Removal of Correlated Noise by Modeling the Signal of Interest in the Wavelet Domain

Bart Goossens, Aleksandra Pižurica and Wilfried Philips

ABSTRACT

Images, captured with digital imaging devices, often contain noise. In literature, many algorithms exist for the removal of white uncorrelated noise, but they usually fail when applied to images with correlated noise. In this paper, we design a new denoising method for the removal of correlated noise, by modeling the significance of the noise-free wavelet coefficients in a local window using a new significance measure that defines the "signal of interest" and that is applicable to correlated noise. We combine the intrascale model with a Hidden Markov Tree model to capture the interscale dependencies between the wavelet coefficients. We propose a denoising method based on the combined model and a less redundant wavelet transform. We present results that show that the new method performs as well as the state-of-the-art wavelet-based methods, while having a lower computational complexity.

I. INTRODUCTION

DIGITAL imaging devices often produce noise, originating from the analogue components (sensors, amplifiers) in the devices. In many cases, it is desirable to remove the noise, not only to improve the visual quality of the images, but also to improve compression performance. Multiresolution representations like those based on wavelets have proven to be powerful tools for image denoising and in literature many techniques have been proposed for this purpose, like [1]–[7]. It is often assumed that wavelet coefficients are uncorrelated and statistically independent, allowing simple and elegant shrinkage rules like soft and hard thresholding. In practice, noise samples are often correlated (for example by demosaicking in digital cameras) and these techniques subsequently fail when applied to images corrupted with correlated noise.

B. Goossens*, A. Pižurica and W. Philips are with the Department of Telecommunications and Information Processing (TELIN-IPI-IBBT), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium. B. Goossens is supported by Special Research Fund (BOF) from Ghent University. A. Pižurica is a postdoctoral research fellow of FWO, Flanders. Email: Bart.Goossens@telin.UGent.be, Aleksandra.Pizurica@telin.UGent.be, Wilfried Philips@telin.UGent.be, Tel: +32 9 264 7966, Fax: +32 9 264 4295. EDICS: RST-DNOI Denoising.

A first approach could be the use of a prewhitening operation for the noise, followed by the wavelet transformation and thresholding. This solution would have the advantage that the noise components are decorrelated and an algorithm designed for white noise could be used. However, the prewhitening operation not only alters the characteristics of the noise, but also the underlying signal, thus the advantages diminish [8]. Therefore, the authors of [8] propose a modified *universal threshold*, that depends on the noise variance on every subband of the wavelet transform. In [2] it is noted that the denoising performance can be improved by doing joint (vector-based) thresholding instead of thresholding each coefficient independently, by looking at correlated coefficients in the spatial neighbourhood.

In [4], a Bayesian Least Squares estimator with a Gaussian Scale Mixture prior (BLS-GSM) has been proposed for this purpose. Local neighbourhoods of coefficients are modeled as the product of a Gaussian vector and a hidden scalar multiplier, both independent of each other. The noise is treated as multivariate Gaussian. The least squares, i.e. minimum mean square error (MMSE) estimator for this model has been derived and this estimator is the weighted average of the local linear (Wiener) estimates over all possible values of the hidden multiplier.

Other authors study interscale models, which characterize the dependencies between wavelet coefficients at *different* scales. In [3], a bivariate distribution is used for a wavelet coefficient and its parent, for the removal of *white* noise. In [9], [10] multiscale stochastic processes on quadtrees are studied for modeling the joint distribution of the wavelet coefficients along the quadtree structure of the wavelet transform. *Hidden Markov Tree* (HMT) models [11]–[13] establish Markovian parent-child relationships among *hidden state variables* rather than among the coefficients themselves. The prior parameters are then estimated iteratively by means of the Baum-Welch algorithm. To overcome the lack of spatial adaptation, a local contextual *Hidden Markov Model* (HMM) [14] offers an improvement. An additional hidden variable is the local spatial activity, calculated as the local average energy of the surrounding wavelet coefficients. Related Markov Random Field models were used in [15]–[17].

The work of [5] introduced an approach of estimating the probability that a given wavelet coefficient represents a signal of interest given its value and knowing the marginal distribution of the noise-free coefficients. This probability was used as a suppression factor for the wavelet coefficients in the so-called ProbShrink estimator. A locally adaptive version of this approach was also introduced in [5] which attempts at making use of spatial correlations that exist between the wavelet coefficients within the same subband. In this case, the probability of signal presence was conditioned not only on the coefficient value but also on a local spatial activity indicator (LSAI) computed from the surrounding coefficients. This LSAI in [5] is practically the locally averaged coefficient magnitude. The rationale behind this approach was: if a noise-free component is large (small) then the majority of the neighbouring coefficients within a local window is also likely to be large (small) because true image discontinuities typically result in spatially clustered wavelet coefficients. This locally adaptive estimator performs quite well given its low complexity but has an intrinstic limitation: it is not applicable to *correlated* noise. In contrast to white noise, correlated noise can result in spatially clustered large wavelet coefficients, and in this case the LSAI of [5] cannot make a difference between the signal and noise.

The main novelty of this work is that we estimate the probability of signal presence given a *vector* of surrounding wavelet coefficients, i.e., given a *structure* of the local neighbourhood. We are now able to estimate how likely it is that a given coefficient represents signal or noise given true correlatedness of wavelet coefficients. Other important contributions are combining the proposed approach with a Hidden Markov Tree model to capture not only intra- but also inter-scale coefficient dependencies and devising a minimum mean squared error estimator for the proposed statistical model.

This work is on the one hand an improvement and generalization of the main ideas of [5] where the estimation of probability of signal presence is now improved, applicable to cases with correlated noise, and where the estimator is combined with a powerful HMT model. On the other hand, this work can also be seen as an improvement and generalization of the HMT approaches of [11]–[14], where we now employ a better likelihood model and a better estimation of the involved state probabilities.

Finally, in relation to GSM based approaches [4], [18], this work applies GSM models to a novel problem: the estimation of the probability that a given wavelet coefficient vector is a signal of interest. The results show that this alternative estimation approach, combined with the HMT model and a less redundant transform, offers some significant savings in complexity (computation time) without losing the performance.

The organization of this paper is as follows: Section II introduces some basic concepts used in this paper. Section III describes the intrascale statistical model that is used in the wavelet domain. In Section V, the intrascale model is combined with the HMT that models interscale dependencies. Results and a discussion are given in Section VII. Finally, Section VIII concludes this paper.

II. Multiscale wavelet analysis of the noisy input signal

A. The spectral noise characteristic

Correlated noise (or coloured noise) is usually specified by its *Energy Spectral Density* (ESD). The ESD describes how the energy (or variance) of a signal is distributed in frequency space and for 2-D signals it is defined as:

$$\Phi(k,l) = |F(k,l)|^2$$

where $F(k,l)$ is the Discrete Fourier Transform (DFT) of the signal. The DFT would be the ideal choice of transform, because it completely decorrelates the noise. However, the DFT cannot recover information on particular positions in the spatial domain. This makes the representation less convenient when it comes to analyzing non-stationary signals.

B. The wavelet transform

To overcome this deficiency, the discrete wavelet transform has been introduced. The orthogonal discrete wavelet transform (ODWT) decomposes a signal over an orthogonal basis of functions that are translates and dilates of the analyzing wavelet, called *mother wavelet*. This provides a non-uniform partitioning of the time (space)-frequency plane, which makes it possible to retrieve information at specific spatial positions. The wavelet coefficients are samples of bandpass filtered versions of the input signal, while the scaling coefficients are samples of lowpass filtered versions. By the linearity of the transform, signal independent additive noise is transformed into signal independent additive noise. For this reason, we will transform our noisy image to the wavelet domain, estimate the noise-free wavelet coefficients using an additive statistical signal–plus–noise model and reconstruct the image by applying the inverse wavelet transform.

Despite the efficiency and the sparsity of the decimated wavelet transform, there are some fundamental problems [19]: 1) positive and negative oscillations of the coefficients around singularities, 2) shift variance, 3) aliasing caused by downsampling operations and 4) poor directional selectivity. For denoising, the second and third problems are the most severe, since the local energy

signature of edges in the transform domain depends on the edge position and the aliasing creates visually disturbing artifacts in the reconstructed signal. In the last decades, many alternative representations have been developed, including redundant wavelet transforms, the Dual-Tree Complex Wavelet transform (DT-CWT) [20], Steerable pyramids [4], the Curvelet transform [21] and the Contourlet transform [22]. In this work, we will adapt the DT-CWT, based on its low redundancy (factor 4 for images[1]), efficiency (by using seperable filters) and better directional selectivity (6 orientations) compared to the orthogonal DWT. For a detailed explanation, see [19].

C. Statistical modeling in the wavelet domain

To build a statistical signal–plus–noise model wavelet domain model that deals with correlated noise, it is useful to relate the noise ESD to the correlation properties of noise components of wavelet coefficients. Therefore we first consider the autocorrelation function of a signal in the spatial domain, denoted as $R(\mathbf{p},\mathbf{q})$ where $\mathbf{p}$, $\mathbf{q}$ are two-dimensional vectors representing the spatial position in the image. Let us first consider *spatially stationary* noise. Then the correlation between two pixel intensities depends only on the difference between their positions:

$$R(\mathbf{p},\mathbf{q}) = R(\mathbf{0},\mathbf{q}-\mathbf{p}) \tag{1}$$

The Wiener−Khintchine theorem [23] states that the autocorrelation function $R(\mathbf{0},\mathbf{q})$ is the inverse Fourier transform of the noise ESD. To obtain the autocorrelation function for a specific wavelet subband, at scale s and orientation o, we apply the wavelet filters and decimations associated with each decomposition stage i on the autocorrelation function of the previous stage $i-1$, starting from the autocorrelation function in the pixel domain $R^{(0)}(\mathbf{0},\mathbf{p})$ as follows:

$$R^{(i+1)}(\mathbf{p},\mathbf{q}) = \sum_{\mathbf{k}\in\mathbb{Z}^2}\sum_{\mathbf{l}\in\mathbb{Z}^2} R^{(i)}(\mathbf{0},2(\mathbf{q}-\mathbf{p})-\mathbf{k}+\mathbf{l})h_{\mathbf{k}}^{(i)}h_{\mathbf{l}}^{(i)} \tag{2}$$

where $h_{\mathbf{k}}^{(i)}$ represents a two-dimensional wavelet filter kernel (either highpass or lowpass) associated with stage i. From the decimation factor 2 in (2) it is clear that the wavelet analysis also decomposes the autocorrelation function and for sufficiently short wavelet filters, the support of the autocorrelation function becomes smaller with each scale. This is illustrated in Fig. 1 for the DT-CWT.[2]. It can be seen that a small square window (e.g. 3×3) suffices to capture most of the noise correlations on each scale and orientation.

[1]compared to a redundancy factor $1+3N_s$ for an undecimated DWT with 3 orientations and N_s scales.

[2]Note that, in order to obtain the Oriented DT-CWT from separable wavelet filters [19], an extra linear transform performed at the output of the trees (see [19]) This also has influence on the orientedness of the autocorrelation functions and is therefore also taken into account here.

III. Statistical image model for denoising

A. Signal and noise model

In this Section, we present a statistical model for one subband (s,o) at scale s and orientation o of the wavelet transform. We assume an input image, corrupted with additive coloured noise. As said before, the linearity of the wavelet transform yields an equivalent additive relationship for subband (s,o) in the transform domain:

$$\mathbf{y}_j = \mathbf{x}_j + \mathbf{w}_j, \quad j = 1,...,N \tag{3}$$

where $\mathbf{y}_j$, $\mathbf{x}_j$ and $\mathbf{w}_j$ are the wavelet coefficients at spatial position j (like in raster scanning) of respectively the observed noisy image, the original image and the noise. In this notation, wavelet coefficients within a small neighbourhood of size $M\times M$ are clustered into a coefficient vector of size $d=M^2$. The neighbourhoods are overlapping and are extended periodically at the image boundaries. $\mathbf{w}_j$ is assumed to be Gaussian noise with zero mean and covariance $\mathbf{C}_w$. If the noise ESD in the spatial domain is known in advance, the noise covariance matrix $\mathbf{C}_w$ can be computed directly from the autocorrelation function for that subband, obtained using (2). Therefore, we ignore correlations between wavelets coefficients that are not in the same neighbourhood of size $M\times M$. If the noise ESD is *not* known in advance, we estimate $\mathbf{C}_w$ for each band from the observed wavelet coefficients.

B. Bayesian spatial prior distribution

Recent statistical studies (e.g. [2], [4]) have shown that distributions of noise-free wavelet coefficients are typically symmetric around the mode, have a highly kurtotic non-Gaussian behaviour and exhibit strong correlations, especially in areas with edges and textures. For many natural images, the histograms of the wavelet coefficients reveal elliptical contours, which suggests the use of the elliptically symmetric family for modeling this behaviour. The family of *elliptically symmetric* distributions is defined by the following class of densities [24]:

$$f_{\mathbf{X}}(\mathbf{x}) = k_d|\mathbf{C}_x|^{-1/2}g[(\mathbf{x}-\mathbf{m})^T\mathbf{C}_x^{-1}(\mathbf{x}-\mathbf{m})] \tag{4}$$

where $\mathbf{m}$ is the mean of the distribution, $g(u)$ is a one dimensional real-valued function, independent of the number of dimensions d and k_d is a proportionality constant. For wavelet coefficients, one usually assumes that $\mathbf{m}=0$. In this work, we use the *Bessel K Form* prior [24]–[26], with:

$$g(u) = \left(\frac{u}{2}\right)^{\frac{\tau}{2}-\frac{d}{4}} K_{\tau-d/2}(\sqrt{2u}),\ k_d = \frac{2(2\pi)^{-d/2}}{\Gamma(\tau)} \tag{5}$$

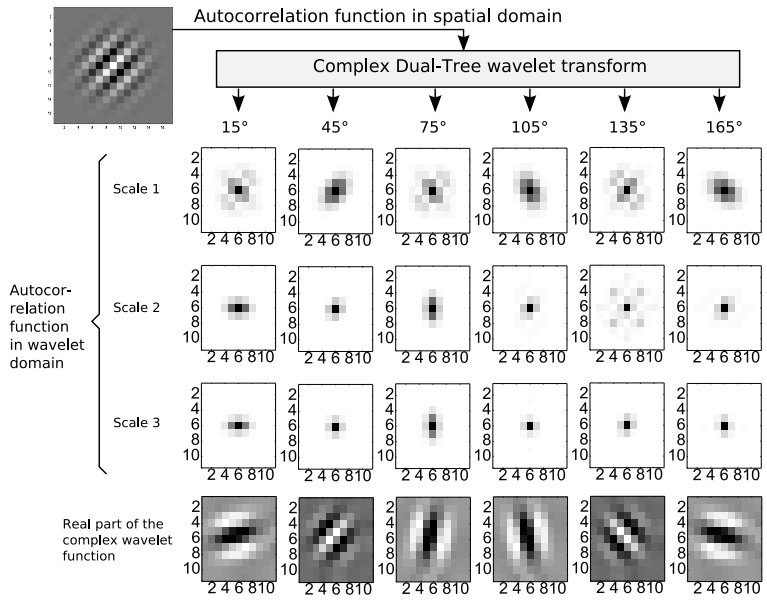


Figure 1. Illustration of the effect of the DT-CWT on a given autocorrelation function in the spatial domain (top-left corner; dark intensities correspond with negative values, dark gray corresponds with 0 and white intensities correspond with positive values). Right below are the autocorrelation functions transformed to the DT-CWT domain (the complex magnitude is shown, dark intensities correspond with large magnitudes). The bottom row contains the real parts of second scale complex wavelets corresponding with each orientation.

where $K_l(u)$ is the modified Bessel function of the second kind and order l (see [24]) and $\Gamma(\tau) = \int_0^\infty z^{\tau-1}e^{-z}dz$ is the Gamma function. In [25], it has been shown that the marginals of this distribution fit well with the observed histograms for a wide variety of images. This density also has the following Gaussian Scale Mixture representation [24]:

$$\mathbf{x} \stackrel{d}{=} z\mathbf{m} + z^{1/2}\mathbf{u} \tag{6}$$

where "$\stackrel{d}{=}$" denotes *equality in distribution*, $\mathbf{u}$ is *Gaussian* $N(\mathbf{0}, \mathbf{C}_x)$ and z, called the *hidden multiplier* [4], is *Gamma* distributed $\Gamma(\alpha = \tau, \beta = 1)$:

$$f_Z(z) = \frac{1}{\Gamma(\tau)} z^{\tau-1}e^{-z} \tag{7}$$

For the special case $\tau = 1$, where z is *exponentially* distributed, we obtain the multivariate Laplace distribution [24], [27] (see Fig. 2). For $\tau \to \infty$, the distribution approaches the Gaussian distribution (see [24]). Thus, the Bessel K Form is a generalisation of the multivariate Laplace distribution, but differs from the generalized Laplace distribution used in [2], [28]. We also note that the Bessel K Form corresponds to the symmetrized Gamma family proposed in [29]. The kurtosis is given by $\kappa = 3 + 3/\tau$, thus for small positive τ, we obtain a highly leptokurtic prior. Furthermore, the parameter τ depends on the frequency of occurrence of particular features in the image, like edges, bands, textures [25]. In Fig. 3 the univariate marginals of the Bessel K Form density are plotted for different values of τ.

Other authors have proposed related GSM distributions like the GSM with a log-normal prior on z [18] and the GSM with Jeffrey's prior on z [4]. Among these priors, the Bessel K Form is the only one that offers explicit control of the kurtosis, which is advantageous when modeling wavelet subbands of natural images (see [25]). In [26] the Bessel K Form prior is compared to the α-stable prior and Generalized Gaussian Distribution in modeling observed histograms by means of the Kullback-Leibler divergence. The authors conclude that the Bessel K prior performs at least as well as the GGD for modeling the statistics of wavelet coefficients of a test set of natural images.

C. Prior parameter estimation

The set of model parameters for the subband at scale s and orientation o is given by $\mathbf{\Theta} = \{\tau, \mathbf{C}_x, \mathbf{C}_w\}$. Adding a Gaussian process to a Bessel K Form process alters the variance, but no other higher order statistics. This allows us to estimate the parameter τ in terms of the second and fourth order cumulants of $\mathbf{y}$ [26]:

$$\hat{k}_2 = \frac{N}{N-1}\hat{m}_2, \hat{k}_4 = \frac{N^2[(N+1)\hat{m}_4 - 3(N-1)\hat{m}_2^2]}{(N-1)(N-2)(N-3)} \tag{8}$$

where $\hat{m}_i$ is an i-th sample central moment of $\mathbf{y}$ (note that for fixed i all i-th sample central moments must be equal, see Section III-B), and N denotes the number of wavelet coefficients in the subband at scale s and

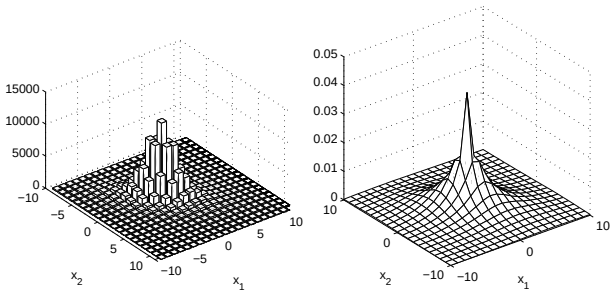


Figure 2. (a) Empirical joint histogram of a wavelet coefficient and its right neighbour (b) histogram modeled using a multivariate Bessel K Form distribution.

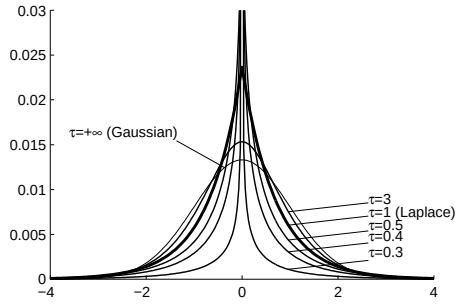


Figure 3. The univariate Bessel K form density, for different values of τ. Special value $\tau = 1$ gives the symmetric Laplace distribution, for $\tau \to +\infty$ the Bessel K form approaches the Gaussian density. Values $\tau < 1$ result in a high kurtosis (sharp peak).

orientation o. An unbiased estimate for τ is given by

$$\hat{\tau} = 3(\hat{k}_4)^{-1} \max(0, \hat{k}_2 - \sigma_w^2) \tag{9}$$

where $\sigma_w^2 = (\mathbf{C}_w)_{11}$ is the noise variance [26]. By noting that $\mathrm{E}(z) = \tau$, we estimate $\mathbf{C}_x$ as:

$$\hat{\mathbf{C}}_x = \hat{\tau}^{-1}(\hat{\mathbf{C}}_y - \hat{\mathbf{C}}_w)_+ \tag{10}$$

Due to estimation errors, usually when N is relatively small, $\hat{\mathbf{C}}_y - \hat{\mathbf{C}}_w$ may not be positive definite. Therefore $(\mathbf{C})_+$ replaces negative eigenvalues of $\mathbf{C}$ with a small positive value, such that the resulting matrix is positive definite. We use the maximum likelihood (ML) estimate for $\mathbf{C}_y$: $\hat{\mathbf{C}}_y = \frac{1}{N}\sum_{j=1}^{N} \mathbf{y}_j \mathbf{y}_j^T$. An alternative *Expectation-Maximisation* (EM) estimate for $\hat{\tau}$ does also exist (see [30], [31]). This estimator has a significantly lower MSE on average and can be equivalently used instead of (9). Because the latter estimate is computationally much more intensive we prefer to use (9) in this work.

IV. "Signal of interest" based denoising

A. Modeling the signal presence

By observing noise-free wavelet coefficients, we notice that there are large regions with *small* coefficients, separated by edges or textures exhibiting *large* coefficients. When adding noise to the image, the large regions are dominated by the noise, while the edges and textures are still visible through the noise, to some degree. The noise reduction results from shrinking the noisy regions towards 0 while keeping the *"wanted"* signal information untouched. This information, called *"signal of interest"*, can be characterized by means of a *significance measure*, based on the magnitude of the considered wavelet coefficient [5]:

$$S(x) = \mathrm{I}(|x|/\sigma_w \geq T) \tag{11}$$

where σ_w is the noise standard deviation, T is a given threshold and $\mathrm{I}(x)$ is the indicator function. The choice of the threshold T will be discussed later. We extend (11) to vectors by the following generalization:

$$S(\mathbf{x}) = \mathrm{I}\left(\left\|\mathbf{C}_w^{-1/2}\mathbf{x}\right\| \geq T\right) \tag{12}$$

where $\mathbf{C}_w^{1/2}$ is the square root of the positive definite matrix $\mathbf{C}_w$ and $||\mathbf{x}||$ is the norm of $\mathbf{x}$. By the positive-definiteness of $\mathbf{C}_w$, $\left\|\mathbf{C}_w^{-1/2}\mathbf{x}\right\|^2 = \mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{x} = T^2$ represents the equation of an ellipsoid in a d-dimensional space. The significance measure (12) then tests whether $\mathbf{x}$ is inside or outside the ellipsoid. This is illustrated in Fig. 4 and Fig. 5.

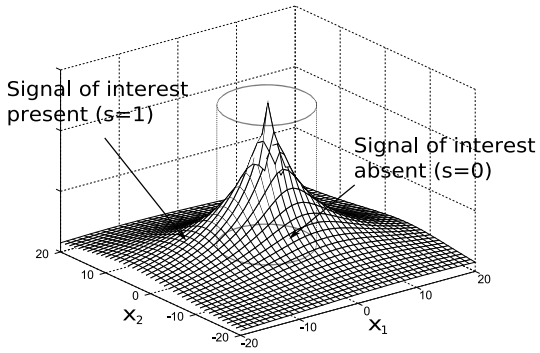


Figure 4. Illustration of "signal of interest" on the spatial prior $f_{\mathbf{X}}(\mathbf{x})$. The ellipse $\mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{x} = T^2$ is extruded to a cylinder, for visibility. Samples $\mathbf{x}$ outside the cylinder are regarded as *significant*, and represent the detail information. *Non-significant* samples inside the cylinder only contain weak signal information and are invisible when noise is added.

B. Bayesian estimation rule

We estimate the noise-free signal according to the probability that it represents *"signal of interest"*, which

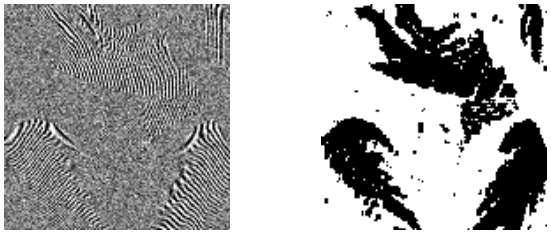


Figure 5. The posterior probability $\mathrm{P}_{S|\mathbf{Y}}(s|\mathbf{y}_j)$ can be used to detect the "signal presence". (*Left*) Cropout of a noisy wavelet band (HH_1) of the Barbara image ($\mathbf{C}_w = 25^2\mathbf{I}_d$) (*Right*) ML estimate of the significance ($\hat{s}_j = \arg\max_s \mathrm{P}_{S|\mathbf{Y}}(s|\mathbf{y}_j)$)

results in the following shrinkage rule:

$$\hat{\mathbf{x}}_j = \mathrm{P}_{S|\mathbf{Y}}(1|\mathbf{y}_j)\mathbf{y}_j \tag{13}$$

$$= \left(1 - \mathrm{P}_{S|\mathbf{Y}}(0|\mathbf{y}_j)\right)\mathbf{y}_j \tag{14}$$

$$= \left(1 - \frac{f_{\mathbf{Y}|S}(\mathbf{y}_j|0)\mathrm{P}(H_0)}{f_{\mathbf{Y}}(\mathbf{y}_j)}\right)\mathbf{y}_j \tag{15}$$

where we applied Bayes' rule in the last step. By exploiting the additivity of the noise in (3), we have

$$f_{\mathbf{Y}|S}(\mathbf{y}|0) = \int_{\mathbb{R}^d} f_{\mathbf{X}|S}(\mathbf{y}-\mathbf{w}|0) f_{\mathbf{W}}(\mathbf{w}) d\mathbf{w} \tag{16}$$

According to the significance measure (12), the conditional density $f_{\mathbf{X}|S}(\mathbf{x}|0)$ is given by:

$$f_{\mathbf{X}|S}(\mathbf{x}|0) = \frac{f_{\mathbf{X}}(\mathbf{x})}{\mathrm{P}(H_0)}\mathrm{I}\left(\left\|\mathbf{C}_w^{-1/2}\mathbf{x}\right\| < T\right) \tag{17}$$

As explained in Section III-B, $f_{\mathbf{X}}(\mathbf{x})$ is a BKF density. Consequently, the convolution in (16) is quite difficult because closed analytical forms for $f_{\mathbf{Y}|S}(\mathbf{y}|0)$ do not exist (as far as the authors are aware of). To solve this problem, we marginalize the density $f_{\mathbf{X}}(\mathbf{x})$ based on the GSM representation in (6) as $f_{\mathbf{X}}(\mathbf{x}) = \int_{z=0}^{+\infty} f_{\mathbf{X}|Z}(\mathbf{x}|z) f_Z(z) dz$ (in a practical implementation, we use numerical integration for this, see further in Section VI). We further remark that if $f_{\mathbf{X}|S}(\mathbf{x}|0)$ is the density of a Gaussian Mixture, the above convolution involves adding the noise covariance matrix $\mathbf{C}_w$ to each component of the mixture. Therefore, we approximate the indicator function in (17) using a Gaussian function:

$$f_{\mathbf{X}|S}(\mathbf{x}|0) \approx C_0 f_{\mathbf{X}}(\mathbf{x}) \exp\left(-\frac{\mathbf{x}^T\mathbf{C}_w^{-1}\mathbf{x}}{2T^2}\right) \tag{18}$$

where C_0 is a density normalization factor. This results in a Gaussian conditional prior density on $\mathbf{x}$:

$$f_{\mathbf{X}|Z,S}(\mathbf{x}|z,0) = N\left(\mathbf{x}; \mathbf{0}, \left((z\mathbf{C}_x)^{-1} + (T^2\mathbf{C}_w)^{-1}\right)^{-1}\right) \tag{19}$$

where $N(\mathbf{x}; \mathbf{0}, \mathbf{C})$ denotes the Gaussian density evaluated in $\mathbf{x}$. Next, the observation density $f_{\mathbf{Y}}(\mathbf{y})$ in (15) is also obtained by marginalizing on z:

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_0^{+\infty} f_{\mathbf{Y}|Z}(\mathbf{y}|z) f_Z(z) dz, \quad \text{with}$$

$$f_{\mathbf{Y}|Z}(\mathbf{y}|z) = N\left(\mathbf{y}; \mathbf{0}, z\mathbf{C}_u + \mathbf{C}_w\right), \tag{20}$$

and again numerical integration is used to evaluate this expression. To simplify the dependency on z in (19)-(20), it is convenient to express $f_{\mathbf{X}|Z}(\mathbf{x}|z)$ and $f_{\mathbf{Y}|Z}(\mathbf{y}|z)$ in a new basis where $\mathbf{C}_x$ and $\mathbf{C}_w$ are diagonal [4] using:

$$z\mathbf{C}_x + \mathbf{C}_w = \mathbf{U}\mathbf{Q}(z\mathbf{\Lambda} + \mathbf{I}_d)\mathbf{Q}^T\mathbf{U}^T \tag{21}$$

where $\mathbf{U}\mathbf{U}^T = \mathbf{C}_w$. $\mathbf{Q}$ and the diagonal matrix $\mathbf{\Lambda}$ are obtained by the diagonalisation $\mathbf{U}^{-1}\mathbf{C}_x\mathbf{U}^{-T} = \mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}$. By applying the linear transform to the observation vectors $\mathbf{y}_j$, i.e. $\mathbf{v}_j = (\mathbf{U}\mathbf{Q})^{-1}\mathbf{y}_j$, the conditional density of $\mathbf{v}_j$ given z is given by [4]:

$$f_{\mathbf{V}|Z}\left(\mathbf{v}_j|z\right) = N\left(\mathbf{y}; \mathbf{0}, z\mathbf{\Lambda} + \mathbf{I}_d\right) \tag{22}$$

In the Appendix, we show that the conditional density $f_{\mathbf{Y}|Z,S}(\mathbf{y}|z,0)$ can also be expressed in this basis as:

$$f_{\mathbf{V}|Z,S}\left(\mathbf{v}_j|z,0\right) = N\left(\mathbf{y}; \mathbf{0}, (z^{-1}\mathbf{\Lambda}^{-1} + T^{-2}\mathbf{I}_d)^{-1} + \mathbf{I}_d\right) \tag{23}$$

Since the linear transform matrix $(\mathbf{U}\mathbf{Q})^{-1}$ only has to be computed once per subband, independent of z, this greatly reduces the computational complexity of the proposed method, since the estimation rule (13) using (22) and (23) only requires the evaluation of Gaussian densities with diagonal covariance matrix in $\mathbf{v}_j$.

Similarly, the probability $\mathrm{P}(S=0)$, which globally estimates the absence of the signal of interest on the *whole* subband, can be efficiently precomputed per wavelet band, using this transformation (see Appendix):

$$\mathrm{P}(S=0) = \int_0^{+\infty} f_Z(z)\left(\prod_{i=1}^{d}\frac{T^2}{T^2 + z\mathbf{\Lambda}_{ii}}\right)^{1/2} dz \tag{24}$$

In case of diagonal covariance matrices ($\mathbf{C}_x = \sigma_x^2\mathbf{I}_d$, $\mathbf{C}_w = \sigma_w^2\mathbf{I}_d$) and for the threshold $T = 1$, we find $\mathbf{\Lambda}_{ii} = \sigma_x^2/\sigma_w^2$ such that:

$$\mathrm{P}(S=0) = \int_0^{+\infty} f_Z(z)\left(\frac{\sigma_w^2}{z\sigma_x^2 + \sigma_w^2}\right)^{d/2} dz,$$

which can be seen as a weighted average of the ratios of the volumes of the hyperspheres with radiuses σ_w and $\sqrt{z\sigma_x^2 + \sigma_w^2}$. It is interesting to note that the weighting function $f_Z(z)$, which is the density of a Gamma distribution, relates the probability $\mathrm{P}(S=0)$ to the frequency of occurrence [25] of image features in the considered subband: if this frequency is low, the signal of interest will be absent and $\mathrm{P}(S=0)$ will be high (and vice versa).

In Fig. 6, the conditional densities $f_{\mathbf{X}|S}(\mathbf{x}|0)$ and

$f_{\mathbf{X}|S}(\mathbf{x}|1)$ are shown for a two-dimensional random vector $\mathbf{x}$, corrupted with positively (in 2-D) correlated Gaussian noise. In this case, the positive correlation between the noise components and negative correlation between the noise-free signal components cause a diagonal cut in $f_{\mathbf{X}|S}(\mathbf{x}|1)$. When there are no correlations between the noise components and between the noise-free signal components, this cut is ring-shaped. The contours of the resulting shrinkage function are generally not elliptical anymore (Fig. 6c and Fig. 6d), although effects of the elliptical contours of the noise probability density function can still be observed in Fig. 6d.

Finally, we remark that equation (13) can also be interpreted as an approximation of Bayes Least Square (BLS, or MMSE) estimator for the model in Section III with $\mathrm{E}_{\mathbf{X}|\mathbf{Y},S}\left(\mathbf{x}_j|\mathbf{y}_j,1\right) \approx \mathbf{y}_j$ and $\mathrm{E}_{\mathbf{X}|\mathbf{Y},S}\left(\mathbf{x}_j|\mathbf{y}_j,0\right) \approx \mathbf{0}$:

$$\begin{aligned}\hat{\mathbf{x}}_j &= \mathrm{E}_{\mathbf{X}|\mathbf{Y}}\left(\mathbf{x}_j|\mathbf{y}_j\right) \\ &= \mathrm{P}_{S|\mathbf{Y}}(1|\mathbf{y}_j)\mathrm{E}_{\mathbf{X}|\mathbf{Y},S}\left(\mathbf{x}_j|\mathbf{y}_j,1\right) + \\ &\quad \mathrm{P}_{S|\mathbf{Y}}(0|\mathbf{y}_j)\mathrm{E}_{\mathbf{X}|\mathbf{Y},S}\left(\mathbf{x}_j|\mathbf{y}_j,0\right) \qquad (25)\end{aligned}$$

In case we are (almost) certain that a given wavelet coefficient vector is purely noise we select $\mathbf{0}$ as the estimate for the noise-free coefficient vector, hence $\mathrm{E}_{\mathbf{X}|\mathbf{Y},S}\left(\mathbf{x}_j|\mathbf{y}_j,\mathbf{0}\right) \approx \mathbf{0}$. On the other hand, using this approximation, significant structures like edges are preserved and no noise is suppressed: $\mathrm{E}_{\mathbf{X}|\mathbf{Y},S}\left(\mathbf{x}_j|\mathbf{y}_j,1\right) \approx \mathbf{y}_j$. This results in the shrinkage rule (13). In the first place, our intention in this paper is not to apply the MMSE estimator for the intrascale model directly (as this offers no notion of "signal presence"). Instead we will combine the shrinkage rule (13) with the HMT model in the next Section.

V. Hidden Markov Tree model for interscale dependencies

It is well known that the wavelet transform does not fully decorrelate the wavelet coefficients. By exploiting dependencies between these coefficients, improvements can be achieved, in denoising [2]–[4], [11] as well as in compression, e.g. the EZW coder of [32]. One way to deal with these dependencies, is to model the joint statistics of the wavelet coefficients in a given wavelet tree, together with their local (spatial) statistics. However, empirical joint histograms of multiscale data generally do not tend to have elliptical iso-probability contours, as assumed for GSM priors. Using high-dimensional probability densities results in larger covariance matrices (with size $d \times d$) and more parameters to estimate. We need to strike a balance between the number of model parameters (if too large, the estimation becomes unreliable) and the number of exploited dependencies. As a solution to this problem, the Hidden Markov Tree [11], [33] models Markovian dependencies between

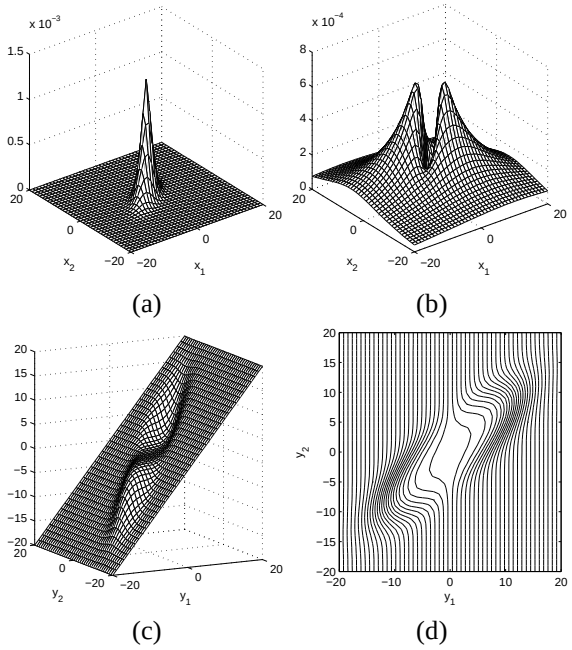


Figure 6. Illustration of the densities, modeling a wavelet coefficient x_1 and its right neighbour x_2. (a) Conditional density $f_{\mathbf{X}|S}(\mathbf{x}|0)$ (b) Conditional density $f_{\mathbf{X}|S}(\mathbf{x}|1)$ (c) The shrinkage function $\mathrm{P}(S = 1|\mathbf{y})y_1$ (d) Isocontours of (c)

hidden state variables of wavelet coefficients that are on different scales, thereby reducing the number of parameters. In the HMT model of Crouse et al. [11], later extended by Romberg et al. [13], the marginal densities of the noise-free wavelet coefficients are modeled as a mixture of two Gaussians. The number of mixture components is directly related to the number of states: one mixture component corresponds to each state. To describe the leptokurtotic behaviour of the noise-free wavelet coefficients more accurately, a larger number of Gaussian mixture components (e.g. 8) may be necessary. This would increase the number of model parameters and subsequently the computational complexity. Another problem that arises is: how to estimate the discrete density of the hidden multiplier z (i.e. the weights and variances of the mixture components). As a general solution, the EM algorithm [34] applies, with the disadvantage that the algorithm may converge to a non-global maximum of the likelihood function instead of a global maximum. In [35], nonparametric HMT models, connecting discrete GSM distributions across states, are trained using a Monte Carlo learning algorithm. The number of states is also learned from the training images. The BLS estimator from [4] can then be used to denoise every wavelet subband. Markov Chain Monte Carlo methods can be designed to escape from local maxima and saddle points of the likelihood function (see e.g. [36]). However the computational cost is often

significant, which makes these methods less practical.

In our approach, the significance measures $S(\mathbf{x})$ are used as hidden nodes for the HMT model, following Malfait's idea from [15] for spatial Markov Random Fields and further elaborated in [17]. Because this only requires two states, independent of the number of Gaussian mixture components, this reduces the computational complexity while the prior distribution is still highly kurtotic. We use independent HMT models for the different *orientations* of the DT-CWT. The HMT structure for the methods of [11], [13] is depicted in Fig. 7a and for our method in Fig. 7b. In the following, we will denote the scale of the wavelet transform by the subscript $k = 1, ..., K$, where $k = K$ represents the finest scale. The number of coefficients on scale k is given by $N_k = L\ 2^{k-K-1}$, where L is the number of pixels in the original image. $\mathbf{x}_j^{(k)}, j = 1, ..., N_k$ represent the noise-free coefficients of a local window at position j and scale k. $\mathbf{x}_j^{(k)}$ are observations of the random vector $\mathbf{x}^{(k)}$. Our HMT model is characterized by:

1) Two possible states for each scale k: $S^{(k)} \in \{0, 1\}$.

2) Two continuous observation densities on each scale: $f(\mathbf{x}^{(k)}|S^{(k)} = 0)$ and $f(\mathbf{x}^{(k)}|S^{(k)} = 1)$ (Section IV). This results in the overall pdf:

$$f(\mathbf{x}^{(k)}) = \mathrm{P}(S^{(k)}=0) f(\mathbf{x}^{(k)}|S^{(k)} = 0) + \mathrm{P}(S^{(k)}=1) f(\mathbf{x}^{(k)}|S^{(k)} = 1)$$

The observation densities are assumed independent for each scale.

3) The state transition probability distributions for modeling state transitions between different scales $\epsilon^{(k)} = \{\epsilon_{m,n}^{(k)}\}$:

$$\epsilon_{m,n}^{(k)} = \mathrm{P}(S^{(k+1)} = n|S^{(k)} = m), \quad m = 0, 1, \quad n = 0, 1 \tag{26}$$

4) The state distribution $\alpha^{(k)} = \{\alpha_n^{(k)}\}$ for scale k, where

$$\alpha_n^{(k)} = \mathrm{P}(S^{(k)} = n). \tag{27}$$

The parameters of the complete HMT model can be grouped in a random vector:

$$\mathbf{\Theta} = \{\tau^{(k)}, \mathbf{C}_x^{(k)}, \alpha^{(k)}, \epsilon^{(k)}\}.$$

$\tau^{(k)}$ and $\mathbf{C}_x^{(k)}$ are estimated once, independently for each scale (see Section III-C). $\alpha^{(k)}, \epsilon^{(k)}$, $k = 2, ..., K$ are estimated iteratively using the Baum-Welch algorithm (also known as the Expectation Maximization (EM) algorithm for HMM's) [11], [33]. Finally, denoising using (13) is quite simple and fast, since the hidden state probabilities $\mathrm{P}(S^{(k)} = 1|\mathbf{y}_j^{(1)}, ..., \mathbf{y}_j^{(k)})$ are already calculated during the *upward-downward* steps of the Baum-Welch algorithm (see [11]):

$$\hat{\mathbf{x}}_j^{(k)} = \mathrm{P}_{S^{(k)}|\mathbf{Y}^{(1)},...,\mathbf{Y}^{(k)}}\left(1|\mathbf{y}_j^{(1)}, ..., \mathbf{y}_j^{(k)}\right) \mathbf{y}_j^{(k)} \tag{28}$$

Alternatively, it is also possible to use the *exact* BLS estimate (25) for this HMT model:

$$\begin{aligned}\hat{\mathbf{x}}_j^{(k)} = &\mathrm{P}_{S^{(k)}|\mathbf{Y}^{(1)},...,\mathbf{Y}^{(k)}}\left(1|\mathbf{y}_j^{(1)}, ..., \mathbf{y}_j^{(k)}\right) \\ &\mathrm{E}_{\mathbf{X}^{(k)}|\mathbf{Y}^{(k)},S^{(k)}}\left(\mathbf{x}_j^{(k)}|\mathbf{y}_j^{(k)}, 1\right) + \\ &\mathrm{P}_{S^{(k)}|\mathbf{Y}^{(1)},...,\mathbf{Y}^{(k)}}\left(0|\mathbf{y}_j^{(1)}, ..., \mathbf{y}_j^{(k)}\right) \\ &\mathrm{E}_{\mathbf{X}^{(k)}|\mathbf{Y}^{(k)},S^{(k)}}\left(\mathbf{x}_j^{(k)}|\mathbf{y}_j^{(k)}, 0\right)\end{aligned} \tag{29}$$

where the conditional expectations can be computed similar as in [4], but based on different conditional prior densities (equation (19)).

The combination of the spatial GSM model and the HMT tree model allows us to capture both spatial and interscale dependencies between wavelet coefficients, which usually improves the denoising performance. For instance, this has been reported for the local contextual HMM in [14], where the LSAI is used to summarize the local context around a given wavelet coefficient. However, because the LSAI cannot make a distinction between signal and noise when noise coefficients are clustered, the method from [14] does not deal correctly with the case of correlated noise.

A. HMT training initialization

The Baum-Welch algorithm requires initialization of the state transition probability matrices $\epsilon_{m,n}^{(k)}$ and the distribution $\alpha^{(1)}$. An intelligent initialization may provide fast convergence of the HMT model training [11], [37]. Our approach consists of detecting the presence of a signal of interest for each position j. If we use a zero cost for the correct decision and equal costs for wrong decisions, then we can apply a MAP decision for this problem [38]:

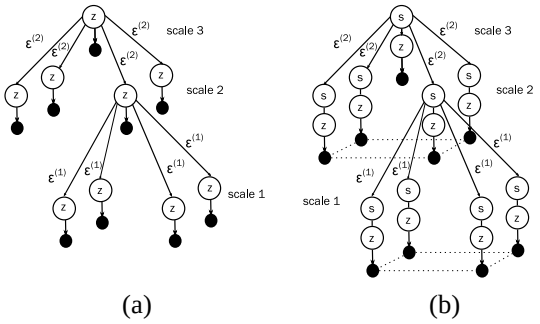
$$\hat{S}_j^{(k)} = \begin{cases} 0 & \lambda(\mathbf{y}_j)\frac{\mathrm{P}(S^{(k)}=1)}{\mathrm{P}(S^{(k)}=0)} < 1 \\ 1 & \lambda(\mathbf{y}_j)\frac{\mathrm{P}(S^{(k)}=1)}{\mathrm{P}(S^{(k)}=0)} \geq 1 \end{cases} \tag{30}$$

where $\lambda(\mathbf{y}_j) = f_{\mathbf{Y}|S}(\mathbf{y}_j|S^{(k)}=1)/f_{\mathbf{Y}|S}(\mathbf{y}_j|S^{(k)}=0)$ is the *likelihood ratio* and $\mathrm{P}(S^{(k)}=1)/\mathrm{P}(S^{(k)}=0)$ is the *prior ratio*, calculated using (24). Using the law of total probabilities $\lambda(\mathbf{y}_j)$ can be written as:

$$\lambda(\mathbf{y}_j) = \frac{f_{\mathbf{Y}}(\mathbf{y}_j) - f_{\mathbf{Y}|S}(\mathbf{y}_j|S^{(k)}=0)\mathrm{P}(S^{(k)}=0)}{f_{\mathbf{Y}|S}(\mathbf{y}_j|S^{(k)}=0)\mathrm{P}(S^{(k)}=1)}$$

with $f_{\mathbf{Y}|S}(\mathbf{y}_j|S^{(k)}=0)$ obtained using equation (16). By counting the number of transitions when passing from scale k to scale $k+1$, we obtain a first reliable estimate of the state transition probability matrix:

$$\epsilon_{m,n}^{(k)} = \frac{\sum_{j=0}^{N_{k+1}} \#\{S_j^{(k+1)} = n \wedge S_{j'}^{(k)} = m\}}{\sum_{j=0}^{N_k} \#\{S_j^{(k)} = m\}} \tag{31}$$



(a) (b)

Figure 7. (a) Hidden Markov Tree structure used in [11], [13] (b) The Hidden Markov Tree structure proposed in this paper. black nodes are wavelet coefficients, z-nodes and s-nodes represent respectively the hidden multiplier (local variance) and the significance associated with the wavelet coefficients. Dotted lines represent spatial correlations, modeled using the independent overlapping window assumption.

with j' the parent index of the coefficient at position j and "#" denotes the cardinality. The state propabilities $\alpha^{(1)}$ of the coarsest scale are estimated using (24). The outline of our algorithm is given in Algorithm 1.

Algorithm 1 Algorithm outline

1: Decompose the image into bands using the DT-CWT
2: **for all** orientations **do**
3: **for all** scales {except the lowpass scale} **do**
4: Estimate the local model parameters $\tau, \mathbf{C}_x, \mathbf{C}_w$ as explained in Section III-C.
5: Estimate the initial state probabilities $\mathrm{P}(S=0)$, $\mathrm{P}(S=1)$ using (24).
6: Estimate the initial state transition probabilities using (31).
7: **end for**
8: **repeat**
9: E-step (*upward-downward algorithm*): estimate probabilities for the hidden state variables.
- Upward step: propagation *upward* the tree
- Downward step: propagation *downward* the tree
10: M-step: update $\alpha^{(k)}, \epsilon^{(k)}$, to maximize the expected likelihood function
11: **until** convergence
12: apply equation (28) or (29) to every wavelet coefficient vector
13: **end for**
14: Reconstruct the image using the scaling coefficients and the modified wavelet coefficients.

VI. Implementation aspects

The threshold T is selected once for all images by minimizing the MSE objective function (or Bayesian risk) defined by $\mathrm{MSE}(T) = \mathrm{E}\left((\hat{\mathbf{x}} - \mathbf{x})^2\right)$, similar to [40]. When the dimension d is low ($d < 3$), the solution can be found using numerical techniques. However, for 3×3 spatial windows (or $d = 9$), Monte-Carlo simulations are needed. To achieve this, we artificially generate a sufficiently large number (4000) of noisy wavelet subbands of size 256×256 according to the BKF prior model with identity covariance matrix and noise covariance matrix given by $\sigma_k^2 \mathbf{I}$, with $\sigma_k = 0.1 + 0.2k, k = 0, ..., 39$ (to cover the range of SNR levels where the algorithm will be used for), with k depending on the subband number. The threshold T is found minimizing the Bayesian risk numerically, this is repeated for all generated subbands. By using the golden section search optimization technique, we obtained as mean $T = 2.4$, with variance given by 0.03. Our experiments show that when small changes to the threshold are made (e.g. within $10\% - 20\%$), the PSNR and visual performance is nearly not affected. The resulting PSNR curves as function of the threshold T are similar to the ones reported in [5] and are therefore omitted here.

Formulas (20) and (24) involve integration over an infinite interval. Instead of approximating $f_Z(z)$ with a discrete density and estimating the discrete points z_p from the data, we evaluate these integrals numerically using the extended trapezoidal rule, by selecting an upper bound for the integration. This is possible since $f_Z(z)$ decays exponentially. Using the exponential sampling $z_p = \exp(-3 + 7p), p = 1, ..., P$, we achieve a good numerical accuracy even with a limited number of samples (typically $P = 4$).

To avoid numerical underflow in the *upward-downward* algorithm, used for likelihood computation in the HMT model, a *scaling procedure* has been proposed [33]. When dealing with high-dimensional local priors (e.g. $d \geq 10$) underflow occurs in the the likelihood computations $f_{\mathbf{Y}|Z,S}(\mathbf{y}_j|z, 1)$ and $f_{\mathbf{Y}|Z,S}(\mathbf{y}_j|z, 0)$, and the HMT scaling procedure fails, because of zero input probabilities. By evaluating $\log f_{\mathbf{Y}|Z}(\mathbf{y}_j|z)$ rather than $f_{\mathbf{Y}|Z,S}(\mathbf{y}_j|z)$ and by adding an extra scaling factor e^{4d} to the Gaussian densities, we can avoid this problem. The constant $4d$ is chosen experimentally such that the numerical values stay within reasonable bounds. The evaluated (denormalized) densities will be renormalized automatically in the subsequent scaling procedure, during the upward/downward steps (see also [33]).

VII. Results and discussion

A. Experimental results for images with white noise

The results for this paper are produced using the Dual Tree Complex Wavelet transform of [20] with 10-tap Q-shift filters. We use overlapping 3×3 spatial windows (as in [4]) in order to keep the computional overhead low. Only $P = 4$ sampling points are used (Section VI). The reported results for white noise are

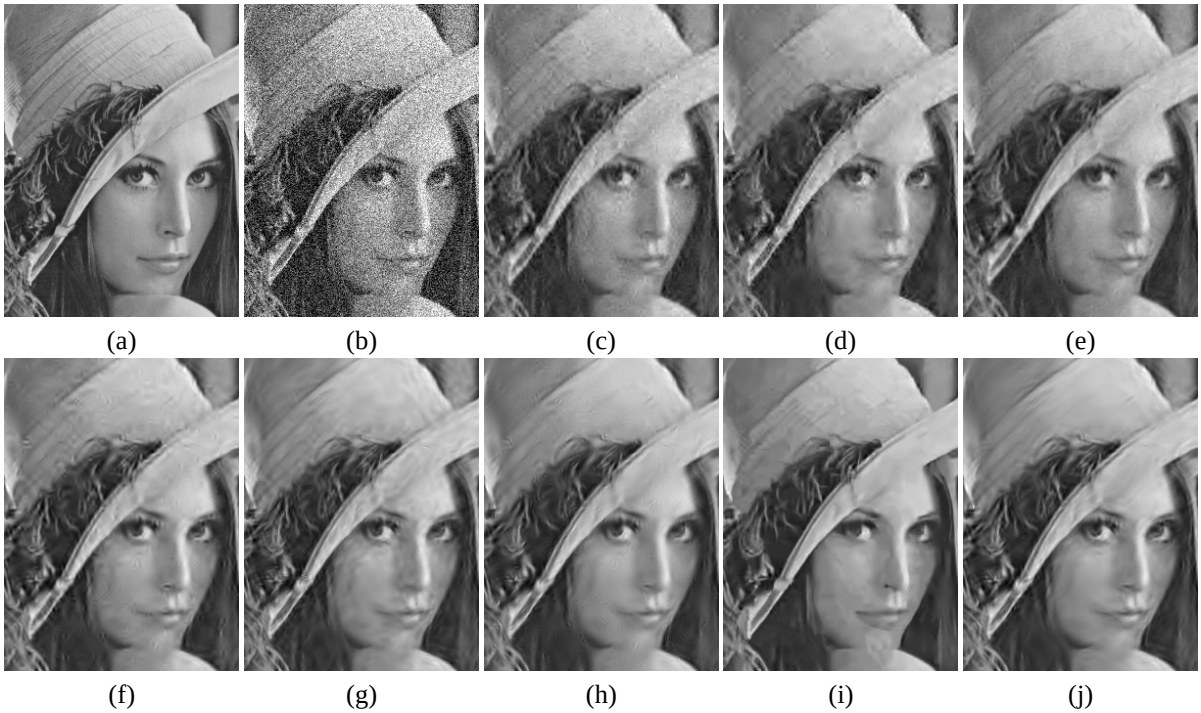


Figure 8. Denoising results for lena corrupted with stationary *white* Gaussian noise ($\sigma = 35$) [cropped out]. (a) Original image, (b) noisy image, (c) Crouse et al. (1998), using orthogonal wavelet transform $\mathrm{PSNR} = 28.30\mathrm{dB}$ [11], (d) Luisier et al. (2007), using orthogonal wavelet transform $\mathrm{PSNR} = 28.81\mathrm{dB}$ [7], (e) Romberg et al. (2000), using DT-CWT $\mathrm{PSNR} = 29.20\mathrm{dB}$ [12], (f) Pižurica et al. (2006), using undecimated wavelet transform $\mathrm{PSNR} = 29.33\mathrm{dB}$ [5], (g) Şendur et al. (2002), using DT-CWT $\mathrm{PSNR} = 29.86\mathrm{dB}$ [3], (h) Portilla et al. (2003), using full steerable pyramids $\mathrm{PSNR} = 30.07\mathrm{dB}$ [4], (i) Dabov et al. (2007), $\mathrm{PSNR} = 30.56\mathrm{dB}$ [39], (j) Proposed, using DT-CWT $\mathrm{PSNR} = 30.20\mathrm{dB}$.

obtained using the estimator in equation (28). The HMT training step is performed on the noisy image and takes in most cases less than 4 iterations before convergence. We found that for images with a high amount of edge and texture information, like *Barbara*, this number is usually higher (up to 10-20). The overall impact of this large number of iterations is limited, because the evaluation of the conditional densities $f(\mathbf{y}_j^{(k)}|S^{(k)}=0)$ and $f(\mathbf{y}_j^{(k)}|S^{(k)}=1)$ takes most part of the computation time, but fortunately this has to be done only once per subband. The noise covariance matrix is assumed known to all the algorithms. A visual result for *Lena* is given in Fig. 8. PSNR results are given in Table I and Fig. 10.

Fig. 9 shows the improvement in PSNR performance obtained by using the Hidden Markov tree model from Section V combined with the spatial model from Section IV, compared to only using the proposed spatial model. The greatest improvement is obtained for the Lena image (see Fig. 8), where sharp features (e.g. in the hat) are preserved very well after denoising. For the Barbara image, there is a small loss ($< 0.15\mathrm{dB}$) for input PSNR around $20\mathrm{dB}$. Here, the model assumptions made the convergence of the EM-algorithm more difficult (in terms of the number of EM-iterations), and the algorithm converges more likely to a local optimum instead of a global optimum. Globally, our combined inter/intrascale method performs equally well as the BLS-GSM method of Portilla (see Table I), but at a lower computational cost, since the redundancy factor of the DT-CWT is 4, while the full steerable pyramid transform with 8 orientations has redundancy factor $56/3 \approx 18.67$. Also, the DT-CWT is generally faster to compute. In Table II, the computation times of the BLS-GSM method and the proposed method are compared. Both methods were implemented in C++ with the same level of optimizations and were run on a Pentium IV 2.4GHz processor. It can be seen that the decrease in redundancy gives a speedup factor of 3-4. In Table III, the proposed method is compared to recent *nonlocal* denoising techniques from [41] and [39]. These methods take advantage of the repetitivity that is present in natural images (which is currently not exploited in our method) and often yield excellent results for images with many repetitive structures. However, at the time of writing, none of these non-local methods can efficiently remove strongly correlated noise from images. Other recent nonlocal techniques that are not included in the comparison are e.g. [42], [43].

Table II
COMPARISON OF THE EXECUTION TIMES OF THE BLS-GSM METHOD AND THE PROPOSED METHOD. TO ALLOW FOR A FAIR COMPARISON, BOTH METHODS ARE IMPLEMENTED IN C++ WITH THE SAME LEVEL OF OPTIMIZATION. REPORTED VALUES ARE THE EXECUTION TIMES AVERAGED OVER 10 RUNS AND THEIR STANDARD DEVIATIONS (BETWEEN PARENTHESES)

	Input image size			
Method	256×256		512×512	
BLS-GSM	6.41s	(0.03s)	25.89s	(0.05s)
Proposed	2.02s	(0.01s)	6.99s	(0.07s)

B. Experimental results for images with correlated noise

In Fig. 11, visual results are given for colour images corrupted with artificial correlated noise. The noise was added independently to the three RGB-colour channels. The algorithm was applied in the YCbCr-colour space to each colour channel individually. Fig. 11 shows the visual performance of the proposed method in comparison to BLS-GSM of [4]. Our results in this figure are slightly better in terms of PSNR and visually, even though we used a much less redundant representation (with redundancy 4 compared to 18.67 of the reference method). In Fig. 12, the algorithm was applied to images captured with a digital video camera, using a low exposure time. The images were processed in the RGB-colour space and the noise covariance matrix was estimated from a flat region in the image with only noise, for each colour channel. In Fig. 12 (right), the difference image between the noisy and denoised image is shown (gray corresponds with difference 0). Experimentally, we found that the estimator in equation (29) offers for white noise slightly worse results than the estimator in (28) (around $0.3\mathrm{dB}$),[3] but for correlated noise with a highly anisotropic character (for example Fig. 11b) the estimator in equation (29) usually gives improvements both visually and in PSNR. In this case, the approximation $\mathrm{E}_{\mathbf{X}|\mathbf{Y},S}\left(\mathbf{x}_j|\mathbf{y}_j,1\right) \approx \mathbf{y}_j$ is not accurate.

VIII. CONCLUSION

A new method for the removal of correlated noise has been presented. An intrascale model, based on the Bessel K Form density, is combined with a Hidden Markov Tree interscale model by modeling the signal presence in a given observed random vector. The signal presence is characterized by a significant measure that quantifies the relevant information in a noisy image and that takes the correlation structure of neighbouring wavelet coefficients into account. When used in combination with the Dual-Tree Complex wavelet transform, we obtain a lower computational cost and memory requirements, while being competitive with recent state-of-the-art wavelet-based denoising methods.

[3]Note that the MMSE estimator in a redundant representation does not necessarily minimize the MSE in the image domain (see e.g. the work of Luisier et al. [7])

Table III
COMPARISON WITH RECENT *nonlocal* METHODS FOR WHITE NOISE: K-SVD-GLOBAL WITH GLOBAL TRAINED DICTIONARY (ELAD ET AL.) [41], K-SVD-ADAPTIVE WITH ADAPTIVE TRAINED DICTIONARY (ELAD ET AL.) [41], BM-3D (DABOV ET AL.) [39], REPORTED ARE PSNR VALUES (AVERAGED OVER 50 RUNS FOR THE PROPSED METHOD AND BM-3D, AND OVER 5 RUNS FOR K-SVD) AND PSNR STANDARD DEVIATIONS (BETWEEN PARENTHESES).

	Standard deviation of the white noise						
	10	15	20	25	35	50	100
	LENA						
Proposed	35.48	33.81	32.59	31.64	30.17	28.60	**25.63**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.05)
K-SVD-global	35.43	33.59	32.28	31.16	29.57	27.78	24.44
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.07)	(0.08)
K-SVD-adaptive	35.50	33.70	32.40	31.28	29.66	27.82	24.44
	(0.01)	(0.02)	(0.04)	(0.02)	(0.04)	(0.07)	(0.06)
BM3D	**35.89**	**34.23**	**33.01**	**32.04**	**30.52**	**28.79**	25.49
	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)	(0.05)	(0.06)
	BARBARA						
Proposed	34.11	31.84	30.23	28.97	27.12	25.29	22.65
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.10)
K-SVD-global	33.36	30.81	29.03	27.71	25.91	24.28	22.05
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)
K-SVD-adaptive	34.83	32.69	31.11	29.82	27.76	25.41	22.19
	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.07)	(0.04)
BM3D	**35.38**	**33.45**	**32.05**	**30.93**	**29.13**	**27.25**	**23.53**
	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.05)	(0.07)

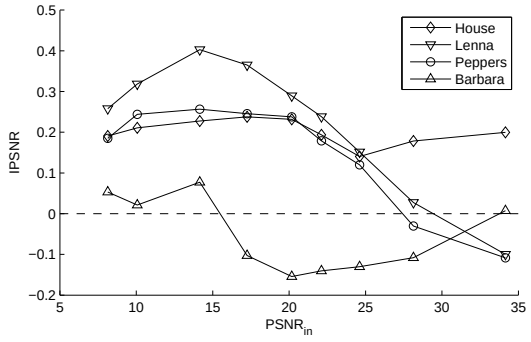


Figure 9. Increase in PSNR (IPSNR) obtained by using the HMT model from Section V (using equation (28)) upon the spatial estimator introduced in Section IV, for different input PSNR levels (PSNR_{in}). Results are averaged over 50 runs.

IX. APPENDIX

First we show that coordinate transform $\mathbf{UQ}$ in (21) also diagonalizes the covariance matrix of the conditional density $f_{\mathbf{Y}|Z,S}(\mathbf{y}|z,0)$ for every z, i.e., $\left((z\mathbf{C}_x)^{-1} + (T^2\mathbf{C}_w)^{-1}\right)^{-1} + \mathbf{C}_w = \mathbf{UQ}((z^{-1}\mathbf{\Lambda}^{-1} + T^{-2}\mathbf{I}_d)^{-1} + \mathbf{I}_d)\mathbf{Q}^T\mathbf{U}^T$:

$$\begin{aligned} & z\mathbf{C}_x^{-1} + (T^2\mathbf{C}_w)^{-1} \\ & = (z\mathbf{C}_x)^{-1} + (T^2\mathbf{U})^{-T}\mathbf{U}^{-1} \\ & = \mathbf{U}^{-T}(z^{-1}\mathbf{U}^T\mathbf{C}_x^{-1}\mathbf{U}^T + T^{-2}\mathbf{I}_d)\mathbf{U}^{-1} \\ & = \mathbf{U}^{-T}\mathbf{Q}^{-T}(z^{-1}\mathbf{\Lambda}^{-1} + T^{-2}\mathbf{I}_d)\mathbf{Q}^{-1}\mathbf{U}^{-1} \end{aligned} \qquad (32)$$

Table I

COMPARISON WITH OTHER METHODS FOR WHITE NOISE, THAT USE DIFFERENT MULTIRESOLUTION REPRESENTATIONS: BLS-GSM (PORTILLA ET AL.) [4] (FULL STEERABLE PYRAMIDS), BISHRINK (ŞENDUR ET AL.) [3] (DT-CWT), PROBSHRINK (PIŽURICA ET AL.) [5] (UNDECIMATED DWT), MBKF-CURVELET (BOUBCHIR) [31] (CURVELETS) [A], CWT-HMT (ROMBERG ET AL.) [12] (DT-CWT), LCHMM-SI (FAN ET AL.) [14] (UNDECIMATED DWT) [B] REPORTED ARE PSNR VALUES AVERAGED OVER 50 RUNS AND PSNR STANDARD DEVIATIONS (BETWEEN PARENTHESES).

	Standard deviation of the white noise							Standard deviation of the white noise						
	10	15	20	25	35	50	100	10	15	20	25	35	50	100
	LENA							HOUSE						
Proposed	35.48	33.81	**32.59**	**31.64**	**30.17**	**28.60**	**25.63**	35.06	33.32	32.09	31.11	29.56	27.96	**24.87**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.06)	(0.06)	(0.07)	(0.09)	(0.12)
BLS-GSM	**35.59**	**33.85**	32.57	31.58	30.05	28.45	25.49	**35.37**	**33.59**	**32.27**	**31.22**	**29.66**	**28.01**	24.83
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.06)	(0.04)	(0.04)	(0.05)	(0.06)	(0.07)	(0.08)	(0.13)
BiShrink	35.29	33.58	32.32	31.35	29.84	28.22	25.16	34.78	33.01	31.74	30.74	29.20	27.60	24.49
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.07)	(0.03)	(0.05)	(0.06)	(0.06)	(0.08)	(0.09)	(0.13)
ProbShrink	35.06	33.23	31.90	30.87	29.33	27.70	24.81	34.61	32.69	31.27	30.18	28.56	26.96	23.99
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.06)	(0.03)	(0.05)	(0.05)	(0.05)	(0.07)	(0.08)	(0.12)
MBKF-Curvelet	35.10	33.28	31.96	30.94	29.27	-	-	34.94	32.61	30.45	28.73	25.72	-	-
	(0.01)	(0.02)	(0.03)	(0.02)	(0.21)			(0.12)	(0.19)	(0.40)	(0.44)	(0.50)		
CWT-HMT	34.91	32.98	31.67	30.72	29.23	27.71	25.01	34.52	32.38	31.06	30.09	28.54	26.94	23.95
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)	(0.07)	(0.09)
LCHMM-SI	35.00	32.50	31.20	30.10	-	-	-	-	-	-	-	-	-	
	BARBARA							MAN						
Proposed	34.11	31.84	30.23	28.97	27.12	25.29	22.65	33.56	31.48	30.09	29.07	27.65	26.25	**23.89**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.10)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)	(0.05)
BLS-GSM	**34.51**	**32.21**	30.56	29.30	27.44	**25.58**	**22.82**	**33.63**	**31.54**	**30.15**	**29.13**	**27.68**	**26.28**	23.81
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.05)
BiShrink	33.51	31.28	29.76	28.64	27.00	25.35	22.71	33.27	31.30	29.97	28.98	27.57	26.16	23.61
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.06)
ProbShrink	33.83	31.46	29.77	28.46	26.47	24.58	22.23	33.26	31.13	29.73	28.73	27.32	25.94	23.53
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)
MBKF-Curvelet	34.33	32.20	**30.72**	**29.59**	**27.72**	-	-	32.89	30.80	29.43	28.44	27.09	-	-
	(0.02)	(0.02)	(0.03)	(0.05)	(0.18)			(0.01)	(0.02)	(0.02)	(0.03)	(0.02)		
CWT-HMT	33.36	31.09	29.54	28.22	26.06	24.32	22.43	33.20	30.96	29.59	28.67	27.31	25.98	23.69
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
LCHMM-SI	33.10	30.80	29.20	28.00	-	-	-	-	-	-	-	-	-	-

[A] This is our implementation of the method presented in Chapter 6 of [31], using the EM-estimation of the hyperparameters of the MBKF prior as decribed in [31]. The wrapping-based implementation of the curvelet transform from [44] was used.
[B] Because an implementation is currently not publicly available, results for Barbara and Lena are copied from [14].

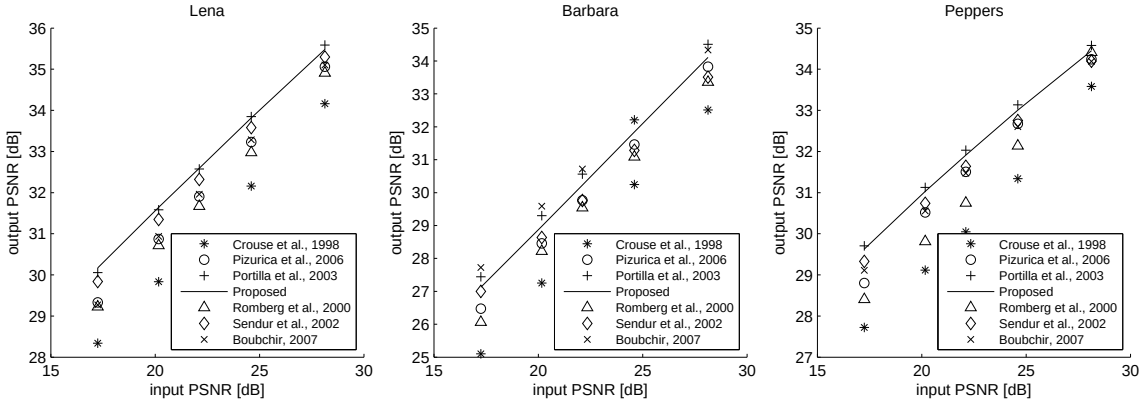


Figure 10. Results of several recent methods that use different multiresolution representations: Crouse et al. [11] (orthogonal DWT), Romberg et al. [12] (DT-CWT), Şendur et al. [3] (DT-CWT), Portilla et al. [4] (full steerable pyramids), Pižurica et al. [5] (undecimated DWT), Boubchir [31] (curvelets), the proposed method (DT-CWT)

Figure 11. Denoising results for images with artificial correlated Gaussian noise. (a) Crop-out of the Barbara image (in colour) (b) Image with artificial noise $\Psi(k,l) \sim 1 - \exp(-0.1(k^2+l^2)) + 300\exp(-4000((k-0.1)^2+(l-0.12)^2))$, uncorrelated in the RGB-colour space, $\mathrm{PSNR}_{in} = 18.59\mathrm{dB}$. (c) BLS-GSM, in the YCbCr-colour space, $\mathrm{PSNR}_{out} = 31.03\mathrm{dB}$. (d) The proposed technique (using equation (29)), in the YCbCr-colour space, $\mathrm{PSNR}_{out} = 31.56\mathrm{dB}$. (e) Crop-out of the boats image (in colour) (f) Image with artificial noise $\Psi(k,l) \sim \mathrm{I}((v^2+(0.1u+7v)^2) < 0.1)$, uncorrelated in the RGB-colour space, $\mathrm{PSNR}_{in} = 18.59\mathrm{dB}$ (g) BLS-GSM, in the YCbCr-colour space, $\mathrm{PSNR}_{out} = 27.22\mathrm{dB}$. (h) The proposed technique (using equation (29)), in the YCbCr-colour space, $\mathrm{PSNR}_{out} = 27.37\mathrm{dB}$.

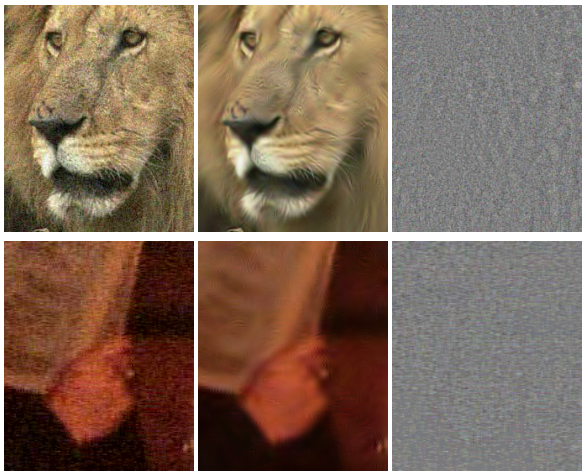


Figure 12. Denoising results for colour images captured with a digital camera, with the noise covariance estimated in a flat region. *(Left)* Noisy image, *(Middle)* Denoised image, using the proposed method (equation (28)), *(Right)* Difference image

where we used $\mathbf{U}\mathbf{U}^T = \mathbf{C}_w$, the SVD $\mathbf{U}^{-1}\mathbf{C}_x\mathbf{U}^{-T} = \mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}$ and $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_d$. This gives:

$$\begin{aligned}
&\left((z\mathbf{C}_x)^{-1} + (T^2\mathbf{C}_w)^{-1}\right)^{-1} + \mathbf{C}_w \\
&= \mathbf{U}\mathbf{Q}(z^{-1}\mathbf{\Lambda}^{-1} + T^{-2}\mathbf{I}_d)^{-1}\mathbf{Q}^T\mathbf{U}^T + \mathbf{U}\mathbf{U}^T \\
&= \mathbf{U}\left(\mathbf{Q}(z^{-1}\mathbf{\Lambda}^{-1} + T^{-2}\mathbf{I}_d)^{-1}\mathbf{Q}^T + \mathbf{I}_d\right)\mathbf{U}^T \\
&= \mathbf{U}\mathbf{Q}((z^{-1}\mathbf{\Lambda}^{-1} + T^{-2}\mathbf{I}_d)^{-1} + \mathbf{I}_d)\mathbf{Q}^T\mathbf{U}^T
\end{aligned}$$

Next we derive an expression for the probability $\mathrm{P}(S=0)$. Using (18), we find:

$$\begin{aligned}
&f_{\mathbf{X}|Z,S}(\mathbf{x}|z,0) \\
&= \frac{f_{\mathbf{X}|Z}(\mathbf{x}|z)}{\mathrm{P}(S=0|z)} \exp\left(-\frac{1}{2}\mathbf{x}^T(T^2\mathbf{C}_w)^{-1}\mathbf{x}\right) \\
&= \frac{|z\mathbf{C}_x|^{-\frac{1}{2}}(2\pi)^{-\frac{d}{2}}}{\mathrm{P}(S=0|z)} \exp\left(\frac{\mathbf{x}^T\left((z\mathbf{C}_x)^{-1}+(T^2\mathbf{C}_w)^{-1}\right)\mathbf{x}}{-2}\right)
\end{aligned} \tag{33}$$

Identification of (33) and (19) leads to:

$$\mathrm{P}(S=0|z) = \frac{\left|(z\mathbf{C}_x)^{-1}\right|^{\frac{1}{2}}}{\left|(z\mathbf{C}_x)^{-1} + (T^2\mathbf{C}_W)^{-1}\right|^{\frac{1}{2}}} \tag{34}$$

Using (32), this can be further simplified to:

$$\mathrm{P}(S=0|z) = \frac{\left|z^{-1}\mathbf{\Lambda}^{-1}\right|^{\frac{1}{2}}}{\left|z^{-1}\mathbf{\Lambda}^{-1} + T^{-2}\mathbf{I}_d\right|^{\frac{1}{2}}} = \left(\prod_{i=1}^{d}\frac{T^2}{T^2 + z\mathbf{\Lambda}_{ii}}\right)^{\frac{1}{2}}$$

Finally, integration over z gives:

$$\begin{aligned}
\mathrm{P}(S=0) &= \int_0^{+\infty} \mathrm{P}(S=0|z) f_Z(z) dz \\
&= \int_0^{+\infty} f_Z(z) \left(\prod_{i=1}^{d}\frac{T^2}{T^2 + z\mathbf{\Lambda}_{ii}}\right)^{\frac{1}{2}} dz
\end{aligned}$$

REFERENCES

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

[2] S. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, pp. 1522–1531, 2000.

[3] L. Şendur and I.W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, vol. 9, pp. 438–441, 2002.

[4] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using Gaussian Scale Mixtures in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, pp. 1338–1351, 2003.

[5] A. Pižurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising.," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 654–665, Mar 2006.

[6] Fei Shi and I. W. Selesnick, "Multivariate Quasi-Laplacian Mixture Models for Wavelet-based Image Denoising," in *Proc. Int. Conf. on Image Processing (ICIP)*, 2006, pp. 2097–2100.

[7] F. Luisier, T. Blu, and M. Unser, "A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Trans. Image Processing*, vol. 16, no. 3, pp. 593–606, Mar. 2007.

[8] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society B*, vol. 59, no. 2, pp. 319–351, 1997.

[9] M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukhah, and A. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 766–784, 1992.

[10] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. Image Processing*, vol. 5, no. 4, pp. 619–634, 1996.

[11] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE. Trans. Signal Processing*, vol. 46, pp. 886–902, 1998.

[12] H. Choi, J. Romberg, R. Baraniuk, and N. G. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *Proc. IEEE Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, 2000.

[13] J.K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree structured image modeling using wavelet-domain Hidden Markov Models," *IEEE Trans. Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[14] G. Fan and X. 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.

[15] 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, April 1997.

[16] M. Jansen and A. Bultheel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *J. of the Amer. Statist. Assoc. (JASA)*, vol. 96, no. 454, pp. 629–639, 2001.

[17] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.

[18] J. Portilla and E.P. Simoncelli, "Adaptive Wiener Denoising using a Gaussian Scale Mixture Model in the Wavelet Domain," *IEEE Int. Conf. on Image Process. (ICIP)*, vol. 2, pp. 37–40, 2001.

[19] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

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

[21] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Processing*, vol. 11, pp. 670–684, 2000.

[22] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[23] H. Baher, *Analog and Digital Signal Processing*, Wiley, Chichester, 2001.

[24] Samuel Kotz, Tomasz Kozubowski, and Krzystof Podgorski, *The Laplace Distributions And Generalizations: A Revisit with Applications to Communications, Economics, Engineering, Finance*, Birkhäuser, 2001.

[25] A. Srivastava, X. Liu, and U. Grenander, "Universal Analytical Forms for Modeling Image Probabilities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1200–1214, Sept. 2002.

[26] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities," *IEEE Trans. Image Processing*, vol. 14, no. 2, pp. 231–240, Feb. 2005.

[27] I. W. Selesnick, "Laplace Random Vectors, Gaussian Noise, and the Generalized Incomplete Gamma Function," in *Proc. Int. Conf. on Image Processing (ICIP)*, 2006, pp. 2097–2100.

[28] S. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 37, no. 12, pp. 2091–2110, Dec 1989.

[29] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Advances in Neural Information Processing Systems*. 2000, vol. 12, pp. 855–861, MIT Press.

[30] L. Boubchir, "Algorithme EM pour l'estimation des hyper-paramètres du debruiteur bayésien d'images basé sur l'a priori des Formes K de Bessel," in *Journées d'Etudes Algéro-Françaises en Imagerie Médicale (JETIM2006)*, Algeria, 2006, pp. 47–54.

[31] L. Boubchir, *Bayesian approaches for image denoising in oriented and non-oriented multiscale sparse transforms*, Ph.D. thesis, University of Caen, France, 2007.

[32] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[33] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[34] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[35] J. J. Kivinen, E. B. Sudderth, and M. I. Jordan, "Image denoising with nonparametric Hidden Markov Trees," in *IEEE Int. Conf. on Image Processing (ICIP)*, San Antonio, Texas, USA, sept 2007.

[36] G. Gamerman, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall/CRC, 1997.

[37] Guoliang Fan and Xiang-Gen Xia, "Improved Hidden Markov Models in the Wavelet-Domain," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 115–120, Jan. 2001.

[38] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Wiley, New York, 1968.

[39] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

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

[41] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries," *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3736–3745, dec 2006.

[42] A. Buades., B. Coll., and J.M Morel, "A non local algorithm for image denoising," in *Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 2, pp. 60–65.

[43] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Processing*, vol. 15, no. 10, pp. 2866–2878, 2006.

[44] E.J. Candes, L. Demanet, D.L. Donoho, and L. Ying, "Fast Discrete Curvelet Transforms," *Multiscale modeling and simulation*, vol. 5, no. 3, pp. 861–899, 2006.