We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

15

Wavelet-Based Analysis and Estimation of Colored Noise

Bart Goossens, Jan Aelterman, Hiêp Luong, Aleksandra Pižurica and Wilfried Philips
Ghent University - TELIN-IPI-IBBT
Belgium

1. Introduction

Digital imaging devices inevitably produce images corrupted with noise. The noise originates from the sensors and analogue circuitry in the camera. In order to have better and sharper images and also for commercial reasons, there is a recent tendency to further increase the image resolution. Nowadays, cameras with more than 20 megapixels are not uncommon. To reach such a high number of megapixels, the area of the sensor elements must be decreased and correspondingly the elements become more sensitive to noise, resulting in a lower image quality due to noise.

During the last decades, the use of image processing techniques has become widespread. The increasing processing power of computers allows for more sophisticated techniques that are better adapted to the classes of images under consideration (e.g. photographic images or medical images). This also allows for new classes of techniques that alleviate the physical limitations of the sensor elements by means of post-processing such as denoising. Because of power and hardware complexity constraints, the post-processing techniques implemented by camera manufacturers are based on simplistic assumptions with respect to the assumed noise model: for example, while it is well known that photon signals are *Poisson* distributed, the techniques most often rely on a white *Gaussian* noise model. In practice, such model mismatches generally lead to inferior denoising results. Also, many factors cause the noise in practice to be colored instead of white (i.e. with a flat power spectrum). For example, the image formation is often a reconstruction process based on an insufficient number of samples, and missing samples need to be estimated using interpolation techniques (e.g. Bayer pattern demosaicing). Doing so, the noise becomes colored. A technique that is designed to remove white Gaussian noise may offers a image quality: either some noise artifacts may be left in the image, or the noise is suppressed too much, leading to an overblurred image.

The obvious solution to this problem is to adapt existing techniques to use a colored noise model that is well matched to the underlying sensor characteristics and/or reconstruction. Therefore, estimation of the noise statistics is indispensable. Stationary colored noise (or correlated noise) is completely described by its Power Spectral Density (PSD). The noise PSD describes the power distribution of the noise in frequency space and can be estimated by using the Discrete Fourier Transform (DFT). However, noisy images also contain information other than noise (e.g. edges and textures), and directly estimating the PSD through the DFT

will yield seriously biased estimates caused by the signal presence. Alternatively, the PSD could be estimated from noise-only patches in the image. However, not all images contain such patches and also the number of noise samples that can be used for this task is often too limited to yield reliable PSD estimates. Hence, more specialized techniques are needed.
The discrete wavelet transform (DWT) is an important tool for developing such techniques. The DWT provides a non-uniform partitioning of the space-frequency plane, which allows positional information of structures to be included in the estimation. This is not possible with the DFT, since the DFT cannot recover information at specified positions in the image.
In this chapter, we investigate the estimation of colored noise. First, we discuss a number of origins for colored noise in images. Next, we explain the importance of wavelets in solving the estimation problem. To proceed, it is necessary to know how the wavelet-domain and spatial-domain autocorrelation functions are related to each other, since we are aiming at estimating the wavelet-domain autocorrelation function. Because the wavelet transform in general does not fully decorrelate signals as we will explain, noise-free wavelet coefficients with significant magnitudes can still be found near high-frequent transitions in the signals (for example, near edges in images). To benefit from prior knowledge in a statistical estimation approach, we will discuss a number of wavelet domain prior models. Two iterative EM-based techniques will be presented, to estimate the wavelet-domain autocorrelation function. Next, we will explain how the parameters of a parametric noise PSD can be estimated using the presented tools. Finally, we will give a number of experimental results for the proposed techniques.

1.1 From white noise to colored noise

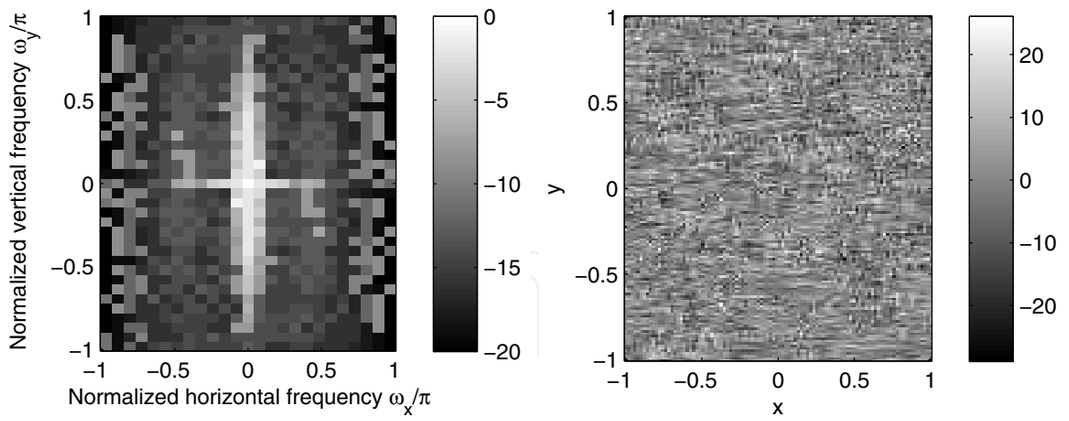
Throughout this chapter, we will consider a stationary additive Gaussian noise process:

$$y(\boldsymbol{p}) = x(\boldsymbol{p}) + w(\boldsymbol{p}) \tag{1}$$

where $x(\boldsymbol{p})$ is a pixel intensity of a noise-free image at position $\boldsymbol{p} \in \mathbb{Z}^2$, $y(\boldsymbol{p})$ is the corresponding observed pixel intensity and $w(\boldsymbol{p})$ is a zero-mean additive noise component. $w(\boldsymbol{p})$ and $x(\boldsymbol{p})$ are mutually statistically independent. We will further assume that the samples $w(\boldsymbol{p})$ are generated by a (wide-sense) spatial stationary process w, in which the correlation between two noise samples only depends on the position difference between the two noise samples, but not on their absolute position. Consequently, w can be completely described by the mean and the autocorrelation function.
A wide-sense stationary random process w obeying the above conditions is called *white* if its autocorrelation function is a Dirac delta function:

$$R_w(\boldsymbol{p}) = \mathrm{E}\left[w(\boldsymbol{p}')\overline{w(\boldsymbol{p}+\boldsymbol{p}')}\right] = \delta(\boldsymbol{p}). \tag{2}$$

For colored noise, neighboring noise samples are *not* statistically independent, hence spatial dependencies exist between these samples. Their dependencies can be characterized by the autocorrelation function of the noise, which is - for colored noise - different from the Dirac delta function.
The PSD is a related descriptor of colored noise. More specifically, the PSD describes how the noise energy is distributed in frequency space. According to the Wiener-Khinchin theorem, the *power spectral density* is the (discrete time) Fourier transform of the autocorrelation function



(a) (b)

Fig. 1. Noise in PAL broadcasting. (a) Power Spectral Density [dB], (b) Noise signal (containing horizontal stripe patterns due to correlations).

$R_w(\boldsymbol{p})$:

$$P_w(\boldsymbol{\omega}) = \sum_{\boldsymbol{p}\in\mathbb{Z}^2} R_w(\boldsymbol{p}) \exp\left(-j\boldsymbol{\omega}^T\boldsymbol{p}\right). \tag{3}$$

White noise has a flat PSD: $P_w(\boldsymbol{\omega}) = 1$. Suppose a filter with frequency response $H(\boldsymbol{\omega}) \neq 1$ is applied to the noise signal, then the resulting PSD $P'_w(\boldsymbol{\omega})$ becomes Baher (2001):

$$P'_w(\boldsymbol{\omega}) = P_w(\boldsymbol{\omega})\,|H(\boldsymbol{\omega})|^2. \tag{4}$$

Clearly, the PSD $P'(\boldsymbol{\omega})$ is subjected to the filter magnitude response $|H(\boldsymbol{\omega})|$. Hence one can think of correlated noise as white noise subjected to linear filtering. In analogy with the term "*white noise*" the resulting term is called "*colored noise*" (or *correlated* noise, because the filtering introduces correlations in the noise samples).
In practical circumstances, there are a number of origins of colored noise in images:

- *Phase Alternating Line (PAL) television*: the noise in PAL television images is a good example of colored noise. The correlations between the noise samples are caused by several mechanisms, such as deinterlacing Kwon et al. (2003), demodulation and filter schemes. In Figure 1, the PSD of a noise patch from a PAL broadcast is shown. Here, there is a high concentration of energy in the lower horizontal frequencies, leading to horizontal stripes and artifacts.
- *Color interpolation (demosaicing)*: modern digital cameras use a rectangular arrangement of photosensitive elements. In this matrix arrangement, photosensitive elements of different color sensitivity are placed in an interleaved way. This allows sampling of full color images without the use of three arrays of photosensitive elements. One popular example is the Bayer pattern Bayer (1976). Color interpolation (or demosaicing) is the process of estimating the values of missing photosensitive elements.
- *Post-processing techniques*: image noise often becomes correlated by the use of post-processing techniques, e.g., image quality enhancement techniques, sharpening filters, digital zoom functions of cameras, JPEG compression...

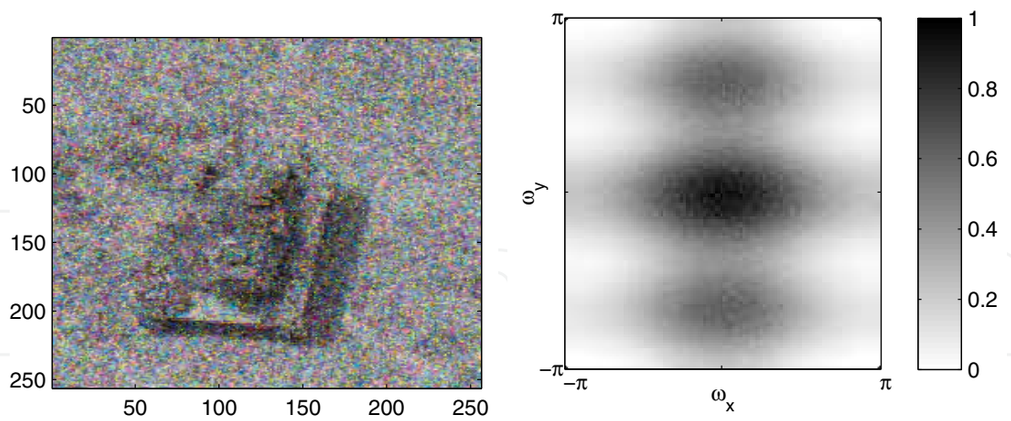


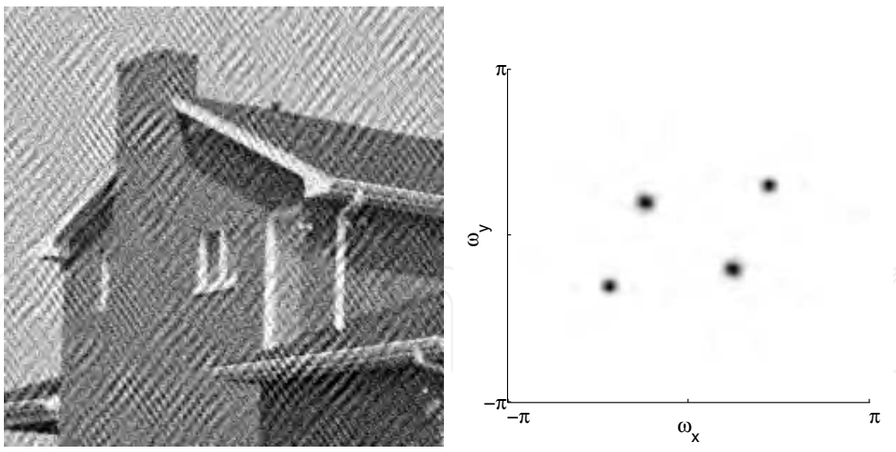
Fig. 2. (a) Image corrupted with colored noise caused by demosaicing (b) PSD of the noise in the green color channel of (a).

- *Thermal cameras*: images captured by thermal cameras of the push broom or whisk broom type often exhibit streaking noise artifacts, mainly caused by detector and sampling circuitry Aelterman, Goossens, Pižurica & Philips (2010). This kind of noise can be approximated using a $1/f$ frequency characteristic (called *pink* noise) Borel et al. (1996). Pink noise also frequently arises in image sensors that acquire pixel data in time.
- *Medical imaging*: in computed tomography (CT), noise correlations are introduced by the specific reconstruction technique that is being used. Noise created by the backprojection algorithm (without reconstruction filter) is called *ramp-spectrum* noise, and has an $1/f$ frequency characteristic. Noise in magnetic resonance imaging (MRI) is traditionally considered *white* Nowak (1999); Pižurica et al. (2003), although many MRI scanner manufacturers have included a wide range of techniques to allow for shorter scanning times (mainly to avoid patient motion artifacts in the images). To name a few: K-space subsampling, partial Fourier, elliptical filtering Aelterman, Deblaere, Goossens, Pižurica & Philips (2010). The use of these techniques results in correlated noise in the reconstructed MRI images.

In Figure 3 another example is shown of an image corrupted with colored noise. The colored noise was artificially generated by subjecting white noise to a filter with magnitude response $\sqrt{P(\boldsymbol{\omega})}$ and subsequently by adding the filtered noise to the images.

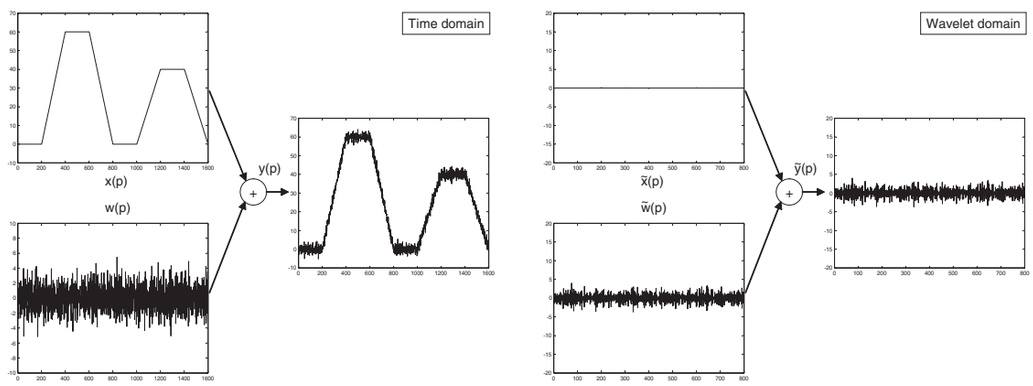
2. Wavelets for the estimation of colored noise

Spatially stationary colored noise can be directly specified through its mean and autocorrelation function and/or power spectral density. Given an observed noise signal $w(\boldsymbol{p})$, the estimation of these parameters is then a relatively simple task by, e.g., using the sample mean and sample autocovariance estimates. However, in practice, it often happens that the observed signal also contains information other than noise, this underlying signal is unknown and it is the signal that we eventually want to estimate. Hence, we are observing $y(\boldsymbol{p})$ instead of $w(\boldsymbol{p})$. The estimation of the noise statistics from the signal $y(\boldsymbol{p})$ is then considerably more difficult.



(a) (b)

Fig. 3. Illustration of the noise PSD: (a) Image with correlated noise, (b) The noise PSD (in frequency domain, the center of the image is the origin of frequency space, white corresponds with low noise powers, black with high noise powers).



(a) (b)

Fig. 4. Example of a piecewise linear signal with correlated noise. Our goal is to estimate the noise power spectrum from the corrupted signal $y(\boldsymbol{p})$. (a) The signals in time domain, (b) The finest scale of the wavelet transform of the signals (Daubechies' wavelet with 2 vanishing moments was used).

This problem is illustrated in Figure 4 for a piecewise linear signal corrupted with correlated Gaussian noise. While the noise statistics can be easily estimated from $w(\boldsymbol{p})$, we only have the degraded signal $y(\boldsymbol{p})$ at our disposal, which also contains an unknown signal component. A straightforward solution is then to first estimate the signal $\hat{x}(\boldsymbol{p})$, to subtract it from $y(\boldsymbol{p})$ and finally to estimate the noise statistics from the difference $y(\boldsymbol{p}) - \hat{x}(\boldsymbol{p})$. However, *optimal* estimation of $x(\boldsymbol{p})$ from $y(\boldsymbol{p})$ requires knowledge of the noise statistics on its own, so we have a chicken-and-egg problem. The common approach is then to use iterative techniques, which first estimate $\hat{x}(\boldsymbol{p})$ and then later refine this estimate $\hat{x}(\boldsymbol{p})$ when better estimates for the noise parameters become available.

In this chapter, we will take a different approach by relying on the properties of *wavelets*. The wavelet transform Daubechies (1992); Mallat (1999) analyzes signals according to different

scales and at different points in time. Starting from a fixed mother wavelet $\psi(t)$, the input signal is correlated with time-shifted and time-stretched (dilated) versions of this wavelet. Correlations with wavelets with a large dilation factor then give the coarse features of the signal, while correlations with wavelets with small dilation factors give the fine signal details. Because the wavelet basis functions are well localized in time or space (this is in contrast to the basis functions of e.g., the Fourier transform), wavelets are ideal candidates for analyzing non-stationary signals, having statistical properties that vary in time (or space).
The Daubechies wavelets are a class of orthogonal wavelets for which the number of vanishing moments for a given support is maximal. More specifically, the n-th moment of a real-valued wavelet function $\psi(t)$ is defined by:

$$\mu_n = \int_{-\infty}^{+\infty} t^n \psi(t) dt. \tag{5}$$

The Daubechies wavelet of support $2N$ (with N vanishing moments) will have moments $\mu_n = 0$ for $0 \leq n < N$. Now, let us denote the time-shifted and dilated basis functions of $\psi(t)$ by:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \tag{6}$$

where a is the dilation factor, b is a time shift, and the constant $1/\sqrt{a}$ is an energy normalization factor. The continuous wavelet transform of a signal $f \in L^2(\mathbb{R})$ is defined by:

$$\mathcal{W}f(a,b) = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}(t) \mathrm{d}t. \tag{7}$$

Now, suppose that a signal is linear on a region larger than the support $S(a)$ of the wavelet function $\psi_{a,b}(t)$:

$$f(t) = c \cdot t \quad \text{if } |t-b| \leq S(a).$$

For Daubechies wavelets with at least two vanishing moments ($N \geq 2$), the corresponding wavelet coefficient $\mathcal{W}f(a,b)$ will be zero:

$$\begin{aligned}
\mathcal{W}f(a,b) &= \int_{-\infty}^{+\infty} c \cdot t \psi_{a,b}(t) \mathrm{d}t \\
&= \frac{c}{\sqrt{a}} \int_{-\infty}^{+\infty} t \psi \left(\frac{t-b}{a} \right) \mathrm{d}t \\
&= c\sqrt{a} \int_{-\infty}^{+\infty} \left(at' + b \right) \psi \left(t' \right) \mathrm{d}t' \\
&= ca^{3/2} \int_{-\infty}^{+\infty} t' \psi \left(t' \right) \mathrm{d}t' + cba^{1/2} \int_{-\infty}^{+\infty} \psi \left(t' \right) \mathrm{d}t' \\
&= 0
\end{aligned}$$

In the remainder of this chapter, for the ease of notation, we will consider one particular wavelet subband (with scale a) at a time and we will denote the corresponding wavelet coefficients by a tilde: for example $\tilde{x}(\boldsymbol{p})$ are the wavelet coefficients for that particular scale of $x(\boldsymbol{p})$. The process can then be repeated for other subbands as well. Let us now apply a

Daubechies wavelet transform to the piecewise linear signal from Figure 4(a). The result is shown in Figure 4(b) for the finest scale of the DWT[1]: because of the vanishing moments of the wavelet, the wavelet coefficients $\tilde{x}(\boldsymbol{p})$ are zero, except at the positions where the derivative of $x(\boldsymbol{p})$ does not exist. At these positions, the wavelet coefficients have a negligibly small magnitude. This nicely illustrates the sparsifying properties of the DWT for this type of signal. Correspondingly, the wavelet coefficients $\tilde{y}(\boldsymbol{p})$ are (approximately) $\tilde{w}(\boldsymbol{p})$, which means that the chicken-and-egg problem is solved: the noise statistics can be directly estimated from $\tilde{y}(\boldsymbol{p})$! More specifically, the wavelet domain autocorrelation function of $w(\boldsymbol{p})$ can in this case be estimated based on the following relationship:

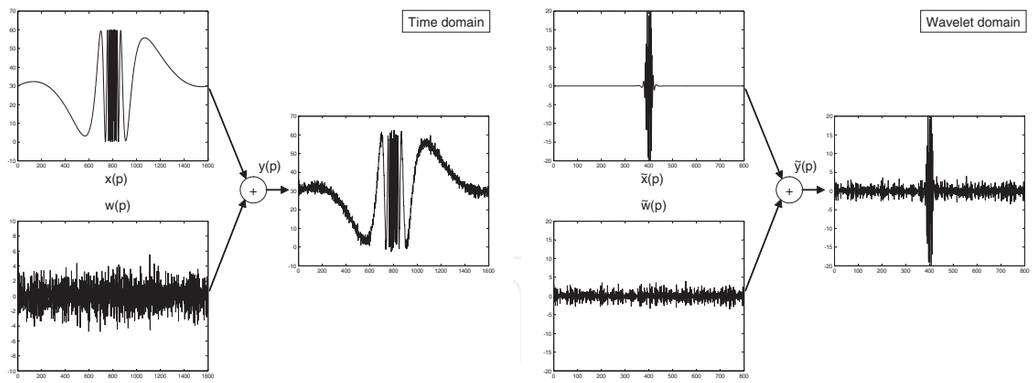
$$R_{\tilde{w}}(\boldsymbol{p}) \approx R_{\tilde{y}}(\boldsymbol{p}) = \mathrm{E}\left[\tilde{y}(\boldsymbol{p}')\overline{\tilde{y}(\boldsymbol{p}+\boldsymbol{p}')}\right]. \tag{8}$$

It then suffices to compute the sample autocorrelation function of $\tilde{y}(\boldsymbol{p})$. There are now two issues remaining, which we will explain in the remainder of this Chapter:

1. The autocorrelation function of a signal in the wavelet domain (e.g. a for particular wavelet subband) is not the same as the autocorrelation function of a signal in time domain. Nevertheless, there exists a simple relation between both, as we will explain in Section 3.
2. Most real-life signals are not piecewise linear functions or piecewise polynomials. For such signals, the wavelet coefficient magnitudes may become non-negligible, causing serious biases to the final noise estimates. An example of a frequency modulated signal with maximal frequency at half length of the signal, is given in Figure 5. Because of the high local bandwidth of the signal at this time position, the wavelet is not able to cancel out the signal, resulting in wavelet coefficients with a large magnitude. Consequently, the approximation $\tilde{y}(\boldsymbol{p}) \approx \tilde{x}(\boldsymbol{p})$ does not hold anymore. However, it can be seen in Figure 5(b) that this phenomenon is well localized in time, hence, because the noise process is assumed to be stationary, a plausible solution would be to estimate the noise statistics from the wavelet coefficients $\tilde{y}(\boldsymbol{p})$ that have a small underlying components $\tilde{x}(\boldsymbol{p})$ (ignoring the outliers in Figure 5(b)). In Section 4 we will discuss solutions that generalize this idea by using a statistical prior model for wavelet coefficients.

So far, we discussed the estimation of colored noise for one dimensional signals. The reasoning can also be extended to higher dimensional signals, such as images. To illustrate this, a noisy image together with its DWT are shown in Figure 6. It can be seen that the wavelet subbands (LH, HL and HH in Figure 6) predominantly contain information on the noise, with exception in the areas of textures and edges (the fine hairs of the mandrill). In these areas, the (noise-free) wavelet coefficients $\tilde{x}(\boldsymbol{p})$ still have a relatively large magnitude, but this phenomenon is localized - in the surrounding smooth regions the wavelet coefficients $\tilde{y}(\boldsymbol{p})$ mostly consist of noise.
For higher dimensional signals, the DWT is usually computed by using basis functions that are tensor products of one dimensional wavelets and one dimensional scaling functions. While this approach can efficiently deal with point-wise singularities (e.g. bumps, dots, ...), most structures in images are line-like singularities with a given direction. However, the DWT can not well adapt to the arbitrary direction of the singularity: for example, the

[1] Note that for other scales the plots are similar.



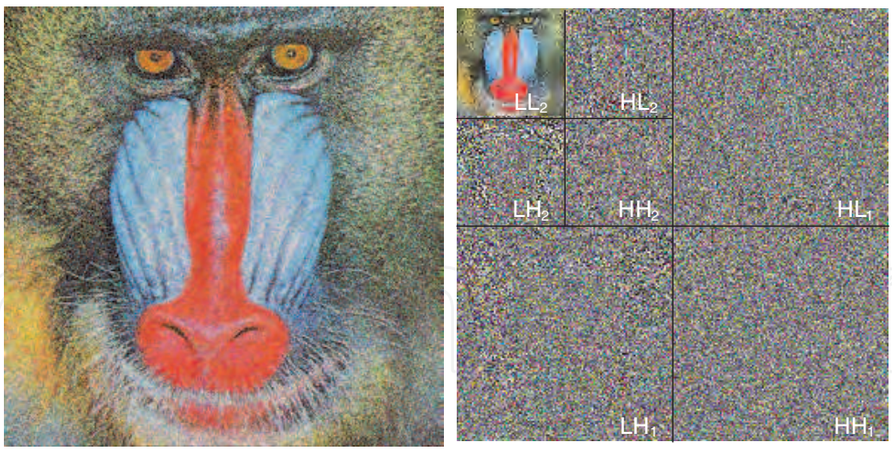
(a) (b)

Fig. 5. Example of a non-piecewise linear signal with correlated noise. Our goal is to estimate the noise power spectrum from the corrupted signal $y(\boldsymbol{p})$. (a) The signals in time domain, (b) The finest scale of the wavelet transform of the signals (Daubechies' wavelet with 2 vanishing moments was used).

transform can not make a distinction between features oriented at +45° and -45°. This is known as the *checkerboard* problem of the DWT: due to the separability of the higher dimensional wavelets, these wavelets appear as a checkerboard pattern which does not have a dominant direction. Consequently, many nonzero wavelet coefficients may be needed to represent a line singularity at an arbitrary orientation. To overcome this limitation there has recently been a lot of interest in transforms that offer a better *directional selectivity*. Examples are steerable pyramids Simoncelli et al. (1992), dual-tree complex wavelets Selesnick et al. (2005a), Marr-like wavelet pyramids Van De Ville & Unser (2008), 2-D (log) Gabor transforms Fischer et al. (2007); Lee (1996), contourlets Do & Vetterli (2005), ridgelets Candès (1998); Do & Vetterli (2003), curvelets Candès et al. (2006) and shearlets Guo & Labate (2007). These transforms are designed to have better sparsifying properties so that our outlier problem in Figure 5(b) is alleviated (but not solved).
In the next subsections we will focus on the DWT as a primary multiresolution decomposition tool, however, the same reasoning can also be applied to more recently developed transforms.

3. From time-domain to wavelet-domain autocorrelation functions

Because our goal is to estimate the autocorrelation function of noise in the wavelet domain, it is very useful to know how the wavelet-domain and time-domain autocorrelation functions are related to each other. When the autocorrelation function of the input signal is known, a simple Monte-Carlo based technique is to generate colored noise with this given autocorrelation function, then to transform the noise to the wavelet domain (or other multiresolution transform domain) and subsequently to estimate the autocorrelation function in this domain Portilla et al. (2003). While such computational method is attractive from an implementation point of view, it does not bring a direct analytical relationship between both autocorrelation functions. We will see in Section 6 that an analytical relationship will prove to be very useful when estimating parametric noise PSDs.
Let us consider the wavelet analysis filterbank shown in Figure 7(a), where a signal with z-transform $\tilde{F}_1(z)$ is filtered by a wavelet filter $G(z)$ and a scaling filter $H(z)$. Both signals



(a) (b)

Fig. 6. (a) *Baboon* image with noise, (b) DWT of the image.

are subsequently decimated by a factor of two. The analysis is iterated on the scaling coefficients $F_2(z)$. Now, the input signal has an autocorrelation function in the z-domain defined by $\tilde{R}_1(z) = E\left[\tilde{F}_1(z)\tilde{F}_1(z^{-1})\right]$. The filtered signals then have autocorrelation functions respectively $\tilde{R}_1(z)G(z)G(z^{-1})$ and $\tilde{R}_1(z)H(z)H(z^{-1})$. Decimating the resulting signals by a factor 2 leads to the signal with autocorrelation function Goossens et al. (2010):

$$\begin{aligned} R_1(z) &= E\left[F_1(z)F_1(z^{-1})\right] \\ &= \frac{1}{2}\left(\tilde{R}_1\left(z^{\frac{1}{2}}\right)G\left(z^{\frac{1}{2}}\right)G\left(z^{-\frac{1}{2}}\right) + \tilde{R}_1\left(-z^{\frac{1}{2}}\right)G\left(-z^{\frac{1}{2}}\right)G\left(-z^{-\frac{1}{2}}\right)\right), \\ R_2(z) &= E\left[F_2(z)F_2(z^{-1})\right] \\ &= \frac{1}{2}\left(\tilde{R}_1\left(z^{\frac{1}{2}}\right)H\left(z^{\frac{1}{2}}\right)H\left(z^{-\frac{1}{2}}\right) + \tilde{R}_1\left(-z^{\frac{1}{2}}\right)H\left(-z^{\frac{1}{2}}\right)H\left(-z^{-\frac{1}{2}}\right)\right). \end{aligned} \tag{9}$$

Hence, the wavelet-domain autocorrelation function $R_1(z)$ can be directly computed from the autocorrelation function of the input signal $\tilde{R}_1(z)$ and the wavelet and scaling filters. This involves two simple convolutions and a decimation operation of the input autocorrelation function $\tilde{R}_1(z)$. For subsequent decompositions (coarser scales of the wavelet transform), this process can be iterated by re-inserting $\tilde{R}_1(z) = R_2(z)$ in (9).

To show that this reasoning also applies to other wavelet transforms, we will briefly discuss the adaptation to the dual-tree complex wavelet transform (DT-CWT) Kingsbury (2001) in one dimension. Extension to higher dimensions is then straightforward. The 1D DT-CWT is implemented using two parallel DWT filter banks, the first filter bank uses the real parts of the complex wavelet and scaling filters (respectively $G_1(z)$ and $H_1(z)$), while in the second filter bank, the imaginary parts of the wavelet and scaling filters (respectively $G_2(z)$ and $H_2(z)$) are applied. Finally, the output of both filter banks are mixed together (see the right square in Figure 7(b)), applying a 45° rotation in the complex plane. This last step is in fact only necessary in 2D (or higher dimensions), where complex wavelets are constructed using tensor products of 1D complex wavelets. The translation of the resulting complex-valued filter banks to parallel real-valued filter banks then automatically results into this phase modulation in the complex plane (for more details, see Selesnick et al. (2005b)). Defining

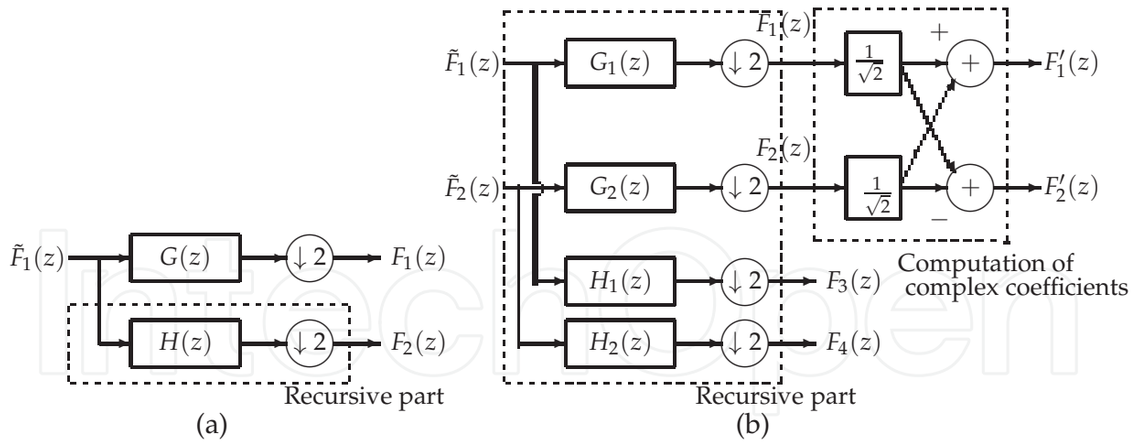


Fig. 7. Analysis filterbank for (a) the DWT, (b) the DT-CWT.

$\tilde{R}_2(z) = E\left[\tilde{F}_2(z)\tilde{F}_2(z^{-1})\right]$, application of (9) to the DT-CWT leads to the following equations:

$$
\begin{aligned}
R_1(z) &= E\left[F_1(z)F_1(z^{-1})\right] = \frac{1}{2}\left(\tilde{R}_1\left(z^{\frac{1}{2}}\right)G_1\left(z^{\frac{1}{2}}\right)G_1\left(z^{-\frac{1}{2}}\right) + \tilde{R}_1\left(-z^{\frac{1}{2}}\right)G_1\left(-z^{\frac{1}{2}}\right)G_1\left(-z^{-\frac{1}{2}}\right)\right)\\
R_2(z) &= E\left[F_2(z)F_2(z^{-1})\right] = \frac{1}{2}\left(\tilde{R}_2\left(z^{\frac{1}{2}}\right)G_2\left(z^{\frac{1}{2}}\right)G_2\left(z^{-\frac{1}{2}}\right) + \tilde{R}_2\left(-z^{\frac{1}{2}}\right)G_2\left(-z^{\frac{1}{2}}\right)G_2\left(-z^{-\frac{1}{2}}\right)\right)\\
R_3(z) &= E\left[F_3(z)F_3(z^{-1})\right] = \frac{1}{2}\left(\tilde{R}_1\left(z^{\frac{1}{2}}\right)H_1\left(z^{\frac{1}{2}}\right)H_1\left(z^{-\frac{1}{2}}\right) + \tilde{R}_1\left(-z^{\frac{1}{2}}\right)H_1\left(-z^{\frac{1}{2}}\right)H_1\left(-z^{-\frac{1}{2}}\right)\right)\\
R_4(z) &= E\left[F_4(z)F_4(z^{-1})\right] = \frac{1}{2}\left(\tilde{R}_2\left(z^{\frac{1}{2}}\right)H_2\left(z^{\frac{1}{2}}\right)H_2\left(z^{-\frac{1}{2}}\right) + \tilde{R}_2\left(-z^{\frac{1}{2}}\right)H_2\left(-z^{\frac{1}{2}}\right)H_2\left(-z^{-\frac{1}{2}}\right)\right)\\
S_{1,2}(z) &= E\left[F_1(z)F_2(z^{-1})\right] = \frac{1}{2}\left(\tilde{S}_{1,2}\left(z^{\frac{1}{2}}\right)G_1\left(z^{\frac{1}{2}}\right)G_2\left(z^{-\frac{1}{2}}\right) + \tilde{S}_{1,2}\left(-z^{\frac{1}{2}}\right)G_1\left(-z^{\frac{1}{2}}\right)G_2\left(-z^{-\frac{1}{2}}\right)\right) \qquad (10)
\end{aligned}
$$

where $\tilde{S}_{1,2}(z)$ is the cross-power spectrum between $\tilde{F}_1(z)$ and $\tilde{F}_2(z)$: $\tilde{S}_{1,2}(z) = E\left[\tilde{F}_1(z)\tilde{F}_2(z^{-1})\right]$. The final autocorrelation functions (after the complex phase modulation) are computed from $R_1(z)$, $R_2(z)$ and $S_{1,2}(z)$, as follows:

$$
\begin{aligned}
R_1'(z) &= E\left[F_1'(z)F_1'(z^{-1})\right] = \frac{1}{2}\left(R_1(z) + R_2(z)\right) + \frac{1}{2}\left(S_{1,2}(z) + S_{1,2}(z^{-1})\right),\\
R_2'(z) &= E\left[F_2'(z)F_2'(z^{-1})\right] = \frac{1}{2}\left(R_1(z) + R_2(z)\right) - \frac{1}{2}\left(S_{1,2}(z) + S_{1,2}(z^{-1})\right). \qquad (11)
\end{aligned}
$$

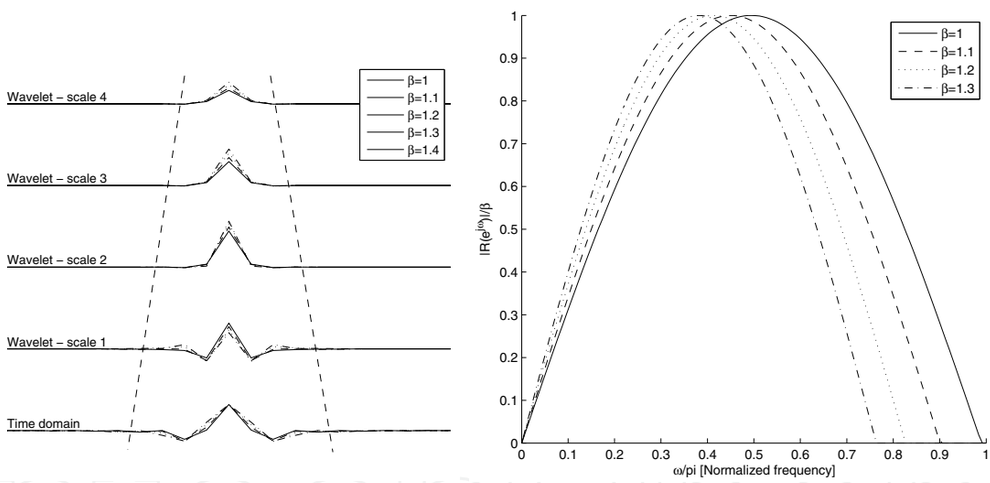
In Algorithm 1, an OCTAVE/MATLAB program is given for computing the autocorrelation functions in case of the DWT and DT-CWT, according to (9) and (10)-(11). In this program, the variables `lo` and `hi` respectively signify the scaling and wavelet coefficients. It can be seen that all operations are linear operations, which makes it possible to express the conversion from time-domain to wavelet-domain as a matrix multiplication applied to the input autocorrelation coefficient vector.

In Figure 8, an example of a parametric autocorrelation function and its DWT decomposition, according to (9), is shown. Due to the cone of influence (Mallat, 1999, p. 174), the support size of the autocorrelation function decreases when increasing the wavelet scale (i.e., when analyzing finer scales). Interesting to note is the envelope of the noise variance in the wavelet domain: the noise variance is identical to the noise autocorrelation function evaluated in the origin (which is in this case also the maximum of the autocorrelation function). When one modifies the center band frequency of the noise PSD in Figure 8(b), this also directly

Algorithm 1 OCTAVE/MATLAB program for computing wavelet domain autocorrelation functions.

```
f = [1 2 1];   % input autocorrelation function
% discrete wavelet transform (DWT)
lo = conv(f, conv(h, h(end:-1:1)));
hi = conv(f, conv(g, g(end:-1:1)));
lo = lo(1:2:end); hi = hi(1:2:end);

% dual-tree complex wavelet transform (DT-CWT)
lo1 = conv(f, conv(h1, h1(end:-1:1)));
hi1 = conv(f, conv(g1, g1(end:-1:1)));
lo2 = conv(f, conv(h2, h2(end:-1:1)));
hi2 = conv(f, conv(g2, g2(end:-1:1)));
cr1 = conv(f, conv(h1, h2(end:-1:1))); % cross-correlation
cr2 = conv(f, conv(h2, h1(end:-1:1)));
lo1 = lo1(1:2:end); lo2 = lo2(1:2:end); % decimations
hi1 = hi1(1:2:end); hi2 = hi2(1:2:end);
cr1 = cr1(1:2:end); cr2 = cr2(1:2:end);
hi1_out = 0.5*(hi1+hi2)+0.5*(cr1+cr2); % complex phase modulation
hi2_out = 0.5*(hi1+hi2)-0.5*(cr1+cr2);
```



(a) (b)

Fig. 8. (a) Wavelet analysis of the autocorrelation function (in z-domain) $R(z) = \sum_n \frac{\beta^2}{\pi(n^2-\beta^2)} \left(1 + \cos\left(\frac{\pi n}{\beta}\right)\right) z^n$ across different scales and for different values of β. Daubechies' wavelet with two vanishing moments was used. (b) Power spectral density $R(e^{j\omega})$ for different values of β.

influences the noise variances of the individual wavelet subbands (see Figure 8(a)), due to the frequency-selective behavior of the wavelets at different scales. For example, increasing the parameter β has as effect that the noise variance at wavelet scale 4 decreases. This also suggests that, when a thresholding strategy (e.g. soft/hardthresholding) would be used to suppress the colored noise process, the thresholds would need to be level-dependent, e.g., as proposed by Johnstone and Silverman Johnstone & Silverman (1997).

4. Statistical priors for noise estimation

As already illustrated in Figure 5, the DWT will in general not fully suppress the signal. Consequently, wavelet-based noise estimation techniques need to take into account that the wavelet coefficients contain a non-negligible signal component. One of the earliest and well-known wavelet-based noise estimation techniques is the MAD estimator from Donoho, which estimates the noise standard deviation as follows Donoho & Johnstone (1995):

$$\hat{\sigma} = \frac{\text{Median}_{\boldsymbol{p}}\left(|\tilde{y}(\boldsymbol{p})|\right)}{0.6745}. \tag{12}$$

The estimator gives level dependent estimates of the noise standard deviation in every wavelet subband. Based on robust statistics, the non-zero signal coefficients are considered to be *outliers*. By computing a median instead of a more traditional mean, the outlier influences in the end result are significantly reduced.

In this chapter, we are interested in estimating the noise *correlations* or *covariances* (next to the noise standard deviation), therefore the estimator (12) can not directly be used. For this purpose, a general class of robust S estimators for the covariance (see, e.g., Campbell et al. (1998); Pena & Prieto (2001)) can be used. These estimators detect outliers after finding projections that maximize the kurtosis of the data. An illustration of such a technique is given in Figure 9: the robust S estimators attempt to estimate the covariance of the noise (the black dots in Figure 9). In this case this is equivalent to determining the sizes of the axes and the orientation of the ellipse shown in the figure (the ellipse can be seen as an isocontour of the probability function of the data). Because of the presence of outliers (the crosses in Figure 9), this is not a trivial task. The robust estimation techniques then try to identify the outliers, in an iterative process.

While robust S estimators are unfamiliar with the structure of the data they are applied to, in our application, we have some more information on the data that we can take into our advantage. In particular, due to the sparsifying properties of the chosen multiresolution transform, the identification of the outliers (signal components) is somewhat easier: the multiresolution transform already performs a projection to maximize the kurtosis. Instead of relying on robust statistics, we will incorporate prior knowledge on the noise-free wavelet coefficients to further improve the estimation performance using Bayesian techniques. Our noise estimation approach will then consist in 1) specifying a statistical prior distribution for the noise-free signal coefficients, 2) maximum likelihood estimation of the unknown noise covariance matrix.

In the next subsections, we will briefly review a number of statistical models for noise-free wavelet coefficients and we will explain how these models can be used to perform noise estimation.

4.1 The generalized Laplace distribution

It has been found in several studies Field (1987); Mallat (1989) that histograms of wavelet coefficients (or generally coefficients of bandpass filtered images) have a highly kurtotic shape. An example is shown in Figure 10(a)-(b) for the *Baboon* image: the wavelet coefficient histogram reveals a sharp peak and a heavy tail. The sample kurtosis of the wavelet coefficients (6.98) is much higher than the theoretical kurtosis of a Gaussian distribution (which is 3). Several authors Antonini et al. (1992); Chang et al. (1998); Mallat

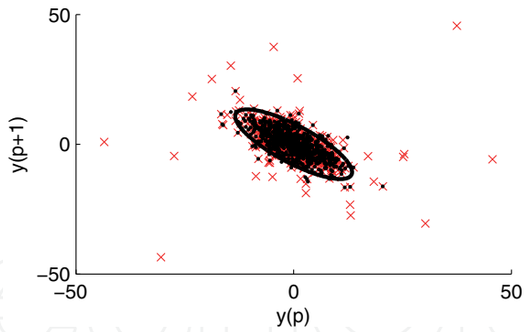
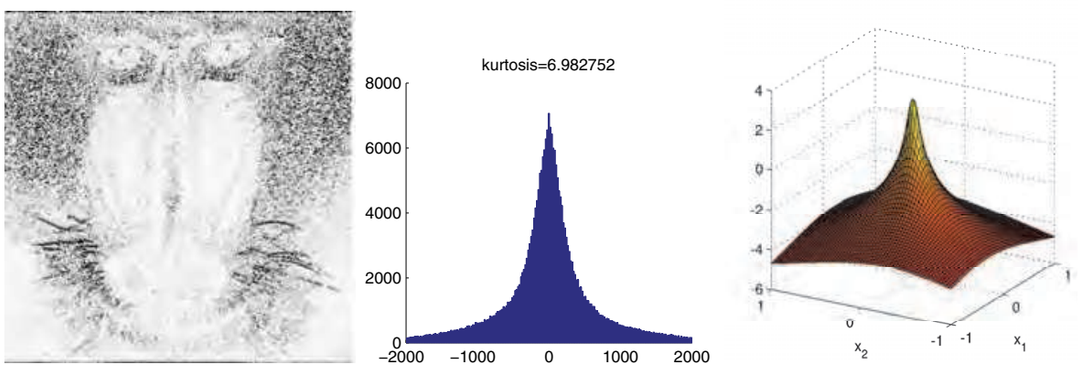


Fig. 9. Joint histogram of neighboring wavelet coefficients for Figure 5(b). Black dots are noise coefficients, crosses are the outliers due to signal presence.



(a) (b) (c)

Fig. 10. (a) wavelet subband LH_1 of the *Baboon* image (*black* corresponds to a large coefficient magnitude, *white* to small magnitudes, the contrast of the image was enhanced to better reveal the details), (b) histogram of the wavelet coefficients in (a), (c) multivariate Gaussian Scale Mixture distribution.

(1989); Moulin & Liu (1999); Simoncelli & Adelson (1996) proposed to use a *generalized Laplace* distribution (GLD, also known as *generalized Gaussian* distribution) to model the kurtotic behavior of wavelet coefficients. The GLD is defined as:

$$f_{\tilde{x}}(\tilde{x}) = \frac{\nu}{2s\Gamma(1/\nu)} \exp\left(-\left|\frac{\tilde{x}}{s}\right|^{\nu}\right), \tag{13}$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t$ is the Gamma function. The parameter s is scale parameter of the distribution, which controls the variance of the distribution. The parameter ν is a shape parameter that is related to the kurtosis of the distribution, given by:

$$\kappa = \frac{\Gamma(5/\nu)\Gamma(1/\nu)}{\Gamma^2(3/\nu)} - 3. \tag{14}$$

The shape parameter ν is typically in the range [0.5, 1]. Because in practice, the actual value of this parameter is unknown, the parameter value is usually estimated from the observed data. This may be done using the maximum likelihood method or the method of moments Srivastava et al. (2003).

4.2 Elliptically symmetric distributions and Gaussian Scale Mixtures

The GLD from (13) is a univariate distribution that can well model highly kurtotic histograms of wavelet coefficients, however this distribution does not allow capture correlations between different observations $\tilde{x}$. This can be achieved by using multivariate distributions, where dependencies between neighboring wavelet coefficients can be modeled. For these densities, a neighborhood of a fixed size (e.g. 3×3 in 2D) is defined around every wavelet coefficient. Next, every neighborhood[2] can be represented by a vector, e.g., by using column stacking. In the following, we will use bold letters $\tilde{\boldsymbol{x}}(\boldsymbol{p})$, $\tilde{\boldsymbol{w}}(\boldsymbol{p})$, $\tilde{\boldsymbol{y}}(\boldsymbol{p})$ to denote neighborhood vectors extracted by column stacking. Statistical studies Portilla et al. (2003); Srivastava et al. (2003) have indicated that, next to the kurtotic behavior, the noise-free wavelet coefficients are typically symmetric around the mode and the joint histograms have elliptical contours. This suggests the use of elliptically symmetric distributions (ESD) to model these characteristics. The ESD is defined by Kotz & Kozubowski (2001):

$$f_{\tilde{\boldsymbol{x}}}\left(\tilde{\boldsymbol{x}}\right) = k_d \left|\boldsymbol{C}_x\right|^{-1/2} g\left(\left|\left(\tilde{\boldsymbol{x}} - \mathbf{m}\right)\boldsymbol{C}_x^{-1}\left(\tilde{\mathbf{x}} - \mathbf{m}\right)\right|^{1/2}\right), \tag{15}$$

where $\mathbf{m}$ is the mean of the distribution (typically $\mathbf{m} = \mathbf{0}$), $g(u)$ is a real-valued function (called density generator function), d is the length of $\tilde{\boldsymbol{x}}$ and k_d is a proportionality constant. A multivariate extension of the GLD is obtained by using the following density generator function Kotz et al. (2000): $g(u) = \exp\left(-|u|^{\nu}\right)$. The resulting distribution is known as the multivariate exponential power distribution (EPD). For our modeling task, the EPD has a number of practical limitations: 1) the marginal densities of the distribution are not EPD-distributed and 2) for estimation purposes, the exponential power ν often leads to integral expressions that are analytically intractable.

Wainwright & Simoncelli (2000) noted that when the wavelet filter responses are normalized by dividing by the square root of the local variance, the statistics of the normalized coefficients are approximately Gaussian. The Gaussian Scale Mixture (GSM), see Figure 10(c), was then proposed to account both for the correlations and the variability in local variance of the wavelet coefficients. A random variable $\tilde{\boldsymbol{x}}$ is GSM distributed if it can be written as the product of a zero mean Gaussian random vector $\tilde{\boldsymbol{u}}$ and a scalar positive random variable $\sqrt{z}$ Andrews & Mallows (1974):

$$\tilde{\boldsymbol{x}} \stackrel{\mathrm{d}}{=} \sqrt{z}\tilde{\boldsymbol{u}} \tag{16}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. The scalar variable z is not observed and is therefore also called 'hidden' multiplier or mixing variable. Because of scaling ambiguity between $\sqrt{z}$ and $\tilde{\boldsymbol{u}}$, the hidden multiplier is often assumed to be normalized such that $\mathrm{E}\left[z\right] = 1$. Prior distributions for z include Jeffrey's *non-informative*[3] prior Portilla et al. (2003), the *log-normal* prior Portilla & Simoncelli (2001), the *exponential* distribution Selesnick (2006) and the *Gamma* distribution Fadili & Boubchir (2005); Srivastava et al. (2002).

[2] Quite often, the neighborhoods are chosen to be overlapping, despite of the fact that this destroys the mutual independence of the different neighborhood vectors. This is done to arrive at a sufficiently large number of neighborhood vectors (for example, for a 3×3 neighborhood, the number of vectors will be multiplied by 9), which will generally result in more reliable estimates.

[3] Note that in this case, the mathematical expectation $\mathrm{E}\left[z\right]$ does not exist.

The GSM also belongs to the family of ESDs. The density generator function is given by:

$$g(x) = \int_0^{+\infty} f_z(z) z^{-\frac{d}{2}} \exp\left(-\frac{1}{2z}x^2\right) \mathrm{d}z. \tag{17}$$

For some hidden multiplier densities $f_z(z)$ a closed-form expression can be found for $g(x)$, although most often, numerical integration is performed over a closed interval. In Gómez et al. (2008) it has been shown that the EPD is also a GSM distribution, for some values of the shape parameter $\nu \in]0,1]$. However, the distribution $f_z(z)$ depends on d and has a complicated analytical expression (see Gómez et al. (2008)).

4.3 Other prior distributions

In literature, several other prior distributions for noise-free wavelet coefficients have been proposed. For example, the *Student-T* distribution Tzikas et al. (2007), *Alpha-stable* distributions Achim et al. (2001); Nikias & Shao (1995) and the *Cauchy* distribution Rabbani et al. (2006). All these heavy tailed distributions have a GSM representation, hence studying general GSMs automatically covers all of these distributions. Next, a complex extension of the Gaussian Scale Mixture density has been proposed for modeling complex-valued wavelet coefficients in Vo et al. (2007). This *complex GSM* distribution is a special case of the GSM distribution, with a special condition imposed to the covariance matrix of the distribution. Next to GSMs, *mixtures of a Gaussian distribution and a point mass at zero* were used in Abramovich et al. (1998); Clyde et al. (1998), *mixtures of two Gaussian distributions* in Crouse et al. (1998); Fan & Xia (2001); Romberg et al. (2001) and *mixtures of truncated or quasi-Laplace distributions* in Pižurica & Philips (2006); Shi & Selesnick (2006).

5. Noise covariance estimation techniques

In this Section, we will use the GSM prior distribution from Section 4 to design a noise covariance estimation technique. We therefore start from the signal-plus-noise model from equation (1). The assumed addititivity of the signal and noise leads to an equivalent expression in the wavelet domain:

$$\tilde{\boldsymbol{y}}(\boldsymbol{p}) = \tilde{\boldsymbol{x}}(\boldsymbol{p}) + \tilde{\boldsymbol{w}}(\boldsymbol{p}), \tag{18}$$

where $\tilde{\boldsymbol{w}}(\boldsymbol{p})$ is spatially stationary Gaussian distributed vector of length d with mean $\boldsymbol{0}$ and covariance $\boldsymbol{C}_{\tilde{w}}$. Due to the assumed noise stationarity, the covariance matrix $\boldsymbol{C}_{\tilde{w}}$ has dimensions $d \times d$ and is directly related to the noise autocorrelation function $R_{\tilde{w}}(\boldsymbol{p})$: the covariance between two coefficients at positions $\boldsymbol{p}$ and $\boldsymbol{q}$ only depends on the difference in location between both positions:

$$\left(\boldsymbol{C}_{\tilde{w}}\right)_{\boldsymbol{p},\boldsymbol{q}} = R_{\tilde{w}}(\boldsymbol{q} - \boldsymbol{p}) \tag{19}$$

where vector-valued indices in $\left(\boldsymbol{C}_{\tilde{w}}\right)_{\boldsymbol{p},\boldsymbol{q}}$ are used as a short notation for their respective column-stacked ordening. By (19), the estimation of the noise autocorrelation function is equivalent to the estimation of the covariance $\boldsymbol{C}_{\tilde{w}}$. Next, the noise-free coefficients are GSM distributed with covariance matrix $\boldsymbol{C}_{\tilde{x}}$. For the GSM model, we have $\tilde{\boldsymbol{x}}|z \sim \mathcal{N}\left(\boldsymbol{0}, z\boldsymbol{C}_{\tilde{u}}\right)$. Consequently, the density of $\tilde{\boldsymbol{y}}$ is a specific case of a Gaussian mixture model:

$$\tilde{\boldsymbol{y}}|z \sim \mathcal{N}\left(\boldsymbol{0}, z\boldsymbol{C}_{\tilde{u}} + \boldsymbol{C}_{\tilde{w}}\right) \tag{20}$$

where the signal covariance is also unknown. We remark that this matrix can be eliminated relying on $\boldsymbol{C}_{\tilde{u}} + \boldsymbol{C}_{\tilde{w}} = \boldsymbol{C}_{\tilde{y}}$ (this directly follows from (1), when $\mathrm{E}[z] = 1$):

$$\tilde{\boldsymbol{y}}|z \sim \mathcal{N}\left(\boldsymbol{0}, z\boldsymbol{C}_{\tilde{y}} + (1-z)\boldsymbol{C}_{\tilde{w}}\right). \tag{21}$$

The signal-plus-noise covariance matrix can be estimated using the method of maximum likelihood: $\widehat{\boldsymbol{C}_{\tilde{y}}} = \frac{1}{N}\sum_{\boldsymbol{p}} \tilde{\boldsymbol{y}}(\boldsymbol{p})\tilde{\boldsymbol{y}}^T(\boldsymbol{p})$, with N the number of coefficients in the considered wavelet subband.

5.1 Generalized Expectation-Maximization algorithm

In Portilla (2004), a Generalized Expectation-Maximization (GEM) algorithm is given to estimate the noise covariance matrix. Based on an initial estimate of the noise covariance (typically chosen as $\boldsymbol{C}_{\tilde{w}}^{(0)} = c\boldsymbol{C}_{\tilde{y}}$, with $0 < c < 1$ a constant), the noise covariance matrix is iteratively updated according to the following rule:

$$\boldsymbol{C}_{\tilde{w}}^{(i+1)} = \frac{\sum_{\boldsymbol{p}} \mathrm{P}\left(z < z_0 | \tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)}\right) \tilde{\boldsymbol{y}}(\boldsymbol{p})\tilde{\boldsymbol{y}}^T(\boldsymbol{p})}{\sum_{\boldsymbol{p}} \mathrm{P}\left(z < z_0 | \tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)}\right)}, \tag{22}$$

where i is the iteration index and $\Theta^{(i)}$ denotes the GSM model parameters at iteration i and where z_0 is a small positive constant. Equation (22) can be motivated by the observation that for z sufficiently small, $\boldsymbol{C}_{\tilde{y}|z} = \boldsymbol{C}_{\tilde{w}}$. The posterior probability that $z < z_0$, conditioned on an observation vector $\tilde{\boldsymbol{y}}(\boldsymbol{p})$, i.e., $\mathrm{P}\left(z < z_0 | \tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)}\right)$ is then used as a weight in the averaging process. We can understand this as follows: $\mathrm{P}\left(z < z_0 | \tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)}\right)$ represents the probability that a given observation vector contains a negligible signal component. The estimated noise covariance is then the average over all sample covariances $\boldsymbol{y}(\tilde{\boldsymbol{p}})\tilde{\boldsymbol{y}}^T(\boldsymbol{p})$, weighted by the probability that the considered sample contains a negligible signal component.

Because the updating rule (22) is not guaranteed to increase the likelihood of the data, at every iteration it is checked if this new covariance estimate results in a higher likelihood: $Q(\Theta^{(i)}, \Theta^{(i+1)}) > Q(\Theta^{(i)}, \Theta^{(i)})$, with $Q(\Theta^{(i)}, \Theta)$ the expected log-likelihood function of the data:

$$\begin{aligned} Q(\Theta^{(i)}, \Theta) &= \mathrm{E}\left[\log f_{z|\tilde{\boldsymbol{y}}}\left(z|\tilde{\boldsymbol{y}}, \Theta\right) | \tilde{\boldsymbol{y}},\ \Theta^{(i)}\right] \\ &= \sum_{\boldsymbol{p}} \int_0^{+\infty} f_{z|\tilde{\boldsymbol{y}}}\left(z|\tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)}\right) \log f_{z|\tilde{\boldsymbol{y}}}\left(z|\tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta\right) \mathrm{d}z. \end{aligned} \tag{23}$$

In case the expected log-likelihood (23) decreases, it is proposed in Portilla (2004) to perform a gradient ascent step:

$$\boldsymbol{C}_{\tilde{w}}^{(i+1)} = \boldsymbol{C}_{\tilde{w}}^{(i)} + \lambda \left. \frac{\partial Q(\Theta^{(i)}, \Theta)}{\partial \boldsymbol{C}_{\tilde{w}}} \right|_{\boldsymbol{C}_{\tilde{w}} = \boldsymbol{C}_{\tilde{w}}^{(i)}}$$

$$= \boldsymbol{C}_{\tilde{w}}^{(i)} + \frac{\lambda}{2} N \int_0^{+\infty} f_z(z)\,(1-z) \boldsymbol{C}_z^{-1} (\boldsymbol{I} - \widehat{\boldsymbol{C}_z} \boldsymbol{C}_z^{-1}) \mathrm{dz}, \tag{24}$$

where

$$\boldsymbol{C}_z = z \boldsymbol{C}_{\tilde{y}} + (1-z) \boldsymbol{C}_{\tilde{w}}, \tag{25}$$

$$\widehat{\boldsymbol{C}_z} = \frac{\sum_{\boldsymbol{p}} f_{z|\tilde{y}} \left(z | \tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)} \right) \tilde{\boldsymbol{y}}(\boldsymbol{p}) \tilde{\boldsymbol{y}}^T(\boldsymbol{p})}{\sum_{\boldsymbol{p}} f_z \left(z | \tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)} \right)}. \tag{26}$$

Although a good fitting to the data was reported in Portilla (2004), the technique requires the relatively costly evaluation of the expected log-likelihood function (23). Another issue is the choice of the constant z_0. In Portilla (2004), this was solved by using a discrete GSM mixture for the hidden multiplier density $f_z(z)$. By assigning a non-zero probability mass at $z = 0$, the probability $\mathrm{P}\left(z = 0 | \tilde{\boldsymbol{y}}(\boldsymbol{p}), \Theta^{(i)} \right)$ is guaranteed to be non-zero.

5.2 Constrained EM algorithm using augmented Lagrangian optimization

In this subsection, we present a novel, alternative estimation method that does not need evaluation of the expected log-likelihood function. First, we assume a discrete hidden multiplier density $\mathrm{P}(z = z_k) = \alpha_k$, with $k = 1, ..., K$. The parameters can be initialized in a manner similar to Portilla et al. (2003):[4]

$$\begin{aligned} z_k &= \exp\left(-3 + 7(k-1)/(K-1)\right), \quad k = 1, ..., K \\ \alpha_k &= 1/K. \end{aligned} \tag{27}$$

In contrast to the GEM algorithm, where $\boldsymbol{C}_{\tilde{w}}$ is optimized directly, we take a slightly different approach. We rely on the fact that the density $f_{\tilde{\boldsymbol{y}}}(\tilde{\boldsymbol{y}})$ corresponds to a Gaussian mixture model. This allows us to use the EM algorithm for Gaussian mixtures, with some modifications that we will describe next. Let us denote by $\boldsymbol{C}_k$ the covariance matrices of the mixture components. Because of (20), the mixture covariance matrices should be subject to the constraint $z_k \boldsymbol{C}_{\tilde{u}} + \boldsymbol{C}_{\tilde{w}} = \boldsymbol{C}_k$. Our method now consists of optimizing the expected log-likelihood function (as in a regular EM algorithm Dempster et al. (1977)), but now subject to the GSM constraint:

$$\max_{\Theta} \mathrm{E}\left[\log f_{z|\tilde{\boldsymbol{y}}}(z | \tilde{\boldsymbol{y}}, \Theta) \,|\tilde{\boldsymbol{y}},\ \Theta^{(i)} \right] \quad \text{s.t.} \quad z_k \boldsymbol{C}_{\tilde{u}} + \boldsymbol{C}_{\tilde{w}} = \boldsymbol{C}_k \tag{28}$$

To solve this constrained problem, we use the augmented Lagrangian (AL) method. In the AL method, a constrained problem is translated to an unconstrained problem with a Lagrange

[4] Here, values $z_{\min}$ and $z_{\max}$ from (Portilla et al., 2003, p. 1343) are slighly modified to have a good sampling of the continuous pdf $f_z(z)$ with a small number of components K (for example, $K = 6$).

multiplier and an extra penalty term. In our case, the unconstrained problem is given by:

$$\max_{\Theta} \mathrm{E}\left[\log f_{z|\tilde{\boldsymbol{y}}}\left(z|\tilde{\boldsymbol{y}},\Theta\right)|\tilde{\boldsymbol{y}},\ \Theta^{(i)}\right] - 2\sum_{k=1}^{K}\mathrm{Vec}\left[\boldsymbol{a}_k\right]^T\mathrm{Vec}\left[\boldsymbol{C}_k\text{-}z_k\boldsymbol{C}_{\tilde{x}}\text{-}\boldsymbol{C}_{\tilde{w}}\right] - \sum_{k=1}^{K}\lambda_k\left\|\boldsymbol{C}_k\text{-}z_k\boldsymbol{C}_{\tilde{x}}\text{-}\boldsymbol{C}_{\tilde{w}}\right\|_F^2 \tag{29}$$

where $\boldsymbol{a}_k, k = 1, ..., K$ are $d \times d$ matrices of Lagrange multipliers, λ_k are penalty factors, $\mathrm{Vec}\left[\cdot\right]$ converts a matrix to a column vector (e.g., using column stacking) and $\|\cdot\|_F$ is the matrix Frobenius norm. Taking the derivatives of (29) with respect to $\boldsymbol{C}_x$ and $\boldsymbol{C}_w$ and setting to zero leads to a linear system of equations, in block matrix form:

$$\begin{pmatrix} \mu_2\boldsymbol{I} & \mu_1\boldsymbol{I} \\ \mu_1\boldsymbol{I} & \boldsymbol{I} \end{pmatrix}\begin{pmatrix} \boldsymbol{C}_{\tilde{x}}^{(i+1)} \\ \boldsymbol{C}_{\tilde{w}}^{(i+1)} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{K} z_k\left(\lambda_k\boldsymbol{C}_k^{(i)} + \boldsymbol{a}_k\right) \\ \sum_{k=1}^{K}\left(\lambda_k\boldsymbol{C}_k^{(i)} + \boldsymbol{a}_k\right) \end{pmatrix} \tag{30}$$

with $\mu_1 = \sum_{k=1}^{K}\lambda_k z_k$ and $\mu_2 = \sum_{k=1}^{K}\lambda_k z_k^2$. Similarly, maximizing (29) with respect to $\boldsymbol{C}_k$ leads to the following update equation:

$$\boldsymbol{C}_k^{(i+1)} = \frac{\sum_{\boldsymbol{p}}\mathrm{P}\left(z = z_k|\tilde{\boldsymbol{y}}(\boldsymbol{p}),\Theta^{(i)}\right)\tilde{\boldsymbol{y}}(\boldsymbol{p})\tilde{\boldsymbol{y}}^T(\boldsymbol{p}) - 2\lambda_k\left(z_k\boldsymbol{C}_{\tilde{x}}^{(i)} + \boldsymbol{C}_{\tilde{w}}^{(i)} - \boldsymbol{a}_k\right)}{\sum_{\boldsymbol{p}}\mathrm{P}\left(z = z_k|\tilde{\boldsymbol{y}}(\boldsymbol{p}),\Theta^{(i)}\right) - 2\lambda_k}, k = 1, ..., K \tag{31}$$

Additionally, the Lagrange multipliers are updated in every iteration:

$$\boldsymbol{a}_k^{(i+1)} = \boldsymbol{a}_k^{(i)} + \frac{\lambda_k}{2}\left(\boldsymbol{C}_k^{(i+1)} - z_k\boldsymbol{C}_{\tilde{x}}^{(i+1)} - \boldsymbol{C}_{\tilde{w}}^{(i+1)}\right). \tag{32}$$

This process is repeated iteratively until a given convergence criterion has been reached (for example $\left\|\boldsymbol{C}_{\tilde{w}}^{(i+1)} - \boldsymbol{C}_{\tilde{w}}^{(i)}\right\|_F < \epsilon$, with ϵ a small positive number). The penalty weights λ_k are chosen in order to speed up the convergence of the algorithm. In our method, we choose λ_k inversely proportional to z_k: $\lambda_k = z_1/z_k$, with $z_1 < z_2 < \cdots < z_K$. The complete algorithm is summarized in Algorithm 2.

Important to mention is that the above algorithm may fail, if the matrix in the update formula (30) is singular, i.e. if $\mu_1^2 = \mu_2$. It is worthful to note that the kurtosis of the wavelet subband coefficients is given by $3\mu_2/\mu_1^2 - 3$ and becomes zero if $\mu_1^2 = \mu_2$. In this case, the probability density function $f_{\tilde{\boldsymbol{y}}}\left(\tilde{\boldsymbol{y}}\right)$ is Gaussian, and every component of the GSM model will have the same hidden multiplier value $z_k = \mu_1$, such that also $f_{\tilde{\boldsymbol{x}}}\left(\tilde{\boldsymbol{x}}\right)$ is Gaussian. Consequently, it becomes impossible to separate the signal from the noise: the highly kurtotic behavior of the noise-free coefficients $\boldsymbol{x}$ can not be exploited. By our specific initialization (27), we actually avoided the latter problem.

The elegance of this algorithm lies in the fact that simple update formulas are being used and that the complete algorithm is guaranteed to converge (albeit to a local maximum of the objective function, as with nearly all EM type of algorithms).

6. Estimation of a parametric noise PSD

In the previous Section, two methods were presented to estimate the noise covariance matrix in the wavelet domain. Although these covariance matrices can be directly used in, e.g., blind

Algorithm 2 Constrained EM algorithm for estimating the noise covariance matrix $\boldsymbol{C}_{\tilde{w}}$.

$\boldsymbol{C}_{\tilde{y}} = \frac{1}{N}\sum_{\boldsymbol{p}} \tilde{\boldsymbol{y}}(\boldsymbol{p})\tilde{\boldsymbol{y}}^T(\boldsymbol{p})$, $\boldsymbol{C}_{\tilde{w}}^{(0)} = \frac{9}{10}\boldsymbol{C}_{\tilde{y}}$, $\boldsymbol{C}_{\tilde{x}}^{(0)} = 0.1\boldsymbol{C}_{\tilde{y}}$, $\boldsymbol{C}_k^{(0)} = z_k\boldsymbol{C}_{\tilde{x}}^{(0)} + \boldsymbol{C}_{\tilde{w}}^{(0)}$, $\alpha_k^{(0)} = \frac{1}{K}$, $\lambda_k = \frac{z_1}{z_k}$.

repeat

$$\hat{\alpha}_k^{(i+1)} = \frac{1}{N}\sum_{\boldsymbol{p}} \mathrm{P}\left(z = z_k | \boldsymbol{y}(\boldsymbol{p}), \Theta\right), \quad \textbf{for } k = 1, ..., K$$

$$\boldsymbol{C}_k^{(i+1)} = \frac{\sum_{\boldsymbol{p}} \mathrm{P}\left(z=z_k|\tilde{\boldsymbol{y}}(\boldsymbol{p}),\Theta^{(i)}\right)\tilde{\boldsymbol{y}}(\boldsymbol{p})\tilde{\boldsymbol{y}}^T(\boldsymbol{p}) - 2\lambda_k\left(z_k\boldsymbol{C}_{\tilde{x}}^{(i)} + \boldsymbol{C}_{\tilde{w}}^{(i)} - \boldsymbol{a}_k^{(i)}\right)}{\sum_{\boldsymbol{p}} \mathrm{P}\left(z=z_k|\tilde{\boldsymbol{y}}(\boldsymbol{p}),\Theta^{(i)}\right) - 2\lambda_k}, \quad \textbf{for } k = 1, ..., K$$

$$\begin{pmatrix} \boldsymbol{C}_{\tilde{x}}^{(i+1)} \\ \boldsymbol{C}_{\tilde{w}}^{(i+1)} \end{pmatrix} = \frac{1}{\mu_2 - \mu_1^2} \begin{pmatrix} \mathbf{I} & -\mu_1\mathbf{I} \\ -\mu_1\mathbf{I} & \mu_2\mathbf{I} \end{pmatrix} \begin{pmatrix} \sum_{k=1}^K z_k\left(\lambda_k\boldsymbol{C}_k^{(i+1)} + \boldsymbol{a}_k^{(i)}\right) \\ \sum_{k=1}^K \left(\lambda_k\boldsymbol{C}_k^{(i+1)} + \boldsymbol{a}_k^{(i)}\right) \end{pmatrix}$$

$$\boldsymbol{a}_k^{(i+1)} = \boldsymbol{a}_k^{(i)} + \frac{\lambda_k}{2}\left(\boldsymbol{C}_k^{(i+1)} - z_k\boldsymbol{C}_{\tilde{x}}^{(i+1)} - \boldsymbol{C}_{\tilde{w}}^{(i+1)}\right) \quad \textbf{for } k = 1, ..., K$$

$i \leftarrow i + 1$

until convergence ($\left\|\boldsymbol{C}_{\tilde{w}}^{(i+1)} - \boldsymbol{C}_{\tilde{w}}^{(i)}\right\|_F < \epsilon$).

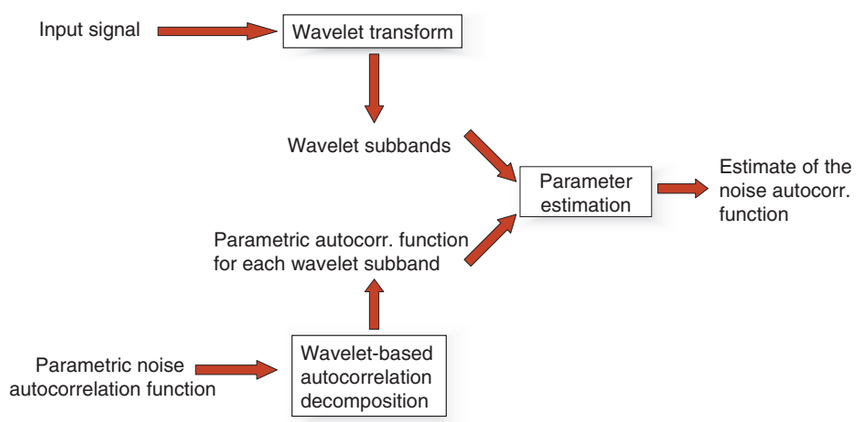


Fig. 11. Overview of the proposed algorithm for the estimation of a parametric noise PSD.

denoising approaches (see Portilla (2004)), the covariance matrices are not directly related to the noise PSD (in the sense that, after estimation of the covariances matrices the noise PSD is still *unknown*). We here present a novel approach to estimate the parameters of a parametric noise PSD based on the covariance matrix estimation methods. As far as the authors are aware of, such a technique does not yet exist. This approach also combines all the different techniques discussed in this Chapter. An overview of our algorithm is given in Figure 11. First, the noise is assumed to have a PSD with an unkown set of parameters $\boldsymbol{\beta}$. Consequently, by the Wiener-Khinchin theorem, the noise autocorrelation function $R_{w,\boldsymbol{\beta}}(\boldsymbol{p})$ is known. The wavelet-domain noise autocorrelation functions can be computed from $R_{w,\boldsymbol{\beta}}(\boldsymbol{p})$, as explained in Section 3. Using the formula (19), the parametric wavelet domain noise covariance matrix

$\boldsymbol{C}_{\tilde{w}}(\boldsymbol{\beta})$ can be found. Defining $\boldsymbol{R}_w(\boldsymbol{\beta}) = \left[R_{w,\boldsymbol{\beta}}(\boldsymbol{p})\right]$, the noise covariance matrix can be expressed in terms of $\boldsymbol{R}_w(\boldsymbol{\beta})$ by using a matrix multiplication:

$$\boldsymbol{C}_{\tilde{w}}(\boldsymbol{\beta}) = \boldsymbol{Q}\boldsymbol{R}_w(\boldsymbol{\beta}) \tag{33}$$

Then, the parameter $\boldsymbol{\beta}$ can be estimated iteratively in every iteration of Algorithm 2. Therefore, we minimize the squared matrix Frobenius norm:

$$\boldsymbol{\beta}^{(i+1)} = \arg\min_{\boldsymbol{\beta}} \left\| \boldsymbol{C}_{\tilde{w}}^{(i+1)} - \boldsymbol{Q}\boldsymbol{R}_w(\boldsymbol{\beta}) \right\|_F^2. \tag{34}$$

Because $\boldsymbol{R}_w(\boldsymbol{\beta})$ is not a linear function in general, this is a non-linear optimization problem, which can be solved using gradient descent or Gauss-Newton techniques. The gradient descent step is given by:

$$\boldsymbol{\beta}^{(i+1)} = \boldsymbol{\beta}^{(i)} + \gamma \left(\boldsymbol{C}_{\tilde{w}}^{(i+1)} - \boldsymbol{Q}\boldsymbol{R}_w\left(\boldsymbol{\beta}^{(i)}\right) \right)^T \boldsymbol{Q} \left| \frac{\partial \boldsymbol{R}_w}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(i)}}. \tag{35}$$

Note that in practice this equation may be iterated several times until convergence in an inner iteration, before the other model parameters are updated. As an example, consider the autocorrelation function from Figure 8, corresponding to the PSD $P(\omega) = \beta \sin(\beta|\omega|) I[\beta|\omega| < \pi]$, with $I[\cdot]$ the indicator function. Application of the inverse DTFT gives the spatial autocorrelation function $R_{w,\beta}(n) = \beta^2 \left(1 + \cos\left(\frac{\pi n}{\beta}\right)\right) / \left(\pi(n^2 - \beta^2)\right)$. Its derivative with respect to β is given by:

$$\frac{\partial R_{w,\beta}(n)}{\partial \beta} = \frac{n}{\pi(n^2-\beta^2)^2} \left(\sin\left(\frac{\pi n}{\beta}\right) \pi\left(n^2 - \beta^2\right) + 2\beta n \left(1 + \cos\left(\frac{\pi n}{\beta}\right)\right)\right). \tag{36}$$

Substitution of (36) into (35) then gives the desired update step.
An interesting special case is the estimation of white Gaussian noise, with autocorrelation function $R_{w,\beta}(n) = s\delta(n)$, with s the unknown noise variance. In this case, (34) comprises a least-squares problem, with a linear solution.

7. Experimental results

In this Section, we will compare the performances of the noise estimation methods from Section 5. For this task, both iterative algorithms (the GEM algorithm and the constrained EM algorithm), are initialized using the same set of parameters. The initial values used are given in Algorithm 2 and in (27). The number of mixture components used is 6: $K = 6$. Five images (Barbara, Baboon, Lena, Boats and Peppers) are transformed to the wavelet domain, using the Daubechies wavelet with two vanishing moments. Artificial Gaussian noise with a known (ground-truth) autocorrelation function is added to each LH_1-subband, which allows us to compute the estimation error afterwards. This ground-truth noise autocorrelation function is given by: $\sigma^2\beta^4 \left(1 + \cos\left(\frac{\pi x}{\beta}\right)\right)\left(1 + \cos\left(\frac{\pi y}{\beta}\right)\right) / \left(\pi(x^2 - \beta^2)(y^2 - \beta^2)\right)$, with $\beta = 3/2$ and with $\sigma \in \{1, 5, 10, 15, 25, 50\}$. Then, after every iteration of both algorithms, the log-likelihood function $\log f_{\tilde{\boldsymbol{y}}|\Theta}(\tilde{\boldsymbol{y}}|\Theta)$ and the quadratic error $\left\| \widehat{\boldsymbol{C}_{\tilde{w}}} - \boldsymbol{C}_{\tilde{w}} \right\|_F^2$ are computed, which allows us to compare the performances of both algorithms as function of the iteration number i. Both

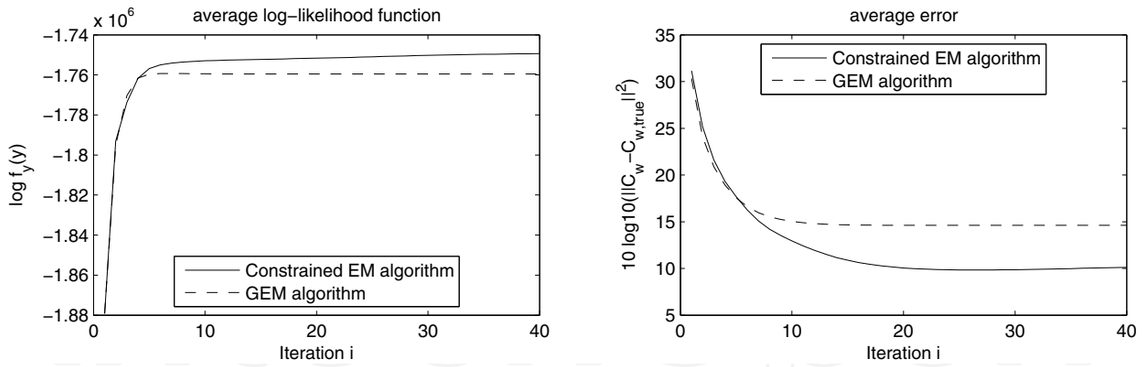


Fig. 12. Comparison of the performance of the GEM algorithm Portilla (2004) and the constrained EM algorithm Subsection 5.2, as a function the iteration number i. Results are averaged over 5 images and 6 noise levels. (*left*) average log-likelihood $\log f_{\tilde{\boldsymbol{y}}|\Theta}\left(\tilde{\boldsymbol{y}}|\Theta\right)$, (*right*) average estimation error in logarithmic scale $10\log_{10}\left(\left\|\widehat{\boldsymbol{C}_{\tilde{w}}}-\boldsymbol{C}_{\tilde{w}}\right\|_F^2\right)$.

Table 1. Comparison of the performance of the GEM algorithm Portilla (2004) and the constrained EM algorithm (CEM) from Subsection 5.2, for 5 images and 6 noise levels. Shown is the estimation error in logarithmic scale $10\log_{10}\left(\left\|\widehat{\boldsymbol{C}_{\tilde{w}}}-\boldsymbol{C}_{\tilde{w}}\right\|_F^2\right)$ after 40 iterations.

	$\sigma=1$		$\sigma=5$		$\sigma=10$		$\sigma=15$		$\sigma=25$		$\sigma=50$	
Image	CEM	GEM	CEM	GEM	CEM	GEM	CEM	GEM	CEM	GEM	CEM	GEM
Barbara	**14.25**	14.56	**-12.98**	-11.78	**-26.63**	-23.21	**-29.29**	-28.08	**-36.50**	-31.83	**-37.93**	-35.50
Baboon	**24.02**	28.74	**-3.79**	2.13	**-14.84**	-7.23	**-21.88**	-12.82	**-25.60**	-18.60	**-30.75**	-28.38
Lena	**9.56**	14.42	**-16.77**	-12.23	-23.25	**-23.34**	-29.54	**-29.68**	**-37.99**	-37.17	-38.38	**-38.93**
Boats	**7.72**	9.66	**-17.77**	-16.55	**-30.35**	-26.31	**-30.09**	-28.59	**-35.83**	-32.65	**-37.90**	-37.06
Peppers	**11.28**	17.06	**-14.34**	-9.92	**-24.71**	-21.25	**-30.10**	-27.86	-31.84	**-34.05**	**-40.29**	-36.51
Average	**13.37**	16.89	**-13.13**	-9.67	**-23.96**	-20.27	**-28.18**	-25.41	**-33.55**	-30.86	**-37.05**	-35.28

algorithms maximize the log-likelihood function, note however that this does not necessarily results in minimizing the quadratic error. The results are shown in Figure 12 and Table 1. It can be seen that while the GEM algorithm converges to its final value, on average the constrained EM algorithm is able to reach a solution with a higher log-likelihood function and a lower error. We remark that the objective function is non-convex, such that both algorithms can get trapped in local maxima. Although both algorithms use the same initialization, in most of the experiments (see Table 1) the constrained EM gives a more accurate estimate of the noise covariance matrix.

In Figure 13 and Figure 14, we used the noise estimation method based on the constrained EM algorithm in combination with the BLS-GSM Portilla et al. (2003) denoising method, in order to perform blind noise removal. An undecimated wavelet transform of 3 levels with the Daubechies wavelet with eight vanishing moments was used. The PSD of the Gaussian noise is in the captions of Figure 13 and Figure 14. Clearly, the combined method is well able to distinguish signal information from noise information, leading to a succesful removal of the noise while preserving signal structures.

(a) Original image (b) With artificial noise (PSNR=20.17dB) (c) Denoised (PSNR=41.21dB)

Fig. 13. Blind denoising results (using the BLS-GSM denoising method and the proposed constrained EM noise estimation technique). Noise PSD $P(\boldsymbol{\omega}) \sim \exp(-4000((\omega_x/\pi - 0.1)^2 + (\omega_y/\pi - 0.12)^2))$.

(a) Original image (b) With artificial noise (PSNR=17.25dB) (c) Denoised (PSNR=29.00dB)

Fig. 14. Blind denoising results (using the BLS-GSM denoising method and the proposed constrained EM noise estimation technique). Noise PSD $P(\boldsymbol{\omega}) \sim \exp(-2000((\omega_x/\pi - 0.1)^2 + (\omega_y/\pi - 0.12)^2)) + \exp(-3000((\omega_x/\pi + 0.15)^2 + (\omega_y/\pi - 0.22)^2)) + 10^{-3}$.

8. Conclusion

In this chapter, we investigated the estimation of stationary colored noise, which is most efficiently described in a Fourier basis using the power spectral density (PSD). Because of the time or spatial locality of signal structures, estimation of colored noise is best performed in a transform domain that allows to adapt to the signal locality. We have shown that wavelets are very good candidates for this task: their vanishing moment properties allow us to complete suppress smoothly varying signals, such that efficient noise estimation can directly be performed on a single wavelet subband. However, in practice, signals are not smoothly varying and may contain transitions (such as edges and textures in images). To take

this into account, we have presented several prior models for noise-free wavelet coefficients. These prior models are then used in an expectation-maximization algorithm, which gives us an estimate of the noise covariance matrix for a given wavelet subband. We have further shown how this covariance matrix is related to the noise autocorrelation function in spatial or time domain. This relationship can then be used, e.g., to estimate parameters of parametric PSDs, yielding reliable and accurate estimates for noise PSDs. Because noise is present in most real-life signals and images, many signal and image processing methods can be further improved by taking advantage of estimated noise characteristics using techniques as described in this chapter.

9. References

Abramovich, F., Sapatinas, T. & Silverman, B. (1998). Wavelet thresholding via a Bayesian approach, *J. of the Royal Statist. Society B* 60: 725–749.

Achim, A., Bezerianos, A. & Tsakalides, P. (2001). Wavelet-based ultrasound image denoising using an alpha-stable prior probability model, *Proc. International Conference on Image Processing*, Vol. 2, pp. 221–224.

Aelterman, J., Deblaere, K., Goossens, B., Pižurica, A. & Philips, W. (2010). Dual Tree Complex Wavelet-Based Denoising of correlated noise in 3D Magnetic Resonance Imaging. Under revision.

Aelterman, J., Goossens, B., Pižurica, A. & Philips, W. (2010). *Recent Advances in Signal Processing*, IN-TECH, chapter Suppression of Correlated Noise.

Andrews, D. & Mallows, C. (1974). Scale mixtures of normal distributions, *J. Royal Stat. Stoc.* 36: 99–102.

Antonini, M., Barlaud, M., Mathieu, P. & Daubechies, I. (1992). Image coding using wavelet transform., *IEEE Trans. Image Process.* 1(2): 205–220.

Baher, H. (2001). *Analog and Digital Signal Processing*, Wiley, Chichester.

Bayer, B. (1976). Color imaging array, United States Patent 3971065.

Borel, C., Cooke, B. & Laubscher, B. (1996). Partial Removal of Correlated noise in Thermal Imagery, *Proceedings of SPIE*, Vol. 2759, pp. 131–138.

Campbell, N. A., Lopuhaä, H. P. & Rousseeuw, P. J. (1998). On the calculation of a robust s-estimator of a covariance matrix., *Stat Med* 17(23): 2685–2695.

Candès, E. (1998). *Ridgelets: Theory and Applications*, PhD thesis, Departement of Statistics, Stanford University.

Candès, E., Demanet, L., Donoho, D. & Ying, L. (2006). Fast Discrete Curvelet Transforms, *Multiscale modeling and simulation* 5(3): 861–899.

Chang, S. G., Yu, B. & Vetterli, M. (1998). Spatially adaptive wavelet thresholding with context modeling for image denoising, *Proc. IEEE Internat. Conf. on Image Proc.*, Chicago, IL, USA.

Clyde, M., Parmigiani, G. & Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets, *Biometrika* 85(2): 391–401.

Crouse, M. S., Nowak, R. D. & Baranuik, R. G. (1998). Wavelet-based statistical signal processing using hidden Markov models, *IEEE Trans. Signal Proc.* 46(4): 886–902.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* 19(1): 1–38.

Do, M. N. & Vetterli, M. (2003). The finite ridgelet transform for image representation, *IEEE Trans. Image Processing* 12(1): 16–28.

Do, M. N. & Vetterli, M. (2005). The contourlet transform: An efficient directional multiresolution image representation, *IEEE Trans. Image Process.* 14(12): 2091–2106.

Donoho, D. L. & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinking, *Journal of the American Statistical Association* 90(432): 1200–1224.

Fadili, J. M. & Boubchir, L. (2005). Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities, *IEEE Trans. on Image Process.* 14(2): 231–240.

Fan, G. & Xia, X. (2001). Image denoising using local contextual hidden Markov model in the wavelet domain, *IEEE Signal Processing Letters* 8(5): 125–128.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Am. A* 4(12): 2379–2394.

Fischer, S., Šroubek, F., Perrinet, L., Redondo, R. & Cristobal, G. (2007). Self-Invertible 2D Log-Gabor Wavelets, *International Journal of Computer Vision* 75(2): 231–246.

Gómez, E., Gómez-Villegas, M. A. & Marín, J. M. (2008). Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications, *Communications in Statistics - Theory and Methods* 37(6): 972–985.

Goossens, B., Aelterman, J., Pižurica, A. & Philips, W. (2010). A Recursive Scheme for Computing Autocovariance functions of complex wavelet subbands, *IEEE Trans. Signal Processing* 58(7): 3907–3912.

Guo, K. & Labate, D. (2007). Optimally Sparse Multidimensional Representation using Shearlets, *SIAM J Math. Anal.* 39: 298–318.

Johnstone, I. M. & Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise, *Journal of the Royal Statistical Society B* 59(2): 319–351.

Kingsbury, N. G. (2001). Complex wavelets for shift invariant analysis and filtering of signals, *Journal of Applied and Computational Harmonic Analysis* 10(3): 234–253.

Kotz, S., Kozubowski, T. J. & Podgorski, K. (2000). An asymmetric multivariate laplace distribution, *Computational Statistics* 4: 531–540.

Kotz, S. & Kozubowski, T.and Podgorski, K. (2001). *The Laplace Distributions And Generalizations: A Revisit with Applications to Communications, Economics, Engineering, Finance*, Birkhäuser, Boston.

Kwon, O., Sohn, K. & Lee, C. (2003). Deinterlacing using Directional Interpolation and Motion Compensation, *IEEE Trans. Consumer Electronics* 49(1): 198–203.

Lee, T. (1996). Image Representation Using 2D Gabor Wavelets, *IEEE Trans. Pattern Analysis and Machine Intelligence* 18(10): 1.

Mallat, S. (1989). Multifrequency channel decomposition of images and wavelet models, *IEEE Trans. Acoust., Speech, Signal Proc.* 37(12): 2091–2110.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing*, Academic Press.

Moulin, P. & Liu, J. (1999). Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors, *IEEE Trans. Info. Theory, Special Issue on Multiscale Analysis* 3(3): 909–919.

Nikias, C. L. & Shao, M. (1995). *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley-Interscience.

Nowak, R. (1999). Wavelet-based rician noise removal for magnetic resonance imaging., *IEEE Trans Image Process* 8(10): 1408–1419.

Pena, D. & Prieto, F. (2001). Multivariate outlier detection and robust covariance matrix estimation, *Technometrics* 43(3): 286–310.

Pižurica, A. & Philips, W. (2006). Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising., *IEEE Trans. Image Process.* 15(3): 654–665.

Pižurica, A., Philips, W., Lemahieu, I. & Acheroy, M. (2003). A versatile wavelet domain noise filtration technique for medical imaging, *IEEE Trans. Medical Imaging* 22(3): 323–331.

Portilla, J. (2004). Full blind denoising through noise covariance estimation using Gaussian Scale Mixtures in the wavelet domain, *IEEE Int. Conf. on Image Process. (ICIP)* 2: 1217–1220.

Portilla, J. & Simoncelli, E. (2001). Adaptive Wiener Denoising using a Gaussian Scale Mixture Model in the Wavelet Domain, *IEEE Int. Conf. on Image Process. (ICIP)* 2: 37–40.

Portilla, J., Strela, V., Wainwright, M. & Simoncelli, E. (2003). Image denoising using scale mixtures of gaussians in the wavelet domain, *IEEE Transactions on image processing* 12(11): 1338–1351.

Rabbani, H., Vafadust, M., Gazor, S. & Selesnick, I. W. (2006). Image Denoising Employing a Bivariate Cauchy Distribution with Local Variance in Complex Wavelet Domain, *12th Digital Signal Processing Workshop - 4th Signal Processing Education Workshop*, pp. 203–208.

Romberg, J., Choi, H. & Baraniuk, R. G. (2001). Bayesian tree structured image modeling using wavelet-domain Hidden Markov Models, *IEEE Trans. Image Process.* 10(7): 1056–1068. Enter text here.

Selesnick, I. W. (2006). Laplace random vectors, Gaussian noise, and the generalized incomplete Gamma function, *Proc. IEEE Int. Conf. on Image Process.*, pp. 2097–2100.

Selesnick, I. W., Baraniuk, R. G. & Kingsbury, N. G. (2005a). The Dual-Tree Complex Wavelet Transform, *IEEE Signal Processing Magazine* 22(6): 123–151.

Selesnick, I. W., Baraniuk, R. G. & Kingsbury, N. G. (2005b). The Dual-Tree Complex Wavelet Transform, *IEEE Signal Processing Magazine* 22(6): 123–151.

Shi, F. & Selesnick, I. W. (2006). Multivariate Quasi-Laplacian Mixture Models for Wavelet-based Image Denoising, *Proc. Int. Conf. on Image Processing (ICIP)*, pp. 2097–2100.

Simoncelli, E., Freeman, W. T., Adelson, E. H. & Heeger, D. J. (1992). Shiftable Multi-scale Transforms, *IEEE Trans. Information Theory* 38(2): 587–607.

Simoncelli, E. P. & Adelson, E. H. (1996). Noise removal via Bayesian wavelet coring, *Proc. IEEE Internat. Conf. Image Proc. ICIP*, Lausanne, Switzerland.

Srivastava, A., Lee, A. B., Simoncelli, E. & Zhu, S.-C. (2003). On Advances in Statistical Modeling of Natural Images, *Journal of Mathematical Imaging and Vision* 18: 17–33.

Srivastava, A., Liu, X. & Grenander, U. (2002). Universal Analytical Forms for Modeling Image Probabilities, *IEEE Trans. Pattern Analysis and Machine Intelligence* 24(9): 1200–1214.

Tzikas, D., Likas, A. & Galatsanos, N. (2007). Variational bayesian blind image deconvolution with student-t priors, *Proc. IEEE International Conference on Image Processing ICIP 2007*, Vol. 1, pp. I–109–I–112.

Van De Ville, D. & Unser, M. (2008). Complex Wavelet Bases, Steerability, and the Marr-like pyramid, *IEEE Trans. Image Processing* 17(11): 2063–2080.

Vo, A., Nguyen, T. & Oraintara, S. (2007). Image denoising using shiftable directional pyramid and scale mixtures of complex gaussians, *Proc. IEEE International Symposium on Circuits and Systems ISCAS 2007*, pp. 4000–4003.

Wainwright, M. J. & Simoncelli, E. P. (2000). Scale mixtures of Gaussians and the statistics of natural images, *Adv. Neural Information Processing Systems (NIPS 1999)* 12: 855–861.

Discrete Wavelet Transforms - Algorithms and Applications

Edited by Prof. Hannu Olkkonen

ISBN 978-953-307-482-5
Hard cover, 296 pages
Publisher InTech
Published online 29, August, 2011
Published in print edition August, 2011

The discrete wavelet transform (DWT) algorithms have a firm position in processing of signals in several areas of research and industry. As DWT provides both octave-scale frequency and spatial timing of the analyzed signal, it is constantly used to solve and treat more and more advanced problems. The present book: Discrete Wavelet Transforms: Algorithms and Applications reviews the recent progress in discrete wavelet transform algorithms and applications. The book covers a wide range of methods (e.g. lifting, shift invariance, multi-scale analysis) for constructing DWTs. The book chapters are organized into four major parts. Part I describes the progress in hardware implementations of the DWT algorithms. Applications include multitone modulation for ADSL and equalization techniques, a scalable architecture for FPGA-implementation, lifting based algorithm for VLSI implementation, comparison between DWT and FFT based OFDM and modified SPIHT codec. Part II addresses image processing algorithms such as multiresolution approach for edge detection, low bit rate image compression, low complexity implementation of CQF wavelets and compression of multi-component images. Part III focuses watermaking DWT algorithms. Finally, Part IV describes shift invariant DWTs, DC lossless property, DWT based analysis and estimation of colored noise and an application of the wavelet Galerkin method. The chapters of the present book consist of both tutorial and highly advanced material. Therefore, the book is intended to be a reference text for graduate students and researchers to obtain state-of-the-art knowledge on specific applications.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元

Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen

IntechOpen