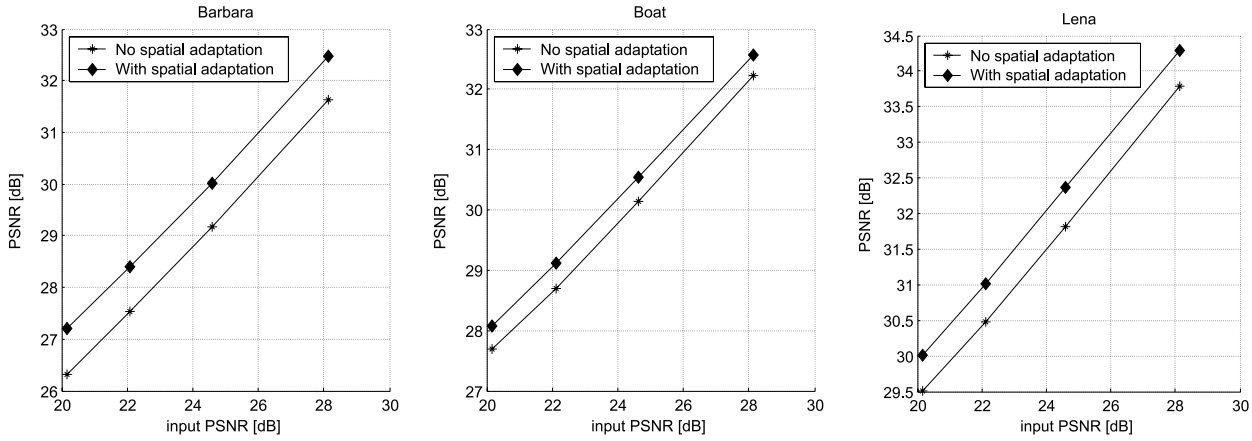


Fig. 4. Performance of the proposed *ProbShrink* method without spatial adaptation (subband adaptive shrinkage) and with spatial adaptation on three test images: (a) *Barbara*, (b) *Boat* and (c) *Lena*.

Fig. 5. 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.

soft thresholding of [6]. Our algorithm was implemented with five decomposition levels, using *symmlet* with eight vanishing moments [1] and the fixed window size 3x3. The results demonstrate that the proposed method yields the results that are among the best available ones with the orthogonal wavelet transform and that it outperforms some of the much more complex recent methods that use Hidden Markov Trees.

B. Results in the redundant wavelet representation

We also implemented the proposed method with a *non-decimated* wavelet representation using the algorithm *à trous* [4]. In Fig. 7, the results are plotted in comparison with seven

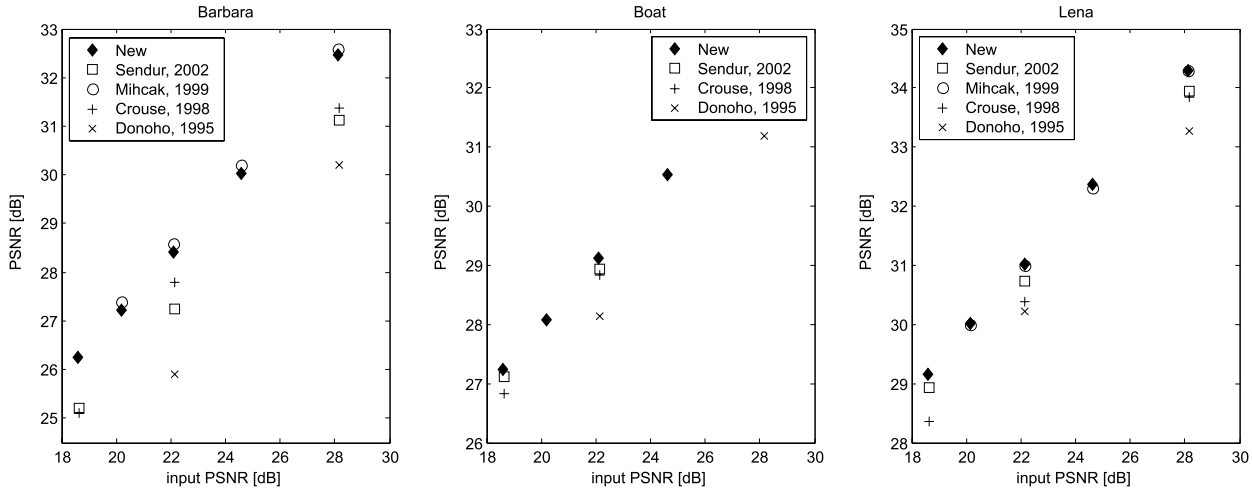


Fig. 6. Performance of the proposed spatially adaptive shrinkage and several recent methods using orthogonal wavelet transform: [6] (Donoho, 1995), [20] (Crouse, 1998), [23] (Mihcak, 1999) and [17] (Sendur, 2002).

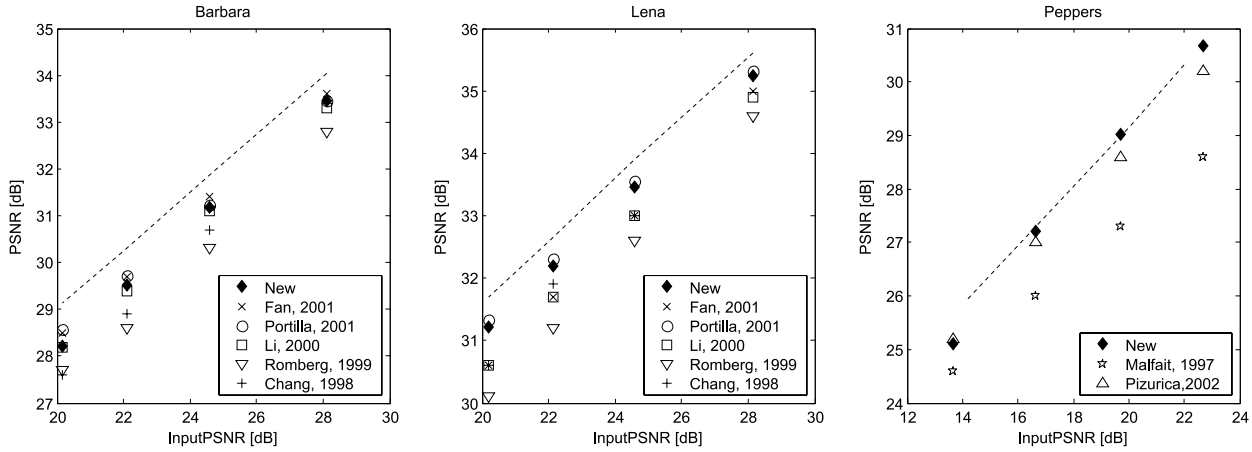


Fig. 7. Results of several recent methods, which use redundant wavelet representation with three orientations: [28] (Malfait, 1997), [19] (Chang, 1998), [21] (Romberg, 1999), [27] (Li, 2000), [22] (Fan, 2001), [25] (Portilla, 2001) and [31] (Pizurica, 2002). Dashed lines show the best available results, obtained with 8-orientation steerable pyramid in [26].

recently published methods, which also use overcomplete wavelet transforms with three orientations per scale: spatially adaptive thresholding of [19], locally adaptive Wiener filtering of [27], MMSE estimation with a Gaussian scale mixture prior of [25], MMSE estimators with two different HMT models [21, 22] and two wavelet shrinkage methods based on MRF priors [28, 31]. Dashed lines in these diagrams show the best published results so far that were obtained with an 8-orientation redundant steerable pyramid in [26]. The results are plotted for three test images: 512x512 *Lena* and *Barbara* and 256x256 *Peppers*. We experimented

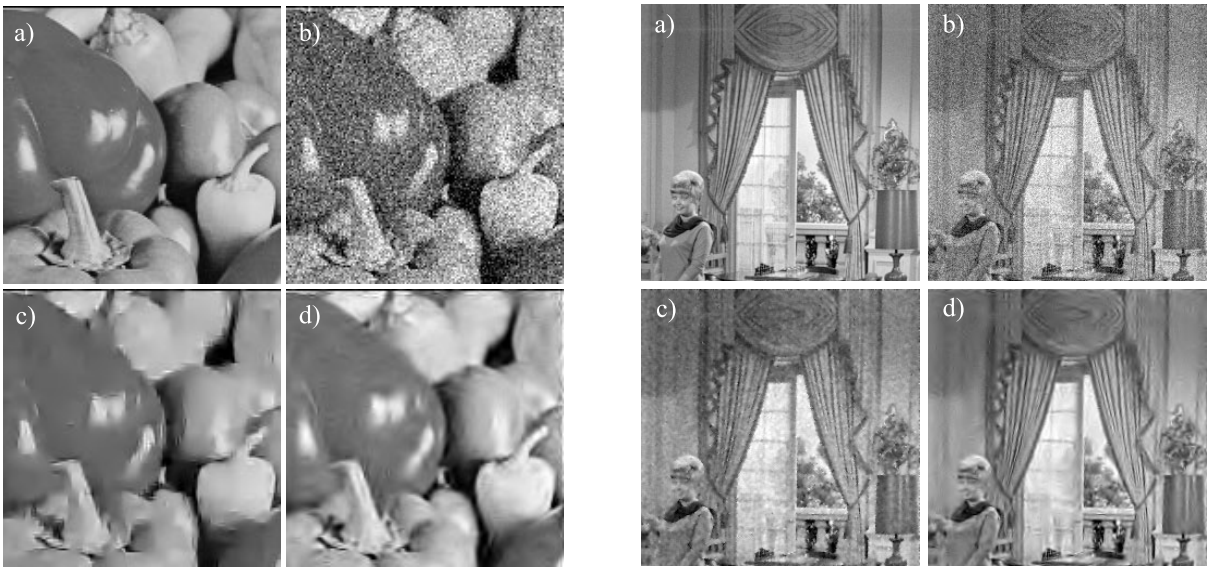


Fig. 8. **Left**: (a) noise-free part of the *Peppers* image, (b) noisy image, $\sigma = 37.5$, (c) the result of the MRF-based wavelet domain method of [31] and (d) the result of the new method. **Right**: noise-free part of the *Couple* image, (b) noisy image, $\sigma = 15$, (c) the result of the spatially adaptive Matlab's Wiener filter in the image domain and (d) the result of the new method.

only with two types of orthogonal wavelets: Daubechies wavelets and symmlets [1, 4]. Among these, on the *Barbara* image the best results were obtained using the symmlet with eight vanishing moments, while for *Lena* and *Peppers* images the best results were obtained using the Daubechies wavelet with two vanishing moments. In all cases, we used four decomposition levels and the square window size 7x7, which was experimentally found optimal.

The results in Fig. 7 demonstrate that the new method outperforms more complex related ones that are based on HMT priors [21] and on MRF priors [28, 31]. Visual improvement is illustrated in Fig. 8. The new, non-recursive method is much faster as compared to our previous MRF-based method from [31]. On a Pentium IV processor with 1.8GHz the new method takes 14s to process 512x512 image; the processing time of the MRF based method [31] is three times as long with the same processor. Here proposed method is also less complex as compared to the sophisticated MMSE approach of [25], while yielding a similar MSE performance (in [25], the reported processing time for a 512x512 image was 12.8 min on a Pentium III processor with 900 MHz). Fig. 7 also shows that in comparison with the approaches of similar complexity [19, 27], the new method yields a significant improvement.

Fig. 9. Parts of the noise-free, noisy (σ=25, PSNR=20.17dB) and the denoised (PSNR=30.73dB) image.

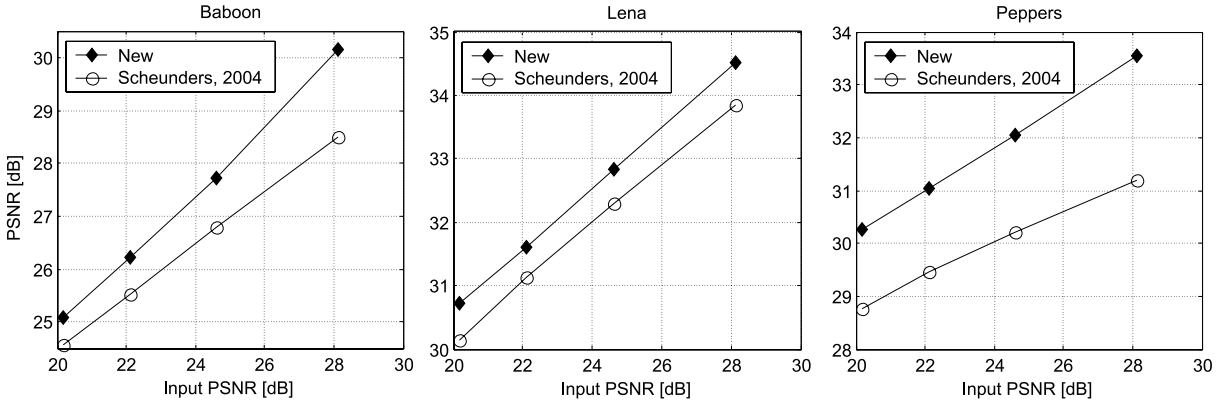


Fig. 10. PSNR results on color images in comparison with the method of [34] (Scheunders, 2004).

C. 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 $\omega^i_{l,s}$ denote the noisy coefficient magnitude in the image band i, wavelet subband s and spatial position l. A possible multiband extension of the LSAI from (13) is:

$$z^b_{l,s} = \frac{1}{NB} \sum_{i=1}^{B} \sum_{k \in \delta(l)} \omega^i_{k,s} \tag{14}$$

TABLE II
PSNR[dB] RESULTS FOR COLOR 512X512 IMAGES.

	Standard deviation of noise			
Image	10	15	20	25
Baboon	30.15	27.73	26.23	25.09
Lena	34.52	32.83	31.62	30.73
Peppers	33.54	32.05	31.05	30.26

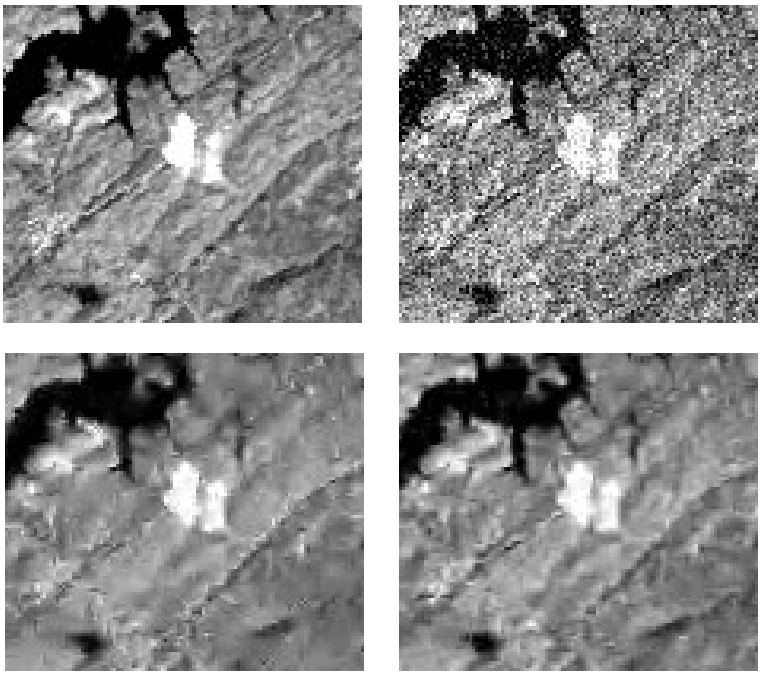


Fig. 11. Left to right: parts of a noise-free band of the *Landsat* image from [34], noisy image ($\sigma = 35, PSNR = 17.24dB$) result of [34] (PSNR=21.16dB) and the result of the proposed method (PSNR=22.42dB).

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.[4]

[4] Actually, we assume 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 $f(y|H_0)$, and that they are *conditionally* independent given that either H_0 or H_1 is true.

For color images, visual results are illustrated in Fig. 9 and the PSNR results are tabulated in Table II. As the reference method, we use the recent multiband wavelet thresholding of [34] that we implemented with the same wavelet transform as our method. Fig. 10 compares the performance of the two methods on color images and Fig. 11 illustrates the results on a multispectral *Landsat* image from [34]. Our results demonstrate a superior performance of the new method on all tested color and multispectral images. Its effectiveness 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 in each image band different even though they are dependant on information from other bands as well as on the measurements from the given band. We believe that further improvements may result from developing a vector based approach, which is currently under investigation.

D. Implementation details

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 [7, 14]. The results in this paper were obtained assuming that the noise standard deviation σ was known (as it is usual for reporting 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]. Matlab implementations of the proposed method for greyscale and for color images are available at http://telin.UGent.be/ ∼sanja.

IV. Conclusion

We developed a new wavelet domain denoising method based on 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, which on natural images yields a better MSE performance than Bayesian soft-thresholding with the MSE-optimum threshold. The proposed spatially adaptive denoising method yields superior results as com-

pared to some much more complex recent approaches based on HMT and MRF models. These results motivate strongly 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 [35], steerable pyramids [26] or curvelets [36].

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 [34] and for the fruitful discussions on multiband image denoising. We also thank to Javier Portilla for providing test images for the performance comparison with [25,26].

V. Appendix

For the generalized Laplacian prior $f(\beta) = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\exp(-|\lambda\beta|^{\nu})$, we have

$$\int_{-T}^{T} f(\beta)d\beta = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\int_{-T}^{T}\exp(-|\lambda\beta|^{\nu})d\beta = \frac{\lambda\nu}{\Gamma(\frac{1}{\nu})}\int_{0}^{T}\exp(-(\lambda\beta)^{\nu})d\beta. \tag{15}$$

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

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

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

$$\mu = \frac{P(H_1)}{P(H_0)} = \frac{1-\int_{-T}^{T} f(\beta)d\beta}{\int_{-T}^{T} f(\beta)d\beta} = \frac{1-\Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)}{\Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)}, \tag{16}$$

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

$$B_0 = \Big(\int_{-T}^{T}\exp(-|\lambda\beta|^{\nu})d\beta\Big)^{-1} = \Big(\frac{2\Gamma(\frac{1}{\nu})}{\lambda\nu}\int_{-T}^{T} f(\beta)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)} \tag{17}$$

and

$$B_1 = \Big(2\int_T^\infty \exp(-|\lambda\beta|^\nu)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\Big(2\int_T^\infty f(\beta)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\Big(1-\int_{-T}^{T} f(\beta)d\beta\Big)^{-1} \quad (18)$$
$$= \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Big[1-\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)\Big]}.$$

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] 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.

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

[9] ——, "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.

[10] 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.

[11] D. Leporini, J. C. Pasquet, 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.

[12] 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.

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

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] M. S. Crouse, R. D. Nowak, and R. G. Baranuik, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. Signal Proc.*, vol. 46, no. 4, pp. 886–902, 1998.

[21] J. K. Romberg, H. Choi, and R. G. 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

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

[28] 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, 1997.

[29] 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.

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

[31] 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.

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

[33] 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.

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

[35] 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.

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