POLITEHNICA UNIVERSITY OF TIMISOARA
TELECOM BRETAGNE

P H D T H E S I S

to obtain the title of

PhD of Science

of the Politehnica University of Timisoara and Telecom Bretagne, Brest

Defended by
Ioana ADAM

Complex Wavelet Transform: application to denoising

Thesis Advisors : Alexandru ISAR
Jean-Marc BOUCHER

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The introductory chapter presents the motivation of the work and ends with a short outline of the thesis.

1.1 Motivation

Wavelet theory is one of the most modern areas of mathematics. Masterfully developed by French researchers, such as Yves Meyer, Stephane Mallat and Albert Cohen, this theory, is now used as an analytical tool in most areas of technical research: mechanical, electronics, communications, computers, biology and medicine, astronomy an so on. In the field of signal and image processing, the main applications of wavelet theory are compression and denoising.

In the context of denoising, the success of techniques based on the wavelet theory is ensured by the ability of decorrelation (separation of noise and useful signal) of the different discrete wavelet transforms [FBB01, ICN02]. Because the signal is contained in a small number of coefficients of such a transform, all other coefficients essentially contain noise. By filtering these coefficients, most of the noise is eliminated. Thus, each method of image denoising based on the use of wavelets follows the classic method, in three steps: computing a discrete wavelet transform of the image to be denoised, filtering in the wavelet domain and the computation of the corresponding inverse wavelet transform.

Throughout recent years, many wavelet transforms (WT) have been used to operate denoising. The first one was the discrete wavelet transform, [DJ94]. It has three main disadvantages [Kin01]: lack of shift invariance, lack of symmetry of the mother wavelet and poor directional selectivity. These disadvantages can be diminished using a complex wavelet transform [Kin01, Kin00]. More than 20 years ago, Grossman and Morlet [GM84] developed the continuous wavelet transform [SBK05]. A revival of interest in later years has occurred in both signal processing and statistics for the use of complex wavelets [BN04], and complex analytic wavelets, particularly in [Kin99, Sel01]. It may be linked to the development of complex-valued discrete wavelet filters [LM95] and the clever dual filter bank [Kin99, SBK05]. The complex WT has been shown to provide a powerful tool in signal and image analysis [Mal99]. In [OM06], the authors derived large classes of wavelets generalizing the concept of 1-D local complex-valued analytic decomposition introducing 2-D vector-valued hyperanalytic decomposition.

The present work is situated in this context, and, by introducing a new version of the Hyperanalytic Wavelet Transform and by combining this transform with various parametric and non-parametric filtering techniques attempts to provide a solution to the denoising problem. The new transform, by allowing the use of all the mother wavelet families that are usually used with the discrete wavelet transform, while achieving the desirable properties of complex wavelet transforms, such as quasi shift-invariance and a good directional selectivity, in association with different filters selected has provided good denoising results both when applied to images affected by additive noise or by multiplicative noise, as is the case of SAR images.

1.2 Thesis outline

The current thesis is organized as follows:

- Chapter 1, *Introduction*, is made of a short presentation of the context the present work relies in, and an overview of the structure of the thesis.
- Chapter 2, *Wavelet Transforms*, introduces the discrete wavelet transform (one-dimensional and two-dimensional), the undecimated wavelet transform and the wavelet packet transform (one-dimensional and two-dimensional), explaining the concept of multiresolution analysis and the limitations of the discrete wavelet transform (shift sensitivity, reduced directional selectivity).
- Chapter 3, *Complex Wavelet Transforms*, constitutes a sequel of the previous chapter by introducing the complex wavelet transforms. The Hyperanalytic Wavelet Transform (HWT) is introduced, this representing the thesis' main contribution and a parallel is drawn between this transform and the dual-tree complex wavelet transform, DTCWT. A special attention is given to the quasi shift-invariance and to the good directional selectivity of the HWT.
- Chapter 4, *Denoising*, is a thorough presentation of denoising. The noise is considered to be additive in this case. The wavelet-based denoising techniques are emphasized. A distinction is being made between the non-parametric and parametric methods, emphasizing the second category. I have insisted on the maximum a posteriori (MAP) parametric methods. The importance of the coefficients' inter-scale dependency is highlighted. Starting from the Gaussian mixture model for characterizing different scales (GSM), the bishrink filter is introduced. The association of the HWT with the bishrink filter is studied and compared with the denoising performances of other methods.
- Chapter 5, *Speckle Reduction*, is a continuation of chapter 4 and its goal is the reduction of multiplicative speckle-type noise, affecting the SAR and SONAR images. I have developed a homomorphic denoising method based on the association of the HWT with the bishrink filter. This method is then compared with the classical despecklisation methods and with other methods proposed in the literature.
- Chapter 6, *Conclusions*, presents the conclusions drawn and future perspectives.

Chapter 2

Wavelet Transforms

2.1 Introduction

2.1.1 Wavelet Definition

The term 'wavelet' refers to an oscillatory vanishing wave with time-limited extend, which has the ability to describe the time-frequency plane, with atoms of different time supports (see fig. 2.1). Generally, wavelets are purposefully crafted to have specific properties that make them useful for signal processing. They represent a suitable tool for the analysis of non-stationary or transient phenomena.

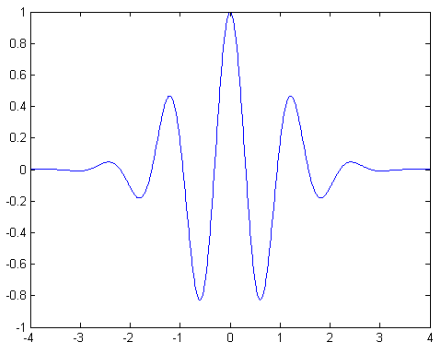


Figure 2.1: Wavelet

2.1.2 Wavelet Characteristics

Wavelets are a mathematical tool, that can be used to extract information from many kinds of data, including audio signals and images. Mathematically, the wavelet ψ, is a function of zero average, having the energy concentrated in time:

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0, \tag{2.1}$$

In order to be more flexible in extracting time and frequency informations, a family of wavelets can be constructed from a function $\psi(t)$, also known as the 'Mother Wavelet', which is confined in a finite interval. 'Daughter Wavelets', $\psi_{u,s}(t)$ are then formed by translation with a factor u and dilation with a scale parameter s:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \cdot \psi\left(\frac{t-u}{s}\right) \tag{2.2}$$

2.1.3 Wavelet Analysis

The wavelet analysis is performed by projecting the signal to be analyzed on the wavelet function. It implies a multiplication and an integration:

$$\langle x(t), \psi_{u,s}(t) \rangle = \int x(t)\, \psi_{u,s}(t)\, dt.$$

Depending on the signal characteristics that we want to analyze, we can use different scales and translations of the mother wavelet. The particularity of the wavelet analysis is that it allows us to change freely the size of the analysis function (window), to make it suitable for the needed resolution, in time or frequency domain. For high resolution in time-domain analysis we want to 'capture' all the sudden changes that appear in the signal, and we do that by using a contracted version of the mother wavelet. Conversely, for high-resolution in the frequency-domain we will be using a dilated version of the same function.

2.1.4 Wavelet History

The development of wavelets can be linked to several works in different domains, starting with the first wavelet introduced by Haar in 1909. In 1946, Denis Gabor, introduced the Gabor atoms or Gabor functions, which are functions used in analysis, a family of functions being built from translations and modulations of a generating function. In 1975, George Zweig, former particle physicist who had turned to neurobiology, has discovered the continuous wavelet transform (named first the cochlear transform and discovered while studying the reaction of the ear to sound). Morlet, studying reflection seismology observed that, instead of emitting pulses of equal duration, shorter waveforms at high frequencies should perform better in separating the returns of fine closely-spaced layers. Grossmann, who was working in theoretical physics, recognised in Morlet's approach some ideas that were close to his own work on coherent quantum states. In 1982, Grossmann and Morlet have given the formulation of the Continuous Wavelet Transform. Yves Meyer recognized the importance of this fundamental mathematical tool and developed this theory with collaborators as Ingrid Daubechies (who introduced the orthogonal wavelets with compact support (1988) [Dau88]) and Stéphane Mallat (who proposed the filter-bank implementation scheme of the Discrete Wavelet Transform).

2.1.5 Wavelet Terminology

Due to Mallat's implementation of the Wavelet transform, the filter-bank theory is closely related to the wavelet theory. Also, the concept of 'Multiresolution Analysis' (MRA) is

connected to the wavelet theory. Besides the classic wavelet transforms (Continuous Wavelet Transform and Discrete Wavelet Transform) we have more 'evolved' transforms connected with the wavelet theory, such as the Complex Wavelet Transform and the Wavelet Packets Transform.

2.2 Evolution of Wavelet Transform

For many years, classical signal processing was concentrated on the characterization of signals and on the designing of time-invariant and space-invariant operators that modify stationary signal properties. But the biggest amount of information is concentrated in the transients rather than in stationary signals. In the following, the evolution of the Wavelet transform will be described, having as departure point the Fourier Transform.

2.2.1 Fourier Transform (FT)

In the first part of the 19th century, Joseph Fourier, a French mathematician and physicist, showed that any periodic function can be decomposed in a series of simple oscillating functions, namely sines and cosines (or complex exponentials). The generalization to the non-periodic signals has come only a century later, and took the name of Fourier Transform (FT), a tribute brought to the original idea. The FT decomposes a signal in complex exponential functions at different frequencies. The equations used in the decomposition and reconstruction part are the following:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt, \tag{2.3}$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \cdot e^{j\omega t} d\omega. \tag{2.4}$$

In the above equations, t stands for time, $\omega = 2\pi f$ for frequency, x denotes the signal in the time domain and X denotes the signal in the frequency domain (also known as the spectrum of the original signal). As can be seen from eq. 2.3, the computation of the FT is done over all times, making no distinction between signals' stationary parts and transient ones (whether the frequency component 'ω' appears at time t_1 or t_2, it will have the same effect at the output of the integration). The scaling property of the FT states that if we have a scaled version of the original $x_s(t)$:

$$x_s(t) = x(st), \tag{2.5}$$

then, its corresponding FT will be $X_s(\omega)$:

$$X_s(\omega) = \frac{1}{|s|} X\left(\frac{\omega}{s}\right). \tag{2.6}$$

We can observe from the last two equations that if we reduce the time spread of x by s (if s>1) than the FT is dilated by s, meaning that if what we have gained in time localization, we have lost in frequency localization. Projecting the signal on complex exponentials leads to good frequency analysis, but no time localization. The poor time localization is the main disadvantage of the Fourier transform, making it not suitable for all kind of applications.

2.2.2 Short Time Fourier Transform (STFT)

To see how the frequency content of a signal changes over time, we can cut the signal into blocks and compute the spectrum of each block. This is the base concept of the Short Time Fourier Transform (STFT) introduced in 1946 by Gabor [Gab46], and again in 1977 by J.B. Allen [All77], the latter giving it a filterbank interpretation. For computing STFT we simply multiply the original signal by a window function, which is non-zero for only a short period of time, and then we compute the Fourier Transform of the obtained signal. The result is a two-dimensional representation of the signal, that can be mathematically written as:

$$STFT\left\{x\left(t\right)\right\} \equiv X\left(\tau,\omega\right) = \int_{-\infty}^{\infty} x\left(t\right) w\left(t-\tau\right) e^{-j\omega t} dt, \tag{2.7}$$

where $w\left(t\right)$ is the window function, commonly a Hann window or a Gaussian centered around zero, and $x\left(t\right)$ is the signal to be analyzed. This equation can be interpreted as an analysis of the signal by a sliding window in time or by a sliding bandpass filter in frequency. A particularity of this transform is the fact that the window is of constant length throughout the whole analysis process, meaning that the transform has a fixed resolution in time and frequency.

Time and frequency energy concentrations are restricted by the Heisenberg uncertainty principle. If we consider a finite energy function, $f \in \mathbf{L}^2\left(\mathbf{R}\right)$ $\left(\int \left|f\left(t\right)\right|^2 dt < \infty\right)$ and we consider it centered around zero in time and its Fourier transform, $F\left(\omega\right)$ centered around zero in frequency, then the temporal variance, σ_t^2 (given in eq. 2.8) and the frequency variance, σ_ω^2 (given in eq. 2.9) of the wave function satisfy the condition (2.10):

$$\sigma_t^2 = \frac{1}{\left\|f\right\|^2} \int_{-\infty}^{\infty} t^2 \left|f\left(t\right)\right|^2 dt, \tag{2.8}$$

$$\sigma_\omega^2 = \frac{1}{8\pi^3 \left\|f\right\|^2} \int_{-\infty}^{\infty} \omega^2 \left|F\left(\omega\right)\right|^2 d\omega, \tag{2.9}$$

$$\sigma_t^2 \sigma_\omega^2 \geq \frac{\pi}{2}. \tag{2.10}$$

By $\left\|f\right\|$ we have denoted the norm of the function f, computed as: $\sqrt{\int_{-\infty}^{\infty} \left|f\left(t\right)\right|^2 dt}$.

Depending on the time localization that is more suitable for our application, we can choose the width of the analysis window, namely a short window for a good time but poor frequency localization (suitable for signals with a high frequency content) or a wide window for good frequency localization with the price of poorer time localization.

2.2.3 Wavelet Transform (WT)

Having in mind the limitations of the Fourier Transform (poor time localization) and of the Short-Time Fourier Transform (fixed time and frequency localisation), Grossman and Morlet gave in 1984 ([GM84]) the formulation of the Continuous Wavelet Transform. Unlike the first two, who were decomposing the signal into a basis of complex exponentials, the Wavelet Transform decomposes the signal over a set of dilated and translated wavelets.

This difference confers to the WT the advantage of performing a multiresolution analysis, meaning that it processes different frequencies in a different way (in contrast with the STFT which analyses in the same way all frequencies). By using this technique, the time resolution is increased when we analyse a high frequency portion of the signal, and the frequencial localisation is increased when analysing a low-frequency part of the same signal. This type of analysis is suitable for signals that have both low-frequency components with long time duration and high-frequency components with short time duration, which is the case of most signals.

If we consider a function $x \in \mathbf{L}^2(\mathbb{R})$ and for analysis we use the mother wavelet ψ (2.1), with its scaled and translated versions in (2.2), we can write the wavelet transform of $x(t)$ at time u and scale s as:

$$Wx(u,s) = \langle x, \psi_{u,s} \rangle = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt \tag{2.11}$$

By looking at eq. 2.11 we can conclude that the Wavelet Transform can be seen as a convolution between the signal to be analyzed and the reverse function , $\frac{1}{\sqrt{s}} \psi^* \left(-\frac{t}{s} \right)$ derived from the Mother Wavelet.

2.2.4 Comparative Visualization

In the following, we will make a visual comparison of the time-frequency resolution cell for the three transforms we have previously mentioned. In fig. 2.2 is represented the Fourier Transform, and it can be observed the very good frequency localization and the non existing time localization of this transform.

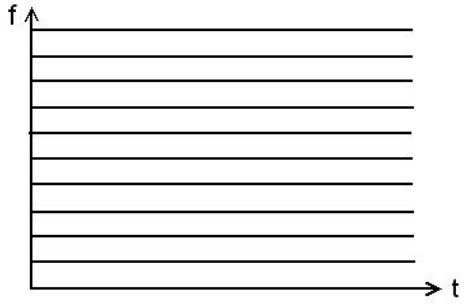
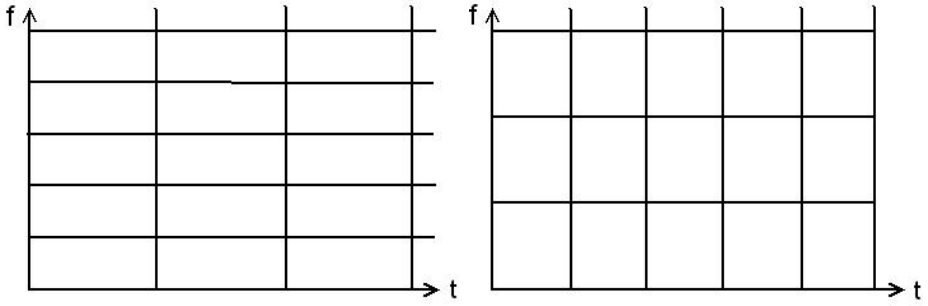


Figure 2.2: Time-frequency representation of the Fourier Transform.

Figure 2.3 presents the time-frequency localization of the Short-Time Fourier Transform. As the Heisenberg principle states, the time and frequency localization are limited to a certain bound which leads to the fact that the time-frequency atoms (the rectangles in our representation) will be of equal surfaces. In 2.3(a) we present a transform that is better localized in frequency, while in fig. 2.3(b) we have a transform with better time localization.

In the case of the Wavelet Transform, we are also limited by the Heisenberg Uncertainty, meaning that the time-frequency atoms will have the same constraints as in



(a) STFT with good frequency localization

(b) STFT with better time localization

Figure 2.3: Time-frequency representation of the Short Time Fourier Transform

the case of the STFT but, instead of using a uniform splitting of the time-frequency plane, it uses a different approach, resulting in a good frequency localization and poor time localization for low-frequencies, and reduced frequency localization with better time localization as the frequency increases (as can be seen in fig. 2.4).

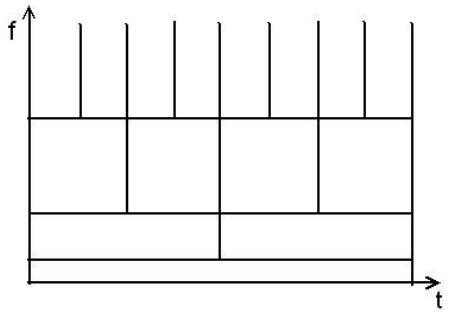


Figure 2.4: Time-frequency representation of the Wavelet Transform.

Because of this particular approach, the Wavelet Transform is suited for most signal and image applications.

2.3 Theoretical Aspects of Wavelet Transform

After we have made a brief presentation of the origins of the Wavelet Transform, we will continue with a more thorough presentation of the theoretical aspects of this transform.

2.3.1 Continuous Wavelet Transform (CoWT)

Let $x(t)$ be the finite energy signal that we want to analyze. The wavelet transform basically decomposes the signal over dilated and translated wavelets. This is why, in order to compute the continuous wavelet transform, we must choose the mother wavelet to be used (in this case denoted by $\psi(t)$). $\psi \in L^2(R)$ has a zero average, it is normalized

($\|\psi\| = 1$) and centered in the neighborhood of $t = 0$. A scaled and translated version of the mother wavelet, $\psi_{u,s}(t)$ can be written as in eq. 2.2, where s denotes the scale parameter ($s = \frac{1}{f}$, where f represents the frequency) and u is the translation parameter. The forward Continuous Wavelet Transform of the signal $x(t)$, using $\psi(t)$ as mother wavelet can be written as:

$$Wx(u,s) = \langle x, \psi_{u,s} \rangle = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt = x * \overline{\psi}_s(u) \tag{2.12}$$

and can be seen as a convolution product between the signal to be analyzed and $\overline{\psi}_s(u)$, where:

$$\overline{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi^* \left(\frac{-t}{s} \right). \tag{2.13}$$

ψ^* represents the complex conjugate of the function ψ. The Fourier transform of $\overline{\psi}_s(t)$ is

$$\overline{\Psi}_s(\omega) = \sqrt{s} \Psi^*(s\omega). \tag{2.14}$$

Because the continuous component of this function is 0, we can say that Ψ is the transfer function of a band-pass filter. Thus, we can say that the wavelet transform is being computed by filtering the original signal with a series of dilated band-pass filters. A wavelet transform is invertible if the mother wavelet satisfies a condition, called the admissibility condition, as results from the following theorem [Mal99]:

Theorem 2.3.1. *(CALDERON, GROSSMAN, MORLET) Let $\psi \in L^2(R)$ be a real function such that*

$$C_\psi = \int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < +\infty \tag{2.15}$$

Any $x \in L^2(R)$ satisfies

$$x(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wx(u,s) \frac{1}{\sqrt{s}} \psi \left(\frac{t-u}{s} \right) du \frac{ds}{s^2}, \tag{2.16}$$

and

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wx(u,s)|^2 du \frac{ds}{s^2}. \tag{2.17}$$

This theorem also states that the wavelet transform preserves the energy of the original signal (eq. 2.17), if the admissibility condition (eq. 2.15) is fulfilled.

As the Continuous Wavelet Transform is computed for a large number of values both for the scale and for the translation, we can conclude that it is a very redundant transform.

2.3.2 Discrete Wavelet Transform (DWT)

Wavelet Frames

Because the CoWT is very redundant, a discretization of the scale and translation variables was introduced. This version of the CoWT is named the '(Continuous Time)

Wavelet Series (WS)' by some authors ([XPZP96], [You93]) but also the 'Discrete Wavelet Transform', and in particular 'Wavelet Frames' by others ([Mal99], [Dau92]).

A wavelet transform that uses the frames, yields a countable set of coefficients in the transform domain. The coefficients correspond to points on a two-dimensional grid or lattice of discrete points in the scale-translation domain. This lattice will be indexed by two integers: the first integer, j, will correspond to discrete scale steps while the second integer, n, corresponds to discrete translation steps (the grid is indexed by j and n). The dilation parameter, s is now $s = s_0^j$ and the translation, u, is now $u = nu_0 s_0^j$, where s_0 and u_0 are the discrete scale and translation steps, respectively,

$$\psi_{j,n}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - nu_0 s_0^j}{s_0^j}\right). \tag{2.18}$$

The necessary conditions imposed on ψ, s_0 and u_0 for $\psi_{j,n}, j, n \in Z^2$ to be a frame of $L^2(R)$ is to fulfill the admissibility condition given in eq. 2.15 and theorem 2.3.2, while the sufficient conditions are given by another theorem (that provides the lower and upper bound for the frame bounds A and B, depending on ψ, s_0 and u_0) introduced also by Daubechies in [Dau92].

Theorem 2.3.2. *(DAUBECHIES) If $\psi_{j,n}, j, n \in Z^2$ is a frame of $L^2(R)$ then the frame bounds satisfy*

$$A \leq \frac{C_\psi}{u_0 log_e s_0} \leq B, \tag{2.19}$$

$$\forall \omega \in R - \{0\}, A \leq \frac{1}{u_0} \sum_{j=-\infty}^{+\infty} \left|\Psi\left(s_0^j \omega\right)\right|^2 \leq B. \tag{2.20}$$

The condition (2.20) imposes that the Fourier axis is completely covered by wavelets dilated by $\left\{s_0^j\right\}_{j \in Z}$.

The WS transform is defined with respect to a continuous mother wavelet, ψ. The wavelet transform maps continuous finite energy signals to a 2-D discrete grid of coefficients, $W_\psi : L^2(R) \to l^2(Z^2)$. The WS transform of a signal $x(t)$ is:

$$Wx(j,n) = \langle x, \psi_{j,n} \rangle = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s_0^j}} \psi^*\left(\frac{t - nu_0 s_0^j}{s_0^j}\right) dt \tag{2.21}$$

These wavelet coefficients represent the original signal but, as in the continuous case, the representation is sensitive to the chosen mother wavelet. Unlike the CoWT, this transform is defined only for positive values of s_0. This constraint is not restrictive as the reflected mother wavelet (a scale of -1) can be used as the new mother wavelet and effectively cover negative scales as well. Wavelet frames offer good localization both in time and frequency but they do not necessary form an orthonormal basis.

The most commonly used values for s_0 and u_0 are 2 and 1, respectively, meaning that the scale is discretized, forming a dyadic sequence while the translation parameter is not

discretized. The transform resulting from this particular case of discretization is called the Dyadic Wavelet Transform. DWT of a signal $x(t)$ can be written as:

$$DWTx\left(2^j, u\right) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-u}{2^j}\right) dt \tag{2.22}$$

When a dyadic wavelet transform is discretized in time with a constant interval, $u = 2^j T$, it leads to the classic Discrete Wavelet Transform (DWT).

In addition, Meyer showed that there exist wavelets $\psi(x)$ such that

$$\left(\sqrt{2^j}\psi\left(2^j t - k\right)\right)_{(j,k)\in Z^2}$$

is an orthonormal basis of $L^2(R)$. Actually, if the frame bounds A and B (from theorem 2.3.2) are equal, that the frame is, in fact, an orthonormal basis. The wavelet orthonormal bases provide an important tool in functional analysis; before them it has been believed that no construction could yield simple orthonormal bases of $L^2(R)$ whose elements had good localization properties in both the spatial and Fourier domains.

2.4 Implementation of DWT

In order to take advantage of the Wavelet Transform's properties, an computation algorithm and an implementation scheme were needed. Mallat [Mal89] solved these problems by discussing the Multi-Resolution Analysis (MRA) which is linked to the Perfect Reconstruction (PR) filterbank structures [SB86].

2.4.1 Multiresolution Analysis (MRA)

A signal's approximation at resolution 2^{-j} is defined as an orthogonal projection on a space $V_j \subset L^2(R)$. The space V_j groups all possible approximations at the resolution 2^{-j}. The orthogonal projection of x on V_j is the function x_j that minimizes distance $\|x - x_j\|$. The details of a signal at resolution 2^{-j} are the difference between the approximations at the resolutions 2^{-j+1} and 2^{-j}.

Multiresolution Approximations

A multiresolution analysis consists of a sequence of successive approximation spaces $\{V_j\}_{j\in Z}$, presented in fig. 2.5, satisfying the following properties:

$$\forall (j,k) \in Z^2,\ x(t) \in V_j \ \Leftrightarrow\ x\left(t - 2^j k\right) \in V_j, \tag{2.23}$$

$$\forall j \in Z,\ V_{j+1} \subset V_j, \tag{2.24}$$

$$\forall j \in Z,\ x(t) \in V_j \ \Leftrightarrow\ x\left(\frac{t}{2}\right) \in V_{j+1}, \tag{2.25}$$

$$lim_{j\to+\infty} V_j = \bigcap_{j=-\infty}^{+\infty} V_j = \{0\}, \tag{2.26}$$

$$lim_{j\to-\infty}V_j = Closure\left(\bigcup_{j=-\infty}^{+\infty} V_j\right) = L^2(R), \tag{2.27}$$

There exists θ *such that* $\{\theta(t-n)\}_{n\in Z}$ *is a Riesz basis of* V_0.

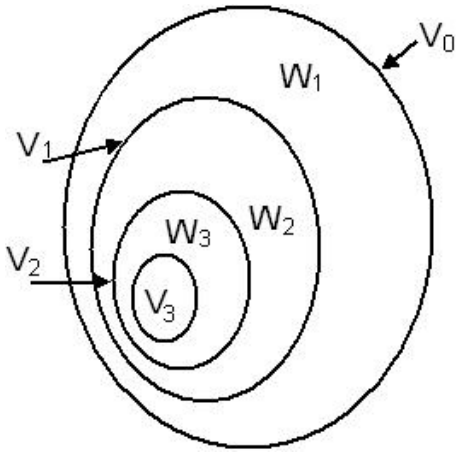


Figure 2.5: Approximation Spaces (V_j) and Detail Spaces (W_j)

For a given multiresolution approximation $\{V_j\}_{j\in Z}$, there exists a unique function $\phi(t)$, called a scaling function, such that $\left(\phi_{j,n}(t) = \sqrt{2^{-j}}\phi\left(2^{-j}t-n\right)\right)_{n\in Z}$ is an orthonormal basis of V_j .

The orthogonal projection on V_j can be computed by decomposing the signal $x(t)$ in the scaling orthonormal basis. Specifically,

$$\forall x(t) \in L^2(R),\ p_{V_j}x(t) = \sum_{n=-\infty}^{+\infty} \langle x, \phi_{j,n}\rangle \phi_{j,n}. \tag{2.28}$$

The inner products

$$a_j[n] = \langle x, \phi_{j,n}\rangle \tag{2.29}$$

represent the discrete approximation of the signal $x(t)$ at scale 2^j. It can also be written as:

$$a_j[n] = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{2^j}}\phi\left(\frac{t-2^jn}{2^j}\right)dt = x * \bar{\phi}_j\left(2^jn\right) \tag{2.30}$$

where $\bar{\phi}_j(t) = \sqrt{2^{-j}}\phi(-2^{-j}t)$. It can be easily proved that $\bar{\phi}_j(t)$ is the impulse response of a low-pass filter, so, the discrete approximation $a_j[n]$ is a low-pass filtering of x, sampled by a factor of 2^j.

The orthonormality condition of the elements of V_0 is:

$$\langle\phi_{0,0}(t), \phi_{0,n}(t)\rangle = \delta[n] \ \Leftrightarrow\ \Gamma_\phi[-n] = \delta[n].$$

The Detail Signal

The difference of information between the approximation of a signal $x(t)$ at scales 2^{j-1} and 2^j is called the detail signal at scale 2^j. It was shown in the previous paragraph that the approximations of a signal at scales 2^{j-1} and 2^j are, respectively equal to its orthogonal projection on V_{j-1} and V_j. It can be easily proved that the detail signal at the scale 2^j is given by the orthogonal projection of the original signal on the orthogonal complement of V_j in V_{j-1}, denoted here by W_j (see fig. 2.5). If W_j is the orthogonal complement, then

$$W_j \text{ is orthogonal to } V_j, \text{ and}$$

$$W_j \oplus V_j = V_{j-1}.$$

Mallat proves in [Mal89] that there exists a function $\psi(t)$, called an orthogonal wavelet, such that, if we denote

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j}\right),$$

for any scale 2^j, $\{\psi_{j,n}\}_{n \in Z}$ is an orthonormal basis of W_j and $\{\psi_{j,n}\}_{(n,j) \in Z^2}$ is an orthonormal basis of $L^2(R)$, for all scales.

The orthonormality condition of the elements of W_0 is:

$$\langle \psi_{0,0}(t), \psi_{0,n}(t) \rangle = \delta[n] \;\Leftrightarrow\; \Gamma_\psi[-n] = \delta[n].$$

Let p_{W_j} be the orthogonal projection on the vector space W_j. The detail signal of $x(t)$ at the resolution 2^j is equal to:

$$p_{W_j} x(t) = \sum_{n=-\infty}^{+\infty} \langle x, \psi_{j,n} \rangle \psi_{j,n}. \tag{2.31}$$

The inner products

$$d_j[n] = \langle x, \psi_{j,n} \rangle, \tag{2.32}$$

represent the wavelet coefficients (or the detail coefficients), calculated at scale 2^j. Analogous to the approximation signal, the detail signal can be implemented as a high-pass filtering of $x(t)$ followed by a sampling at rate 2^j. A signal $x(t)$ can be fully characterized by its wavelet decomposition, and can be written as a sum between the projection on the approximation space at level L and the projections on the detail spaces at all the other levels, as in eq. 2.33

$$x(t) = p_{V_L} x(t) + \sum_{j=L}^{0} p_{W_j} x(t). \tag{2.33}$$

2.4.2 Filter-bank Implementation of the Discrete Wavelet Transform

As previously mentioned, both approximation and detail coefficients can be obtained by filtering and sub-sampling of the original signal. It is proved that any scaling function

is specified by a discrete filter called a 'conjugate mirror filter'. Its impulse response is given by:

$$h[n] = \left\langle \frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right), \phi(t-n) \right\rangle, \tag{2.34}$$

where $\phi(t)$ denotes the scaling function. Its Fourier transform, denoted by $H(\omega)$ is given by:

$$H(\omega) = \sum_{n=-\infty}^{+\infty} h[n] e^{-j\omega n}.$$

With the definitions from above, the following theorem can be introduced:

Theorem 2.4.1. *(MALLAT, MEYER) $H(\omega)$ satisfies the following conditions:*

$$\forall\ \omega \in \mathbb{R},\ |H(\omega)|^2 + |H(\omega+\pi)|^2 = 2, \tag{2.35}$$

$$|H(0)| = \sqrt{2},\ \textit{and}\ h[n] = O\left(n^{-2}\right)\ \textit{at infinity}. \tag{2.36}$$

Conversely, let $H(\omega)$ be a Fourier transform satisfying 2.35 and 2.36 and such that

$$|H(\omega)| \neq 0\ for\ \omega \in [0, \pi/2]. \tag{2.37}$$

The function defined by

$$\Phi(\omega) = \prod_{p=1}^{+\infty} H\left(2^{-p}\omega\right) \tag{2.38}$$

is the Fourier transform of a scaling function.

The filters that satisfy property 2.35 are called *conjugate mirror filters*. Relation 2.37 implies that $H(\omega)$ is a low-pass filter.

As $\{\phi_{j,n}\}_{n\in Z}$ is an orthonormal basis of V_j (see 2.4.1), any $\phi_{j+1,p} \in V_{j+1} \subset V_j$ can be decomposed as follows:

$$\phi_{j+1,p} = \sum_{n=-\infty}^{\infty} \langle \phi_{j+1,p}, \phi_{j,n} \rangle \phi_{j,n}. \tag{2.39}$$

The inner products can be further processed and, taking into account relation 2.34, we obtain:

$$\langle \phi_{j+1,p}, \phi_{j,n} \rangle = \left\langle \frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right), \phi(t-n+2p) \right\rangle = h[n-2p]. \tag{2.40}$$

Hence

$$\phi_{j+1,p} = \sum_{n=-\infty}^{\infty} h[n-2p]\, \phi_{j,n}. \tag{2.41}$$

Using 2.29 we can write:

$$a_{j+1}[p] = \langle x, \phi_{j+1,p} \rangle = \sum_{n=-\infty}^{\infty} h[n-2p] \langle x, \phi_{j,p} \rangle = \sum_{n=-\infty}^{\infty} h[n-2p]\, a_j[n] = a_j[p] * h_d[2p], \tag{2.42}$$

where $h_d[n]$ is the reverse filter associated to $h[n]$, $h_d[n] = h[-n]$.

From equation 2.42 we can observe that the approximation coefficients from one iteration can be computed from the approximation coefficients from the previous iteration through low-pass filtering and subsampling with a factor of 2.

As previously discussed, orthonormal wavelets carry the details necessary to increase the resolution of a signal approximation. Theorem 2.4.2 proves that one can construct an orthonormal basis of W_j by scaling and translating a wavelet.

Theorem 2.4.2. *(MALLAT, MEYER) Let ϕ be a scaling function and h the corresponding conjugate mirror filter. Let ψ be the function whose Fourier transform is*

$$\Psi(\omega) = \frac{1}{\sqrt{2}} G\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right), \tag{2.43}$$

with

$$G(\omega) = e^{-j\omega} H^*(\omega + \pi). \tag{2.44}$$

let us denote

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j}\right).$$

For any scale 2^j, $\{\psi_{j,n}\}_{n \in Z}$ is an orthonormal basis of W_j. For all scales, $\{\psi_{j,n}\}_{(j,n) \in Z^2}$ is an orthonormal basis of $L^2(R)$.

The necessary and sufficient conditions imposed on G for designing an orthogonal wavelet are:

$$|G(\omega)|^2 + |G(\omega + \pi)|^2 = 2, \tag{2.45}$$

and

$$G(\omega) H^*(\omega) + G(\omega + \pi) H^*(\omega + \pi) = 0. \tag{2.46}$$

From theorem 2.4.2 we can prove that $G(\omega)$ is the Fourier transform of:

$$g[n] = \left\langle \frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right), \phi(t - n) \right\rangle, \tag{2.47}$$

which are the decomposition coefficients of

$$\frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{\infty} g[n] \phi(t - n),$$

and

$$g[n] = (-1)^{1-n} h[1 - n]. \tag{2.48}$$

Because H is a low-pass filter we can state that G is a high-pass filter (see relation 2.44). Also, due to relation 2.45, G can be called *conjugate* filter.

Let us consider $\psi_{j+1,p} \in W_{j+1} \subset V_j$. We can write:

$$\psi_{j+1,p} = \sum_{n=-\infty}^{\infty} \langle \psi_{j+1,p}, \phi_{j,n} \rangle \phi_{j,n}. \tag{2.49}$$

It can be proved that:

$$\langle \psi_{j+1,p}, \phi_{j,n} \rangle = \left\langle \frac{1}{\sqrt{2}} \psi \left(\frac{t}{2} \right), \phi \left(t - n + 2p \right) \right\rangle = g \left[n - 2p \right], \tag{2.50}$$

and, consequently,

$$\psi_{j+1,p} = \sum_{n=-\infty}^{\infty} g \left[n - 2p \right] \phi_{j,n}. \tag{2.51}$$

But, the detail coefficients from scale $j+1$ can be computed with (see also 2.32):

$$d_{j+1} \left[p \right] = \langle x, \psi_{j+1,n} \rangle . \tag{2.52}$$

By replacing 2.51 in 2.52 and having in mind relation 2.29, we get:

$$\begin{aligned} d_{j+1} \left[p \right] &= \left\langle x, \sum_{n=-\infty}^{\infty} g \left[n - 2p \right] \phi_{j,n} \right\rangle = \sum_{n=-\infty}^{\infty} g \left[n - 2p \right] \langle x, \phi_{j,n} \rangle \\ &= \sum_{n=-\infty}^{\infty} g \left[n - 2p \right] a_j \left[n \right] = a_j \left[p \right] * g_d \left[2p \right]. \end{aligned} \tag{2.53}$$

Analyzing 2.53 we can conclude that the detail coefficients from one scale can be computed from the approximation coefficients from the previous scale by convolution with the high-pass reverse filter g_d, $g_d \left[n \right] = g \left[-n \right]$, followed by a subsampling with a factor of 2.

To resume the previous results we present in figure 2.6, the decomposition scheme corresponding to one-level decomposition.

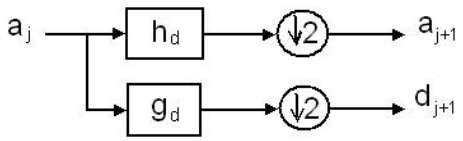


Figure 2.6: One-level DWT decomposition scheme

If we consider level 0 as the starting level, namely $x \left[n \right] = a_0 \left[n \right]$, we obtain for a three-level decomposition the 'tree' in figure 2.7.

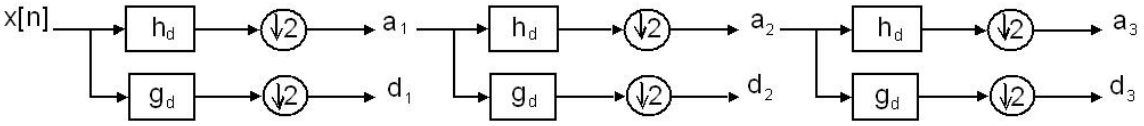


Figure 2.7: DWT decomposition tree

The implementation presented above was first proposed by Stephane Mallat and is also called 'Mallat's implementation'. Due to the downsamplers, the number of coefficients

from one scale is equal to the number of approximation coefficients from the previous scale, $(length\,(a_{j+1}) + length\,(d_{j+1}) = length\,(a_j))$. It is defined the 'orthogonal wavelet representation' of x as all the wavelet coefficients at scales $1 < 2^j < 2^J$ plus the remaining approximation at the largest scale 2^J:

$$\left[\{d_j\}_{1<j\leq J}, a_J\right]. \tag{2.54}$$

It can be easily observed that the size of the orthogonal wavelet representation (2.54) is equal to the size of the original signal, implying that this implementation of the Discrete Wavelet Transform is non-redundant.

2.4.3 Perfect Reconstruction

In the reconstruction stage we want to synthesize the original signal from the coefficients obtained in the decomposition stage. This transform is also known as the 'Inverse Discrete Wavelet Transform' (IDWT).

Since W_{j+1} is the orthogonal complement of V_{j+1} in V_j, the union of the two bases $\{\psi_{j+1,n}\}_{n\in Z}$ and $\{\phi_{j+1,n}\}_{n\in Z}$ is an orthonormal basis of V_j. Consequently, any $\phi_{j,p}$ can be decomposed in this basis:

$$\phi_{j,p} = \sum_{n=-\infty}^{\infty} \langle \phi_{j,p}, \phi_{j+1,n} \rangle \, \phi_{j+1,n} + \sum_{n=-\infty}^{\infty} \langle \phi_{j,p}, \psi_{j+1,n} \rangle \, \psi_{j+1,n}. \tag{2.55}$$

Inserting 2.40 and 2.50 in 2.55 yields:

$$\phi_{j,p} = \sum_{n=-\infty}^{\infty} h\,[p-2n]\,\phi_{j+1,n} + \sum_{n=-\infty}^{\infty} g\,[p-2n]\,\psi_{j+1,n}. \tag{2.56}$$

Using 2.56 and the properties of the inner product, we can write the approximation coefficients at level j, $a_j\,[n]$ as:

$$\begin{aligned} a_j\,[p] &= \langle x, \phi_{j,p} \rangle = \\ &= \sum_{n=-\infty}^{\infty} h\,[p-2n]\,\langle x, \phi_{j+1,n} \rangle + \sum_{n=-\infty}^{\infty} g\,[p-2n]\,\langle x, \psi_{j+1,n} \rangle \\ &= \sum_{n=-\infty}^{\infty} h\,[p-2n]\,a_{j+1}\,[n] + \sum_{n=-\infty}^{\infty} g\,[p-2n]\,d_{j+1}\,[n] \\ &= \breve{a}_{j+1} * h_r\,[p] + \breve{d}_{j+1} * g_r\,[p] \end{aligned} \tag{2.57}$$

where, $h_r = h$, $g_r = g$, and with $\breve{x}\,[n]$ we have denoted the signal:

$$\breve{x}\,[n] = \begin{cases} x\,[p] & \text{,if } n = 2p, \\ 0 & \text{,if } n = 2p+1. \end{cases}$$

This reconstruction can be seen as an interpolation with a factor of 2, that inserts zeros to expand a_{j+1} and d_{j+1}, followed by a filtering of these signals.

In order to achieve perfect reconstruction, one can use orthogonal filters, that satisfy:

$$|H(\omega)|^2 + |G(\omega)|^2 = 2,$$

or biorthogonal filters, satisfying the condition Vetterli ([KV92]) gave in 2.4.3:

Theorem 2.4.3. *(Vetterli) The filter bank performs an exact reconstruction for any input signal if and only if*

$$H^*(\omega + \pi) H_r(\omega) + G^*(\omega + \pi) G_r(\omega) = 0, \tag{2.58}$$

and

$$H^*(\omega) H_r(\omega) + G^*(\omega) G_r(\omega) = 2. \tag{2.59}$$

A one-level reconstruction scheme is presented in figure 2.8.

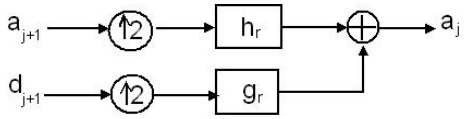


Figure 2.8: One-level DWT reconstruction scheme

A three-level reconstruction tree can be seen in figure 2.9, where we have considered the approximation at level zero equal to the signal to be reconstructed.

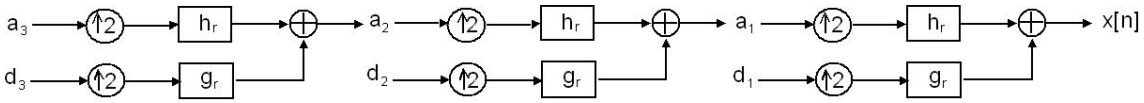


Figure 2.9: Three-level DWT reconstruction scheme

2.5 Extensions of DWT

The DWT transform, to which we will refer to as 'classical DWT', presented in section 2.3.2, is the most commonly used as it is fast, non-redundant and assures the perfect reconstruction. Despite all these properties, classical DWT might not be good enough for some specific applications. In the following, we will present some transforms, derived from DWT, that can be encountered in practical applications.

2.5.1 Two Dimensional DWT (2D DWT)

DWT is largely used for one-dimensional discrete signals. If we want apply the discrete wavelet transform to two-dimensional signals (images, for instance), we need to use the DWT's extension to two dimensions, namely the 2D DWT.

At the base of the 2D DWT's implementation relies the concept of separable multiresolutions and of two-dimensional wavelet bases, notions that we will define further.

Let $\{V_j\}_{j\in Z}$ be a multiresolution of $L^2(R)$. A separable two-dimensional resolution is composed of the tensor product spaces

$$V_j^2 = V_j \otimes V_j.$$

The space V_j^2 is a set of finite energy functions $x(t_1, t_2)$ that are linear expansions of separable functions

$$x(t_1, t_2) = \sum_{m=-\infty}^{\infty} a[m] f_m(t_1) g_m(t_2) \text{ with } f_m, g_m \in V_j.$$

If $\{V_j\}_{j\in Z}$ is a multiresolution approximation of $L^2(R)$, then $\left\{V_j^2\right\}_{j\in Z}$ is a multiresolution approximation of $L^2(R^2)$.

A separable wavelet orthonormal basis of $L^2(R^2)$ is constructed with separable products of a scaling function ϕ and a wavelet ψ, being associated to a one-dimensional multiresolution approximation $\{V_j\}_{j\in Z}$. Let W_j^2 be the detail space equal to the orthogonal complement of the lower resolution approximation space V_j^2 in V_{j-1}^2,

$$V_{j-1}^2 = V_j^2 \oplus W_j^2$$

To construct a wavelet orthonormal basis of $L^2(R^2)$, the following theorem builds a wavelet basis of each detail space W_j^2.

Theorem 2.5.1. *Let ϕ be a scaling function and ψ be the corresponding wavelet, generating a wavelet orthonormal basis of $L^2(R)$. We define three wavelets:*

$$\begin{aligned} \psi^1(t_1, t_2) &= \phi(t_1)\psi(t_2), \\ \psi^2(t_1, t_2) &= \psi(t_1)\phi(t_2), \\ \psi^3(t_1, t_2) &= \psi(t_1)\psi(t_2). \end{aligned} \tag{2.60}$$

and denote for $1 \le k \le 3$

$$\psi_{j,(n_1,n_2)}^k(t_1, t_2) = \frac{1}{2^j}\psi^k\left(\frac{t_1 - 2^j n_1}{2^j}\frac{t_2 - 2^j n_2}{2^j}\right).$$

The wavelet family

$$\left\{\psi_{j,n}^1, \psi_{j,n}^2, \psi_{j,n}^3\right\}_{n\in Z^2},$$

with $n = (n_1, n_2)$ is an orthonormal basis of W_j^2 and

$$\left\{\psi_{j,n}^1, \psi_{j,n}^2, \psi_{j,n}^3\right\}_{(j,n)\in Z^2}$$

is an orthonormal basis of $L^2(R^2)$.

The orthonormality condition of the elements of W_0^2 (the orthogonal complement of V_0^2) is:

$$\left\langle \psi_{0,0,0}^k(t_1,t_2), \psi_{0,n_1,n_2}(t_1,t_2) \right\rangle = \delta[n_1,n_2] \;\Leftrightarrow\; \Gamma_{\psi^k}[n_1,n_2] = \delta[n_1,n_2].$$

The separable wavelet expressions (2.60) imply that, in frequency, we will have the following relations:

$$\begin{aligned} \Psi^1(\omega_1,\omega_2) &= \Phi(\omega_1)\,\Psi(\omega_2), \\ \Psi^2(\omega_1,\omega_2) &= \Psi(\omega_1)\,\Phi(\omega_2), \\ \Psi^3(\omega_1,\omega_2) &= \Psi(\omega_1)\,\Psi(\omega_2). \end{aligned} \tag{2.61}$$

If we have images in mind, where lines represent the horizontal component and columns represent the vertical component, by analyzing relation 2.61 we can affirm that $|\Psi^1(\omega_1,\omega_2)|$ is large at low horizontal frequencies ω_1 and high vertical frequencies ω_2, $|\Psi^2(\omega_1,\omega_2)|$ is large at high horizontal frequencies and low vertical frequencies, whereas $|\Psi^3(\omega_1,\omega_2)|$ is large at high horizontal and vertical frequencies.

The approximation and details coefficients can be computed with:

$$a_j[n] = \left\langle f, \phi_{j,n}^2 \right\rangle \text{ and } d_j^k[n] = \left\langle f, \psi_{j,n}^k \right\rangle \text{ for } 1 \le k \le 3.$$

Let $h[n]$ and $g[n]$ be the conjugate mirror filters associated to the wavelet ψ. We denote with $f_d[n]$ the mirror filter associated to $f[n]$, $f_d[n] = f[-n]$.

The wavelet coefficients at the scale 2^{j+1} are calculated from the approximation coefficients at scale 2^j, a_j, with two-dimensional separable convolutions and subsamplings. The decomposition formula are obtained by applying the one-dimensional convolution formula 2.42 and, respectively, 2.53, to the separable two-dimensional wavelets ans scaling functions:

$$\begin{aligned} a_{j+1}[n_1,n_2] &= a_j[n_1,n_2] * h_d[2n_1]\,h_d[2n_2] \\ d_{j+1}^1[n_1,n_2] &= a_j[n_1,n_2] * h_d[2n_1]\,g_d[2n_2] \\ d_{j+1}^2[n_1,n_2] &= a_j[n_1,n_2] * g_d[2n_1]\,h_d[2n_2] \\ d_{j+1}^3[n_1,n_2] &= a_j[n_1,n_2] * g_d[2n_1]\,g_d[2n_2]. \end{aligned} \tag{2.62}$$

The convolution equations form 2.62 are computed with only six groups of one-dimensional convolutions. The rows of a_j are first convolved with h_d and g_d and subsampled by 2. Then, the columns of these two output images are convolved with h_d and g_d and subsampled, resulting four subsampled images a_{j+1}, d_{j+1}^1, d_{j+1}^2 and d_{j+1}^3. Figure 2.10 illustrates one level 2D DWT decomposition.

Similar to classical DWT, 2D DWT is a non-redundant transform, the wavelet image representation of x, at resolution J:

$$\left[a_j, \left\{ d_j^1, d_j^2, d_j^3 \right\}_{0<j\le J} \right]$$

having the same size as the original two-dimensional signal, x. In the coefficients image, the coefficients' repartition is presented in figure 2.11, where 'LL' (a) are the approxima-

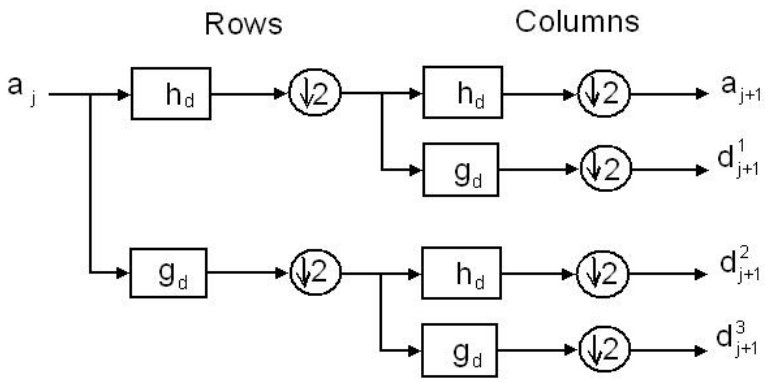


Figure 2.10: One-level 2D DWT decomposition scheme

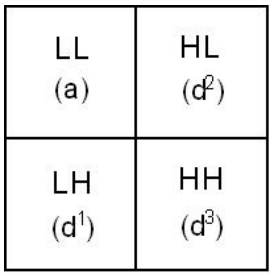


Figure 2.11: 2D DWT coefficients' image

tion coefficients, 'LH' (d^1) the horizontal details, 'HL' (d^2) the vertical details and 'HH' (d^3) the diagonal details. A second order statistical analysis of 2D DWt is presented in [NII+10].

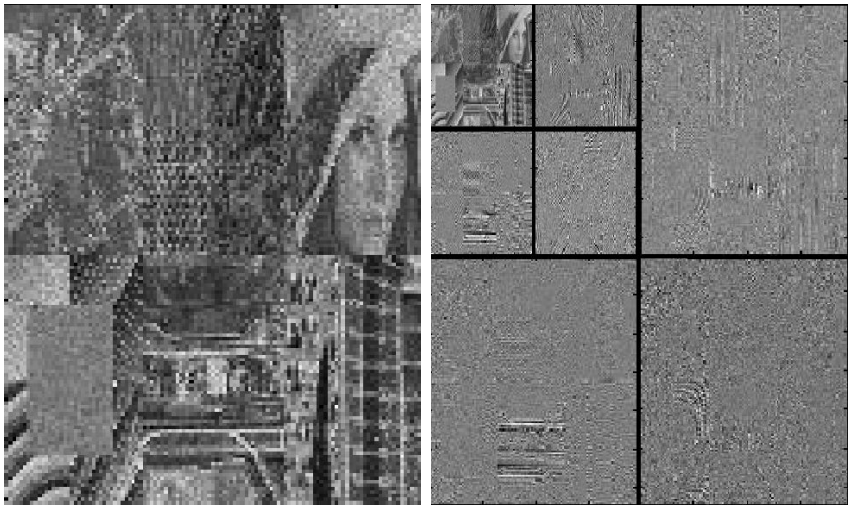
An example of a wavelet 2D DWT two-level decomposition is presented in figure 2.12(b), having as input image the one presented in fig 2.12(a).

In what concerns the reconstruction part, the 2D DWT, same as the DWT, ensures perfect reconstruction if the conditions given by theorem 2.4.3 are fulfilled. The reconstruction formula for the approximation coefficients at level j, a_j from the coarser scale approximations a_{j+1} and the wavelet coefficients d^k_{j+1}, $1 \leq k \leq 3$ is:

$$\begin{aligned} a_j[n_1,n_2] &= \check{a}_{j+1}[n_1,n_2] * h[n_1]\,h[n_2] + \check{d}^1_{j+1}[n_1,n_2] * h[n_1]\,g[n_2] + \\ &+ \check{d}^2_{j+1}[n_1,n_2] * g[n_1]\,h[n_2] + \check{d}^3_{j+1}[n_1,n_2] * g[n_1]\,g[n_2], \end{aligned} \tag{2.63}$$

where, with $\check{x}[n_1,n_2]$ we have denoted the image twice the size of $x[n_1,n_2]$, obtained by inserting a row of zeros and a column of zeros between pairs of consecutive rows and columns.

The corresponding implementation scheme is presented in fig. 2.13



(a) Original Image (b) 2-Level 2D DWT's coefficients

Figure 2.12: Example of a 2D DWT decomposition

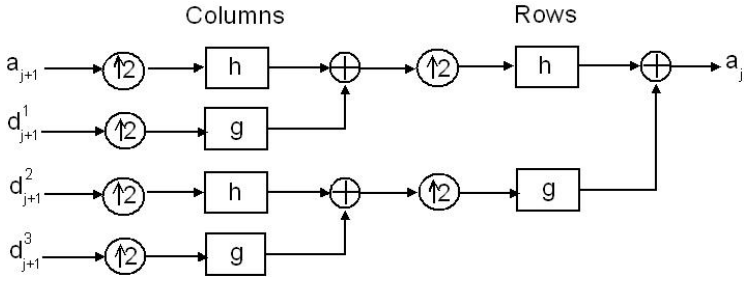


Figure 2.13: One-level 2D DWT reconstruction scheme

2.5.2 Wavelet Packet Transform

Different types of time-frequency structures are encountered in more complex signals such as speech recordings and real images. This motivates the design of bases whose time-frequency properties may be adapted. Wavelet bases are one particular family of bases that represent piecewise smooth signals effectively. Other bases are constructed to approximate different types of signals, such as highly oscillatory waveforms.

Wavelet packets were introduced by Coifman, Meyer and Wickerhauser [RC92] by generalizing the link between multiresolution approximation and wavelets.

In section 2.5.2.1 we will introduce one-dimensional packets and we will define the main notions that characterize the packets, while in section 2.5.2.2 we will briefly introduce the two-dimensional wavelet packets.

2.5.2.1 One-Dimensional Wavelet Packet Transform (1D WPT)

In section 2.4.1 we have seen that a space V_j of a multiresolution approximation can be decomposed in a lower resolution space V_{j+1} plus a detail space W_{j+1}. This is done

by dividing the orthogonal basis $\{\phi_j(t-2^j n)\}_{n\in Z}$ of V_j into two new orthogonal bases: $\{\phi_{j+1}(t-2^{j+1}n)\}_{n\in Z}$ of V_{j+1} and $\{\psi_{j+1}(t-2^{j+1}n)\}_{n\in Z}$ of W_{j+1}. The decompositions 2.41 and 2.51 of ϕ_{j+1} and ψ_{j+1} in the basis $\{\phi_j(t-2^j n)\}_{n\in Z}$ are specified by a pair of conjugate mirror filters, h and g (see 2.48).

The following theorem generalizes this result to any space U_j that admits an orthogonal basis of functions translated by $n2^j$, for $n \in Z$:

Theorem 2.5.2. *(COIFMAN, MEYER, WICKERHAUSER) Let $\{\theta_j(t-2^j n)\}_{n\in Z}$ be an orthonormal basis of space U_j. Let h and g be a pair of conjugate mirror filters. Define*

$$\theta^0_{j+1}(t)=\sum_{n=-\infty}^{\infty} h[n]\,\theta_j\left(t-2^j n\right) \text{ and } \theta^1_{j+1}(t)=\sum_{n=-\infty}^{\infty} g[n]\,\theta_j\left(t-2^j n\right). \tag{2.64}$$

The family

$$\left\{\theta^0_{j+1}\left(t-2^{j+1}n\right),\theta^1_{j+1}\left(t-2^{j+1}n\right)\right\}_{n\in Z}$$

is an orthonormal basis of U_j.

Theorem 2.5.2 proves that conjugate mirror filters transform an orthogonal basis $\{\theta_j(t-2^j n)\}_{n\in Z}$ in two orthogonal families: $\left\{\theta^0_{j+1}(t-2^{j+1}n)\right\}_{n\in Z}$ and, respectively, $\left\{\theta^1_{j+1}(t-2^{j+1}n)\right\}_{n\in Z}$. Let U^0_{j+1} and U^1_{j+1} be the spaces generated by each of these families, they are orthogonal and

$$U^0_{j+1}\oplus U^1_{j+1}=U_j.$$

Computing the Fourier transform of 2.64 we obtain:

$$\Theta^0_{j+1}(\omega)=H\left(2^j\omega\right)\Theta_j(\omega) \text{ and } \Theta^1_{j+1}(\omega)=G\left(2^j\omega\right)\Theta_j(\omega). \tag{2.65}$$

Since the transfer functions $H(2^j\omega)$ and $G(2^j\omega)$ have their energy concentrated in different frequency intervals, this transform can be interpreted as a division of the frequency support of θ_j.

Binary Wavelet Packet Tree

Instead of dividing only the approximation spaces V_j to construct the detail spaces W_j and wavelet bases, theorem 2.5.2 proves that we can set $U_j = W_j$ and divide these detail spaces to derive new bases. The recursive splitting of vector spaces is represented in a binary tree. If the signals are approximated at the scale 2^L, to the root of the tree we associate the approximation space V_L. This space admits an orthogonal basis of scaling functions $\left\{\phi_L\left(t-2^L n\right)\right\}_{n\in Z}$, with $\phi_L(t)=2^{-L/2}\phi\left(2^{-L}t\right)$.

Any node of the binary tree is labeled by (j,p), where $j-L\geq 0$ is the depth of the node in the tree, and p is the number of nodes that are on its left at the same depth $j-L$. Such a tree is presented in figure 2.14. To each node (j,p) we associate a space W^p_j, which admits an orthonormal basis $\left\{\psi^p_j(t-2^j n)\right\}_{n\in Z}$, by going down the tree. At the root, we have $W^0_L=V_L$ and $\psi^0_L=\phi_L$. The two wavelet packet orthogonal bases at the children nodes are defined by the splitting relations (see 2.64):

$$\psi^{2p}_{j+1}(t)=\sum_{n=-\infty}^{\infty} h[n]\,\psi^p_j\left(t-2^j n\right), \tag{2.66}$$

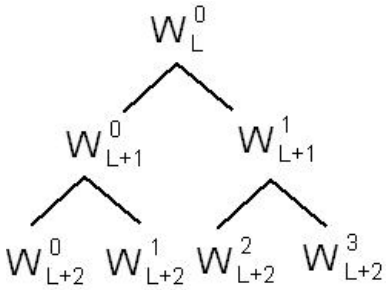


Figure 2.14: Binary tree of wavelet packet spaces

and

$$\psi_{j+1}^{2p+1}(t) = \sum_{n=-\infty}^{\infty} g[n]\,\psi_j^p\left(t - 2^j n\right). \tag{2.67}$$

Since $\left\{\psi_j^p\left(t-2^j n\right)\right\}_{n\in Z}$ is orthonormal,

$$h[n] = \left\langle \psi_{j+1}^{2p}(u), \psi_j^p\left(u - 2^j n\right)\right\rangle,\ g[n] = \left\langle \psi_{j+1}^{2p+1}(u), \psi_j^p\left(u - 2^j n\right)\right\rangle. \tag{2.68}$$

Theorem 2.5.2 proves that $\left\{\psi_{j+1}^{2p}\left(t-2^{j+1}n\right)\right\}_{n\in Z}$ and $\left\{\psi_{j+1}^{2p+1}\left(t-2^{j+1}n\right)\right\}_{n\in Z}$ are orthonormal bases of two orthogonal spaces W_{j+1}^{2p} and W_{j+1}^{2p+1} such that

$$W_{j+1}^{2p} \oplus W_{j+1}^{2p+1} = W_j^p. \tag{2.69}$$

This recursive splitting (2.69) defines a binary tree of wavelet packet spaces where each parent node is divided in two orthogonal subspaces.

Admissible Tree

It is called 'admissible tree', any binary tree where each node has either zero or two children, as shown in figure 2.15.

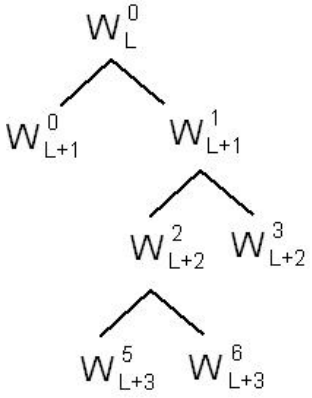


Figure 2.15: Example of admissible wavelet packet binary tree

Let $\{j_i, p_i\}_{1 \leq i \leq I}$ be the leaves of an admissible binary tree. By applying the recursive splitting 2.69 along the branches of an admissible tree we verify that the spaces $\left\{W_{j_i}^{p_i}\right\}_{1 \leq i \leq I}$ are mutually orthogonal and add up to W_L^0:

$$W_L^0 = \oplus_{i=1}^{I} W_{j_i}^{p_i}.$$

The union of the corresponding wavelet packet bases $\left\{\psi_{j_i}^{p_i}\left(t - 2^{j_i} n\right)\right\}_{n \in Z, 1 \leq i \leq I}$ thus defines an orthogonal basis of $W_L^0 = V_L$.

Number of Wavelet Packet Bases

The number of different wavelet packet orthogonal bases of V_L is equal to the number of different admissible binary trees. It can be proved that there are more than $2^{2^{j-1}}$ different wavelet packet ortonormal bases included in a full wavelet packet binary tree of depth j.

Time-Frequency Localization

Time support If the conjugate mirror filters h and g have a finite impulse response of size K, it can be proved ([Mal99]) that ϕ has a support of size $K-1$, so $\psi_L^0 = \phi_L$ has a support of $(K-1)\,2^L$. Due to the decomposition relations 2.66 and 2.67, one can show that the size of the support of ψ_j^p is $(K-1)\,2^j$, where j specifies the scale 2^j of the support.

Frequency localization The Fourier transform of 2.66 and 2.67 proves that the spectra of wavelet packet children are related to their parent's spectrum by:

$$\Psi_{j+1}^{2p}(\omega) = H\left(2^j \omega\right) \Psi_j^p(\omega), \;\; \Psi_{j+1}^{2p+1}(\omega) = G\left(2^j \omega\right) \Psi_j^p(\omega).$$

The energy of Ψ_j^p is mostly concentrated over a frequency band and the two filters $H\left(2^j\omega\right)$ and $G\left(2^j\omega\right)$ select the lower and, respectively, the higher frequency component within this band.

Best bases

Particular wavelet bases are the 'best bases'. Application of orthogonal bases often rely on their ability to efficiently approximate signals with only a few non-zero vectors. Choosing a wavelet packet basis that concentrates the signal energy over a few coefficients also reveals its time-frequency structures. The best basis algorithm finds a set of wavelet bases that provide the most desirable representation of the data relative to a particular cost function, chosen to best fit the application the wavelet packet decomposition is used for.

Wavelet Packet Filter Banks

Wavelet packet coefficients are computed with a filter bank algorithm that generalizes the fast discrete wavelet transform. Let us take the signal to be decomposed $x[n]$ equal to $d_L^0[n]$.

At the decomposition we will have:

$$d_{j+1}^{2p}[k] = d_j^p * h_d[2k] \text{ and } d_{j+1}^{2p+1}[k] = d_j^p * g_d[2k]. \tag{2.70}$$

The reconstruction formula is:

$$d_j^p[k] = \check{d}_{j+1}^{2p} * h[k] + \check{d}_{j+1}^{2p+1} * g[k]. \tag{2.71}$$

We have denoted with $x_d[n] = x[-n]$ and by $\check{x}$ the signal obtained by inserting a zero between each two consecutive sample of x.

From 2.70 we observe that the coefficients of the wavelet packet children d_{j+1}^{2p} and d_{j+1}^{2p+1} are obtained by subsampling the convolutions of d_j^p with h_d and g_d. Iterating these equations along the branches of a wavelet packet tree computes all wavelet coefficients, as illustrated by fig. 2.16. From the wavelet packet coefficients at the leaves $\{j_i, p_i\}_{1 \le i \le I}$

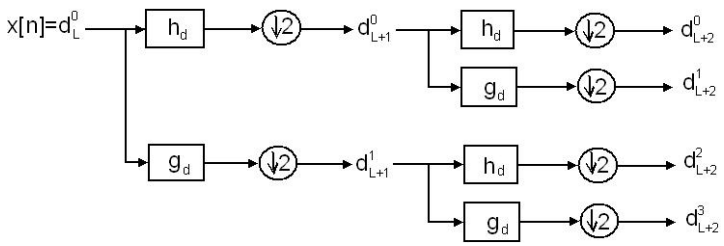


Figure 2.16: One-dimensional Wavelet Packet Decomposition

of an admissible tree, we recover $x[n] = d_L^0[n]$ to the top of the tree by computing 2.71 for each node inside the tree as illustrated in 2.17.

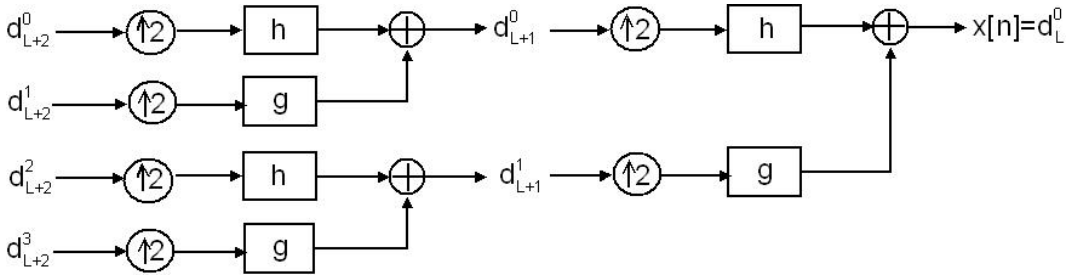


Figure 2.17: One-dimensional Wavelet Packet Reconstruction

2.5.2.2 Two-Dimensional Wavelet Packet Transform (2D WPT)

As for the two-dimensional discrete wavelet transform, the concept of separable wavelet bases, or two-dimensional wavelet packet bases in this case, was used when the two-dimensional wavelet packets were introduced. There can be constructed wavelet packet bases of $L^2(R^2)$ whose elements are separable products of wavelet packets that can be

written $\psi_j^p(x_1 - 2^j n_1)\,\psi_j^q(x_2 - 2^j n_2)$ having the same scale along x_1 and x_2. These separable wavelet packet bases are associated to quad-trees (the generalization to binary trees in the two-dimensional case), and divide the two-dimensional Fourier plane (ω_1, ω_2) into square regions of varying sizes.

If images are approximated at the scale 2^L, to the root of the quad-tree we associate the approximation space $V_L^2 = V_L \otimes V_L \subset L^2(R^2)$. The two-dimensional wavelet packet quad-tree is composed of separable wavelet packet spaces. Each node of this quad-tree is labeled $\{j, p, q\}$, where 2^j represents the scale and the two integers $0 \le p < 2^{j-L}$ and $0 \le q < 2^{j-L}$ correspond to a separable space:

$$W_j^{p,q} = W_j^p \otimes W_j^q \tag{2.72}$$

that can be written as the direct sum of the four orthogonal subspaces corresponding to the four children nodes in the quad-tree:

$$W_j^{p,q} = W_{j+1}^{2p,2q} \oplus W_{j+1}^{2p+1,2q} \oplus W_{j+1}^{2p,2q+1} \oplus W_{j+1}^{2p+1,2q+1}. \tag{2.73}$$

An example of a quad-tree is presented in figure 2.18

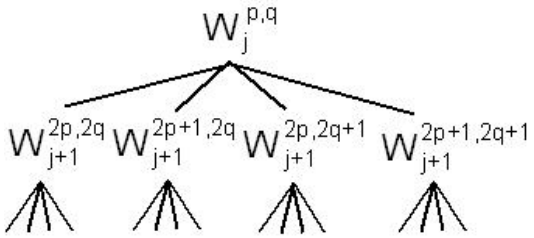


Figure 2.18: Example of a wavelet packet quad-tree

The filter-bank implementation of the 2D WPT is straightforward, the decomposition coefficients being computed by iterating equation 2.74 along the branches of the quad-tree.

$$\begin{aligned}
d_{j+1}^{2p,2q}[n_1, n_2] &= d_j^{p,q}[n_1, n_2] * h_d[n_1]\, h_d[n_2] \\
d_{j+1}^{2p+1,2q}[n_1, n_2] &= d_j^{p,q}[n_1, n_2] * g_d[n_1]\, h_d[n_2] \\
d_{j+1}^{2p,2q+1}[n_1, n_2] &= d_j^{p,q}[n_1, n_2] * h_d[n_1]\, g_d[n_2] \\
d_{j+1}^{2p+1,2q+1}[n_1, n_2] &= d_j^{p,q}[n_1, n_2] * g_d[n_1]\, g_d[n_2]
\end{aligned} \tag{2.74}$$

The calculations are performed with separable convolutions along the rows and columns of the image, as illustrated in figure 2.19.

The reconstruction relation is:

$$\begin{aligned}
d_j^{p,q}[n_1, n_2] = {} & \check{d}_{j+1}^{2p,2q}[n_1, n_2] * h[n_1]\, h[n_2] + \check{d}_{j+1}^{2p+1,2q}[n_1, n_2] * g[n_1]\, h[n_2] + \\
& + \check{d}_{j+1}^{2p,2q+1}[n_1, n_2] * h[n_1]\, g[n_2] + \check{d}_{j+1}^{2p+1,2q+1}[n_1, n_2] * g[n_1]\, g[n_2]
\end{aligned} \tag{2.75}$$

The original image $x[n_1, n_2] = d_L^{0,0}$ is reconstructed from the wavelet packet coefficients stored at the leaves of any admissible quad-tree, by repeating the partial reconstruction 2.75 in the inside nodes of the quad-tree. The reconstruction scheme is presented in figure 2.20.

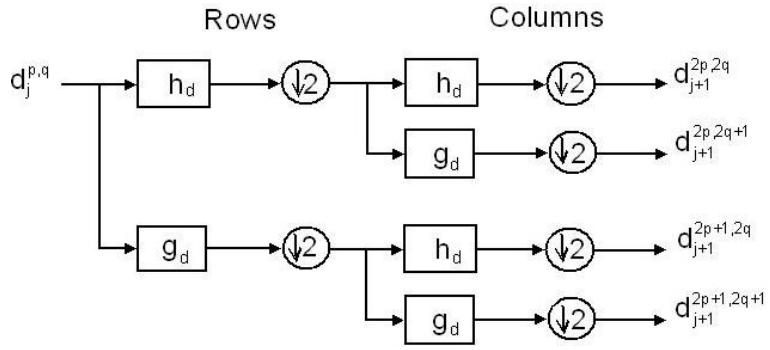


Figure 2.19: One Level 2D WPT Decomposition Scheme

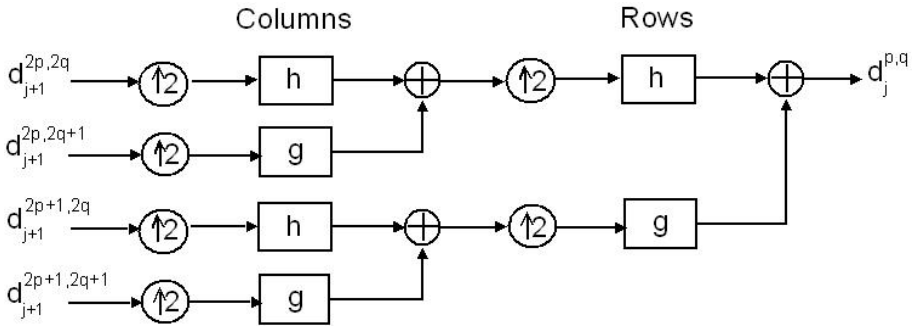


Figure 2.20: One Level 2D WPT Reconstruction Scheme

2.5.3 Undecimated Discrete Wavelet Transform (UDWT)

A particular wavelet transform is the Undecimated Wavelet Transform (UWT), also known as the 'Stationary Wavelet Transform' and can be implemented using the '*Algorithme à Trous*'. The Undecimated Discrete Wavelet Transform (UDWT) refers to the discrete implementation of this transform.

The '*Algorithme à Trous*', first introduced by Holschneider et al. [HKMMT89] in 1989 for music synthesis applications, is similar to a nonorthonormal multiresolution algorithm for which the discrete wavelet transform is exact.

The dyadic wavelet decomposition coefficients at level $j > 0$ of a continuous-time signal $x(t)$ are:

$$a_j[n] = \langle x(t), \phi_{j,n}(t) \rangle \tag{2.76}$$

and

$$d_j[n] = \langle x(t), \psi_{j,n}(t) \rangle \tag{2.77}$$

For any filter f it is denoted by f_j the filter obtained by inserting $2^j - 1$ zeros between each sample of $f[n]$, this zero-insertion leading to holes ('trous' in French) in the filter. The Fourier transform of $f_j[n]$ is $F(2^j\omega)$, so that if f is of finite bandwidth, f_j has a bandwidth 2^j times narrower. Let $f_{d_j}[n] = f_j[-n]$.

The filter-bank decomposition formula of the UDWT, considering the original signal

$x[n] = a_0[n]$, are presented in eq. 2.78:

$$\begin{aligned} a_{j+1}[n] &= a_j[n] * h_{d_j}[n], \\ d_{j+1}[n] &= a_j[n] * g_{d_j}[n], \end{aligned} \tag{2.78}$$

whereas the reconstruction will be given by eq. 2.79:

$$a_j[n] = \frac{1}{2}\left(a_{j+1}[n] * h_{r_j}[n] + d_{j+1}[n] * h_{r_j}[n]\right), \tag{2.79}$$

where h_r and g_r satisfy the conditions given in theorem 2.4.3.

The decomposition implementation scheme is presented in fig. 2.21, while the relation between the filters from one level and those from the previous one is illustrated in fig. 2.22.

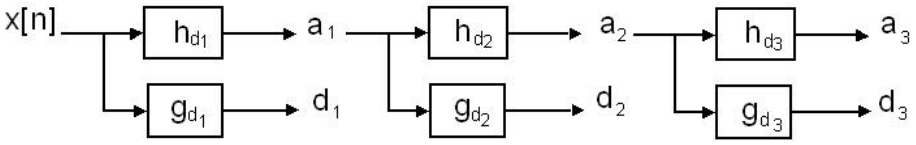


Figure 2.21: Three-Level UDWT Decomposition Scheme

Figure 2.22: Relation between the filters corresponding to two consecutive levels of UDWT decomposition

Due to the absence of downsamplers in the UDWT's implementation, each coefficient sequence from any level of decomposition has the same length as the original, in other words, if the original signal has N samples, the UDWT L-level representation

$$\left\{a_L[n], d_j[n]_{0<j\leq L}\right\},$$

is of size $N \cdot (L+1)$, making from the UDWT a highly redundant transform.

In the reconstruction stage we will have the implementation scheme in figure 2.23, the

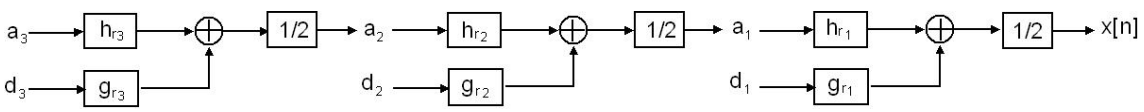


Figure 2.23: Three-Level UDWT Reconstruction Scheme

relation between the filters in this case being presented in fig. 2.24.

Figure 2.24: Relation between the filters corresponding to two consecutive levels of UDWT reconstruction

Two-Dimensional UDWT

In what concerns the images' undecimated wavelet decomposition and reconstruction, the separable wavelet bases are, again, the starting point for the filter-bank implementations, presented in figure 2.25 and figure 2.26.

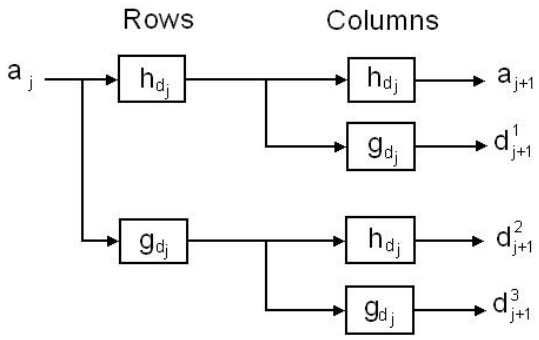


Figure 2.25: One Level 2D UDWT Decomposition Scheme

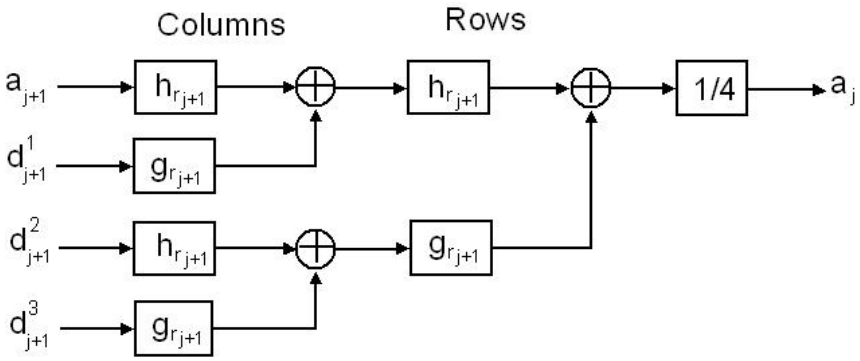


Figure 2.26: One Level 2D UDWT Reconstruction Scheme

2.6 Applications of Wavelet Transforms

Wavelet Transforms are adopted for a vast number of applications, often replacing conventional Fourier Transform.

Due to DWT's lack of redundancy and to the property to concentrate important data in a small number of coefficients (also known as 'sparsity'), it is very appropriate for

signal and image compression. Compared to techniques previously used in compression, such as those using the Discrete Cosine Transform, wavelet-based coding [VK95] provides substantial improvements in picture quality at higher compression rates. One example of image compression is JPEG 2000, which uses biorthogonal wavelets.

Another use is that of denoising data based on wavelet coefficients filtering [DJ94]. By adaptively thresholding the wavelet coefficients that correspond to undesired frequency components smoothing and/or denoising operations can be performed.

Besides compression and denoising, we can also use wavelets in image watermarking ([Naf08]).

Apart from signal and image processing, wavelet transforms are starting to be used in communication applications where traditional FFT OFDM systems are loosing field in favor of newer, more performant, Wavelet OFDM systems ([Olt09]).

2.7 Limitations of Wavelet Transforms

The wavelet transform comes in many forms. The most common of them, the Discrete Wavelet Transform, provides the most compact representation, but has several limitations. We will further discuss only about two of them: shift-sensitivity and poor directional selectivity [Kin01].

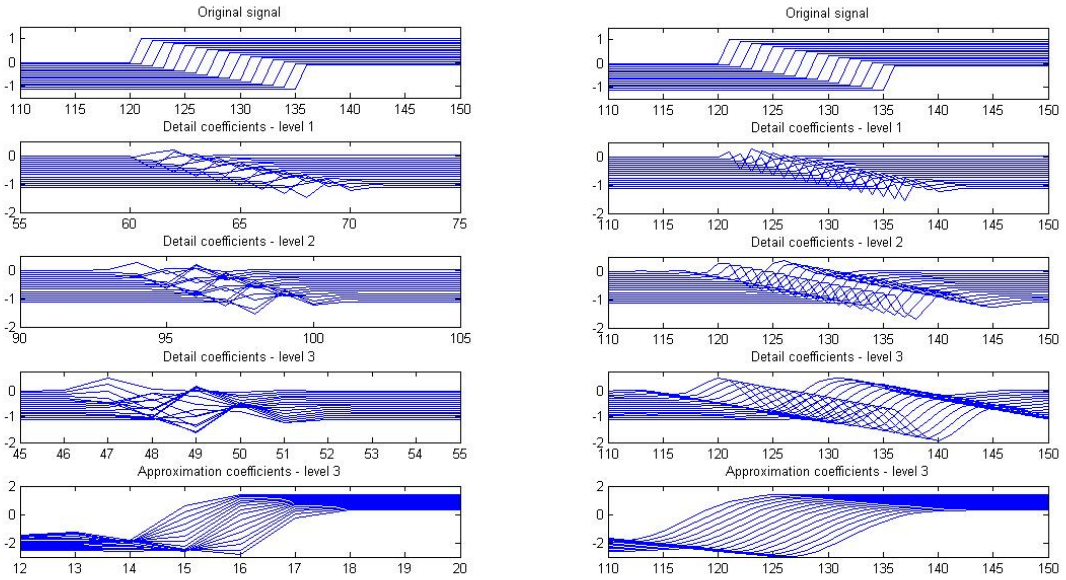
2.7.1 Shift Sensitivity

Shift sensitivity, the opposite of 'translation invariance', implies that an input signal change (or translation) causes an unpredictable change in the output coefficients. In other words, small shifts in the input signal can cause major variations in the distribution of energy between DWT coefficients at different scales [Kin01]. This is an undesirable property, making DWT not suited for some applications (e.g. pattern recognition).

Continuous Wavelet Transform and Short-Time Fourier Transform provide shift invariant representations, while the uniform sampling of the translation parameter destroys this translation invariance [Mal99]. The shift-sensitivity of the DWT is a consequence of the aliasing introduced by the down-samplers followed by up-samplers from the DWT's filter-bank implementation [Kin01].

In order to illustrate the shift-sensitivity of the DWT, we have taken an input signal consisting of 16 different step-functions, each at one-sample delay from the others (2.27). We have pursued with a 3-level decomposition, using as mother wavelet a function from the Daubechies family, having 5 vanishing moments, and have taken the detail coefficients from the three levels and the approximation coefficients from the 3rd level and plotted them. In figure 2.27(a) we have used classical DWT, while in figure 2.27(b) we have used the UDWT. By analyzing the two figures we can observe that the DWT is far from being shift-invariant, whereas UDWT reaches translation invariance with the cost of higher redundancy.

In the following chapter we will discuss this aspect and other methods to improve a transform's shift sensitivity with a lower redundancy cost.



(a) DWT (b) UDWT

Figure 2.27: Shift-sensitivity

2.7.2 Directional selectivity

In section 2.5.1 we have introduced the two-dimensional Discrete Wavelet Transform and we have seen that separable filtering along the rows and columns of an image produces four images at each level. The LH and HL bandpass subimages can select mainly horizontal or vertical edges respectively, but the HH subimage contains components from diagonal features of either orientation. This means that the separable real DWT has 'poor directional selectivity'. This is illustrated in fig. 2.28, figure generated by prof. Selesnick by applying the inverse 2D DWT to an image of coefficients, where each type of coefficients were considered 2D unit pulses. From left to right we can observe the vertical details (HL), the horizontal details (LH) and the diagonal details (HH).

Prof. Kingsbury, in [Kin99] explains this limitation by the fact that real highpass row filters select both positive and negative horizontal high frequencies and, consequently, the combined HH filter must have passbands in all four quadrants of the 2-D frequency plane. On the other hand, a directionally selective filter for diagonal features with positive gradient must have bandpass filters only in quadrants 2 and 4 of the frequency plane, while a filter for diagonals with negative gradient must have passbands only in quadrants 1 and 3. The poor directional properties of real separable filters make it difficult to generate steerable or directionally selective algorithms, based on the separable real DWT.

The two-dimensional Wavelet Packet Transform manages to have better directional selectivity due to the different bandwidth filters have but, as it is implemented with real filters, it still can not distinguish between diagonal directions.

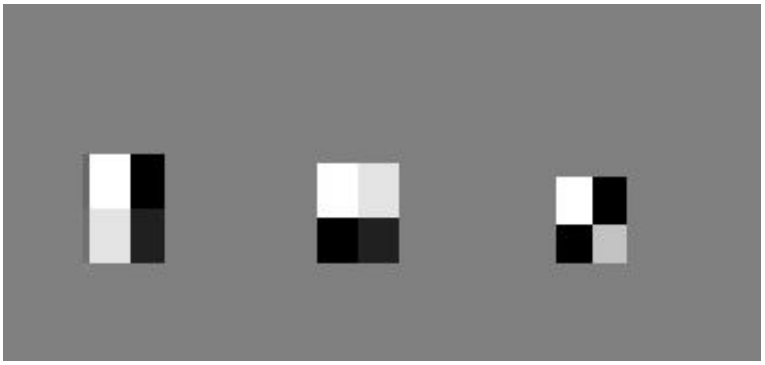


Figure 2.28: DWT's Directional selectivity

2.8 Summary

Wavelet analysis is a powerful tool in signal and image processing, and due to its time-frequency localization it gained over the Fourier transform, being preferred in many applications from signal and image compression to image denoising and even communications. Its most common form is the Discrete Wavelet Transform (2.3.2), being the most compact as well. Due to its filter-bank implementation it offers a high flexibility in implementation, offering the possibility of using a wide number of wavelet families.

Besides all the benefits of this transform, it has also a series of limitations such as its shift-sensitivity and its poor directional selectivity. These limitations can be somewhat overcome by using some of DWT's extensions, such as the UDWT which is translation invariant or the Discrete Wavelet Packet Transform, that offers a better directional selectivity. Another way of overcoming these limitations is given by the use of complex mother wavelets, or complex wavelet transforms. We will discuss in detail about this type of transforms in the following chapter.

Chapter 3

Complex Wavelet Transforms (CWT)

3.1 Introduction

As discussed in the previous chapter, multiscale transforms are powerful tools for signal and image processing. One-dimensional wavelet transform, which acts as a multiresolution version of an Nth-order derivative operator, where N is the number of vanishing moments of the wavelet ([Mal99]), is a prominent example in this direction. Its extension to multiple dimensions, and to 2D, in particular, is typically achieved by forming tensor-product basis functions (see 2D DWT). However, it was observed that such separable wavelets are not well matched to the singularities occurring in images, such as lines and edges which can be arbitrarily oriented and even curved.

The Complex Wavelet Transforms (CWT) were introduced in the attempt to solve this problem and also to overcome the limitations of the classical DWT (see Limitations of DWT).

In this chapter I will present the early work performed in this field (section 3.2) and I will mention also some of the recent developments (section 3.3). I will continue with some examples of CWTs, including the Dual-tree based CWT (DT CWT) and the Hyperanalytical Wavelet Transform (HWT), the latter representing my contribution to this field. A presentation of the advantages and the applications of Complex Wavelet Transforms will conclude this chapter.

3.2 Earlier Work

A review of the work in the domain of complex wavelets can be found in [VU08], [Shu03] and [SBK05].

Over more than twenty years ago, Grossman and Morlet [GM84] developed the Continuous Wavelet Transform (CoWT) [Mal99], using complex-valued mother wavelets. The Morlet's mother wavelet has the expression:

$$\psi_M(\tau) = \frac{c_M}{\sqrt[4]{\pi}} \left(e^{j\omega_0 \tau} - e^{-\frac{\omega_0^2}{2}} \right) \cdot e^{-\frac{\tau^2}{2}} \tag{3.1}$$

where the normalization constant is:

$$c_M = \frac{1}{\sqrt{1 + e^{-\omega_0^2} - 2e^{-\frac{3}{4}\omega_0^2}}}.$$

Generally $\omega_0 \geq 5$ to assure a sufficient temporal localization. Initial analysis based on wavelet decompositions was implemented using such mother wavelets. Both magnitude and phase descriptions of non-stationary signals were determined, and an early example of analysis includes wavelet ridge methods proposed by Delprat et al. [DEG+92].

Lawton [Law93] gives a method for constructing complex valued linear phase FIR conjugate quadrature filters and the associated wavelet bases. He achieves this by replacing certain zeros of a real valued FIR conjugate filter by their reciprocal conjugates.

Lina and Gagnon [LG95], [GLG] propose the use of some complex symmetric Daubechies wavelets that provide a natural way to calculate zero-crossings because of a hidden 'Laplacian operator', in the imaginary part of the scaling function. They propose the use of such wavelets in the context of a multiscale sharpening enhancement algorithm. In [LM95] are presented other complex mother wavelets that can be used to implement complex DWT. A possible implementation of a complex wavelet transform is the one presented by Kingsbury in [Kin99] (fig. 3.1). In this figure are presented four levels of

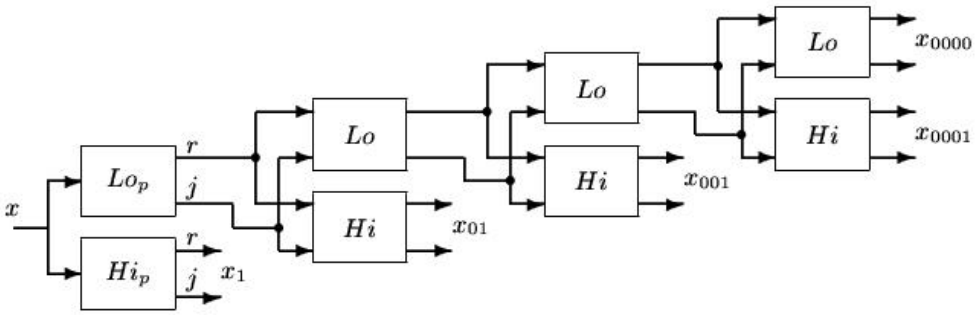


Figure 3.1: Complex wavelet tree

the complex wavelet tree for a 1-D input signal x. The real and imaginary parts (r and j) of the inputs and outputs are shown separately. Where there is only one input to a block, it is a real signal.

The extension of complex wavelets to 2-D is achieved by separable filtering along rows and then columns. In fig. 3.2 are presented two levels of the complex wavelet tree for a 2-D input image x, giving six directional bands at each level (the directions are shown for level 1). Components of 4-element 'complex' vectors are labeled r, j_1, j_2 and $j_1 j_2$.

However, subsequently, for many years interest focused on the Discrete Wavelet Transform (DWT) and signal estimation. As seen in section 2.7, DWT has a number of limitations. If we take a closer look to the Fourier Transform (sec. 2.2.1), we will see that it does not suffer from the same problems. First, the magnitude of the Fourier transform is perfectly shift invariant, with a simple linear phase offset encoding the shift and secondly, the sinusoids of the M-D Fourier basis are highly directional plane waves. This difference in behavior between the DWT and the FT comes from the fact that, unlike the DWT,

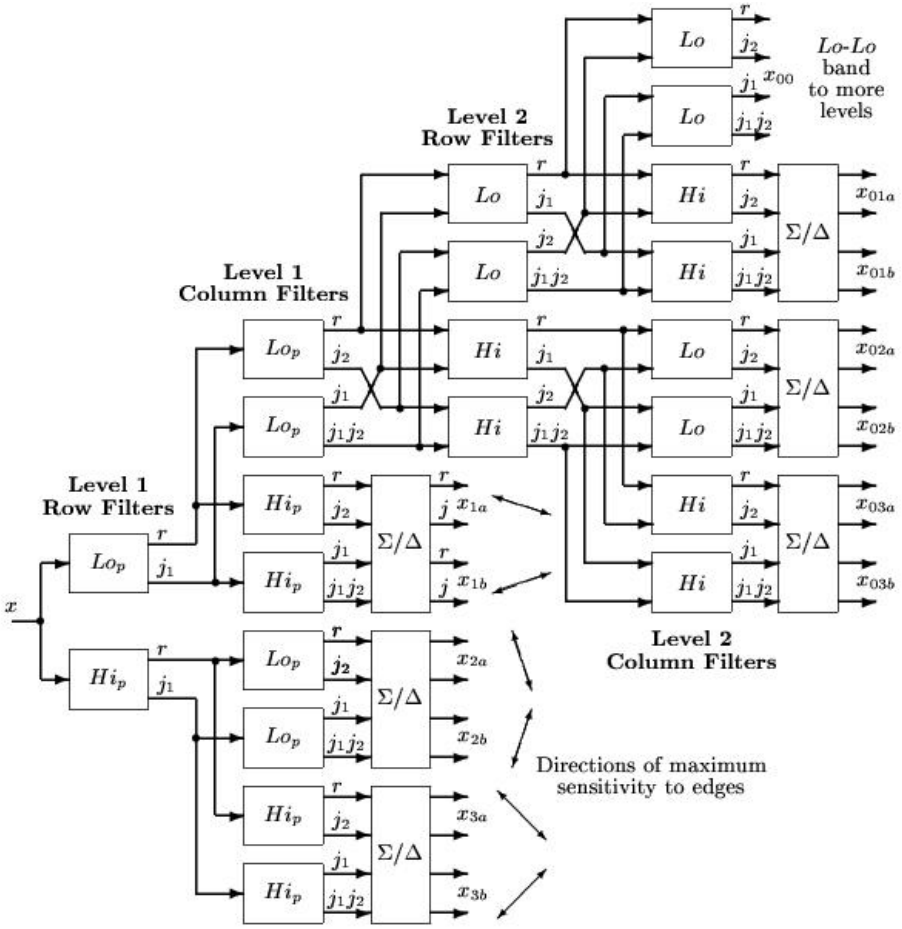


Figure 3.2: 2-D Complex wavelet tree

which is based on real-valued oscillating wavelets, the FT is based on complex-valued oscillating sinusoids:

$$e^{j\Omega t} = cos\,(\Omega t) + j sin\,(\Omega t)$$

with $j = \sqrt{-1}$. The oscillating cosine and sine components (the real and imaginary parts, respectively) form a Hilbert transform pair; i.e., they are 90 ° out of phase with each other. Together they constitute an analytic signal $e^{j\Omega t}$ that is supported on only one half of the frequency axis ($\Omega > 0$).

Mallat, in his book [Mal99], has a section dedicated to analytic wavelets and analytic wavelet transform.

Analytic Signal and Analytic Wavelet Transform

We can associate an analytic part $f_a\,(t)$ to every real signal $f\,(t)$ with:

$$f_a\,(t) = f\,(t) + jH\{f\,(t)\} \tag{3.2}$$

where, we have denoted with $H\{f\,(t)\}$ the Hilbert transform of the signal $f\,(t)$.

A possible definition of the analytic signal can be done in relation to its spectrum. A function $f_a \in L^2(R)$ is said to be analytic if its Fourier transform is zero for negative frequencies:

$$F_a(\omega) = 0 \text{ if } \omega < 0. \tag{3.3}$$

An analytic function is necessary complex but it is entirely characterized by its real part:

$$F_a(\omega) = \begin{cases} 2F(\omega) & \text{if } \omega \geq 0 \\ 0 & \text{if } \omega < 0 \end{cases} \tag{3.4}$$

When talking about discrete-time signals, the analytic part $f_a[n]$ of a discrete signal $f[n]$ of size N is obtained by setting to zero the negative frequency components of its discrete Fourier transform. The spectrum's values at $k = 0$ and $k = N/2$ must be carefully adjusted so that $Real\{f_a\} = f$:

$$F_a[k] = \begin{cases} F[k] & \text{if } k = 0, N/2, \\ 2F[k] & \text{if } 0 < k < N/2, \\ 0 & \text{if } N/2 < k < N \end{cases} \tag{3.5}$$

$f_a[n]$ is obtained through the computation of the inverse discrete Fourier transform.

An analytic wavelet transform is equivalent to a continuous wavelet transform (eq. 2.11) computed with an analytic wavelet, ψ:

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{\infty} f(t) \cdot \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt. \tag{3.6}$$

An analytic wavelet transform of f depends only on its analytic part, f_a:

$$Wf(u,s) = \frac{1}{2} Wf_a(u,s).$$

An analytic wavelet can be constructed with a frequency modulation of a real and symmetric window g. The Fourier transform of

$$\psi(t) = g(t) e^{j\eta t} \tag{3.7}$$

is

$$\Psi(\omega) = G(\omega - \eta).$$

If $G(\omega) = 0$ for $|\omega| > \eta$, then $\Psi(\omega) = 0$ for $\omega < 0$, resulting that $\psi(t)$ is analytic.

A *Gabor wavelet* is a particular type of such an analytic wavelet (presented in eq. 3.7), obtained with a Gaussian window:

$$g(t) = \frac{1}{(\sigma^2 \pi)^{1/4}} e^{-\frac{t^2}{2\sigma^2}}.$$

The Fourier transform of this window is $G(\omega) = (4\pi\sigma^2)^{-1/4} exp(-\sigma^2\omega^2/2)$. If $\sigma^2\eta^2 >> 1$, then $G(\omega) \approx 0$ for $|\omega| > \eta$. Such Gabor wavelets are thus considered to be approximately analytic.

Auscher, in [Aus92] proved that compact discrete wavelet filters cannot be exactly analytic.

In [Hig84], the author provides a way to build new complete orthonormal sets of the Hilbert space of finite-energy band-limited functions with a bandwidth of π, named the Paley-Wiener (PW) space. He proved the following proposition:

Proposition 3.2.1. *Let χ denote the characteristic function of the interval $[-\pi, \pi]$ and let $\mu(x)$ be real valued and piecewise continuous there. Then, the integer translations of the inverse Fourier transform of $\chi e^{j\mu}$ constitute a complete orthonormal set in PW.*

Following this proposition, some new orthonormal complete sets of integer translations of a generating function can be constructed in the PW space. The scaling function and the mother wavelets of the standard multiresolution analysis of PW generate through integer shifts such complete orthonormal sets. Proposition 3.2.1 was generalized in [Isa93] to give a new mechanism of mother wavelet construction. In this reference, the following two propositions were formulated.

Proposition 3.2.2. *If $A_m = \{\alpha_{m,n}(t)\}_{n\in Z}$ is a complete orthonormal set generating a Hilbert space H_m, then the set $\hat{A}_m = \left\{\left(\frac{1}{\sqrt{2\pi}}\right)\hat{\alpha}_{m,n}(\omega)\right\}_{n\in Z}$ is a complete orthonormal set of $\hat{H}_m$ and vice versa.*

We have denoted with $\widehat{\alpha}$, the FT of α.

Proposition 3.2.3. *If $\mu(\omega)$ is a real-valued and piecewise-continuous function and $\hat{A}_m = \left\{\left(\frac{1}{\sqrt{2\pi}}\right)\hat{\alpha}_{m,n}(\omega)\right\}_{n\in Z}$ is a complete orthonormal set of $\hat{H}_m$, then*

$$ {}_{\mu}\hat{A}_m = \left\{\left(\frac{1}{\sqrt{2\pi}}\right)e^{j\mu(\omega)}\hat{\alpha}_{m,n}(\omega)\right\}_{n\in Z} $$

is another complete orthonormal set of the same space.

These two propositions can be used to build new mother wavelets if we identify the subspaces of an orthogonal decomposition of the Hilbert space $L^2(R)$, W_m, $m \in Z$, with the Hilbert spaces $\hat{H}_m$. With respect to this, the function $\mu(\omega)$ must satisfy the following constraint: $\mu(\omega) = \mu(2^m\omega)$, $\forall m \in Z$. An example of a function that satisfies this constraint is $\mu(\omega) = \frac{\pi}{2}(\text{sgn}\omega + 1)$. In this case, $e^{j\mu(\omega)} = \text{sgn}(\omega)$. Therefore, the function generating the set ${}_{\mu}A_m$ (which corresponds to the new mother wavelets) is obtained by applying the Hilbert transform to the function generating the set A_m (which corresponds to the initial mother wavelets) multiplied by j. Consequently, if the function ψ is a mother wavelet, then the functions $jH\{\psi\}$ and $\psi_a = \psi + jH\{\psi\}$ are also mother wavelets. Using the analytical mother wavelets we can implement an analytical DWT (ADWT). The pair $(\psi, jH\{\psi\})$ defines a complex DWT (CDWT), presented in fig. 3.3.

3.3 Recent Developments

The revival or interest in complex wavelets may be linked to the development of complex-valued discrete wavelet filters [LM95] and the clever dual filter bank [Kin99]. The complex

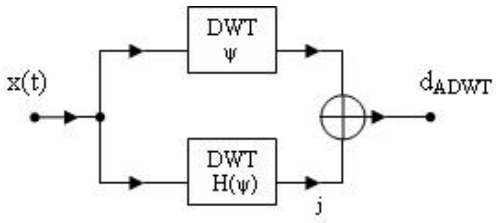


Figure 3.3: Implementation of an analytical DWT

wavelet transform has been shown to provide a powerful tool in signal and image analysis [SBK05], where most of the properties of the transform follow from the analyticity of the wavelet function. In [OM06] were derived large classes of wavelets generalizing the concept of 1-D local complex-valued analytic decomposition to 2-D vector-valued hyperanalytic decomposition.

Recent research in the development of CWTs follows, mainly, two directions. One direction regards the redundant transforms, meaning that if the input signal has N samples, then we will obtain M output coefficients, with $M > N$. From this class of transforms the most important are those relying on the dual-tree implementation. In this case, the original signal is passed through 2 real DWT filter-bank trees, the resulting coefficients being complex, having for real part the output values of the first tree and for imaginary part the output values of the second tree. From this type of transform we can mention Kingsbury's Dual-Tree Complex Wavelet Transform (DT CWT) [Kin99], [Kin01] and another, quite similar, named as Selesnick's DT CWT [Sel04].

Another class of transforms includes the non-redundant ones, meaning that if the input signal has N samples, the transformation will provide us N output coefficients. This non-redundancy is achieved by mapping the signal into a complex-function space. An example of such a transform is given in [FvSB03] and will be further presented in this chapter.

3.4 Examples of Complex Wavelet Transforms

In the following I will present in detail some complex wavelet transforms, starting with the DT CWT, continuing with projection-based complex wavelet transform and finishing with my own contribution to this field, namely the Hyperanalytical Wavelet Transform.

3.4.1 Dual-Tree based Complex Wavelet Transforms

For many applications it is important that the transform be perfectly invertible. Some authors, including Lawton [Law93], have experimented with complex factorization of the standard Daubechies polynomials and obtained perfect reconstruction (PR) complex filters, but these do not give filters with good frequency selectivity properties. DT CWT comes with a different approach in order to overcome this drawback.

3.4.1.1 One-dimensional DT CWT

In 1998, Kingsbury first introduced the DT CWT [Kin98], that relies on the observation that approximate shift invariance can be achieved with a real DWT by doubling the sampling rate at each level of the tree. For this to work, the samples must be evenly spaced. The sampling rates can be doubled by eliminating the down-sampling by 2 after the level 1 filters. This is equivalent to having two parallel fully-decimated trees a and b, like in fig. 3.4, provided that the delays of H_{0b} and H_{1b} are one sample offset from H_{0a} and H_{1a}. Kingsbury found that, to get uniform intervals between samples from the two trees below level 1, the filters in one tree must provide delays that are half a sample different (at each filter input rate) from those in the other tree. This statement is also supported by Selesnick who, in [Sel01], gives an alternative derivation and explanation of the same result.

The implementation of such a transform is done using two mother wavelets, one for each tree, one of them being (approximately) the Hilbert transform of the other. On one hand, the dual-tree DWT can be viewed as an overcomplete wavelet transform with a redundancy factor of two. On the other hand, the dual-tree DWT is also a complex DWT, where the first and second DWTs represent the real and imaginary parts of a single complex DWT.

The first implementation proposed had the constraint of linear phase, and to accomplish this, the implementation required *odd-length* filters in one tree and *even-length* filters in the other. Greater symmetry between the two trees occurs if each tree uses odd and even filters alternately from level to level, but this is not essential.

In another implementation proposed in [Kin01], the condition of linear phase is dropped, resulting the so-called *Q-shift* dual tree. In this case all the filters beyond level 1 are even length and are designed to have a group delay of approximately $\frac{1}{4}$ sample $(+q)$. The required delay difference of $\frac{1}{2}$ sample $(2q)$ is then achieved by using the time reverse of the tree a filters in tree b so that the delay then becomes $3q$. Furthermore, because the filter coefficients are no longer symmetric, it is now possible to design the perfect-reconstruction filter sets to be orthonormal, resulting that all filters beyond level 1 are derived from the same orthonormal prototype set. The design of such Q-shift filters and of odd/even filters is quite complicated; it can be done only through approximations and is largely presented in [Kin01] and, respectively [Kin98].

In order to have a visual aspect of the DT CWT, we present in figure 3.4, the Q-shift version of the DT CWT as it is given in [Kin01].

Shift Invariance

In order to examine the shift invariance properties of a transform, Kingsbury [Kin01] proposes a method based on the retention of just one type (details or approximations), from just one level of the decomposition tree. For example one might choose to retain only the level-3 detail coefficients and set all the others to zero. If the signal y reconstructed from just these coefficients, is free of aliasing then it can be said that the transform is shift invariant at that level.

The degree of shift invariance of two implementation schemes (one for the DT CWT and the other for the classical DWT) is presented in fig. 3.5.

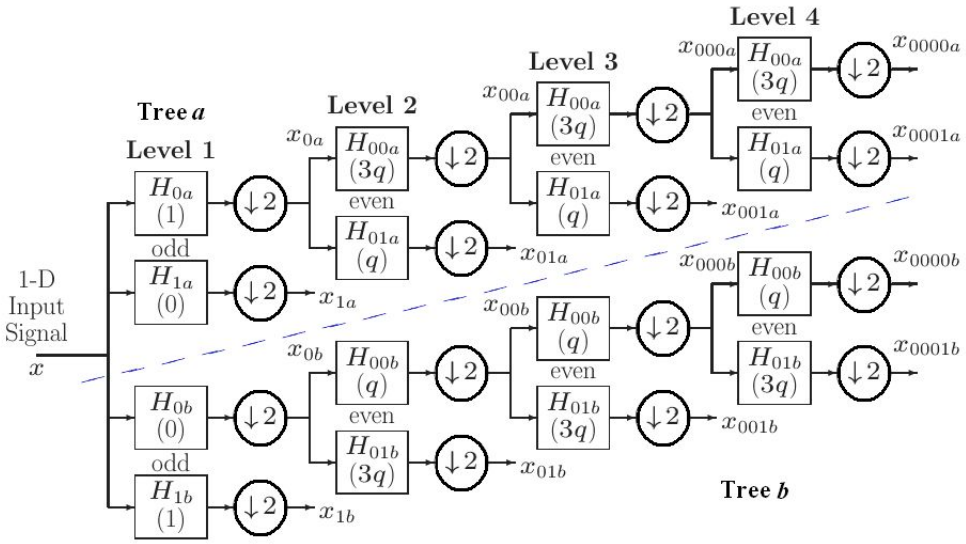


Figure 3.4: The Q-shift version of the DT CWT, giving real and imaginary parts of complex coefficients from tree a and tree b respectively.

In each case the input is a unit step, shifted to 16 adjacent sampling instants in turn. Each unit step is passed through the forward and inverse version of the chosen wavelet transform. The figure shows the input steps and the components of the inverse transform output signal, reconstructed from the wavelet coefficients at each of levels 1 to 4 in turn and from the scaling function coefficients at level 4. Summing these components reconstructs the input steps perfectly. Good shift invariance is shown when all the 16 output components from a given level have the same shape, independent of shift. It is easily observed that the DT CWT has outstanding performances in this direction compared to the severe shift dependence of the normal DWT.

3.4.1.2 Two-dimensional DT CWT

Extension of the DT CWT to two dimensions is achieved by separable filtering along columns and then rows. However, if column and row filters both suppress negative frequencies, then only the first quadrant of the 2-D signal spectrum is retained. It is well known, from 2-D Fourier transform theory, that two adjacent quadrants of the spectrum are required to represent fully a real 2-D signal. Therefore in the DT CWT it is also filtered with complex conjugates of the row (or column) filters in order to retain a second (or fourth) quadrant of the spectrum. This then gives 4:1 redundancy in the transformed 2-D signal.

A schematic representation of the 2D DT CWT based on the even-odd implementation was given by Jalobeanu et. al. [JBFZ03]. At level $m = 1$, the 2D DT CWT is simply a non-decimated wavelet transform (using a pair of odd-length filters h^o and g^o) whose coefficients are re-ordered into 4 interleaved images by using their parity. This defines the 4 trees $T = A$, B, C and D. If a and d denote approximation and detail coefficients

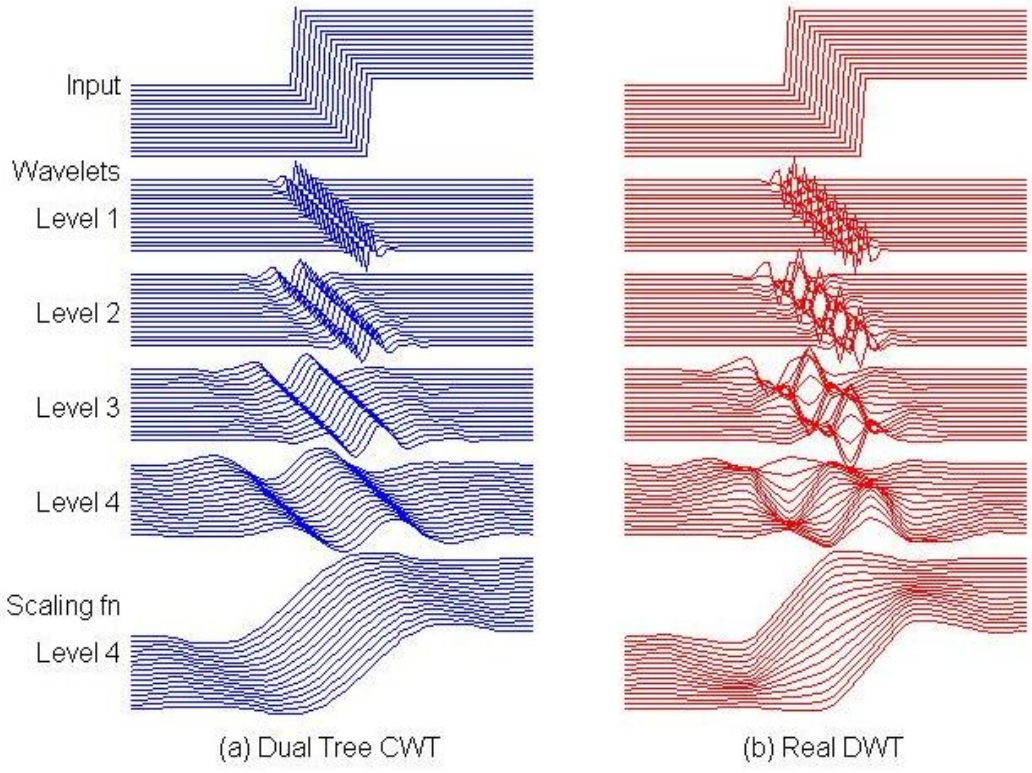


Figure 3.5: Detail and approximation components at levels 1 to 4 of 16 shifted step responses of the DT CWT (a) and real DWT (b)

($a^0 = x$, the input image), we have:

Tree T	A	B	C	D
$(a_T^1)_{x,y}$	$(a^0 * h^o h^o)_{2x,2y}$	$(a^0 * h^o h^o)_{2x,2y+1}$	$(a^0 * h^o h^o)_{2x+1,2y}$	$(a^0 * h^o h^o)_{2x+1,2y+1}$
$(d_T^{1,1})_{x,y}$	$(a^0 * g^o h^o)_{2x,2y}$	$(a^0 * g^o h^o)_{2x,2y+1}$	$(a^0 * g^o h^o)_{2x+1,2y}$	$(a^0 * g^o h^o)_{2x+1,2y+1}$
$(d_T^{1,2})_{x,y}$	$(a^0 * h^o g^o)_{2x,2y}$	$(a^0 * h^o g^o)_{2x,2y+1}$	$(a^0 * h^o g^o)_{2x+1,2y}$	$(a^0 * h^o g^o)_{2x+1,2y+1}$
$(d_T^{1,3})_{x,y}$	$(a^0 * g^o g^o)_{2x,2y}$	$(a^0 * g^o g^o)_{2x,2y+1}$	$(a^0 * g^o g^o)_{2x+1,2y}$	$(a^0 * g^o g^o)_{2x+1,2y+1}$

For all other scales ($m > 1$), the transform involves an additional pair of filters, even-length, denoted h^e and g^e. There must be a half-sample shift between the trees to achieve the approximate shift invariance. Therefore, different length filters are used for each tree, i.e. it is necessary to combine h^e, g^e with h^o, g^o, the 4 possible combinations corresponding to the 4 trees:

Tree T	A	B	C	D
$(a_T^{m+1})_{x,y}$	$(a_A^m * h^e h^e)_{2x,2y}$	$(a_B^m * h^e h^o)_{2x,2y+1}$	$(a_C^m * h^o h^e)_{2x+1,2y}$	$(a_D^m * h^o h^o)_{2x+1,2y+1}$
$(d_T^{m+1,1})_{x,y}$	$(a_A^m * g^e h^e)_{2x,2y}$	$(a_B^m * g^e h^o)_{2x,2y+1}$	$(a_C^m * g^o h^e)_{2x+1,2y}$	$(a_D^m * g^o h^o)_{2x+1,2y+1}$
$(d_T^{m+1,2})_{x,y}$	$(a_A^m * h^e g^e)_{2x,2y}$	$(a_B^m * h^e g^o)_{2x,2y+1}$	$(a_C^m * h^o g^e)_{2x+1,2y}$	$(a_D^m * h^o g^o)_{2x+1,2y+1}$
$(d_T^{m+1,3})_{x,y}$	$(a_A^m * g^e g^e)_{2x,2y}$	$(a_B^m * g^e g^o)_{2x,2y+1}$	$(a_C^m * g^o g^e)_{2x+1,2y}$	$(a_D^m * g^o g^o)_{2x+1,2y+1}$

The trees are processed separately, as in a real transform. The combination of odd and even filters depends on each tree. The transform is achieved by a fast Filter Bank (FB) technique, of complexity $O(N)$. The reconstruction is done in each tree independently, by using the dual filters. To obtain a^0, the results of the 4 trees are averaged. This ensures the symmetry between them, thus enabling the desired shift invariance. The complex coefficients are obtained by combining the different trees together. If the subbands are indexed by k, the detail subbands $d^{m,k}$ of the parallel trees A, B, C and D are combined to form complex subbands $z_+^{m,k}$ and $z_-^{m,k}$, by the linear transform:

$$\begin{aligned} z_+^{m,k} &= \left(d_A^{m,k} - d_D^{m,k}\right) + j\left(d_B^{m,k} + d_C^{m,k}\right) \\ z_-^{m,k} &= \left(d_A^{m,k} + d_D^{m,k}\right) + j\left(d_B^{m,k} - d_C^{m,k}\right) \end{aligned} \tag{3.8}$$

Shift Invariance

The main property of the 2D DT CWT is the quasi shift invariance, as shown by Kingsbury [Kin01] i.e. the magnitudes $|z_\pm|$ are nearly invariant to shifts of the input image. The shift invariance is perfect at level 1, and approximately achieved beyond this level: the transform algorithm is designed to optimize this property.

In fig. 3.5, the shift-dependence properties of the DT CWT were compared with the DWT for one-dimensional step functions. A similar comparison in the 2-D is presented in fig. 3.7. The input is now an image of a light circular disc on a dark background (see fig. 3.6). This circular form is suited for the analysis of the shift dependence in 2D as neighbor pixels from the contour of the disc can be interpreted as obtained through 2D shifts. The upper row of images, from left to right in fig. 3.7, show the components of the output image, reconstructed from the DT CWT wavelet coefficients at levels 1, 2, 3 and 4 and from the scaling function coefficients at level 4. The lower row of images show the equivalent components when the fully decimated DWT is used instead. In the lower row, we see substantial aliasing artifacts, manifested as irregular edges and stripes that are almost normal to the edge of the disc in places. Contrast this with the upper row of DT CWT images, in which artifacts are virtually absent. The smooth and continuous images here demonstrate good shift invariance because all parts of the disc edge are treated equivalently; there is no shift dependence.

Directional Selectivity

Complex filters in multiple dimensions can provide true directional selectivity, despite being implemented separably, because they are still able to separate all parts of the m-D frequency space. For example a 2D DT CWT produces six bandpass subimages of complex coefficients at each level, which are strongly oriented at angles $\pm 15°$, $\pm 45°$, $\pm 75°$, as illustrated by the level 4 impulse responses in fig. 3.8(a). In order to obtain these directional responses, it is necessary to interpret the scaling function (lowpass) coefficients from the two trees as complex pairs (rather than as purely real coefficients at double rate) so that they can be correctly combined with wavelet (highpass) coefficients, which are also complex, to obtain the filters oriented at $\pm 15°$ and $\pm 75°$ (see 3.8). The

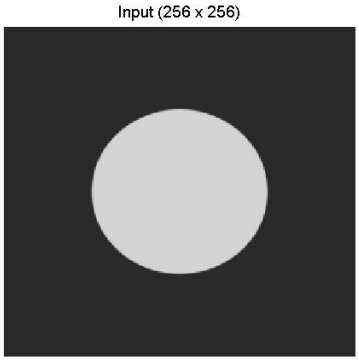


Figure 3.6: Input image used for the 2D shift sensitivity test

2D DT CWT is directionally selective because the complex filters can separate positive and negative frequency components (due to the analyticity of the transform) in 1D, and hence separate adjacent quadrants of the 2D spectrum.

In figure 3.9 is presented an example of directionally selective subbands obtained at four decomposition levels of the 2D DT CWT applied on the image 'Lena'. The positive orientations are grouped in the left part while the negative orientations are in the right part. The decomposition levels are represented in the 2D DWT's traditional manner; at each decomposition level there are three detail subbands for positive orientations and other three detail subbands for negative orientations. The similarity between the subimages corresponding to the same type of details of a given orientation from successive decomposition levels can be observed. The increasing of the absolute value of the details with the decomposition level can be also noticed. The image Lena is not symmetrical, the orientation of $45°$ being better represented than the orientation $-45°$ due to the contours of the hat. This asymmetry can be observed analyzing the corresponding subbands in figure 3.9.

Rotational Invariance

The directionality of the 2D DT CWT renders it nearly rotation invariant in addition to nearly shift invariant. Figure 3.7 illustrates the image obtained by reconstruction from only one level coefficients of the real DWT and of the DT CWT for a test image with a sharp edge on a hyperbolic trajectory. The ringing and aliasing artifacts in the DWT coefficients that change with the edge orientation are not present in the CWT coefficients. This may be due also to the fact that each image in 3.7 is using coefficients from all six directional subbands at the given wavelet level. The only rotational dependence is a slight thinning of the rings of the bandpass images near orientations of $\pm 45°$ and $\pm 135°$, due to the diagonal subbands having higher center frequencies than the others.

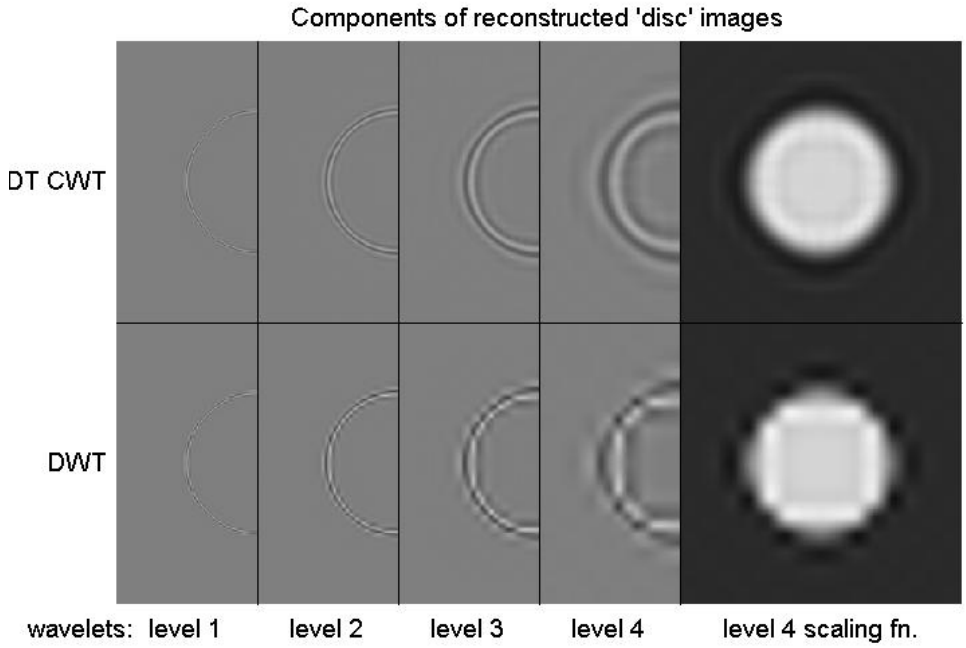


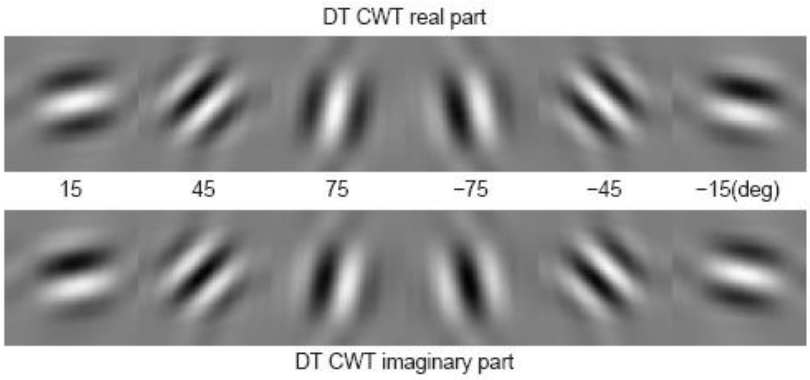
Figure 3.7: Wavelet and scaling function components at levels 1 to 4 of an image using the 2D DT CWT (upper row) and 2D DWT (lower row). Only half of each wavelet image is shown in order to save space.

3.4.2 Projection-based CWTs

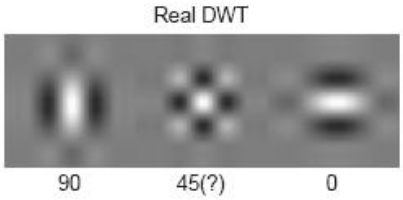
Another class of Complex Wavelet Transforms is represented by the set of projection-based non-redundant complex wavelet transforms (NRCWT), introduced by Fernandes et. al., [FvSB03].

The projection (mapping) represents the conversion of a real signal to an analytic (complex) form through digital filtering. NRCWT is basically the DWT of the complex-valued projection. While this type of transforms are restricted to IIR filters, they produce orthogonal solutions. Fernandes' projection-based CWT (PCWT) uses flexible design techniques to trade-off between redundancy and shift-invariance. The implementation of the mapping-based complex wavelet transform is presented in fig. 3.10. The forward CWT consists of an arbitrary DWT filter-bank preceded by a mapping stage. The CWT is then inverted by appending an inverse-mapping stage after an inverse-DWT filterbank. The independence of the two stages in the PCWT implementation (the complex mapping and the DWT) allows them to be performed separately and alternatively, leading toward a greater flexibility through the implementation.

In true sense, all of the class of NRCWT, which are designed to mitigate all three disadvantages of standard DWT, are not exactly non-redundant CWT. For instance, Fernandes' implementation has a redundancy factor of 2.67 in two dimensions. These complex wavelet transforms are considered NRCWTs because of their two main design constraints. The first condition that must be fulfilled by these transforms is to offer



(a) DT CWT



(b) DWT

Figure 3.8: Basis functions of 2D Q-shift complex wavelets (top) and 2D real wavelet filters (bottom), all illustrated at level 4 of the transforms. The complex wavelets provide 6 directionally selective filters, while real wavelets provide 3 filters, only two of which have a dominant direction

controllable redundancy, equal or less to the redundancy of the DT CWT, while preserving all the benefits introduced by the DT CWT. Second constraint is to offer improved directionality and phase information with perfect non-redundancy at the cost of increased shift-sensitivity. Unlike other redundant complex wavelet transforms, such as DT CWT, that also mitigate DWT shortcomings, the decoupled implementation of PCWT (from a class of NRCWT) (see fig. 3.10) has two important advantages. First, the controllable redundancy of the mapping stage offers a balance between the degree of shift sensitivity and the transform's redundancy. The second advantage of PCWT is the flexibility to use any DWT in the transform implementation.

Analyzing the DT CWT, Magarey and Kingsbury [KM97] showed that the complex coefficient obtained at the output of this transform can also be obtained using a single DWT tree based on a quasi-analytic wavelet. Fernandes et al. [FvSB03] pointed out that there is yet another interpretation for the DT CWT coefficients: they may be interpreted as the DWT of the analytic signal associated with the input signal. For an $L^2(R)$ input, the analytic signal is defined in this case as the Hardy-space image of the input signal.

The classical Hardy-space $H^2(R \to C)$ is defined by:

$$H^2(R \to C) \equiv \left\{ f \in L^2(R \to C) : F(\omega) = 0 \text{ for a. e. } \omega < 0 \right\}.$$

Let the superscript 'H' denote the isomorphic Hardy-space image (or Hardy-space map-

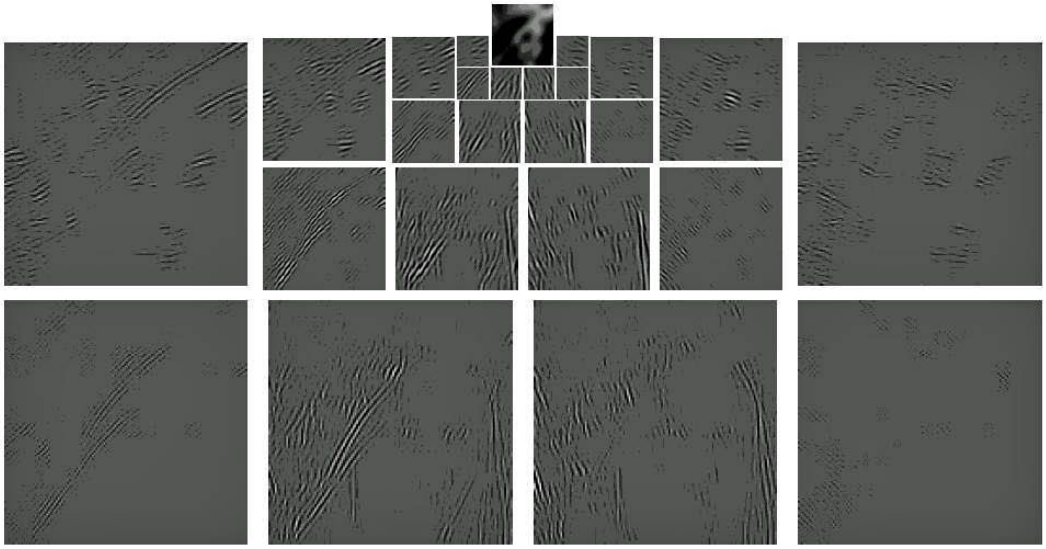


Figure 3.9: Components from each subband of the reconstructed output image for a 4-level 2D DT CWT decomposition of Lena (central part (128x128) only)

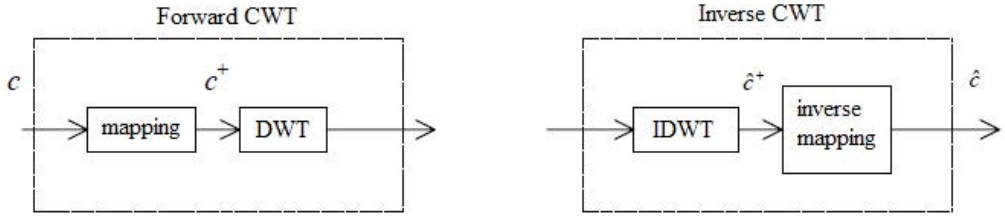


Figure 3.10: Projection-based CWT and its inverse

ping) of an $L^2(R)$ function. Then, the Hardy-space image of a function f is given by

$$F^H(\omega) = F(\omega)\,\chi_{[0,\infty)}(\omega),$$

resulting that f^H has half the bandwidth of f.

The other important term related to this complex mapping is 'unitary map'. The unitary map is a linear, bijective, inner-product preserving map. The function spaces that are related through unitary map are called isomorphic to each other. The unitary map or isomorphism is realizable if it can be implemented through digital filter. Thus, complex or Hardy space projection of a real signal is said to be realizable if it is possible to use some form of digital filtering for the mapping.

If we consider f to be the projection of an $L^2(R)$ function onto the scaling space V_1 so that $f(x) = \sum_n c(n)\,\phi(x-n)$ or, equivalently, $F(\omega) = C(\omega)\,\Phi(\omega)$ then, the Hardy-space image f^H is given by $F^H(\omega) = \chi_{[0,\infty)}(\omega)\,C(\omega)\,\Phi(\omega)$. Unfortunately, since $\chi_{[0,\infty)}(\omega)$ is not 2π-periodic, it cannot be applied to the scaling-coefficient sequence c using a digital filter, meaning that the Hardy-space mapping of an $L^2(R)$ signal is impossible to compute.

To circumvent this obstacle, Fernandes [FvSB03] defined a new function space called the Softy-space, S^+. The Softy-space is an approximation to Hardy-space and, moreover, the Softy-space mapping of an $L^2(R)$ signal may be computed using a digital mapping filter h^+ with magnitude response shown in fig. 3.11.

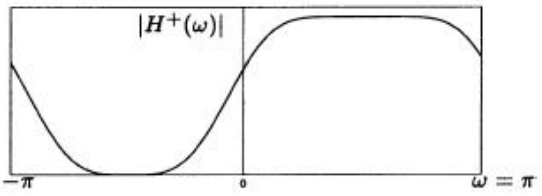


Figure 3.11: $|H^+(\omega)|$, the magnitude response of the mapping filter h^+

By suppressing the negative frequencies of an input signal, h^+ maps the signal onto the Softy-space. The triangle in fig. 3.12 depicts the relationship between these function spaces.

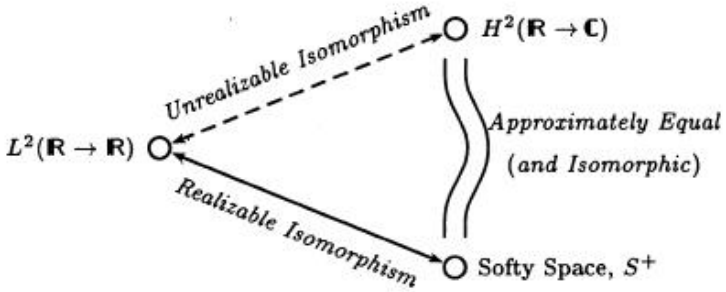


Figure 3.12: Relationship between $L^2(R)$, Hardy-space and Softy-space

The PCWT is defined to be the DWT of the Softy-space image of an $L^2(R)$ function, as shown in fig. 3.13. The mapping-filter h^+ is used to project an $L^2(R)$ function f onto the Softy-space. Since h^+ is a complex-coefficient filter, it introduces a redundancy factor of two when applied to a real-valued scaling-coefficient sequence. It is then computed the DWT of the Softy-space mapping f^+ to obtain the CWT coefficients that will undergo application-specific processing. After the IDWT, an inverse-mapping filter g^+ computes the $L^2(R)$ processed function f. Since $f^+ \approx f^H$, the CWT will have approximate shiftability.

In order to eliminate the data redundancy, the non-redundant mapping scheme shown in fig. 3.14 is proposed. As depicted in the figure, the non-redundant mapping is defined as the concatenation of a mapping filter and a downsampler (elimination of odd-indexed scaling coefficients). The downsampler eliminates the redundancy created by the mapping filter that generates complex scaling coefficients from real scaling coefficients. It can be observed that the scaling-coefficient sequences c and $\widetilde{c}^+$ can both be represented by N real numbers within a digital computer; therefore, there is no data redundancy in the scaling-coefficient sequence and $\widetilde{c}^+$.

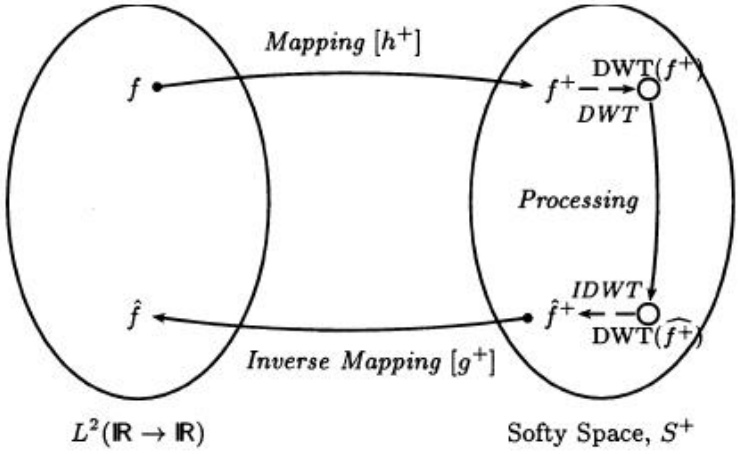


Figure 3.13: PCWT

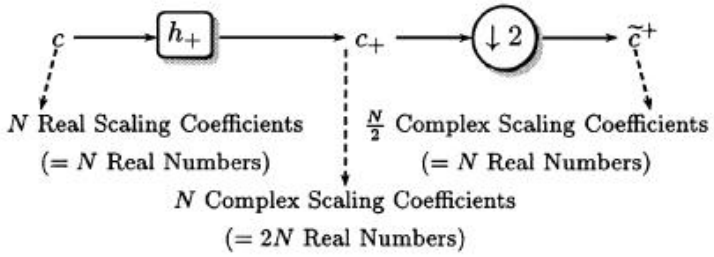


Figure 3.14: Non-redundant mapping

3.4.3 Hyperanalytic Wavelet Transform

In this section, a very simple implementation of the HWT is introduced. I will begin by presenting the 1D case, highlighting the shift-invariance properties and I will continue with the 2D case and insist both on the quasi shift-invariance and directional selectivity.

3.4.3.1 Analytical Discrete Wavelet Transform (ADWT)

As discussed in the previous chapter, a 1D wavelet transform (WT), is shift-sensitive if an input signal shift causes an unpredictable change of the transform coefficients. Shift-sensitivity is an undesirable property because it implies the impossibility to distinguish between wavelet transform coefficients corresponding to input signal shifts. The shift-sensitivity of the DWT is generated by the down-samplers used for its computation.

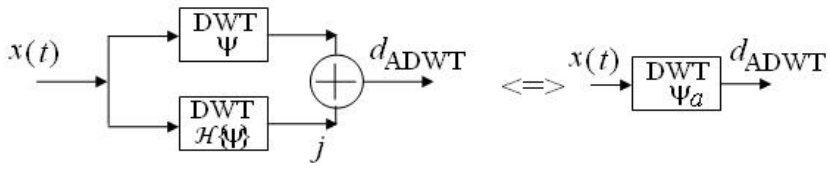
Throughout time, there were numerous attempts to overcome this disadvantage. A straight-forward solution is to remove the downsamplers from the DWT's implementation, resulting the UDWT [Mal99], [LGO+96]. Although the UDWT is shift-insensitive, it has high redundancy (equal to 2^J, where J denotes the number of iterations), caused by the absence of down-samplers. Unfortunately, the high redundancy incurs a massive storage requirement that makes the UDWT inappropriate for most signal processing

applications. Another disadvantage of the UDWT comes from the fact that it requires the implementation of a large number of different filters.

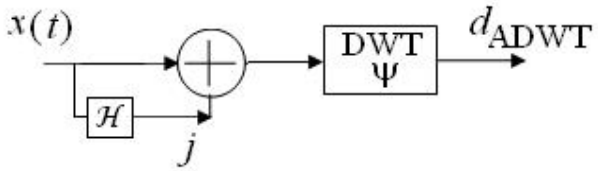
Lang, Guo, Odegard, Burrus and Welles [LGO+96] have proposed a new shift-invariant but very redundant wavelet transform, named Shift Invariant Discrete Wavelet Transform, SIDWT. Their proposition is based on a translation invariant algorithm proposed by Coifman and Donoho [CD95]. The computation of this transform implies the consideration of all circular shifts of the input signal. After the computation of the DWT of every shifted version of the signal, this method requires the shifting back (or unshifting) and averaging over all results obtained.

The method introduced by Coifman and Donoho in [CD95] and called Cycle Spinning (CS) was conceived to suppress the artifacts in the neighborhood of discontinuities introduced by the classical DWT in denoising applications, and it implies the rejection of the translation dependence. For a range of shifts, data (time samples of a signal) is shifted (right or left as the case may be), the DWT of shifted data is computed, and than the result is un-shifted. Doing this for a range of shifts, and averaging the several results so obtained, a quasi shift-invariant discrete wavelet transform is obtained. The degree of redundancy of this transform is proportional to the number of shifts of the input signal produced. Cycle spinning over the range of all circular shifts of the input signal is equivalent to SIDWT.

Abry [Abr94], first demonstrated that approximate shiftability is possible for the DWT with a small, fixed amount of transform redundancy. He designed a pair of real wavelets such that one is approximately the Hilbert transform of the other. This wavelet pair defines a CWT. For explaining that such a transform is complex, consider the pair of DWT trees associated with the wavelet pair already mentioned. A complex wavelet coefficient is obtained by interpreting the wavelet coefficient from one DWT tree as being its real part, whereas the corresponding coefficient from the other tree is interpreted as its imaginary part. This transform is represented in figure 3.15(a).



(a) Previous ADWT implementations



(b) Proposed ADWT implementation

Figure 3.15: Equivalent implementations of the ADWT

In 3.15(b) we propose a new implementation of the complex wavelet transform, similar to the DTCWT but easier to implement, [AOB06]. It involves computing a single

DWT but, instead of applying it to the original signal we apply it to the analytical signal associated with our input signal. The analytic signal is defines as in 3.2. These implementations of the ADWT, presented in fig. 3.15 are equivalent because, taking into consideration the properties of the inner product and the 'anti-self adjointness' of the Hilbert transform, namely

$$\langle H\{u\}, v\rangle = \langle u, -H\{v\}\rangle \text{ with } u \in L^p(R) \text{ and } v \in L^q(R),$$

we can write the wavelet coefficients of the ADWT, at level m, d^m_{ADWT} as follows:

$$\begin{aligned} d^m_{ADWT}[n] &= \left\langle x(t), \psi^m_{a,n}(t)\right\rangle \\ &= \langle x(t), \psi^m_n(t) + jH\{\psi^m_n(t)\}\rangle \\ &= \langle x(t), \psi^m_n(t)\rangle - j\langle x(t), H\{\psi^m_n(t)\}\rangle \\ &= \langle x(t), \psi^m_n(t)\rangle + j\langle H\{x(t)\}, \psi^m_n(t)\rangle \\ &= \langle x(t) + jH\{x(t)\}, \psi^m_n(t)\rangle \\ &= \langle x_a(t), \psi^m_n(t)\rangle \end{aligned} \tag{3.9}$$

In fact neither the DTCWT, nor the proposed implementation of ADWT correspond to perfect analytic mother wavelets, because the exact digital implementation of a Hilbert transform pair of mother wavelets with good performance is not possible in the case of the first transform (due to the fractional delay between the two trees required) and because the digital Hilbert transformer is not a realizable system in the case of the second transform, in practice this transformation being done only through approximations. Our proposal requires a implementation of the Hilbert transformer acting in the domain of the Fast Fourier Transform, as in figure 3.16. In simulations we have used the processing

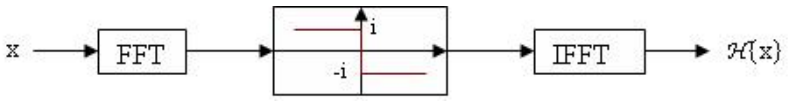


Figure 3.16: The implementation of the Hilbert transformer

scheme in 3.17.

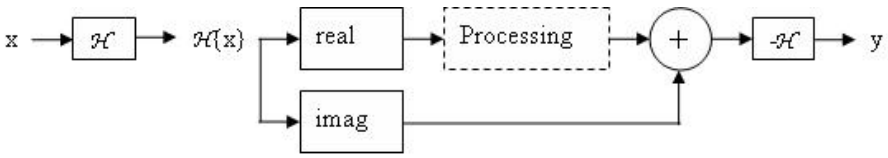


Figure 3.17: The use of the Hilbert transform in simulations

The DTCWT requires special mother wavelets but can assure a higher degree of shift invariance whereas the implementation of the ADWT proposed in figure 3.15(b) can use any classical mother wavelets (e.g. those from the Daubechies family, Symmlet family and others). Both transforms have in the 1D case a redundancy of 2.

Shift Invariance

In order to evaluate the shift-invariance performance of ADWT, two types of simulations are performed. First, we make a visual evaluation of the degree of shift invariance by taking the same test prof. Kingsbury has proposed in [Kin01]. We have used this test to make a comparison between ADWT, DT CWT, and the classical DWT. As can be seen, we obtain results comparable with those obtained by Kingsbury, even though we have used Daubechies, 10-taps filter. From fig. 3.18 it can be observed that the DWT is not

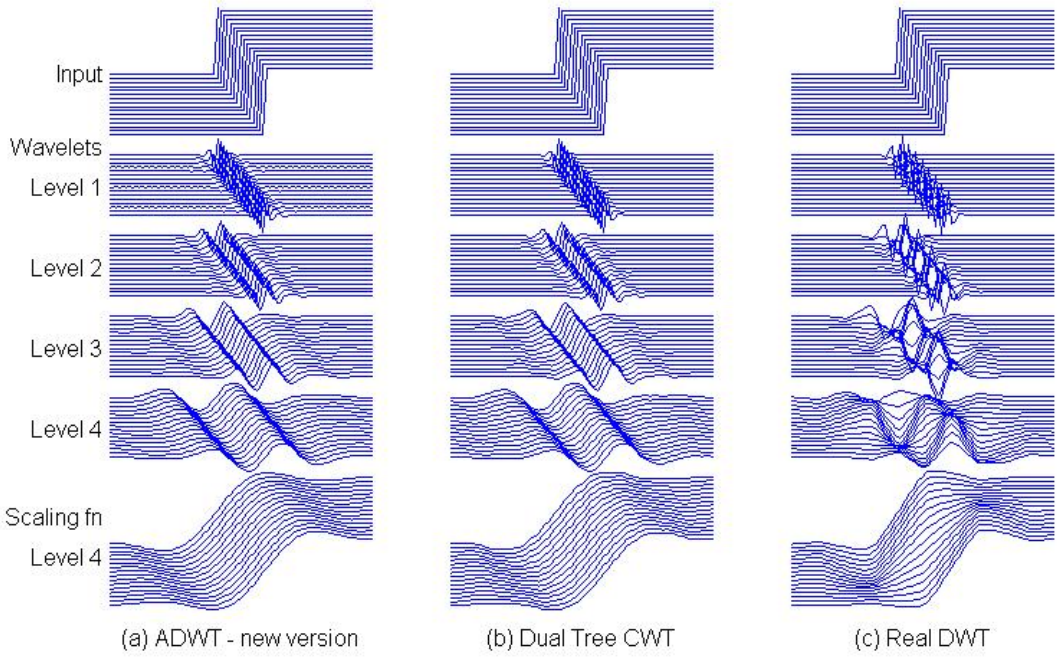


Figure 3.18: A visual comparison ment to illustrate the shiftability of ADWT, DT CWT and DWT

shift-invariant; the lines of coefficients corresponding to different shift are not parallel, while the ADWT and DT CWT are quasi shift-invariant.

The second series of tests involves the computation of a 'degree of shift-invariance' that I have introduced. In [SFAH92], Simoncelli has defined a new measure of the shift-invariance, called 'shiftability'. According to their definition, a transform is shiftable if and only if any subband energy of the transform is invariant under input-signal shifts. Although weaker than shift invariance, shiftability is important for applications because it is equivalent to interpolability, which is a property ensuring the preservation of transform-subband energy under input-signal shifts. Based on this observation, we introduced a new criterion: the degree of shift invariance. In order to calculate this measure, we compute the energies of every set of detail coefficients (at different decomposition levels) and of the approximation coefficients, corresponding to a certain delay (shift) of the input signal samples. This way, I obtain a sequence of energies at each decomposition level, each

sample of this sequence corresponding to a different shift. Then the mean m and the standard deviation d of every energy sequence are computed. My degree of invariance is defined as:

$$Deg = 1 - d/m \tag{3.10}$$

I perform the normalization with respect to the mean of the energy sequence because I want the values of the degree of invariance to be within the interval $[0, 1]$, in order to facilitate its interpretation.

If the transform is shift-invariant, then the value of its degree of invariance is 1 because the standard deviation of the energy sequence is zero in this case. The reciprocity is not guaranteed. There are quasi shift invariant wavelet transforms with the degree of shift-invariance equal to 1 that are not perfectly shift-invariant. However, generally, when the transform is not shift-invariant the value of this degree of invariance is smaller than 1. This observation is also sustained by experimental work.

We consider that the degree of shift invariance is an objective way of analyzing the shift invariance of a transform.

In simulations, I have used as input signal a unitary step, like in 3.18. In fact, 16 different unitary steps were used. They were generated one from another by delaying with a sample. The number of iterations used for the computation of the DWT was 3. I repeated the simulations for several mother wavelets commonly used in the literature (Daubechies, Symmlet and Coiflet). In order to isolate the coefficients corresponding to each level, after the computation of the corresponding forward WT, all the coefficients corresponding to the other levels are set to zero, by applying a 'mask' on the sequence obtained. For a better understanding of this procedure, I illustrate in figure 3.19 the system used for the analysis of the shift invariance at the 3rd decomposition level of the ADWT.

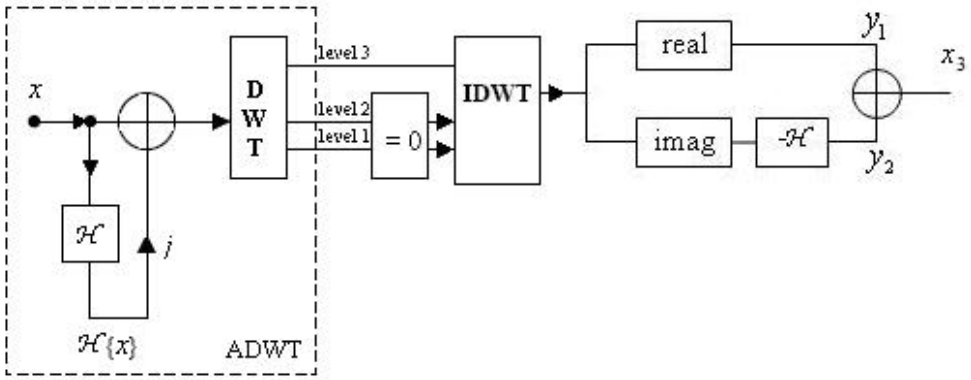


Figure 3.19: The system used for the shift-invariance analysis of the third level of the wavelet decomposition. In this example is considered the case of the proposed implementation of ADWT

The inverse ADWT (IADWT) can be implemented using the systems proposed in [FIB09], [ANBI07b], [FIBI09] and [FNBI09].

In the first test involving the values of the degree of shift invariance I have compared its values obtained for our transform with the values obtained in the case of the CS with

a various number of cycle spins and for a variety of spinning steps (a spinning step is the number of samples the signal is shifted once). In table 3.1 I present a comparison between my transform and the CS. This comparison is based on the values of the degree of shift invariance calculated for the approximation coefficients obtained after the 3rd iteration of the corresponding WT computation algorithm (Scaling fn., level 3), for the detail coefficients obtained after the 3rd iteration (Wavelets level 3), for the detail coefficients obtained after the 2nd iteration (Wavelets level 2) and for the detail coefficients obtained after the 1st iteration (Wavelets level 1). It can be observed, analyzing this table, that the ADWT is equivalent to the CS with redundancy 64, from the degree of shift-invariance point of view.

Symmlet, 10	ADWT	CS step=1,64 delays	CS step=1, 512 delays
Redundancy	2	64	512
Scaling fn. level 3	0,8594	0,7551	0,7551
Wavelets level 3	0,9962	0,9962	0,9995
Wavelets level 2	0,9963	0,9965	0,9996
Wavelets level 1	0,9992	0,9985	0,9998
Daubechies, 10	ADWT	CS step=1, 64 delays	CS step=1, 512 delays
Scaling fn. level 3	0,8594	0,7551	0,7551
Wavelets level 3	0,9981	0,9965	0,9996
Wavelets level 2	0,9982	0,9968	0,9996
Wavelets level 1	0,9992	0,9985	0,9998

Table 3.1: A comparison between two quasi shift-invariant WTs, the ADWT and the CS

Another test involved the dependency of the degree of shift invariance of the proposed WT with respect to the regularity of the mother wavelet used for its computation. I investigated the Daubechies family, each element being indexed by its number of vanishing moments. As the curve illustrated in figure 3.20 indicates, the degree of shift-invariance increases with the regularity of the mother wavelets used.

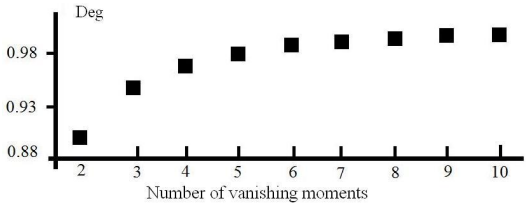


Figure 3.20: The dependency of the degree of shift-invariance of HWT on the regularity of the mother wavelet used for its computation

Finally, in table 3.2 I have put the values of this degree computed for the signals deployed in fig. 3.18. As it can be seen, my new implementation is comparable with the DT CWT both from the visual point of view and by analyzing the values of the degree of invariance.

WT	Degree of invariance				
	W. lev. 1	W. lev. 2	W. lev. 3	W. lev. 4	Sn. lev. 4
ADWT	0.9967	0.9995	0.9986	0.9988	0.9990
DT CWT	1.0000	0.9811	0.9749	0.9734	0.9997
DWT	0.9236	0.8265	0.7878	0.8149	0.9958

Table 3.2: A comparison of ADWT, DT CWT and DWT

3.4.3.2 Hyperanalytic Wavelet Transform (HWT)

All the 1D WTs already mentioned have simpler or more complicated 2D generalizations. The generalization of the analyticity concept in 2D is not obvious, because there are multiple definitions of the Hilbert transform in this case. In the following I will use the definition of the analytic signal associated to a 2D real signal named hypercomplex signal. So, the hypercomplex mother wavelet associated to the real mother wavelet, $\psi(x,y)$ is defined as:

$$\psi_a(x,y) = \psi(x,y) + iH_x\{\psi(x,y)\} + jH_y\{\psi(x,y)\} + kH_x\{H_y\{\psi(x,y)\}\} \tag{3.11}$$

where $i^2 = j^2 = -k^2 = -1$, $ij = ji = k$, $jk = kj = -i$, $ki = ik = -j$ and $ijk = 1$, [Dav].

The HWT of an image $f(x,y)$ is:

$$HWT\{f(x,y)\} = \langle f(x,y), \psi_a(x,y)\rangle. \tag{3.12}$$

Replacing 3.11 in 3.12, we will obtain:

$$\begin{aligned}
HWT\{f(x,y)\} &= \langle f(x,y), \psi(x,y)\rangle + \langle f(x,y), iH_x\{\psi(x,y)\}\rangle + \\
&\quad + \langle f(x,y), jH_y\{\psi(x,y)\}\rangle + \langle f(x,y), kH_x\{H_y\{\psi(x,y)\}\}\rangle = \\
&= \langle f(x,y), \psi(x,y)\rangle - i\langle f(x,y), H_x\{\psi(x,y)\}\rangle - \\
&\quad - j\langle f(x,y), H_y\{\psi(x,y)\}\rangle + k\langle f(x,y), H_x\{H_y\{\psi(x,y)\}\}\rangle = \\
&= \langle f(x,y), \psi(x,y)\rangle - i\langle -H_x\{f(x,y)\}, \psi(x,y)\rangle - \\
&\quad - j\langle -H_y\{f(x,y)\}, \psi(x,y)\rangle + k\langle -H_x\{f(x,y)\}, H_y\{\psi(x,y)\}\rangle = \\
&= \langle f(x,y), \psi(x,y)\rangle + i\langle H_x\{f(x,y)\}, \psi(x,y)\rangle + \\
&\quad + j\langle H_y\{f(x,y)\}, \psi(x,y)\rangle - k\langle -H_y\{H_x\{f(x,y)\}\}, \psi(x,y)\rangle = \\
&= DWT\{f(x,y)\} + iDWT\{H_x\{f(x,y)\}\} + \\
&\quad + jDWT\{H_y\{f(x,y)\}\} + kDWT\{H_y\{H_x\{f(x,y)\}\}\} = \\
&= \langle f_a(x,y), \psi(x,y)\rangle = DWT\{f_a(x,y)\}.
\end{aligned} \tag{3.13}$$

Based on 3.13 we can conclude that the HWT of the image $f(x,y)$ can be computed with the aid of the 2D DWT of its associated hyper complex image. In consequence the HWT implementation uses four trees, each one implementing a 2D DWT, thus having a redundancy of four. The first tree is applied to the input image. The second and the third trees are applied to 1D Hilbert transforms computed across the lines (H_x) or columns (H_y) of the input image. The fourth tree is applied to the result obtained

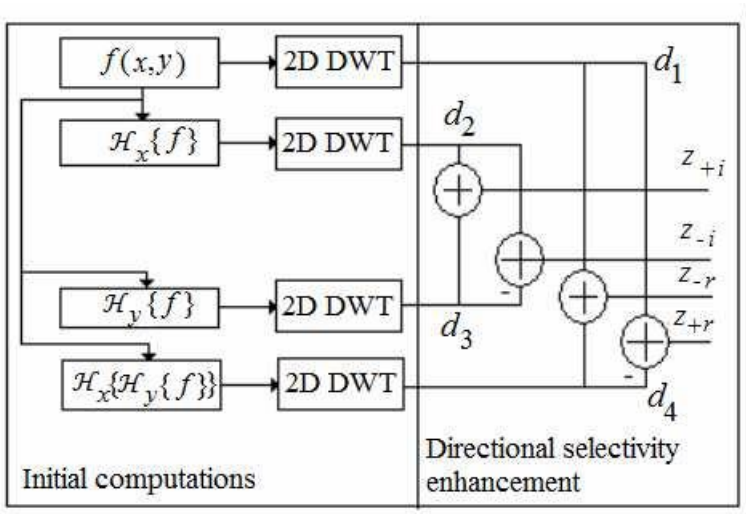


Figure 3.21: HWT implementation scheme

after the computation of the two 1D Hilbert transforms of the input image. The HWT implementation is presented in figure 3.21, [ANBI07a].

In what concerns the reconstruction part, a possible method would be to reverse the Hilbert transform applied at the input (both Hilbert transforms in the last case) for each of the three trees mentioned above, and the final result to be the mean of the four images computed on the four trees.

Shift Invariance

To illustrate the quasi shift-invariance properties of the HWT, I have, once again, retaken one of Kingsbury's tests, presented in fig. 3.7, the results obtained being depicted in fig. 3.22. In this case the wavelet used in the computation of the HWT has been Daubechies with 10 vanishing moments. Analyzing the results in 3.22 we can conclude that the two complex WTs outperform the 2D DWT, the behavior of the HWT being quite similar to the behavior of the 2D DTCWT.

Directional Selectivity

The enhancement of the directional selectivity of the HWT is made, as in the case of the 2D-DTCWT, [Kin98], [SBK05], through linear combinations of detail coefficients belonging to each subband of each of the four 2D-DWTs (see the right side of fig. 3.21).

Let us consider, for example, the case of the diagonal detail subbands, (HH), presented in figure 3.23. I have selected a particular input image, $f(x,y)) = \delta(x,y)$, to appreciate the frequency responses associated to different transfer functions represented in figure 3.21 [ANBI07b]. More precisely, the example in figure 3.23 refers to the transfer functions that relate the input f with the outputs z_{-r} and z_{+r}.

The spectrum of the input image, $F\{\delta(x,y)\}(f_x, f_y)$ is constant. The wavelet coefficients belonging to the subband HH are obtained by lines and columns high-pass filtering.

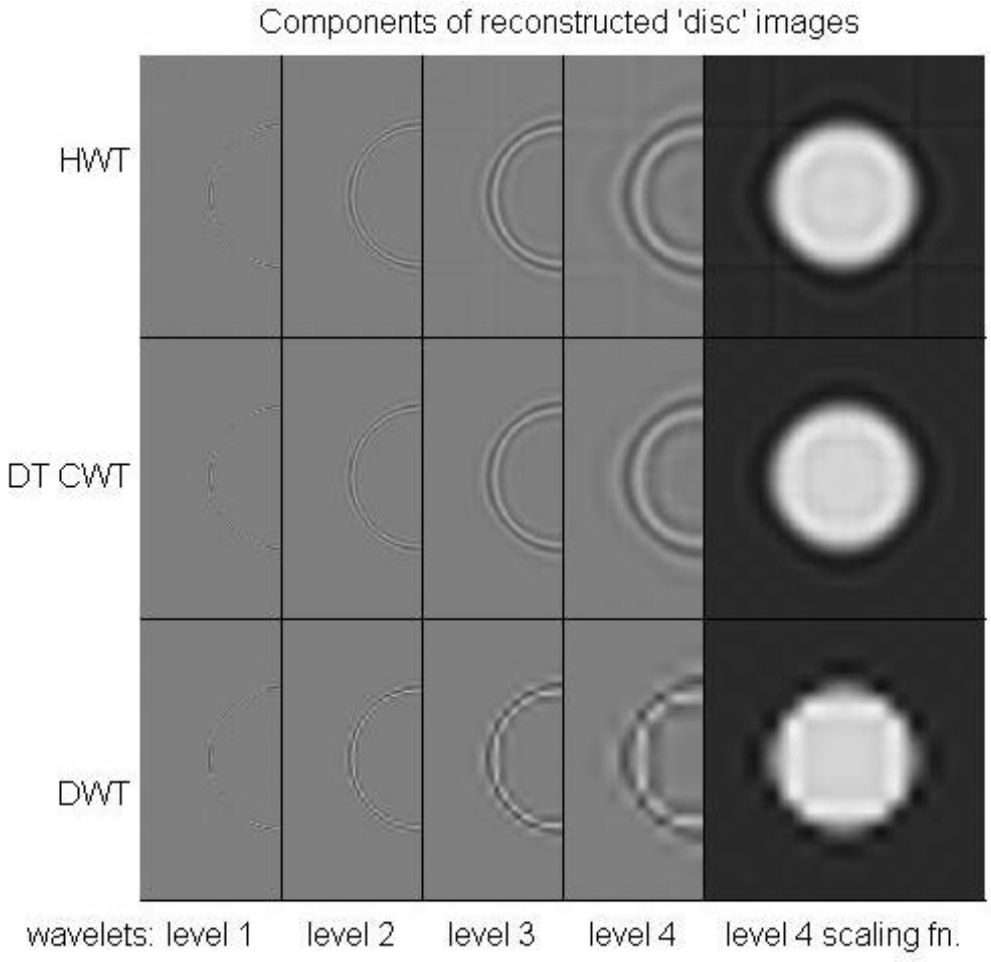


Figure 3.22: Comparison in the 2D case between the HWT, the DT CWT and the DWT

I have assumed ideal high-pass filters. The spectra of the wavelet coefficients d_1, d_2, d_3, d_4 belonging to the subband HH, denoted by $F\left\{DWT_{HH}\left\{\delta\left(x,y\right)\right\}\right\}$, $F\left\{DWT_{HH}\left\{H_x\left\{\delta\left(x,y\right)\right\}\right\}\right\}$, $F\left\{DWT_{HH}\left\{H_y\left\{\delta\left(x,y\right)\right\}\right\}\right\}$ and, respectively, $F\left\{DWT_{HH}\left\{H_y\left\{H_x\left\{\delta\left(x,y\right)\right\}\right\}\right\}\right\}$, have two preferential orientations, corresponding to the two diagonals $\pm\pi/4$. These directions are the result of the fact that 2D DWT cannot separate these two orientations.

After the linear combinations proposed, we can observe that the spectra of the coefficients so obtained, for example z_{-r} and z_{-r}, $F\left\{HH_{z_{-r}}\right\}\left(f_x,f_y\right)$ and $F\left\{HH_{z_{+r}}\right\}\left(f_x,f_y\right)$, have only one preferential direction, namely the second diagonal, respectively the first one. In conclusion, by using the HWT these directions can be separated. The same strategy can be used to enhance the directional selectivity in the other two subbands: LH and HL, obtaining the preferential orientations at $\pm atan\left(2\right)$ and $\pm atan\left(1/2\right)$.

A comparison of the directional selectivity of the 2D DWT and HWT, implemented as proposed in figure 3.21, is presented in figure 3.24. I have conceived a special input image, in the frequency domain, to conduct this simulation. Its spectrum, represented in figure 3.24, is oriented following the directions: 0, $\pm atan\left(1/2\right)$, $\pm\pi/4$, $\pm atan\left(2\right)$ and π.

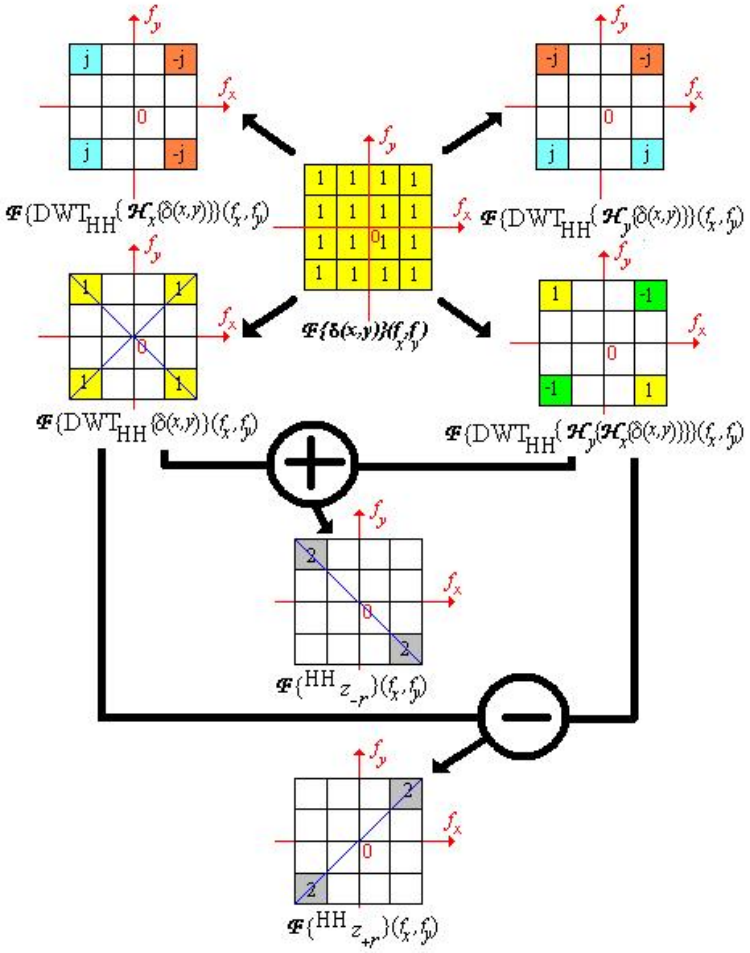


Figure 3.23: The strategy of directional selectivity enhancement in the HH subband illustrated through the transfer functions of the systems used in the HWT implementation

Like the 2D DTCWT, the HWT implemented as proposed in figure 3.21, has six preferential orientations: $\pm atan\,(1/2)$, $\pm\pi/4$ and $\pm atan\,(2)$. The 2D-DWT has only three preferential orientations: 0, $\pi/4$ and $\pi/2$, it does not make the difference between the two principal diagonals. The better directional selectivity of the proposed implementation of HWT versus the 2D DWT can be easily observed, comparing the corresponding detail sub-images in figure 3.24. For the diagonal detail sub-images, for example, the imaginary part of the HWT rejects the directions: $-atan\,(1/2)$, $-\pi/4$ and $-atan\,(2)$, whereas the 2D DWT conserves these directions.

3.5 Advantages and Applications of Complex Wavelet Transforms

The advantages of the complex wavelet transforms lay in the reason they were introduced in the first place, namely decreased shift sensitivity and increased directional selectivity.

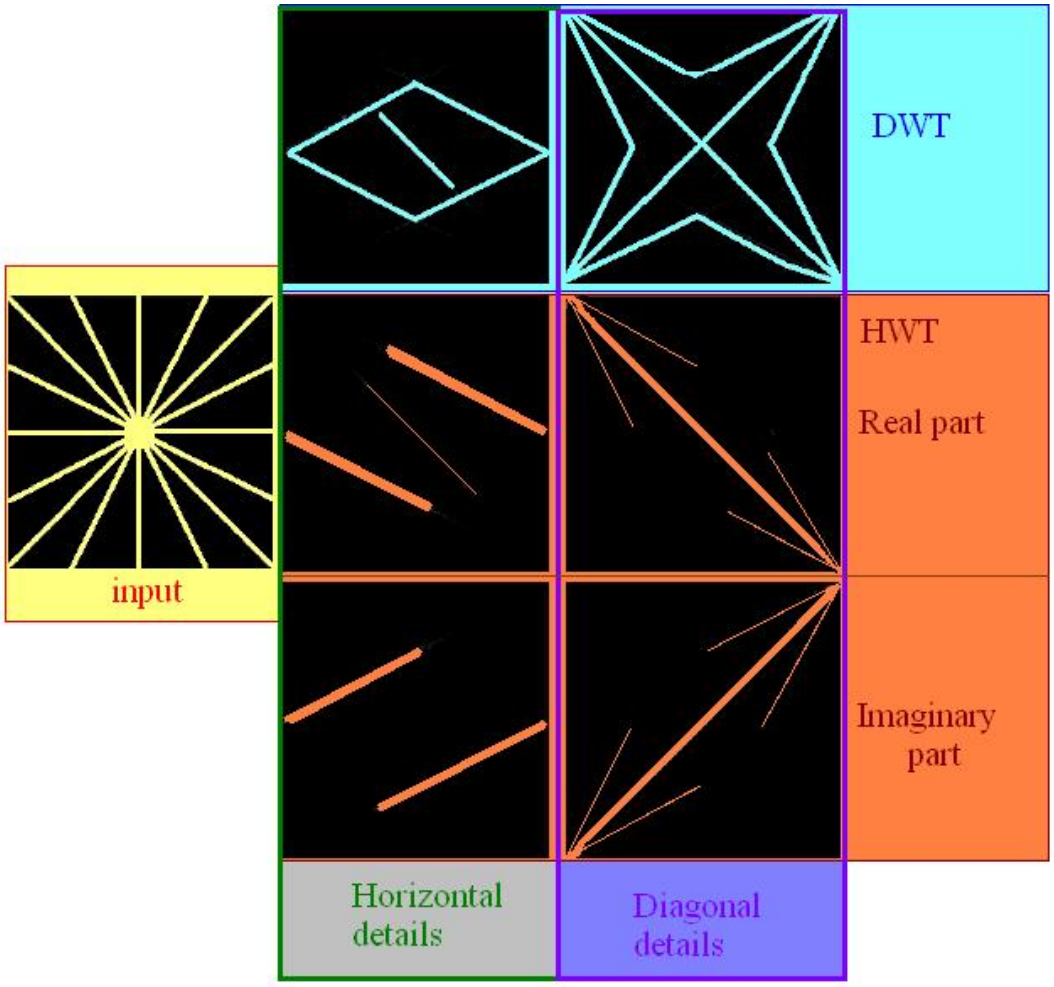


Figure 3.24: The absolute values of the spectra of horizontal and diagonal detail sub-images obtained after the first iterations of 2D DWT and HWT. In the HWT case, the real and imaginary parts of complex coefficients are separated

Although these goals might be achieved with real wavelet transforms (UDWT is shift invariant and DWPT offers good directional selectivity), complex wavelets transforms manage to obtain good results in these directions while preserving a reasonable amount of redundancy.

When talking about the specific complex wavelet transforms presented above, each of them has advantages and limitations, the latter resulting mainly from the fact that there is impossible to compute an analytic signal using digital filters.

In what concerns the applications of the complex wavelet transforms, they cover mainly the area of signal and image processing were the advantages introduced with the price of increased redundancy are significant. A description of these advantages is nicely presented in [Kin01], [Shu03] and [SBK05].

One of the most important applications in signal and image processing is denoising. In what concerns signals, apparently many denoising methods perform better when associated with shift-invariant wavelet transforms. In the case of images, the increased di-

rectional selectivity in addition to shift-invariance leads to better performances especially when it comes to images containing significant diagonal edges. Directional selectivity is a desirable feature when dealing with textures, in applications like texture analysis and synthesis and also in image segmentation and classification.

Other possible applications include motion estimation and compensation, watermarking, deconvolution, coding, traffic applications [SI10] and so on.

3.6 Summary

In this chapter I have introduced the complex wavelet transform, focusing on the theoretical approach (the analytic wavelet transform - not realizable in practice), continued with the presentation of two classes of CWTs, namely the dual-tree based complex wavelet transforms and the projection-based complex wavelet transforms. For each of these classes I have taken a representative complex wavelet transform and presented it in large. Finally I have presented my own contribution to this domain, namely the implementation of the Hyperanalytic Wavelet Transform, which permits the exploitation of the mathematical results and of the algorithms previously obtained in the evolution of wavelets theory. This implementation has a very flexible structure, as we can use any orthogonal or bi-orthogonal real mother wavelets for the computation of the HWT. I have also illustrated its performances with the results of a few simulation tests.

In the following chapter I will focus on the denoising of images based on the theory of wavelets and I will present some simulations based on the use of the HWT.

Chapter 4

Denoising

4.1 Introduction

In common use, the word 'noise' refers to any unwanted sound. In both analog and digital electronics, noise is an unwanted perturbation to a wanted signal; it is called noise as a generalization of the audible noise heard when listening to a weak radio transmission. Signal noise is heard as acoustic noise if played through a loudspeaker and it manifests as 'snow' on a television or video image. Noise can block, distort, change or interfere with the meaning of a message in both human and electronic communication. In signal processing or computing it can be considered unwanted data without meaning; that is, data that is not being used to transmit a signal, but is simply produced as an unwanted by-product of other activities.

The term 'denoising', also known as 'noise reduction' refers to the removal of noise.

4.1.1 Digital images and noise

Images taken with both digital cameras and conventional film cameras will record noise from a variety of sources. Many further uses of these images require that the noise will be (partially) removed - for aesthetic purposes as in artistic work or marketing, or for practical purposes such as computer vision.

Depending on the model used to characterize the noise, we can encounter 'salt and pepper' noise, gaussian noise and many others. Another possible classification takes into consideration the way the noise affects the image, resulting in additive noise and multiplicative noise. The case of additive noise will be presented in this chapter, while the multiplicative noise will be discussed in chapter 5.

In Gaussian noise, each pixel in the image will be changed from its original value by a (usually) small amount. A histogram, a plot of the amount of distortion of a pixel value against the frequency with which it occurs, shows an estimation of the distribution of noise. While other distributions are possible, the Gaussian (normal) distribution is usually a good model, due to the central limit theorem that says that the sum of independent noises tends to approach a Gaussian distribution.

The case of Additive White Gaussian Noise (AWGN) will be considered. The acquired

image is expressed in this case in the following form:

$$x = s + n \tag{4.1}$$

where x is the observed/acquired image, s is the noiseless input image and n is the AWGN component.

In salt and pepper noise (sparse light and dark disturbances), pixels in the image are very different in color or intensity from their surrounding pixels; the defining characteristic is that the value of a noisy pixel bears no relation to the color of surrounding pixels. Generally this type of noise will only affect a small number of image pixels. When viewed, the image contains dark and white dots, hence the term salt and pepper noise. Typical sources include flecks of dust inside the camera, or with digital cameras, faulty CCD elements.

The speckle noise is modeled as multiplicative noise, the observed image having the expression:

$$x = s \cdot n \tag{4.2}$$

where n represents the speckle.

In either case, the noises at different pixels can be either correlated or uncorrelated; in many cases, noise values at different pixels are modeled as being independent and identically distributed, and hence uncorrelated.

4.1.2 Denoising algorithms

In his book, [Mal99], Mallat stated that the estimation of signals in additive noise is optimized by finding a representation that discriminates the signal from the noise. An estimation is calculated by an operator that attenuates the noise while preserving the signal.

There is a big diversity of estimators used as denoising systems. A review of such systems can be found in [Bua07]. One may classify these systems in two categories: those directly applied to the signal and those who use a wavelet transform before processing. From the first category, we must mention the denoising systems proposed in [FKE07] and [WD00]. The first one is based on the shape-adaptive discrete cosine transform (SA-DCT), transform that can be computed on a support of arbitrary shape. The second one is a maximum a posteriori (MAP) filter that acts in the spatial domain. Linear operators have long predominated because of their simplicity, despite their limited performance. It is possible to keep the simplicity while improving the performance with non-linearities in a sparse representation.

Optimizing an estimator requires taking advantage of prior information. Bayes theory uses a probabilistic signal model to derive estimators that minimize the average risk. These models are often not available for complex signals such as natural images. An alternative is offered by the minimax approach, which only requires knowing a prior set where the signal is guaranteed to be. The quasi minimax optimality of wavelet thresholding estimators is proved for piecewise regular signals and images.

Performance evaluation criteria

The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are the two error metrics used to compare image processing (denoising, compression) quality. The MSE represents the cumulative squared error between the denoised and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the higher the PSNR.

To compute the PSNR, it is first calculated the mean-squared error using the following equation:

$$MSE = \frac{\sum_{m=1,n=1}^{M,N} \left[I_o\left(m,n\right) - I_i\left(m,n\right) \right]^2}{M \cdot N} \tag{4.3}$$

In the previous equation, I_i is the input, noise-free image, I_o is the output image (might be noisy or already denoised) and M and N are the number of rows and columns in the input image, respectively. Then PSNR is being computed using the following equation:

$$PSNR = 10log_{10}\left(\frac{R^2}{MSE}\right) \tag{4.4}$$

In the previous equation, R is the maximum fluctuation in the input image data type. For example, if the input image has a double-precision floating-point data type, then R is 1. If it has an 8-bit unsigned integer data type, R is 255, etc.

4.1.3 Local averaging and PDEs

Many classical image denoising methods are based on a local average. The restored value at a pixel is obtained as an average of its neighboring pixels. The most classical algorithm is the Gaussian filtering. In that case, the restored value is obtained as a weighted average where the weight at each pixel depends on the distance to the restored one. This low pass filter tends to blur the image. This is known as the 'local average treatment effect'.

The neighborhood filters avoid the blurring effect by restricting the average to pixels having a similar gray level value. The idea is that gray level values inside a homogeneous region slightly fluctuate while pixels belonging to different regions have a larger gray level difference. The neighborhood filter takes an average of the values of pixels which are simultaneously close in gray level value and spatial distance.

One can involve calculus to interpret more in depth the neighborhood filters and improve them. In fact, expert photographers object to the creation of irregularities, in particular the flat tints and false details. In general, image filters can be better understood by establishing their asymptotic action when they are made more and more local. This action is then described by a Partial Differential Equation (PDE).

A further inquiry has led us to discover that a very simple version of the neighborhood filter has a well posed subjacent PDE, namely the mean curvature motion. This variant simply consists of replacing the average by a linear regression in the formula defining the neighborhood filter. The improvement has been accepted by professionals.

4.1.4 The total variation minimization

Total variation (TV) based filtering was introduced by Rudin et al. [ROF92]. TV denoising is an effective, robust filtering method for recovering piecewise-constant signals. Many algorithms have been proposed to implement total variation filtering.

The total-variation of a signal measures how much the signal changes between signal values. Specifically, the total-variation of an N-point signal $x(n)$, $1 \leq n \leq N$ is defined as:

$$TV(x) = \sum_{n=2}^{N} |x(n) - x(n-1)| . \tag{4.5}$$

The total variation of x can also be written as:

$$TV(x) = \|Dx\|_1 \tag{4.6}$$

where

$$D = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ldots & \\ & & & -1 & 1 \end{pmatrix}, \tag{4.7}$$

is a matrix of size $(N-1) \times N$.

We assume we observe the signal x corrupted by additive white Gaussian noise,

$$y = x + n,\ y, x, n \in R^N$$

One approach to estimate x is to find the signal x that minimizes the objective function

$$J(x) = \|y - x\|_2^2 + \lambda \|Dx\|_1 .$$

This approach is called TV denoising. The regularization parameter, λ, controls how much smoothing is performed. Larger noise levels call for larger λ. Many algorithms have been proposed to implement total variation filtering, and it still represents an active topic of research.

4.1.5 Properties of natural images

In this section are presented axioms and properties of images justified by the processing of observed natural images. Structural properties are defined, enabling the description of image contents with general statistical rules. These properties have precise consequences on image modeling.

1. Axiom A1: self-similarity.

   Natural phenomena are often self-similar, from a statistical point of view. As a result, natural images display scale invariance. This means that changing the scale of a signal (by studying only a small part of it for instance) does not change the statistical properties of the signal. When studying a statistic (e.g. the expectation of the power spectrum), the spatial shape of this statistic is not affected by scaling. The self-similarity can be associated with the notion of fractal, widely used to model objects such as mountains or coast lines.

2. Axiom A2: spatial adaptivity.

   Natural images generally exhibit a strong spatial adaptivity, since they are composed of different textures, homogeneous areas, sharp edges and small features. The importance of having a spatially variant model is shown by the failure of stationary approaches to correctly model images, especially when dealing with inverse problems such as denoising or deblurring.

   However, taking into account the space-varying characteristics of a natural scene is a difficult task, since it requires the definition (followed by the estimation) of a large number of additional parameters, at least one per pixel. This can inevitably lead to over fitting issues, since the non stationary parameters can adapt more to the noise than to the underlying signal, if they are not subject to any constraint. The use of a prior probability density function helps to constrain these parameters and take advantage of the spatial adaptivity.

3. Property P1: inter-scale persistence.

   When dealing with scale invariant data, it is natural to use a multiresolution decomposition (e.g. wavelets) to analyze them. This enables us to get a sparser representation, and to separate the scale invariance from other properties. The underlying spatial structure of the image persists through the different scales, i.e. high/low value of the coefficient magnitudes often lead to high/low values at the next scale. This can be seen as the inability of any multiscale transform to perfectly decorrelate natural images, since such a transform exhibits some inter-scale redundancy.

4. Property P2: intra-scale dependencies.

   Now we consider two different subbands of the same scale, but with different orientations (they are generated by linear filtering with the same filter with respect to size and shape, but oriented differently). Homogeneous areas and details are found approximately at the same spatial location in the different subbands. Isotropic details, such as smooth areas and round features, are naturally found at all orientations. This shows again that multiscale transforms do not perfectly decorrelate the information contained in natural images.

The best basis to represent such a process is the Karhunen-Loeve (K-L) basis. The image coefficients in this basis are independent. It is optimal, since it provides the minimum mean error when approximating an image by its projection on orthogonal basis vectors chosen a priori (this relates to image denoising by thresholding).

For stationary processes, the K-L basis is the Fourier basis. However, the Fourier basis is not well suited to the nonstationarity of natural images. In the case of natural images a wavelet basis is the closest approximation to a K-L basis, since it provides both a frequency and a spatial representation. The fact that it is only an approximation is illustrated by the properties P1 and P2 (residual dependencies). The aim is not only to diagonalize a Gaussian process, but also to provide the shortest description of the useful signal. Modeling can be seen as finding a simple description of nature, and in this case it can be related to image compression. It is well-known that wavelets provide a compact representation of natural images, and are widely used in coding.

An interesting consequence of the self-similarity is that high order statistics are scale invariant. It means that they obey a function f of the scale, and the shape of f is not affected by scaling (it is only multiplied by a scale factor). A generalization of this law is the scale invariance of the wavelet subband histograms, they relate to the full coefficient distribution instead of its order n moments.

4.1.6 Frequency domain filters

The effect of averaging the spatially closest pixels can also be achieved in the Fourier domain. The convolution of an image by a kernel k is equivalent to the multiplication of its Fourier transform by K. The average of the spatially closest pixels is then equivalent to the cancellation of the high frequencies. As the analogous spatial filter, this cancellation leads to the blurring of the image and a Gibbs effect. The optimal filter in the Fourier domain is the Wiener filter which does not cancel the high frequencies but attenuates them all. The coefficients with a magnitude over the noise standard deviation are nearly untouched while the coefficients with a lower value are reduced. By the use of the Fourier basis, global image characteristics may prevail over local ones. The presence of any boundary makes all the Fourier coefficients to have a meaningful value and therefore they are less attenuated. In that case, the flat zones which should be easily restored preserve a regular amount of noise. In order to avoid these effects, the basis must take into account more local features, as the wavelet transform and the local discrete cosine transform (DCT) do.

David Donoho actually introduced the word denoising in association with the wavelet theory [DJ94]. Thresholding estimators are studied in wavelet and wavelet packet bases, where they are used to suppress additive noises and restore signals degraded by low-pass filters. Non-linear estimations from sparse representations are also studied for operators, with an application to power spectrum estimation.

The multi-resolution analysis performed by the WT has been shown to be a powerful tool in order to achieve good denoising. In the wavelet domain, the noise is uniformly spread throughout the coefficients, while most of the image information is concentrated in the few largest ones (sparsity of the wavelet representation) [FBB01, SS02, PP06, AK05, GD06, LBU07, Shu05, ZS07, Olh07]. The most straightforward way of distinguishing information from noise in the wavelet domain consists of thresholding the wavelet coefficients. Soft-thresholding filter (stf) is the most popular strategy and has been theoretically justified by Donoho and Johnstone [DJ94]. They propose a three steps denoising algorithm:

1. the computation of the forward WT,

2. the filtering of the wavelet coefficients,

3. the computation of the IWT of the result obtained.

They use the Discrete Wavelet Transform (DWT) and the stf. Because it is not made any explicit hypothesis on the noise-free image it results in a non-parametric method. Their unique statistical hypothesis refers to the noise, considered additive white and Gaussian (AWGN). The stf is used to put to zero all the wavelet coefficients with the

absolute value smaller than a threshold. This threshold is selected to minimize the min-max approximation error. Consequently, regarding the three steps denoising algorithm, there are two tools to be chosen: the WT and the filter. In what concerns the first choice, in this chapter will be used the new implementation of the HWT. In [FBB01, PP06] was used the UDWT, in [SS02, AK05, Shu05] the DTCWT, and in [GD06, LBU07] the DWT. Concerning the second choice, numerous non-linear filter types can be used in the WT domain.

A possible classification is based on the nature of the noise-free component of the image to be processed. Basically, there are two categories of filters: those built assuming only the knowledge of noise statistics (a non-parametric approach), and those based on the knowledge of both signal and noise statistics (a parametric approach). From the first category can be mentioned: the hard-thresholding filter, [DJ94], the stf [DJ94, Mal99], that minimizes the Min-Max estimation error and the Efficient SURE-Based Inter-scales Point-wise Thresholding Filter [LBU07], that minimizes the Mean Square Error (MSE).

To the second category belong filters obtained by minimizing a Bayesian risk under a cost function, typically a delta cost function (MAP estimation [FBB01, SS02, AK05]) or the minimum mean squared error (MMSE estimation [PP06]). The denoising algorithms proposed in [SS02, PP06, AK05, GD06, LBU07] exploit the inter-scale dependence of wavelet coefficients. The method proposed in [PP06] takes into account the intra-scale dependence of wavelet coefficients as well. The statistical distribution of the wavelet coefficients changes from scale to scale. The coefficients of the WT have a heavy tailed distribution. To deal with this mobility, there are two solutions. The first one assumes the use of a fixed simple model, risking a decrease of accuracy across the scales. This way, there is a chance to obtain a closed form input-output relation for the MAP filter. Such an input-output relationship has two advantages: it simplifies the implementation of the filter and it allows the sensitivity analysis. The second solution assumes the use of a generalized model, defining a family of distributions and the identification of the best fitting element of this family to the distribution of the wavelet coefficients at a given scale (e.g. the family of Pearson's distributions in [FBB01], the family of $S\alpha S$ distributions in [AK05] and the model of Gauss-Markov random field in [GD06]). The use of a generalized model makes the treatment more accurate but requires implicit solutions for the MAP filter equation, which can often be solved only numerically.

4.1.7 Non local averaging

Non-local (NL) means algorithms are proposed in [Bua07]. The NL-means algorithm tries to take advantage of the high degree of redundancy of any natural image. Every small window in a natural image has many similar windows in the same image. In a very general sense, one can define as 'neighborhood of a pixel i' any set of pixels j in the image so that a window around j looks like a window around i. All pixels in that neighborhood can be used for predicting the value at i. Given a discrete noisy image $x = \{x(i) \,|i \in I\}$, the estimated value, $NL_x(i)$ is computed as a weighted average of all the pixels in the image,

$$NL_x(i) = \sum_{j \in I} \beta(i,j)\, x(j),$$

where the weights $\{\beta(i,j)\}_j$ depend on the similarity between the pixels i and j and satisfy the usual conditions $0 \leq \beta(i,j) \leq 1$ and $\sum_j \beta(i,j) = 1$. In order to compute the similarity between the image pixels, a neighborhood system on I is defined.

While producing state-of-the-art denoising results, this method is computationally impractical [MS05]. Its high computational complexity is due to the cost of weights calculation for all pixels in the image during the process of denoising. For every pixel being processed, the whole image is searched, and differences between corresponding neighborhoods are computed. The complexity is then quadratic in the number of image pixels.

In [MS05] the computational complexity of the algorithm proposed in [Bua07] is addressed in a different fashion. The basic idea proposed in [MS05] is to pre-classify the image blocks according to fundamental characteristics, such as their average gray values and gradient orientation. This is performed in a first path in [MS05], and while, denoising in the second path, only blocks with similar characteristics are used to compute the weights. Accessing these blocks can be efficiently implemented with simple lookup tables. The basic idea is then to combine ideas from [Bua07], namely, weighted average based on neighborhoods similarity, with concepts that are classical in information theory and were introduced in image denoising context.

Images with much finer texture and details will not benefit that much from the denoising; while reducing most of the noise, this type of processing will inevitably degrade important image features [GSZ06]. The first problem is to distinguish between good and bad candidates for denoising. Many natural images exhibit a mosaic of piecewise smooth and texture patches. This type of image structure calls for position (spatial)-varying filtering operation. Textured regions are characterized by high local variance. In order to preserve the detailed structure of such regions, the level of filtering should be reduced over these regions. The basic concept amounts to a reduction in the extent of filtering in regions where signal power exceeds that of the noise. So, the solution proposed in [GSZ06] supposes the anisotropic treatment of the acquired image taking into account the local variance values of its regions. This procedure can be seen like an NL-means algorithm where the classification of the image blocks is done on the basis of their local variance.

4.2 Non-parametric Denoising

The term 'non-parametric' is used to define all the methods developed without an a priori knowledge of the distribution of the signal to be processed. This non-parametric approach is motivated by the fact that in most practical situations, the distribution of the signals forming the observation are not known and is difficult to estimate them precisely.

4.2.1 Basic Concept

Even though we do not have any information on the distribution of the signal to be estimated, it is possible to satisfactory estimate this signal if we have a basis that permits the noise discrimination. The idea is to consider the noise additive, white and Gaussian. It is known that the projection of this type of noise on an orthonormal basis does not change

the nature of the noise. Let there be an observation made up of a signal perturbed by additive white Gaussian noise. If we dispose of an orthonormal basis able to 'concentrate' this signal's energy over a small number of high-amplitude coefficients, than the projection of the observation on such a basis will be made of a small class of relatively large coefficients (containing mainly the information on the signal) and a large class of small coefficients which are mainly due to the noise. The estimation implies recovering and, eventually, processing the large coefficients, after setting to zero the small ones. Such an estimation is realized with a so called 'thresholding function', named after the threshold underneath which every coefficients will be forced to zero.

It is said that a basis able to concentrate a big part of the signal's energy on a small number of coefficients does a sparse representation of the signal [DJ94]. This definition is not really precise as the terms 'small' and 'big' are not clearly defined, resulting in a sparse representation in a wide sense. It can be said about a basis that it realizes a sparse representation in a strict sense if, for a given signal, the energy of this signal is 'almost entirely' contained in a small number of high-amplitude coefficients. In this type of sparsity, almost all the coefficients qualified as 'small' are, in fact, almost nulls or with very low amplitudes, consequently not containing any significant information on the signal. In the case of the sparse representation in a wide sense, and not strictly sparse, it might be useful to process in equal manner the small coefficients. In this case it is being considered not a thresholding function but a so called attenuation function, in the sense that it does not force to zero all the small coefficients but it imposes them a higher attenuation than on the large coefficients.

4.2.2 Shrinkage Strategies

In order to realize the non-parametric estimation described before, one must choose a sparse representation and an appropriate attenuation function. When the attenuation function sets to zero all the coefficients inferiors to a threshold, than the function is called thresholding. Some of the most used thresholding function are the Hard Thresholding and the Soft Thresholding,

Soft Thresholding

The three-step denoising algorithm proposed by Donoho and Johnstone [DJ94], already mentioned, uses the DWT and the stf. They do not make any explicit hypothesis on the noise-free image and, consequently, this method can be considered non-parametric. Their unique statistical hypothesis refers to the noise, considered AWGN. The stf is used to put to zero all the wavelet coefficients with the absolute value smaller than a threshold. This threshold is selected to minimize the min-max approximation error.

The input-output relation of the stf is given by:

$$\hat{W}_s\left(W_x\right) = \begin{cases} W_x - t, & \text{for} \quad W_x > t \\ 0, & \text{for} \quad |W_x| \leq t \\ W_x + t, & \text{for} \quad W_x < -t. \end{cases} \tag{4.8}$$

where $W_x = W_s + W_n$ and t represents a threshold. W_x represents the sequence of WT coefficients of the noisy image, W_s represents the sequence of the WT coefficients of

the noiseless component of the acquired image and W_n represents the sequence of WT coefficients of the noise in the acquired image.

The interval $(-t, t)$ is named dead zone, because all the detail WT coefficients having the values in this interval are set to zero by the stf. Because the magnitudes of all the detail WT coefficients are reduced by the stf, it is named shrinkage operator.

If the threshold t is chosen accordingly with the value of the noise's standard deviation, then this filter is called adaptive. The noise's variance can be estimated using the relation:

$$\hat{\sigma}^2_{W_n} = \frac{median\left(|W_x|\right)}{0.6745} \; W_x \in D_1^3 \tag{4.9}$$

In the case of the minimization of the Min-Max approximation error, the constant that multiplies the noise's standard deviation in order to obtain the threshold value t is equal to $\sqrt{2lnN}$, where N represents the number of pixels of the observed image. Another value, specific for the criteria that demands the minimization of the output noise power, is 3 (having in mind the three sigmas rule, specific for the Gaussian laws).

In the following we present an experiment that associates the Diversity Enriched Discrete Wavelet Transform (DE DWT - a transform whose result is the mean of 9 different DWTs, each one corresponding to a different mother wavelet from the Daubechies family) [IIA06] and the soft-thresholding filter. The optimization criteria (in this case the maximization of the output peak signal-to-noise ratio) demands the use of another value for the proportionality constant, between the threshold of the stf filter and the noise's standard deviation. In the following simulations the value of this constant is chosen such that the PSNR associated to each DWT is maximum. The results obtained for the image Lena are presented in the folowing table. The values in this table were obtained by selecting each time a number of 6 iterations for the DWTs.

	$\boldsymbol{\sigma_n}$	10	15	20	25	30	35
PSNRi		28.17	24.62	22.15	20.17	18.63	17.30
PSNRo	**Daub,4**	30.71	28.64	26.99	25.77	24.80	23.97
	Daub,6	30.96	28.77	27.10	25.87	24.88	24.11
	Daub,8	30.99	28.82	27.17	25.89	24.98	24.23
	Daub,10	30.96	28.90	27.22	25.91	25.03	24.24
	Daub,12	30.77	28.79	27.17	25.91	25.02	24.21
	Daub,14	30.70	28.65	27.12	25.86	24.94	24.16
	Daub,16	30.68	28.71	27.06	25.83	24.92	24.16
	Daub,18	30.64	28.66	27.08	25.78	24.91	24.09
	Daub,20	30.72	28.64	27.07	25.76	24.84	24.07
	DE DWT	**31.80**	**29.77**	**28.14**	**26.87**	**25.94**	**25.18**

Table 4.1: PSNRs obtained using the soft-thresholding filter in the 2D DWTs domain, computed using the mother wavelets from the Daubechies family and in the DE DWT domain, for the image Lena perturbed with AWGN with different variances

Hard Thresholding

Another type of non-parametric filter is the hard-thresholding filter (Htf). This is also a shrinkage operator. The input-output relation of this filter is:

$$hard\,(y,t) = (|y| - t)_+ = \begin{cases} y, & \text{for} \quad |y| \geq t \\ 0, & \text{for} \quad |y| < t \end{cases} \tag{4.10}$$

where

$$(X)_+ = \begin{cases} X, & \text{for } X > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.11}$$

This filter also has a dead zone. Next, the results of some simulations on the association of the DWT and ED DWT with a hard thresholding filter are presented. The optimization criterion (in this chapter it is the PSNR maximization) demands for another value of the proportionality constant, between the threshold value and the standard deviation of the noise. In the following simulations the value of this constant is chosen to insure that the PSNR value associated to each DWT is maximum. The results obtained are presented in the following table. The values in the table correspond to a number of six iterations for each DWT.

$\boldsymbol{\sigma_n}$		10	15	20	25	30	35
PSNRi		28.17	24.66	22.11	20.21	18.62	17.29
PSNRo	**Daub,4**	30.46	28.02	26.39	25.22	24.20	23.49
	Daub,6	30.63	28.21	26.49	25.23	24.28	23.59
	Daub,8	30.69	28.20	26.61	25.33	24.49	23.73
	Daub,10	30.81	28.25	26.56	25.39	24.48	23.64
	Daub,12	30.77	28.21	26.53	25.34	24.51	23.75
	Daub,14	30.69	28.14	26.34	25.23	24.41	23.51
	Daub,16	30.56	28.04	26.29	25.20	24.24	23.58
	Daub,18	30.53	28.03	26.29	25.15	24.22	23.49
	Daub,20	30.58	28.04	26.35	25.26	24.25	23.47
	DE DWT	**33.37**	**30.91**	**29.13**	**27.81**	**26.75**	**25.83**

Table 4.2: PSNRs obtained using the hard-thresholding filter in the 2D DWTs domain, computed using the mother wavelets from the Daubechies family and in the DE DWT domain, for the image Lena perturbed with AWGN with different variances

The results obtained using the hard-thresholding are better than the results obtained using the soft-thresholding, as can be seen by comparing tables 4.1 and 4.2.

4.2.2.1 A New Class of Shrinkage Functions Based on Sigmoid

The standard WaveShrink functions considered are the hard and soft thresholding functions (proposed in [DJ94]); the Non-Negative Garrote (NNG) function (proposed for wavelet shrinkage in [Gao98]) and the Smoothly Clipped Absolute Deviation (SCAD) function of [AF01]). The graphical representations of these functions are given in the following figure.

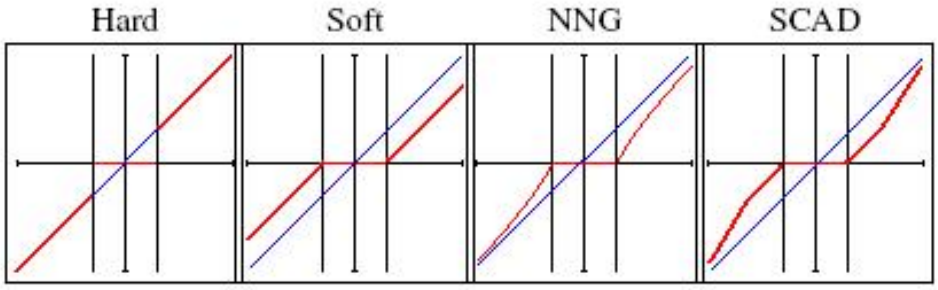


Figure 4.1: Examples of standard WaveShrink functions

On one hand, the hard thresholding function is not everywhere continuous and its discontinuities generate a high variance in the estimated signal; on the other hand, the soft thresholding function is continuous, but creates an attenuation on large coefficients, which results in an oversmoothing and an important bias for the estimated signal [Gao98]. The NNG and SCAD functions achieve a certain compromise between the hard and the soft thresholding functions as can be seen in figure 4.1. However, all the standard WaveShrink functions presented above include zero-forcing. This zero-forcing induces singularities of the thresholding function. As a consequence, it results a significant variance of the estimation due to the sensitivity of the inverse wavelet transform. In addition, thresholding rules assume that the wavelet representation is sparse. Note that smooth signals yield sparse wavelet representations in the sense given by [DJ94]: for such signals, large coefficients are very few in number. In contrast, wavelet representations of natural images, which tend to be piecewise regular rather than smooth, fail to be sparse enough since large coefficients are not very few. This justifies the introduction of more flexible WaveShrink methods for correcting the drawbacks of thresholding rules. As an example, a WaveShrink parameterization such as that proposed in [LBU07] results in non zero-forcing and is shown to be more efficient than soft thresholding estimation. Therefore, a suitable shrinkage function should satisfy the following two properties [APM08],

Proposition 4.2.1. *Smoothness: a smooth shrinkage function induces small variability among coefficients with close values.*

Proposition 4.2.2. *Penalized shrinkage: a strong (resp. a weak) attenuation for small (resp. large) coefficients is required because small (resp. large) coefficients contain less (resp. more) information about the signal.*

In [APM08] are reported very good denoising results obtained applying a class of shrinkage operators, called the Smooth Sigmoid-Based Shrinkage (SSBS) functions, satisfying propositions 4.2.1 and 4.2.2, associated with the UDWT. This WT was preferred for its translation invariance. The absence of translation invariance in denoising applications produces artifacts in the neighborhood of discontinuities resembling with the manifestation of Gibbs phenomenon.

4.2.2.2 Translation invariant wavelet thresholding

Another possibility to obtain a translation invariant denoising is to use the DWT associated with the Cycle Spinning (CS) algorithm proposed by R. Coifman and D. Donoho [CD95]. They improved the wavelet thresholding methods by averaging the estimations of all translations of the degraded signal. The CS algorithm was conceived to suppress the artifacts in the neighborhood of the discontinuities introduced by the DWT, and it implies the rejection of the translation dependency. For a range of delays, data is shifted, its DWT is computed, and then the result is unshifted. Doing this for a range of shifts, and averaging the several results so obtained, a quasi shift-invariant DWT is implemented. The degree of redundancy of this transform is proportional to the number of shifts of the input signal produced. CS over the range of all circular shifts of the input signal is equivalent with a translation-invariant WT.

Calling $y_p(i)$ the translated signal $y(i-p)$, the wavelet coefficients of the original and translated signals can be very different, and they are not related by a simple translation or permutation,

$$y_B^p(\alpha) = \langle y(n-p), g_\alpha(n)\rangle = \langle y(n), g_\alpha(n+p)\rangle$$

The vectors $g_\alpha(n+p)$ are not, generally, in the basis $B = \{g_\alpha\}_{\alpha\in A}$, and therefore the estimation of the translated signal is not related to the estimation of y. This new algorithm yields an estimate $\widehat{u}^p$ for every translated version, y^p, of the original signal,

$$\widehat{u}^p = Dy^p = \sum_{\alpha\in A} a^p(\alpha)\, y_B^p(\alpha)\, g_\alpha.$$

The translation invariant thresholding based on CS, denoted by $TIWT(y)$, is obtained by averaging all these estimations after being translated in the inverse sense,

$$TIWT(y) = \frac{1}{|I|}\sum_{p\in I} \widehat{u}^p(i+p).$$

The Gibbs effect is considerably reduced by the translation invariant wavelet thresholding, since the average of different estimations of the signal reduces the oscillations.

4.2.2.3 A semi parametric denoising method using a MMSE estimator

In [LBU07] is proposed a new denoising method acting in the field of orthogonal WTs, lying on the border between the non-parametric and parametric methods. It has the usual structure in three steps for a denoising method based on wavelets, computation of the direct WT, filtering in the wavelets domain and computation of the IWT.

- The MSE in the space domain is a weighted sum of the MSE of each individual subband:

$$\left\langle |\widehat{w} - w|^2 \right\rangle = \sum_{j=0}^{J} \frac{N_j}{N} \left\langle \left|\widehat{w}^j - w^j\right|^2 \right\rangle$$

  where the expectation is denoted by

$$\langle u\rangle = \frac{1}{N}\sum_{k=1}^{N} u_k$$

- The noise remains white and Gaussian, having the same statistics in the orthonormal wavelet domain as in the spatial domain, i.e. $n^j \cong N(0, \sigma_n^2)$. This makes it possible to apply a new denoising function, independently, in every high pass subband, leading to a subband-adaptive solution as in the case of most of the successful wavelet denoising approaches.

In denoising applications, the performance is often measured in terms of PSNR. Since the noise is a random process, in [LBU07] is introduced an expectation operator $E\{\}$ to estimate the potential results obtained after processing the noisy data y. Note that the noise-free data w is not modeled as a random process; thus, $E\{w\} = w$. This is the reason why the proposed denoising method can be considered as non-parametric. The aim of image denoising is naturally to maximize the PSNR and, thus, to minimize the MSE. Such an estimator is of minimum MSE (MMSE) type. In [LBU07], the authors have chosen to estimate w^j by a point wise function of y^j

$$\left(\widehat{w}_k^j\right)_{k\in[1,N_j]} = \left(\theta^j\left(y_k^j\right)\right)_{k\in[1,N_j]}.$$

The subband index can be dropped since a new denoising function is independently applied in each individual subband. The goal in [LBU07] is to find a function θ which minimizes:

$$MSE = \left\langle \left|\theta(y) - w\right|^2 \right\rangle = \left\langle \theta(y)^2 \right\rangle - 2\left\langle w\theta(y) \right\rangle + \left\langle w^2 \right\rangle.$$

In practice, only the noisy signal $y = w + n$ can be accessed. So it is necessary to remove the explicit dependence on w in the last relation. The solution taken by Luisier and al. consists of thresholding the coefficients with a pointwise thresholding function depending on only two parameters: a_1 and a_2

$$\theta_0(y, \mathbf{a}) = \left(a_1 + a_2 e^{-\frac{y^2}{12\sigma_n^2}}\right) y$$

This denoising method is considered semi parametric, as it requires only the identification of the coefficients a_1 and a_2, without any prior information about the noiseless component of the acquired image.

4.3 Parametric Denoising

The aim of the denoising process is to improve the visual quality of the image and, ideally, to recover the original one. Mathematically, it is necessary to minimize the error between the restored and the original images. This section deals with the definition and the minimization of such an error. Mathematically, this error is written as a distance and, for simplicity, the Euclidean distance can be used. In order to optimize the restoration process, the definition of such an error should incorporate the a priori information about the original image and the noise, when available.

4.3.1 The Bayesian Approach

In this section we will begin by presenting the general features of the Bayesian estimation both in the spatial and in the wavelet coefficients' domain and we will continue with some examples of filters using this approach.

Spatial Domain

Let $x = s + n$, where s is the original signal, n is the noise and x is the received or observed signal.The problem is to estimate s when the only available data is x.

The Bayes theorem gives the relation between one conditional probability and its inverse. This relation can also be applied in the probability density function (pdf) domain, for example to the probability density function of s, knowing x and the pdf of x knowing s:

$$p_{S|X}(s \mid x) = \frac{p_{X|S}(x \mid s) \cdot p_S(s)}{p_X(x)} \tag{4.12}$$

In 4.12, $p_S(s)$ is the prior or marginal probability density function of s. It is 'prior' in the sense that it does not take into consideration any information about x. $p_{S|X}(s \mid x)$ is the conditional pdf of s given x, and is also called the posterior pdf because it depends on the specified value of x. $p_{X|S}(x \mid s)$ is the conditional pdf of x given s, also known as the likelihood of s, $L(s) = p_{X|S}(x \mid s)$. The $p_X(x)$ is the prior or marginal probability of x and acts as a normalizing factor. However, once the observation x has been made, it is a constant and often ignorable in applications.

If we want to take into account the noise statistics, than we can replace $p_{X|S}(x \mid s)$ by $p_{N|S}(x - s \mid s)$ as they are equivalent.

Further more, if the noise and the signal are independent, as is the case in general, $p_{N|S}(x - s \mid s) = p_N(x - s)$ and relation 4.12 becomes:

$$p_{S|X}(s \mid x) = \frac{p_N(x - s) \cdot p_S(s)}{p_X(x)} \tag{4.13}$$

The estimated value of s, $\hat{s}$, is the value S which maximizes the posterior distribution, $p_{S|X}(s \mid x)$. This is called the Maximum A Posteriori estimate. If there is a range of values of s with high probability, then a confidence interval or 'error bar' can be defined.

Consequently, the MAP estimation of s based on the observation $x = s + n$ is given by the MAP filter equation:

$$\hat{s} = \underset{s}{argmax}\left(p_{X|S}(x \mid s) \cdot p_S(s)\right) \tag{4.14}$$

or, equivalently,

$$\hat{s} = \underset{s}{argmax}\left(p_N(x - s) \cdot p_S(s)\right) \tag{4.15}$$

The computation of $\hat{s}$ is usually based on searching the zeros of the derivative of either the function to be maximized (eq. 4.16) or of its logarithm (eq. 4.17).

$$\frac{\partial}{\partial s}\left(p_N(x - s) \cdot p_S(s)\right) = 0 \tag{4.16}$$

$$\frac{\partial}{\partial s}\left(ln\left(p_N(x - s)\right) + ln\left(p_S(s)\right)\right) = 0 \tag{4.17}$$

Wavelet coefficients domain

The MAP criterion can be applied in the spatial domain, as seen in the previous paragraph, but it can be applied as well in the wavelet coefficients' domain. If we consider the observed signal $x = s + n$, than, as the wavelet transform is a linear transformation, we will obtain the following relations between the approximation (A) and wavelet (W) coefficients from level j:

$$\begin{aligned} A_x^j &= A_s^j + A_n^j \\ W_x^j &= W_s^j + W_n^j. \end{aligned} \tag{4.18}$$

In the following there will be considered only the wavelet coefficients. By repeating the same considerations previously exposed, the a posteriori probability density function of the noise-free signal's wavelet coefficients conditional to the observed signal's wavelet coefficients is $p_{W_X|W_S}(W_x \mid W_s)$. Hence, by applying the Bayes rule and considering the noise independent of the signal, we have:

$$\begin{aligned} p_{W_X|W_S}(W_x \mid W_s) &= \frac{p_{W_X|W_S}(W_x \mid W_s) \cdot p_{W_S}(Ws)}{p_{W_X}(W_x)} \\ &= \frac{p_{W_N}(W_x - W_s) \cdot p_{W_S}(W_s)}{p_{W_X}(W_x)}. \end{aligned} \tag{4.19}$$

The MAP estimate of the process W_s will be:

$$\begin{aligned} \hat{W}_s &= \underset{W_s}{argmax}\left(p_{W_X|W_S}(W_x \mid W_s)\right) \\ &= \underset{W_s}{argmax}\left(p_{W_N}(W_x - W_s) \cdot p_{W_S}(W_s)\right) \end{aligned} \tag{4.20}$$

If the pdfs p_{W_s} and p_{W_n} do not take into account the inter-scale dependency of the wavelet coefficients, the obtained filter is called marginal. For the MAP filters that take into account the inter-scale dependency, the pdfs are multivariate functions. Generally, the equation 4.20 has not an analytical solution.

4.3.1.1 The Wiener Filter

The Wiener filter reduces the noise affecting the signal by comparing the observed signal with an estimation of the noiseless signal and minimizing the mean square error, in this sense the Wiener filter being considered an MMSE filter.

In the following I will consider, for simplicity, time-domain sequences, but the equations deduced are also available in the wavelet coefficients' domain.

Let $x = s + n$ be the observed sequence applied at the input of the Wiener filter with impulse response h. The estimation given by the filter will be:

$$\hat{s} = x * h = \sum_{i=0}^{N} h_i x[n-i] \tag{4.21}$$

and the estimation error is:

$$e[n] = \hat{s}[n] - s[n] = \sum_{i=0}^{N} h_i x[n-i] - s[n]. \tag{4.22}$$

The filter coefficients, namely the values of the vector h are the solutions of the equation:

$$h_i = argmin\left\{E\left\{e^2\left[n\right]\right\}\right\} \tag{4.23}$$

The solution of equation 4.23 can be obtained by solving the following equation:

$$\underbrace{\begin{bmatrix} R_x\left[0\right] & R_x\left[1\right] & \dots & R_x\left[N\right] \\ R_x\left[1\right] & R_x\left[0\right] & \dots & R_x\left[N-1\right] \\ \vdots & \vdots & \ddots & \vdots \\ R_x\left[N\right] & R_x\left[N-1\right] & \dots & R_x\left[0\right] \end{bmatrix}}_{T} \times \underbrace{\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_N \end{bmatrix}}_{h} = \underbrace{\begin{bmatrix} R_{xs}\left[0\right] \\ R_{xs}\left[1\right] \\ \vdots \\ R_{xs}\left[N\right] \end{bmatrix}}_{V}, \tag{4.24}$$

where T is the Toeplitz matrix, having for elements the autocorrelation of the observed signal, h is the Wiener filter's vector of coefficients and V is the inter correlation vector whose elements are the values of the inter correlation between the received and the original signal.

Zero-Order Wiener Filter

Generally, the following assumptions are made:

1. The signal s and the noise n are not correlated, namely $R_{sn}\left[k\right] = 0, \forall k$.
2. The noise, n, is a random additive white gaussian (AWG) process, with zero-mean and variance σ_n^2, $N(0, \sigma_n^2)$.
3. The useful signal, s, is a random AWG process with zero-mean and variance σ_s^2, $N(0, \sigma_s^2)$.

Using these assumptions, we can write:

$$\begin{aligned} \mu_x &= E\left\{x\right\} = E\left\{s+n\right\} = E\left\{s\right\} + E\left\{n\right\} = 0, \\ \sigma_x^2 &= E\left\{\left(x-\mu_x\right)^2\right\} = E\left\{x^2\right\} = E\left\{\left(s+n\right)^2\right\} = E\left\{s^2\right\} + 2E\left\{sn\right\} + E\left\{n^2\right\} \\ &= \sigma_s^2 + 2R_{sn} + \sigma_n^2 = \sigma_s^2 + \sigma_n^2, \\ R_x\left[k\right] &= \sigma_x^2 \cdot \delta\left[k\right] = \left(\sigma_s^2 + \sigma_n^2\right) \cdot \delta\left[k\right], \forall k, \\ R_{xs}\left[k\right] &= E\left\{x\left[i\right] \cdot s\left[i+k\right]\right\} = E\left\{s\left[i\right] \cdot s\left[i+k\right]\right\} + E\left\{n\left[i\right] \cdot s\left[i+k\right]\right\} \\ &= \sigma_s^2 \cdot \delta\left[k\right], \forall k. \end{aligned} \tag{4.25}$$

Based on 4.25, we can rewrite 4.24 as follows:

$$\begin{bmatrix} \sigma_s^2+\sigma_n^2 & 0 & \dots & 0 \\ 0 & \sigma_s^2+\sigma_n^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_s^2+\sigma_n^2 \end{bmatrix} \times \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_N \end{bmatrix} = \begin{bmatrix} \sigma_s^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{4.26}$$

leading to the system of equations:

$$\begin{cases} \left(\sigma_s^2+\sigma_n^2\right) \cdot h_0 &= \sigma_s^2 \\ \left(\sigma_s^2+\sigma_n^2\right) \cdot h_1 &= 0 \\ \qquad \dots & \\ \left(\sigma_s^2+\sigma_n^2\right) \cdot h_N &= 0 \end{cases} \tag{4.27}$$

Solving 4.27 we obtain the coefficients:

$$h_0 = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_n^2}, \text{ and } h_1 = h_2 = \ldots = h_N = 0. \tag{4.28}$$

As this filter has only one non-null coefficient (4.28), namely h_0, it is called 'Zero-order Wiener filter'. Its impulse response, in the 2-dimensional case, is

$$h(i,j) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_n^2} \cdot \delta(i,j), \tag{4.29}$$

similar to the impulse response of the Kuan filter used in speckle filtering.

The filtering equation can be put in the following form:

$$\hat{s} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_n^2} \cdot x = \frac{\frac{\sigma_s^2}{\sigma_n^2}}{1 + \frac{\sigma_s^2}{\sigma_n^2}} \cdot x. \tag{4.30}$$

The quantity $\frac{\sigma_s^2}{\sigma_n^2}$ represents the Signal to Noise Ratio (SNR) of the acquired image. Consequently, the input-output relation of the zero order Wiener filter can be written in the equivalent form:

$$\hat{s} = \frac{SNR}{1 + SNR} \cdot x. \tag{4.31}$$

For good quality acquired images (SNR » 1), equation 4.31 becomes:

$$\hat{s} \cong x,$$

while for poor quality acquired images (SNR « 1), the input-output relation of the zero order Wiener filter has the expression:

$$\hat{s} \cong 0.$$

In conclusion, the zero order Wiener filter is not a good solution for the denoising of images with low SNR.

Generally, the variances in eq. 4.30, especially σ_s^2, vary in space. If we choose to estimate only once σ_s^2 for the whole image, we can say we have a global filter. If we estimate σ_s^2 locally, for each pixel using a neighborhood, we call the filter a local filter.

In order to evaluate the space domain performances of the zero-order Wiener filter we have applied it directly to the image corrupted with additive noise and we have computed the PSNR ratio. In table 4.3 we present the input and output PSNR values obtained for the input image 'Lena' corrupted with additive white gaussian noise with variance σ_n^2 presented on the first line of the table and the input PSNR value on the second line of the table. We have estimated σ_s^2 locally using a square analysis window of 7x7 pixels and the output PSNRs are presented on the bottom line of the table.

The relations deduced for time-domain signals are also available in the wavelet coefficients domain, under the same hypotheses, the estimated coefficient being computed with:

$$\hat{W}_s(i,j) = \frac{\sigma_{W_s}^2}{\sigma_{W_s}^2 + \sigma_{W_n}^2} \cdot W_x(i,j). \tag{4.32}$$

$\boldsymbol{\sigma_n}$	10	15	20	25	30	35	50	100
PSNRi	28.16	24.64	22.15	20.20	18.62	17.28	14.17	8.16
PSNRo	31.38	30.31	29.35	28.48	27.75	27.10	25.37	21.23

Table 4.3: Denoising using zero-order Wiener filter directly on the image

In the following set of simulations we associate 2D DWT to the zero-order Wiener filter. In the case of the 2D DWT, the noise coefficients' variance, $\sigma^2_{W_n}$ is considered equal to the variance of the detail sub-image obtained at the first level of decomposition [DJ94].

The useful image coefficients' variance can be estimated either globally or locally.

Globally, we compute in each subband the standard deviation, $\sigma^2_{W_x}$ and compute

$$\sigma_{W_s} = \sqrt{max\left(\sigma^2_{W_x} - \sigma^2_{W_n}, 0\right)}$$

Then, we replace these values in 4.32 and compute the denoised coefficients.

Locally, we use an analysis window, square for instance, centered in the pixel to be processed (i,j) and of dimension $2p+1$, $F_p(i,j)$. To estimate the useful image's variance we must compute its local mean

$$_l\hat{\mu}_{W_s}(i,j) = \frac{1}{(2p+1)^2} \sum_{(m,n)\in F_p(i,j)} W_x(m,n), \tag{4.33}$$

and the variance of the acquired image's coefficients inside the analysis window:

$$_l\hat{\sigma}^2_{W_x}(i,j) = \frac{1}{(2p+1)^2} \sum_{(m,n)\in F_p(i,j)} \left(W_x(m,n) -_l \hat{\mu}_{W_s}(i,j)\right)^2. \tag{4.34}$$

Using these relations, the local variance of the useful image's coefficients is given by:

$$_l\hat{\sigma}^2_{W_s}(i,j) = max\left(0, \left(_l\hat{\sigma}^2_{W_x}(i,j) - \hat{\sigma}^2_{W_n}\right)\right). \tag{4.35}$$

The impulse response of the filter for the considered pixel is:

$$\hat{h}(i,j) = \frac{_l\hat{\sigma}^2_{W_s}(i,j)}{_l\hat{\sigma}^2_{W_s}(i,j) + \hat{\sigma}^2_{W_n}} \cdot \delta(i,j). \tag{4.36}$$

Next we filter the coefficient considered and we shift the center of the analysis window to the next pixel and retake steps 4.33 - 4.36. In this way we perform an adaptive filtering. The Wiener filter can be considered a MAP filter, based on the hypotheses that the probability density functions of the useful image and of the noise are gaussian functions. The input-output relation of the equivalent MAP filter is:

$$\hat{W}_s(i,j) = \frac{_l\hat{\sigma}^2_{W_s}(i,j)}{_l\hat{\sigma}^2_{W_s}(i,j) + \hat{\sigma}^2_{W_n}} \cdot W_x(i,j) = \frac{\frac{_l\hat{\sigma}^2_{W_s}(i,j)}{\hat{\sigma}^2_{W_n}}}{1 + \frac{_l\hat{\sigma}^2_{W_s}(i,j)}{\hat{\sigma}^2_{W_n}}} \cdot W_x(i,j). \tag{4.37}$$

This is a first example of local marginal MAP filter. The significance of the variances' ratio $\frac{{}_l\hat{\sigma}^2_{W_s}(i,j)}{\hat{\sigma}^2_{W_n}}$ is, in this case, of local SNR. The input-output relation of the local marginal zero order Wiener filter can be put in the form:

$$\hat{W}_s(i,j) \cong W_x(i,j) \cdot \begin{cases} 1, & \text{if } SNR(i,j) >> 1, \\ \frac{SNR(i,j)}{1+SNR(i,j)}, & \text{if } SNR(i,j) \text{ is moderate}, \\ 0, & \text{if } SNR(i,j) << 1. \end{cases} \tag{4.38}$$

The local marginal zero order Wiener filter has a dead zone (just like the shrinkage filters) because the values of all the pixels with small local SNR values are set to zero.

The denoising methods based on the use of zero-order Wiener filter in the wavelet domain can not completely remove the noise when the signal to noise ratio of the acquired image is low.

One can observe that relation 4.35 introduced non-linearities. This is the reason why the Wiener filter, presented in this section, is not perfectly linear.

Simulations in which the zero-order Wiener filter is associated to the 2D DWT are presented in table 4.4. The first line of the table presents the noise standard deviation and the second line the values of the input image's PSNR. The third line presents the output PSNRs obtained after a zero-order global Wiener filtering, while the fourth line presents the PSNRs obtained after using a zero-order local Wiener filter, the analysis window in this case being square, of 7x7 pixels. In all the cases we have performed a 4-level wavelet decomposition using Daubechies with 10 vanishing moments as mother wavelet. The input image considered was Lena.

$\boldsymbol{\sigma_n}$	10	15	20	25	30	35	50	100
PSNRi	28.16	24.63	22.15	20.21	18.62	17.29	14.19	8.16
PSNRo global	31.65	29.66	28.77	27.92	27.26	26.75	25.51	23.28
PSNRo local	34.40	32.36	30.90	29.76	28.84	28.06	26.25	22.46

Table 4.4: Denoising using zero-order Wiener filters, both global and local, in the 2D DWT domain

From table 4.4 we can observe that the local estimation outperforms the global one for almost all the cases considered, the exception being at very high noise, or, in other words, very low SNR. By comparing the results obtained with the filter applied directly on the image (table 4.3) and with the filter applied in the wavelet domain (table 4.4), we can observe that the latter gives better results.

These results can be slightly improved if we replace the 2D DWT by the HWT. The hypotheses used for the construction of the local marginal zero order Wiener filter in the 2D DWT domain can be applied in the case of the four components of the detail HWT coefficients. The denoising is presented in figure 4.2.

The results obtained using the processing scheme in fig. 4.2, simulations performed in the same conditions as those in 4.4, are presented in table 4.5.

4.3.1.2 The adaptive soft-thresholding filter

To highlight the heavy tailed distributions which characterize the HWT coefficients resulted after the directional selectivity enhancement, as shown in fig. 4.3, the following

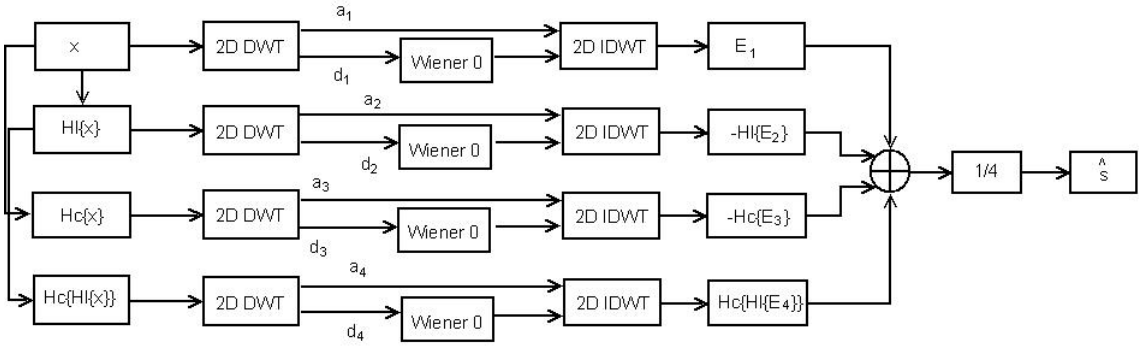


Figure 4.2: Denoising scheme using HWT and Zero-Order Wiener filter

$\boldsymbol{\sigma_n}$	10	15	20	25	30	35	50	100
PSNRi	28.15	24.65	22.15	20.21	18.62	17.28	14.19	8.16
PSNRo	34.81	32.80	31.37	30.21	29.29	28.47	26.61	22.92

Table 4.5: Denoising using zero-order local Wiener filter, in the HWT domain

experiment was done. The histograms of the different subbands obtained applying the HWT to the image Lena are computed. The results are represented in figure 4.4.

The linear dependencies of the two branches of the logarithms of the histograms prove that the pdfs of the real and imaginary parts of the HWT coefficients correspond to exponential laws (which are heavy tailed):

$$log\left(K_1 \cdot e^{-|x|}\right) = K_2 - |x| \tag{4.39}$$

where K_1 and K_2 represent two constants. So, the hypothesis that the real and imaginary parts of the useful HWT coefficients are distributed following Laplace laws can be made.

If W_n is Gaussian distributed and W_s has a Laplacian distribution (this is a heavy tailed one) then the MAP filter becomes an adaptive stf [SS02]. The hypotheses for this type of marginal MAP filter are:

$$p_{W_s}(W_s) = \frac{1}{\sqrt{2}\sigma_{W_s}} \cdot e^{-\frac{\sqrt{2}|W_s|}{\sigma_{W_s}}} \text{ and } p_{W_n}(W_n) = \frac{1}{\sqrt{2\pi}\sigma_{W_n}} \cdot e^{-\frac{W_n^2}{2\sigma_{W_n}^2}} \tag{4.40}$$

The MAP filter equation 4.20 becomes:

$$\begin{aligned} \hat{W}_s &= \underset{W_s}{argmax}\left\{ln\left(\frac{1}{\sqrt{2\pi}\sigma_{W_n}} \cdot e^{-\frac{(W_x-W_s)^2}{2\sigma_{W_n}^2}} \cdot \frac{1}{\sqrt{2}\sigma_{W_s}} \cdot e^{-\frac{\sqrt{2}|W_s|}{\sigma_{W_s}}}\right)\right\} \\ &= \underset{W_s}{argmax}\left\{ln\left(\frac{1}{2\sqrt{\pi}\sigma_{W_s}\sigma_{W_n}}\right) - \frac{(W_x-W_s)^2}{2\sigma_{W_n}^2} - \frac{\sqrt{2}\,|W_s|}{\sigma_{W_s}}\right\}. \end{aligned} \tag{4.41}$$

To maximize the argument of the right hand side of 4.41, the following equation must be

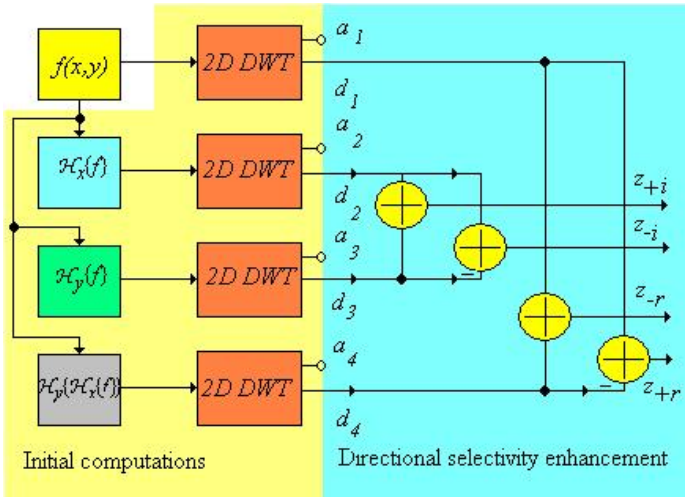


Figure 4.3: The architecture of HWT with directional selectivity enhancement

solved:

$$\begin{aligned}
&\frac{d}{dW_s}\left(ln\left(\frac{1}{2\sqrt{\pi}\sigma_{W_s}\sigma_{W_n}}\right)-\frac{(W_x-W_s)^2}{2\sigma_{W_n}^2}-\frac{\sqrt{2}\,|W_s|}{\sigma_{W_s}}\right)=0 \;\Leftrightarrow\\
&\Leftrightarrow\;\frac{W_x-W_s}{\sigma_{W_n}^2}-\frac{\sqrt{2}}{\sigma_{W_s}}\cdot\frac{|W_s|}{W_s}=0 \;\Leftrightarrow\\
&\Leftrightarrow\;\frac{W_x-W_s}{\sigma_{W_n}^2}-\frac{\sqrt{2}}{\sigma_{W_s}}\cdot\mathrm{sgn}\,(W_s)=0. \qquad (4.42)
\end{aligned}$$

For $W_s > 0$, eq. 4.42 becomes:

$$\begin{aligned}
&\frac{W_x-W_s}{\sigma_{W_n}^2}-\frac{\sqrt{2}}{\sigma_{W_s}}=0 \;\Leftrightarrow\\
&\Leftrightarrow\;\frac{W_x-W_s}{\sigma_{W_n}^2}=\sqrt{2}\frac{1}{\sigma_{W_s}} \;\Leftrightarrow\\
&\Leftrightarrow\;W_x-W_s=\sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}} \;\Leftrightarrow\\
&\Leftrightarrow\;W_s=W_x-\sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}}, \qquad (4.43)
\end{aligned}$$

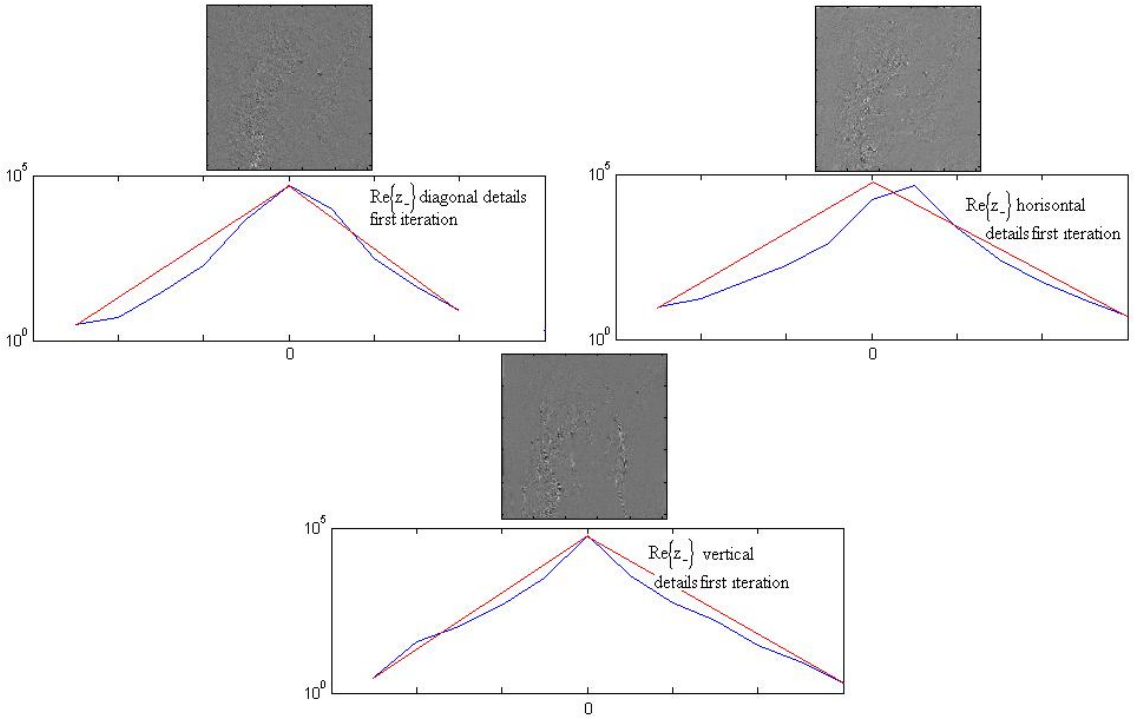


Figure 4.4: The histograms of some subbands of the HWT of the image Lena computed using the mother wavelets 'Daub, 20' are represented semi logarithmically (on the vertical axis are represented the logarithms of the values of the histograms) in blue. The corresponding linear dependencies are represented in red.

while for $W_s < 0$, eq. 4.42 becomes:

$$\begin{aligned}
&\frac{W_x - W_s}{\sigma_{W_n}^2} + \frac{\sqrt{2}}{\sigma_{W_s}} = 0 \Leftrightarrow \\
&\Leftrightarrow \frac{W_x - W_s}{\sigma_{W_n}^2} = -\sqrt{2}\frac{1}{\sigma_{W_s}} \Leftrightarrow \\
&\Leftrightarrow W_x - W_s = -\sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}} \Leftrightarrow \\
&\Leftrightarrow W_s = W_x + \sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}},
\end{aligned} \tag{4.44}$$

The condition $W_s > 0$ implies the condition $W_x > \sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}}$ and the condition $W_s < 0$ implies $W_x < -\sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}}$. Taking into account the fact that any real number, x can be written as $x = \text{sgn}(x) \cdot |x|$, the solution to equation 4.42 can be put under the following form:

$$\hat{W}_s = \begin{cases} \text{sgn}(W_x)\left(|W_x| - \sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}}\right), & |W_x| > \sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}}, \\ 0, & |W_x| < \sqrt{2}\frac{\sigma_{W_n}^2}{\sigma_{W_s}}. \end{cases} \tag{4.45}$$

equivalent with the final form:

$$\hat{W}_s = \operatorname{sgn}(W_x)\left(|W_x| - \sqrt{2}\frac{\sigma^2_{W_n}}{\sigma_{W_s}}\right)_+. \tag{4.46}$$

Comparing the input-output relations from 4.8 and 4.46 it can be observed their identity if $t = \sqrt{2}\frac{\sigma^2_{W_n}}{\sigma_{W_s}}$. So, the last MAP filter is a stf. Because the quantity σ_{W_s} varies from subband to subband, the threshold's value is variant. In this respect the last MAP filter can be considered adaptive. Of course the two variances in equation 4.46 must be estimated.

If we assume all the detail coefficients of the noise component, W_{n_1}, W_{n_2}, W_{n_3} and W_{n_4}, in fig. 4.3 are distributed following the same Gaussian law:

$$p_{W_{n_1}}(W_n) = p_{W_{n_2}}(W_n) = p_{W_{n_3}}(W_n) = p_{W_{n_4}}(W_n) = \frac{1}{\sqrt{2\pi}\sigma_{W_{n_d}}} \cdot e^{-\frac{W_n^2}{2\sigma^2_{W_{n_d}}}},$$

the pdfs of the real and imaginary parts of the detail HWT coefficients of the noise z_{+i_n}, z_{-i_n}, z_{-r_n}, z_{+r_n} can be computed by convolutions:

$$\begin{aligned}
p_{z_{+i_n}}(W_n) &= p_{W_{n_2}}(W_n) * p_{W_{n_3}}(W_n);\\
p_{z_{-i_n}}(W_n) &= p_{W_{n_2}}(W_n) * p_{W_{n_2}}(-W_n);\\
p_{z_{-r_n}}(W_n) &= p_{W_{n_1}}(W_n) * p_{W_{n_4}}(W_n);\\
p_{z_{+r_n}}(W_n) &= p_{W_{n_1}}(W_n) * p_{W_{n_4}}(-W_n);
\end{aligned}$$

All these four pdfs will be Gaussians (by convolving two Gaussians a new Gaussian is obtained) with the same standard deviation:

$$\sigma_{W_n} = \sqrt{\sigma^2_{W_{n_d}} + \sigma^2_{W_{n_d}}}. \tag{4.47}$$

The noise's variance σ_{W_n} can be estimated using the relations 4.9 and 4.47. The local version of this MAP filter is described by:

$$\hat{W}_s(i,j) = \operatorname{sgn}(W_x(i,j))\left(|W_x(i,j)| - \sqrt{2}\frac{\hat{\sigma}^2_{W_n}}{{}_l\hat{\sigma}_{W_s}(i,j)}\right)_+ \tag{4.48}$$

To estimate the local variance of the real and imaginary parts of the useful component, the relations 4.33, 4.34 and 4.35 can be used in each subband of the form z_{+i}, z_{-i}, z_{-r}, z_{+r}. The association of the HWT computed with the mother wavelets Daub,20 with the adaptive stf was tested to denoise the image Lena perturbed by AWGN with different variances. The results obtained are compared in table 4.6 with the results obtained using the association of the real and imaginary parts of the HWT with the local marginal zero order Wiener filter presented in table 4.5.

The superiority of the association HWT-adaptive stf, observed mainly for low PSNRi, is justified by the fact that the Laplace law is more suited to model the repartition of the

σ_n	10	15	20	25	30	35	50	100
PSNRi	28.15	24.65	22.15	20.21	18.62	17.28	14.19	8.16
PSNRo Adaptive Stf	34.81	32.97	31.64	30.55	29.67	28.98	27.28	24.01
PSNRo Wiener 0	34.81	32.80	31.37	30.21	29.29	28.47	26.61	22.92

Table 4.6: A comparison of the results obtained with the associations HWT-adaptive stf and HWT-local zero order Wiener filter used to denoise the image Lena perturbed with AWGN with different variances

wavelet coefficients than the Gauss law, as it is heavier tailed. The Gaussian hypothesis holds only asymptotically.

Instead of applying filters on the real and imaginary parts of the HWT coefficient, a filtering can be applied on the magnitudes of the HWT coefficients, but, although the simulations results are good, they do not overtake the results previously mentioned, this being the reason why I will not further insist on this type of filtering in this section.

4.3.1.3 Local vs global

In the previous sections were presented some marginal MAP filters. In each case there can be considered global and local approaches. To implement the local versions, the local variances were estimated in sliding windows. For the experimental results already presented, rectangular estimation windows with the size of 7x7 pixels were used. In each experiment already reported, the performance of the local version were better than the performance of the global one. The shape of the estimation windows was rectangular (isotropic) in all subbands. So, the preferential direction of each subband was not exploited. Elliptically estimation windows with the principal axis parallel with the preferential direction of each subband (anisotropic) will be considered in a following section. All the pixels of the estimation windows already used in the simulation reported in previous sections belonged to the same decomposition level of the WT (same scale) because the corresponding MAP filters were marginal. Better results can be obtained using three dimensional neighborhoods of the pixel which is currently estimated. Such neighborhoods will be considered in following sections.

4.3.1.4 Inter-scale dependency

If we talk about the dependency of wavelet coefficients we can consider four types of dependencies. This is illustrated in figure 4.5, where the coefficients, spatially located at (n, p), $(2n, 2p)$ and, respectively, $(4n, 4p)$, from three consecutive decomposition levels $m = 1 : 3$, are represented.

When we speak about the dependence of coefficients situated at the same scale and in the same sub-band (represented in yellow in the figure), we have intra-scale and intra-band dependency. When we refer at the dependency between coefficients situated at the same decomposition level, but in two different sub-bands (in blue), we have intra-scale inter-band dependency. If we consider coefficients from different scales but belonging to the same sub-band (green), we have inter-scale, intra-band dependency while, in the case when the coefficients belong to different scales and different sub-bands, we talk about

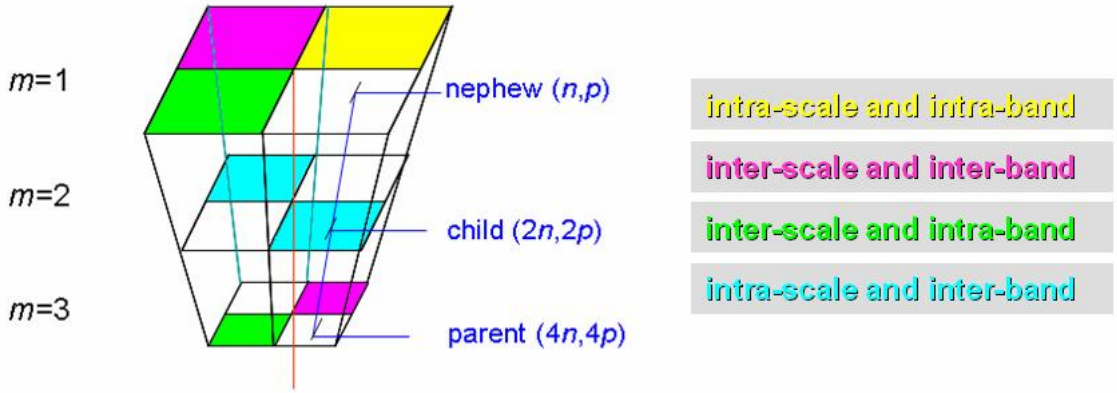


Figure 4.5: Different types of wavelet coefficients' dependencies

inter-scale inter-band dependency.

In the case of DWT a strong inter-scale dependency was observed. The marginal distributions are typically long tailed with high kurtosis (fourth moment divided by the squared variance). This shape, including the sharp peak at zero and the long tails, is the statistical manifestation of the sparseness property of wavelet coefficients. The wavelet coefficients of natural images display a self-reinforcing characteristic such that if one coefficient is large in magnitude, then other coefficients in its neighborhood are also likely to be large in magnitude (intra-scale dependence). The intuitive explanation for this is that localized image structures such as edges have substantial power across many scales and nearby spatial locations at a given orientation. The wavelet coefficients that represent the image will also have large magnitudes at these scales, locations and orientations (inter-scale intra-band dependency). However, the signs and relative magnitudes of these coefficients will depend on the exact shape, location and orientation of the structure they represent.

Each wavelet coefficient at level j, $W^j(2n, 2p)$, has a parent $W^{j+1}(n, p)$ and four children $W^{j-1}(4n, 4p)$, $W^{j-1}(4n+1, 4p)$, $W^{j-1}(4n, 4p+1)$ and $W^{j-1}(4n+1, 4p+1)$. As the relation between the acquired image's coefficients and the useful image's coefficients is given by: $W_x^j = W_s^j + W_n^j$, if the current level is considered $j = 1$ then the parent level is $j = 2$ and then we have the vectorial relationship $\mathbf{W_x} = \mathbf{W_s} + \mathbf{W_n}$, where:

$$\mathbf{W_x} = \left(W_x^1, W_x^2\right), \ \mathbf{W_s} = \left(W_s^1, W_s^2\right) \text{ and } \mathbf{W_n} = \left(W_n^1, W_n^2\right).$$

The inter-scale dependence implies that, if a parent coefficient is large, then its child coefficient is also large. By taking into consideration the parent coefficients, we capture the 'persistence across scale' characteristic.

If we want to take into account the inter-scale dependency in the construction of a MAP filter, we must use bivariate pdf's in equation 4.20. A technique for the construction of such bivariate pdf's will be presented in the following paragraph.

Gaussian Scale Mixtures (GSM)

The construction of the bivariate pdf can be done with the aid of Gaussian Scale Mixtures (GSM). This simple statistical model has been used to model natural signals such as speech and more recently the wavelet coefficients of natural images and is given in equation 4.49. It assumes that each vector of coefficients $\mathbf{W_s}$ is specified by a stationary bivariate zero mean Gaussian process $\mathbf{u}$ and a spatially fluctuating variance z.

$$\mathbf{W_s} = \sqrt{z} \cdot \mathbf{u},\ z \in \mathbb{R},\ \mathbf{W_s}, \mathbf{u} \in \mathbb{R}^d \tag{4.49}$$

The multiplier z is usually a function of the surrounding coefficient values (like the local variance of the coefficients within the same scale or a more complex function of the neighboring coefficients within the same and adjacent scales). The result is always leptokurotic (kurtosis ≥ 3), its distribution having long tails. The MMSE estimate with such priors takes the form of a locally adaptive Wiener-like estimator. The number of elements of the vectors $\mathbf{W_s}$ and $\mathbf{u}$, when it is considered only the parent-child dependence, is $d = 2$.

To model the self-reinforcing property of the coefficients, z must be slowly varying but does not need to be symmetric in all directions. It has been shown that, for slowly varying z, this model can successfully simulate the high kurtosis and long tails of the marginal distributions. The stationary part of the model $\mathbf{u}$ is Gaussian distributed over a small neighborhood of wavelet coefficients. It is generally assumed that z varies slowly enough to be considered constant over that neighborhood of coefficients. Under this assumption the model is now a particular form of a spherically invariant random process called a GSM. For a small neighborhood of coefficients at nearby spatial locations and scale, a GSM vector W_s is the product of two independent random variables: a positive scalar z referred to as the hidden multiplier or mixing variable and a Gaussian random vector $\mathbf{u}$. The pdf of the Gaussian vector $\mathbf{u}$ is given by:

$$p_u(u) = \frac{1}{(2\pi\sigma^2)^{d/2}} \cdot e^{-\frac{\|u\|^2}{2\sigma^2}}. \tag{4.50}$$

Setting $a = \sqrt{z}$ we will have $W_s = a \cdot u$ and the pdf of the random vector W_s is given by:

$$p_{W_s}(W_s) = \int_0^\infty p_a(a) \cdot \frac{1}{a^d} \cdot p_u\left(\frac{W_s}{a}\right) da. \tag{4.51}$$

Taking into account the relation between a and z, the pdf of a can be expressed as a function of the pdf of z:

$$p_a(a) = p_z(z) \cdot \frac{dz}{da} = 2a \cdot p_z\left(a^2\right). \tag{4.52}$$

Substituting 4.52 in 4.51, the expression of the pdf of W_s becomes:

$$p_{W_s}(W_s) = \int_0^\infty 2a \cdot p_z\left(a^2\right) \cdot \frac{1}{a^d} \cdot p_u\left(\frac{W_s}{a}\right) da. \tag{4.53}$$

It remains to specify the prior probability function of the multiplier z, $p_z(z)$.

Bivariate probability density functions Some proposals of prior probability functions are made in [Sel08]. The first of these models that I will further present is the Gamma law:

$$p_z(z) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot z^{\alpha-1} \cdot e^{-\beta z} \text{ with } \alpha = \beta = \frac{d+1}{2}. \tag{4.54}$$

Substituting 4.50 and 4.54 in 4.53, the expression of the pdf of W_s becomes:

$${}^1p_{W_s}(W_s) = \frac{2\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty a \cdot a^{2(\alpha-1)} \cdot e^{-\beta a^2} \cdot \frac{1}{a^d} \cdot \frac{1}{(2\pi\sigma^2)^{d/2}} \cdot e^{-\frac{\|W_s\|^2}{2a^2\sigma^2}} da, \tag{4.55}$$

or, taking into account the values of the parameters α and β:

$${}^1p_{W_s}(W_s) = \frac{2\left(\frac{d+1}{2}\right)^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^\infty a \cdot a^{2\left(\frac{d-1}{2}\right)} \cdot e^{-\frac{d+1}{2}a^2} \cdot \frac{1}{a^d} \cdot \frac{1}{(2\pi\sigma^2)^{d/2}} \cdot e^{-\frac{\|W_s\|^2}{2a^2\sigma^2}} da. \tag{4.56}$$

Finally:

$${}^1p_{W_s}(W_s) = \frac{2}{\Gamma\left(\frac{d+1}{2}\right)} \cdot \left(\frac{d+1}{2}\right)^{\frac{d+1}{2}} \cdot \frac{1}{(2\pi\sigma^2)^{d/2}} \cdot \int_0^\infty e^{-\frac{d+1}{2}a^2 - \frac{\|W_s\|^2}{2a^2\sigma^2}} da. \tag{4.57}$$

The last integral can be computed using the following identity:

$$\int_0^\infty e^{-K_1 a^2 - \frac{K_2}{a^2}} da = \frac{1}{2}\sqrt{\frac{\pi}{K_1}} e^{-2\sqrt{K_1 K_2}}. \tag{4.58}$$

By comparing 4.58 with 4.57 we can identify the constants $K_1 = \frac{d+1}{2}$ and $K_2 = \frac{\|W_s\|^2}{2\sigma^2}$, and the pdf of W_s can be put in the form:

$${}^1p_{W_s}(W_s) = \frac{2}{\Gamma\left(\frac{d+1}{2}\right)} \cdot \left(\frac{d+1}{2}\right)^{\frac{d+1}{2}} \cdot \frac{1}{(2\pi\sigma^2)^{d/2}} \cdot \frac{1}{2}\sqrt{\frac{\pi}{\frac{d+1}{2}}} \cdot e^{-\frac{\sqrt{d+1}}{\sigma}\|W_s\|}. \tag{4.59}$$

This is a d-dimensional spherically-contoured multivariate pdf [Sel08].

For $d = 2$, this pdf is bivariate:

$${}^1p_{W_s}(W_s) = \frac{3}{2\pi\sigma^2} \cdot e^{-\frac{\sqrt{3}}{\sigma} \cdot \sqrt{W_s^1 + W_s^2}}. \tag{4.60}$$

With the aid of this bivariate distribution, the bishrink MAP filter [SS02] can be constructed. This construction will be explained in the following section.

For $d = 1$, the pdf in 4.59 is univariate:

$${}^1p_{W_s}(W_s) = \frac{\sqrt{2}}{\sigma} \cdot e^{-\frac{\sqrt{2}}{\sigma} \cdot |W_s|}. \tag{4.61}$$

This univariate pdf is of the form of the Laplace law (see the hypotheses of the adaptive stf, 4.41). In consequence the GSM hypothesis, even for the case of a unique scale, is useful for modeling the repartition of the wavelet coefficients.

The joint pdf from 4.60 is proportional with the product of two univariate pdfs of Laplace type:

$$\begin{aligned} {}^{1}p_{W_s}\left(W_s^1, W_s^2\right) &= \frac{3}{2\pi\sigma^2}\cdot e^{-\frac{\sqrt{3}}{\sigma}\cdot\sqrt{W_s^1+W_s^2}} \\ &= \sqrt{\frac{3}{2\pi}}\frac{1}{\sigma}e^{-\frac{\sqrt{3}}{\sigma}|W_s^1|}\cdot\sqrt{\frac{3}{2\pi}}\frac{1}{\sigma}e^{-\frac{\sqrt{3}}{\sigma}|W_s^2|} \\ &= K^1p_{W_s^1}\left(W_s^1\right)\cdot K^1p_{W_s^2}\left(W_s^2\right), \end{aligned} \tag{4.62}$$

which can be regarded as marginal distributions of the corresponding wavelet coefficients at the considered successive scales. K represents a constant. Because:

$${}^{1}p_{W_s}\left(W_s^1, W_s^2\right) \neq^{1} p_{W_s^1}\left(W_s^1\right)\cdot^{1} p_{W_s^2}\left(W_s^2\right), \tag{4.63}$$

the random variables W_s^1 and W_s^2 are not independent. This observation reflects the inter-scale dependence. But it can be affirmed that the hypothesis of the univariate Laplace law 4.41 for the marginal distribution of the wavelet coefficients already analyzed is in accord with the model in 4.60.

Another proposition for the prior probability function, made in [Sel08], is the exponential law: $p_z(z) = e^{-z}$ for $z \geq 0$. In the following, two wavelet coefficients bivariate distribution models corresponding to different pdfs for the vector $\mathbf{u}$, constructed with the exponential prior are presented:

$$\begin{aligned} p_u(u) &= \frac{1}{2\pi\sigma^2}\cdot e^{-\frac{u_1^2+u_2^2}{2\sigma^2}}, \\ {}^{2}p_{W_s}(W_s) &= \frac{1}{\pi\sigma^2}K_\lambda\left(\frac{\sqrt{2}}{\sigma}\sqrt{(W_s^1)^2+(W_s^2)^2}\right), \text{ with} \\ K_\lambda(a) &= \frac{1}{2}\cdot\left(\frac{a}{2}\right)^\lambda\cdot\int_0^\infty t^{-\lambda-1}\cdot e^{-t-\frac{u^2}{4t}}dt \end{aligned} \tag{4.64}$$

and

$$\begin{aligned} p_u(u) &= \frac{1}{2\pi\sigma_1\sigma_2}\cdot e^{-\frac{1}{2}\left(\frac{u_1^2}{\sigma_1^2}+\frac{u_2^2}{\sigma_2^2}\right)}, \\ {}^{3}p_{W_s}(W_s) &= \frac{1}{\pi\sigma_1\sigma_2}K_0\left(\sqrt{2\left(\frac{(W_s^1)^2}{\sigma_1^2}+\frac{(W_s^2)^2}{\sigma_2^2}\right)}\right), \text{ with} \\ K_0(a) &= \frac{1}{2}\cdot\int_0^\infty t^{-1}\cdot e^{-t-\frac{a^2}{4t}}dt \end{aligned} \tag{4.65}$$

The model in 4.64 is also a spherically-contoured bivariate pdf. The model in 4.65 represents an elliptically-contoured bivariate pdf.

The GSM model was also used for the conception of one of the best denoising methods [PSWS03]. The prior probability function proposed in this reference is $p_z(z) = \frac{1}{z}$, for $z \geq 0$. It must be noticed that this is not a pdf in the strict sense. By substituting this

function in 4.53 and supposing that the pdf of the Gaussian vector is that given in 4.50, it can be written:

$$ {}^{4}p_{W_s}\left(W_s\right) = \int_0^{\infty} 2a \cdot p_z\left(a^2\right) \cdot \frac{1}{a^d} \cdot p_u\left(\frac{W_s}{a}\right) da, \tag{4.66} $$

or:

$$ {}^{4}p_{W_s}\left(W_s\right) = \int_0^{\infty} 2a \cdot \frac{1}{a^2} \cdot \frac{1}{a^d} \cdot \frac{1}{\left(2\pi\sigma^2\right)^{d/2}} \cdot e^{-\frac{\|W_s\|^2}{2a^2\sigma^2}} da. \tag{4.67} $$

The las equation can be put in the following form:

$$ {}^{4}p_{W_s}\left(W_s\right) = \frac{2}{\left(2\pi\sigma^2\right)^{d/2}} \cdot \int_0^{\infty} \frac{1}{a^{d+1}} \cdot e^{-\frac{\|W_s\|^2}{2a^2\sigma^2}} da. \tag{4.68} $$

By changing the variable $a = \frac{1}{t}$, the pdf of W_s becomes:

$$ {}^{4}p_{W_s}\left(W_s\right) = \frac{-2}{\left(2\pi\sigma^2\right)^{d/2}} \cdot \int_{\infty}^{0} t^{d+1} \cdot e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}} \frac{dt}{t^2}. \tag{4.69} $$

For $d = 2$, the bivariate pdf becomes

$$ {}^{4}p_{W_s}\left(W_s^1, W_s^2\right) = \frac{2}{2\pi\sigma^2} \cdot \int_0^{\infty} t \cdot e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}} dt. \tag{4.70} $$

But:

$$ \frac{d}{dt}\left(e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}}\right) = -\frac{\|W_s\|^2 t}{\sigma^2} \cdot e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}}. \tag{4.71} $$

Hence:

$$ t \cdot e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}} = -\frac{\sigma^2}{\|W_s\|^2} \cdot \frac{d}{dt}\left(e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}}\right). \tag{4.72} $$

Substituting 4.72 in 4.70, this equation becomes:

$$ {}^{4}p_{W_s}\left(W_s^1, W_s^2\right) = \frac{2}{2\pi\sigma^2} \cdot \int_0^{\infty} -\frac{\sigma^2}{\|W_s\|^2} \cdot \frac{d}{dt}\left(e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}}\right) dt, \tag{4.73} $$

or:

$$ {}^{4}p_{W_s}\left(W_s^1, W_s^2\right) = -\frac{1}{\pi\|W_s\|^2} \cdot e^{-\frac{\|W_s\|^2 t^2}{2\sigma^2}}|_0^{\infty} = \frac{1}{\pi\|W_s\|^2} = \frac{1}{\pi\left(\left(W_s^1\right)^2 + \left(W_s^2\right)^2\right)}. \tag{4.74} $$

It results in a third spherically-contoured bivariate pdf.

A last bivariate pdf for the wavelet coefficients was proposed in [AK05]:

$$ {}^{5}p_{W_s}\left(W_s^1, W_s^2\right) = \frac{\gamma}{2\pi\left(\left(W_s^1\right)^2 + \left(W_s^2\right)^2 + \gamma^2\right)^{\frac{3}{2}}}. \tag{4.75} $$

4.3.1.5 The Bishrink Filter

For the construction of the Bishrink filter, the noise is assumed i.i.d. Gaussian,

$$p_{W_n}(W_n) = \frac{1}{2\pi\sigma_{W_n}^2} \cdot e^{-\frac{\left(W_n^1\right)^2+\left(W_n^2\right)^2}{2\sigma_{W_n}^2}}. \tag{4.76}$$

The model of the noise-free image is given in 4.60:

$$p_{W_s}(W_s) = \frac{3}{2\pi\sigma_{W_s}^2} \cdot e^{-\frac{\sqrt{3}}{\sigma_{W_s}}\cdot\sqrt{(W_s^1)^2+(W_s^2)^2}},$$

a heavy tailed distribution. Substituting these two pdfs in the equation of the MAP filter, 4.20, this equation becomes:

$$\begin{aligned}\hat{W}_s(W_x) &= \underset{W_s}{argmax}\left\{ln\left(\frac{1}{2\pi\sigma_{W_n}^2}e^{-\frac{\left(W_x^1-W_s^1\right)^2+\left(W_x^2-W_s^2\right)^2}{2\sigma_{W_n}^2}}\frac{3}{2\pi\sigma_{W_s}^2}e^{-\frac{\sqrt{3}}{\sigma_{W_s}}\sqrt{(W_s^1)^2+(W_s^2)^2}}\right)\right\} \\ &= \underset{W_s}{argmax}\left\{ln\frac{3}{4\pi^2\sigma_{W_s}^2\sigma_{W_n}^2} - \frac{(W_x^1-W_s^1)^2+(W_x^2-W_s^2)^2}{2\sigma_{W_n}^2} - \frac{\sqrt{3}}{\sigma_{W_s}}\sqrt{(W_s^1)^2+(W_s^2)^2}\right\}\end{aligned} \tag{4.77}$$

The system of equations to be solved in order to obtain the values that give the maximum of the right side of eq. 4.77 is:

$$\begin{cases} \frac{d}{dW_s^1}\left(ln\frac{3}{4\pi^2\sigma_{W_s}^2\sigma_{W_n}^2} - \frac{\left(W_x^1-W_s^1\right)^2}{2\sigma_{W_n}^2} - \frac{\left(W_x^2-W_s^2\right)^2}{2\sigma_{W_n}^2} - \frac{\sqrt{3}}{\sigma_{W_s}}\cdot\sqrt{(W_s^1)^2+(W_s^2)^2}\right) &= 0 \\ \frac{d}{dW_s^2}\left(ln\frac{3}{4\pi^2\sigma_{W_s}^2\sigma_{W_n}^2} - \frac{\left(W_x^1-W_s^1\right)^2}{2\sigma_{W_n}^2} - \frac{\left(W_x^2-W_s^2\right)^2}{2\sigma_{W_n}^2} - \frac{\sqrt{3}}{\sigma_{W_s}}\cdot\sqrt{(W_s^1)^2+(W_s^2)^2}\right) &= 0 \end{cases}, \tag{4.78}$$

or:

$$\begin{cases} \frac{W_x^1-W_s^1}{\sigma_{W_n}^2} - \frac{\sqrt{3}}{\sigma_{W_s}}\cdot\frac{W_s^1}{\sqrt{(W_s^1)^2+(W_s^2)^2}} &= 0 \\ \frac{W_x^2-W_s^2}{\sigma_{W_n}^2} - \frac{\sqrt{3}}{\sigma_{W_s}}\cdot\frac{W_s^2}{\sqrt{(W_s^1)^2+(W_s^2)^2}} &= 0 \end{cases}. \tag{4.79}$$

The system can be put in the following form:

$$\begin{cases} \frac{W_x^1}{\sigma_{W_n}^2} &= \frac{\sqrt{3}}{\sigma_{W_s}}\cdot\frac{W_s^1}{\sqrt{(W_s^1)^2+(W_s^2)^2}} + \frac{W_s^1}{\sigma_{W_n}^2} \\ \frac{W_x^2}{\sigma_{W_n}^2} &= \frac{\sqrt{3}}{\sigma_{W_s}}\cdot\frac{W_s^2}{\sqrt{(W_s^1)^2+(W_s^2)^2}} + \frac{W_s^2}{\sigma_{W_n}^2} \end{cases}, \tag{4.80}$$

or, equivalently:

$$\begin{cases} W_s^1 &= \frac{\sigma_{W_s}\sqrt{(W_s^1)^2+(W_s^2)^2}}{\sigma_{W_s}\sqrt{(W_s^1)^2+(W_s^2)^2}+\sqrt{3}\sigma_{W_n}^2}W_x^1 \\ W_s^2 &= \frac{\sigma_{W_s}\sqrt{(W_s^1)^2+(W_s^2)^2}}{\sigma_{W_s}\sqrt{(W_s^1)^2+(W_s^2)^2}+\sqrt{3}\sigma_{W_n}^2}W_x^2 \end{cases} \tag{4.81}$$

Computing the square of each equation it results:

$$\begin{cases} \left(W_s^1\right)^2 &= \frac{\sigma_{W_s}^2\left(\left(W_s^1\right)^2+\left(W_s^2\right)^2\right)}{\left(\sigma_{W_s}\sqrt{\left(W_s^1\right)^2+\left(W_s^2\right)^2}+\sqrt{3}\sigma_{W_n}^2\right)^2}\left(W_x^1\right)^2 \\ \left(W_s^2\right)^2 &= \frac{\sigma_{W_s}^2\left(\left(W_s^1\right)^2+\left(W_s^2\right)^2\right)}{\left(\sigma_{W_s}\sqrt{\left(W_s^1\right)^2+\left(W_s^2\right)^2}+\sqrt{3}\sigma_{W_n}^2\right)^2}\left(W_x^2\right)^2 \end{cases} \tag{4.82}$$

By adding the two equations in 4.82 it can be obtained:

$$\begin{aligned} \left(W_s^1\right)^2+\left(W_s^2\right)^2 &= \frac{\sigma_{W_s}^2\left(\left(W_s^1\right)^2+\left(W_s^2\right)^2\right)}{\left(\sigma_{W_s}\sqrt{\left(W_s^1\right)^2+\left(W_s^2\right)^2}+\sqrt{3}\sigma_{W_n}^2\right)^2}\left(\left(W_x^1\right)^2+\left(W_x^2\right)^2\right) \Leftrightarrow \\ \left(W_x^1\right)^2+\left(W_x^2\right)^2 &= \frac{\left(\sigma_{W_s}\sqrt{\left(W_s^1\right)^2+\left(W_s^2\right)^2}+\sqrt{3}\sigma_{W_n}^2\right)^2}{\sigma_{W_s}^2} \Leftrightarrow \\ \sqrt{\left(W_s^1\right)^2+\left(W_s^2\right)^2} &= \left(\sqrt{\left(W_x^1\right)^2+\left(W_x^2\right)^2}-\frac{\sqrt{3}\sigma_{W_n}^2}{\sigma_{W_s}}\right)_+ \end{aligned} \tag{4.83}$$

Substituting the result obtained in 4.83 in the two equations of the system in 4.81, it can be written:

$$\begin{cases} \hat{W}_s^1 &= \frac{\left(\sqrt{\left(W_x^1\right)^2+\left(W_x^2\right)^2}-\frac{\sqrt{3}\sigma_{W_n}^2}{\sigma_{W_s}}\right)_+}{\sqrt{\left(W_x^1\right)^2+\left(W_x^2\right)^2}}\cdot W_x^1 \\ \hat{W}_s^2 &= \frac{\left(\sqrt{\left(W_x^1\right)^2+\left(W_x^2\right)^2}-\frac{\sqrt{3}\sigma_{W_n}^2}{\sigma_{W_s}}\right)_+}{\sqrt{\left(W_x^1\right)^2+\left(W_x^2\right)^2}}\cdot W_x^2 \end{cases}. \tag{4.84}$$

Finally, the input-output relation of the bishrink filter is:

$$\hat{W}_s^1 = \frac{\left(\sqrt{\left(W_x^1\right)^2+\left(W_x^2\right)^2}-\frac{\sqrt{3}\sigma_{W_n}^2}{\sigma_{W_s}}\right)_+}{\sqrt{\left(W_x^1\right)^2+\left(W_x^2\right)^2}}\cdot W_x^1. \tag{4.85}$$

It can be observed that the bishrink filter is an adaptive stf type estimator, the threshold being $t=\frac{\sqrt{3}\sigma_{W_n}^2}{\sigma_{W_s}}$. So, the bishrink filter has a dead zone as well.

This estimator requires prior knowledge of the marginal variance of the noise-free image for each wavelet coefficient. To estimate the noise variance from the noisy wavelet coefficients, a robust median estimator from the finest scale wavelet coefficients is used [DJ94]:

$$\hat{\sigma}_{W_n}^2=\frac{median\left(|W_x|\right)}{0.6745},\ W_x\in \text{subband } HH^1. \tag{4.86}$$

In [SS02] the marginal variance of the k^{th} coefficient is estimated using neighboring coefficients in the region $N(k)$, a squared shaped window of 7x7 pixels, centered in this coefficient. To make this estimation, we take into account the fact that $\sigma_{W_x}^2=\sigma_{W_s}^2+\sigma_{W_n}^2$,

where $\sigma^2_{W_x}$ represents the marginal variance of noisy observations W_x^1 and W_x^2. In order to estimate the marginal variance of the noisy observations, the following relation is proposed in [SS02]:

$$\hat{\sigma}^2_{W_x} = \frac{1}{M} \sum_{W_x^i \in N(k)} \left(W_x^i\right)^2, \tag{4.87}$$

where M is the size of the neighborhood $N(k)$. Then σ_{W_s} can be estimated as:

$$\hat{\sigma}_{W_s^i} = \sqrt{\left(\hat{\sigma}^2_{W_x^i} - \hat{\sigma}^2_{W_n}\right)_+} \; i = 1, 2. \tag{4.88}$$

For the estimation of the local standard deviation of the useful component of a parent coefficient, $\hat{\sigma}_{W_s^2}$ in a given subband, we first interpolate the subband by repeating each line and column and then we apply relations 4.87 and 4.88. Finally, the local standard deviation of the useful component of the child coefficients taking into consideration the parent is computed with the relation:

$$\hat{\sigma}_{W_s} = \frac{\hat{\sigma}_{W_s^1} + 0.5\hat{\sigma}_{W_s^2}}{2}. \tag{4.89}$$

Sensitivity analysis of the Bishrink filter

The sensitivity of the bishrink filter with the estimation of the noise standard deviation, $\hat{\sigma}_{W_n}$ can be computed with the relation:

$$S^{\hat{\sigma}_{W_n}}_{\hat{W}_s^1} = \frac{d\hat{W}_s^1}{d\sigma_{W_n}} \cdot \frac{\sigma_{W_n}}{\hat{W}_s^1}. \tag{4.90}$$

The input-output relation of the bishrink, 4.85, can be put in the following form:

$$\hat{W}_s^1 = \begin{cases} \frac{\sqrt{(W_x^1)^2+(W_x^2)^2} - \frac{\sqrt{3}\sigma^2_{W_n}}{\sigma_{W_s}}}{\sqrt{(W_x^1)^2+(W_x^2)^2}} \cdot W_x^1, & \text{if } \sqrt{\left(W_x^1\right)^2 + \left(W_x^2\right)^2} > \frac{\sqrt{3}\sigma^2_{W_n}}{\sigma_{W_s}}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.91}$$

Taking into account 4.91, we can rewrite 4.90 as:

$$S^{\hat{\sigma}_{W_n}}_{\hat{W}_s^1} = \begin{cases} \frac{-2\sqrt{3}\sigma^2_{W_n}}{\sigma_{W_s}\sqrt{(W_x^1)^2+(W_x^2)^2} - \sqrt{3}\sigma^2_{W_n}}, & \text{if } \sqrt{\left(W_x^1\right)^2 + \left(W_x^2\right)^2} > \frac{\sqrt{3}\sigma^2_{W_n}}{\sigma_{W_s}}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.92}$$

The absolute value of this sensitivity is an increasing function of $\hat{\sigma}_{W_n}$. When the value of the estimation of the noise standard deviation increases, the performance of the bishrink filter decreases.

Another very important parameter of the bishrink filter is the local estimation of the marginal variance of the noise-free image, $\hat{\sigma}_{W_s}$. The sensitivity of the estimation $\hat{W}_s^1$ with $\hat{\sigma}_{W_s}$ is given by:

$$S^{\hat{\sigma}_{W_s}}_{\hat{W}_s^1} = \begin{cases} \frac{\sqrt{3}\sigma^2_{W_n}}{\sigma_{W_s}\sqrt{(W_x^1)^2+(W_x^2)^2} - \sqrt{3}\sigma^2_{W_n}}, & \text{if } \sqrt{\left(W_x^1\right)^2 + \left(W_x^2\right)^2} > \frac{\sqrt{3}\sigma^2_{W_n}}{\sigma_{W_s}}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.93}$$

This is a decreasing function of $\hat{\sigma}_{W_s}$. The precision of the estimation based on the use of the bishrink filter decreases with the decreasing of $\hat{\sigma}_{W_s}$.

The local variance of a pixel, $\hat{\sigma}_{W_s}$, can be interpreted in two ways.

First, it represents a homogeneity degree measure for the region to which the considered pixel belongs. This behavior can be observed in figure 4.6, where the Barbara image, the associated local variances image and the images composed by the local variances of the pixels of the original image belonging to a given interval (classes) are presented together.

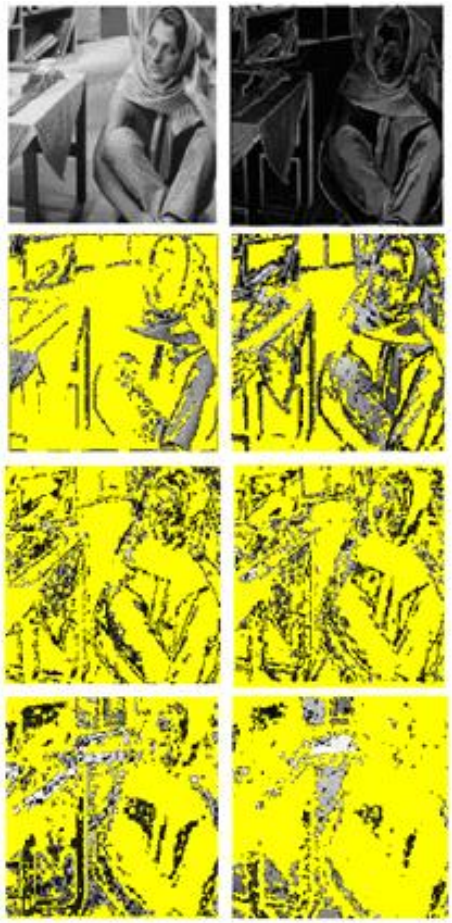


Figure 4.6: From left to right and up to bottom: original Barbara image; the image of local variances, the correspondent classes (obtained comparing the local variances with decreasing thresholds) - the first four classes contain textures and contours; the last two classes contain textures and homogeneous regions. For each of the last six pictures, the pixels belonging to a different class are represented in yellow.

The regions with high homogeneity in the Barbara image correspond to the dark regions in the image of local variances. All the pixels belonging to a perfect homogeneous region have the same value. So, their local variances are equal with zero. The space localization of a pixel belonging to a homogeneous region is not very important because it has a value very similar with the values of its neighbors.

The values of the pixels belonging to a textured region oscillate in space and they have non-null local variances. The space localization of a pixel belonging to a textured region is more important than the space localization of one from a homogeneous region because of those oscillations (its neighbors could have different values). Finally, the pixels belonging to an edge have the highest local variances. Their space localization is more important than the space localization of a pixel belonging to a textured region because their values differ from those of their neighbors. So, the bishrink filter treats the edges very well (because their pixels have high local variance and the sensitivity of the bishrink filter is inverse proportional with the local variance), the estimation of the textured regions must be corrected and the worst treatment corresponds to the homogeneous regions. The denoising quality of pixels with slightly different σ_s will be very different in the homogeneous regions. The sensitivity $S_{\hat{W}_s^1}^{\hat{\sigma}_{W_s}}$ increases with the increasing of σ_{W_n}. So, the degradation of the homogeneous and textured zones of the noise-free image is amplified by the increasing of the noise standard deviation. Consequently the most difficult regime of the bishrink filter corresponds to the treatment of homogeneous regions of highly noisy images. This drawback can be diminished by the fusion of some partial results of denoising. These partial results can be obtained through diversification.

Similar sensitivity analysis can be accomplished for the zero order Wiener filter or for the adaptive soft-thresholding filter, concluding that their worst behavior corresponds to the homogeneous regions of their noise-free input image component.

Secondly, the local variance of a pixel gives some information about the frequency content of the region to which the considered pixel belongs. If the pixels of a given region have low local variances then the considered region contains low frequencies. If these pixels have high local variances then the considered region contains high frequencies.

In conclusion, the local variance of a given pixel could be regarded as a measure of the space-frequency localization of that pixel.

Associating the Bishrink filter with different WTs

Some experimental results, obtained using the bishrink filter, are presented in the following. First the bishrink filter is associated with the 2D DWT and with the Diversity Enhanced Discrete Wavelet Transform (DEDWT) respectively. The construction of the DEDWT and its association with the MAP filter is presented in figure 4.7. The nine members of the Daubechies family of mother wavelets with the shortest supports, having a number of vanishing moments between 2 and 10, denoted with 'Daub, par', with $par = 2 \cdot$ no. of vanishing moments are used for the computation of the DEDWT.

The denoising system makes, first, a diversification of the input image with the aid of the N DWTs computed with different mother wavelets associated to the same bishrink filter and with the N IDWTs, producing N partial results, $\hat{s}_1$, $\hat{s}_2$, ..., $\hat{s}_N$, and second, it synthesizes the final result by averaging the partial results ($\beta_1 = \beta_s = \ldots = \beta_N = \frac{1}{N}$). So, it does not make any difference between different mother wavelets. All of them are weighted with the same weight. The results obtained after this association are presented in table 4.7

The best results, corresponding to the use of the DE DWT, are marked with bold fonts. The values corresponding to the second place concerning the quality of denoising

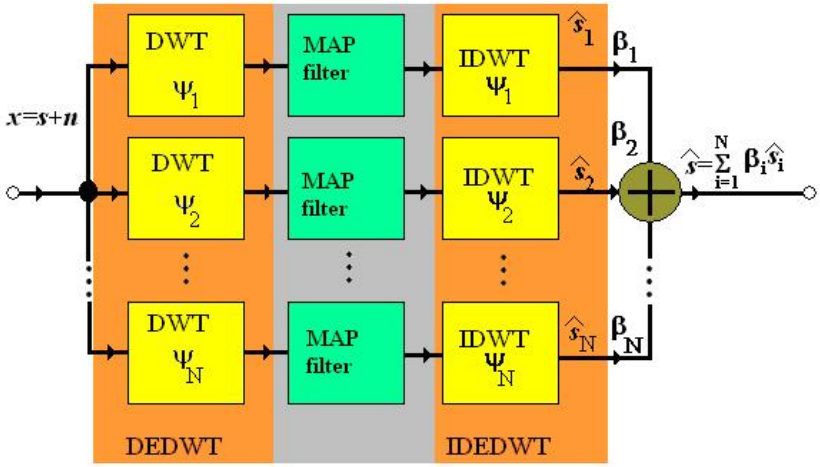


Figure 4.7: The architecture of the denoising system based on the association of the DE DWT with the bishrink filter

σ_n		10	15	20	25	30	35
PSNRi		28.15	24.65	22.15	20.21	18.62	17.28
PSNRo	**Daub,4**	33.92	31.93	30.58	29.51	28.62	27.91
	Daub,6	34.11	32.17	30.79	29.7	28.81	28.06
	Daub,8	34.17	32.21	30.87	29.75	28.90	28.18
	Daub,10	34.21	32.29	30.96	*29.84*	*28.99*	*28.31*
	Daub,12	34.23	*32.29*	*30.96*	29.83	28.96	28.28
	Daub,14	*34.23*	32.27	30.95	29.83	28.94	28.21
	Daub,16	34.19	32.26	30.94	29.83	28.97	28.22
	Daub,18	34.21	32.27	30.95	29.82	28.97	28.25
	Daub,20	34.21	32.28	30.95	29.82	28.95	28.25
	DE DWT	**35.01**	**33.14**	**31.84**	**30.72**	**29.83**	**29.11**

Table 4.7: PSNRs obtained using the bishrink filter in the 2D DWTs domain, computed using the mother wavelets from the Daubechies family and in the DE DWT domain, for the image Lena perturbed with AWGN with different variances

are marked with italic fonts. The utilization of the diversification mechanism brings a denoising gain around 0.8 dB. There is no best mother wavelet for this experiment. The best results, given by an association 2D DWT - bishrink filter are obtained for high PSNRs using the Daubechies mother wavelets with 6 vanishing moments and for low PSNRs using the Daubechies mother wavelets with 7 vanishing moments. This is why a combination of all the mother wavelets belonging to the Daubechies family like that used in the case of the DE DWT improves the denoising.

The superiority of the DE DWT versus the 2D DWT can be also observed comparing the values on the last column of the table with the corresponding values presented in [SS02], where the 2D DWT was computed using a mother wavelets proposed by Kingsbury.

The following experiments refer to the association HWT - bishrink filter. In the first experiment the bishrink filter is applied separately to the real and imaginary parts of the

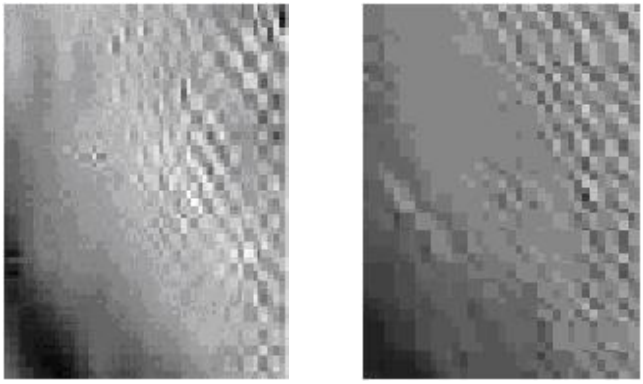
HWT coefficients computed using the mother wavelets 'Daub,20'.

$\sigma_{\mathbf{n}}$	10	15	20	25	30	35	50	100
PSNRi	28.15	24.65	22.15	20.21	18.62	17.28	14.19	8.16
PSNRo DWT - Bishrink	34.21	32.28	30.95	29.82	28.95	28.25	26.78	24.03
PSNRo HWT - Bishrink	34.76	32.97	31.70	30.65	29.81	29.17	27.55	24.42

Table 4.8: A comparison of the performance obtained associating the bishrink filter with the DWT and with the HWT respectively for denoising the image Lena perturbed with AWGN with different variances. The mother wavelets 'Daub,20' was used in both experiments.

The superiority of the HWT versus the DWT can be observed from the enhancement of PSNR point of view. The importance of the diversification can be appreciated comparing the last line of table 4.8 with the last line of table 4.7. Despite the superiority of the HWT versus the DWT, the DEDWT produces better results than the HWT computed with the mother wavelets Daub, 20.

To compare the proposed version of HWT with the 2D DWT, we applied the same denoising procedure based on both WTs in similar conditions (input image, mother wavelets) obtaining the results in fig. 4.8. Both images represent a zoom on a leg with a regular



(a) 2D DWT (b) HWT

Figure 4.8: A comparison of the directional selectivity of 2D DWT (a) and HWT (b).

texture from Barbara image. This illustrates that, compared with 2D DWT, the HWT leads to better visual results. Fig. 4.8(a) corresponding to the 2D DWT is strongly blurred. It clearly appears that the texture with an apparent angle of $-\pi/4$ is heavily corrupted by patterns in the opposite direction, due to the mixing in the "diagonal" subband produced in the 2D DWT case. Details are better preserved in the HWT case, fig. 4.8(b). There is much less directional mixture in the HWT case.

In the second experiment the bishrink filter is associated with the magnitudes values of the HWT coefficients. This association can work well due to the existence of the dead zone of the bishrink filter. The results obtained applying this association to denoise the image Lena perturbed with AWGN with different variances are presented in table 4.9.

σ_n	10	15	20	25	30	35	50	100
PSNRi	28.15	24.65	22.15	20.21	18.62	17.28	14.19	8.16
PSNRo Re + Im HWT - Bishrink	34.76	32.97	31.70	30.65	29.81	29.17	27.55	24.42
PSNRo Abs HWT - Bishrink	34.73	32.93	31.62	30.63	29.79	29.10	27.47	24.34

Table 4.9: A comparison of the performance obtained associating the bishrink filter with the real and imaginary parts of HWT and with the magnitude of the HWT respectively for denoising the image Lena perturbed with AWGN with different variances. The mother wavelets 'Daub,20' was used in both experiments.

It can be observed that the results obtained from the association magnitude of HWT - bishrink are similar to those obtained from the association of the real and imaginary parts of HWT - bishrink filter, slightly inferior.

Another WT that can be used in denoising is the Diversity Enhanced HWT, DE HWT. A denoising system based on this WT has the architecture from figure 4.9. The DE HWT is a redundant WT. Its redundancy is quit high.

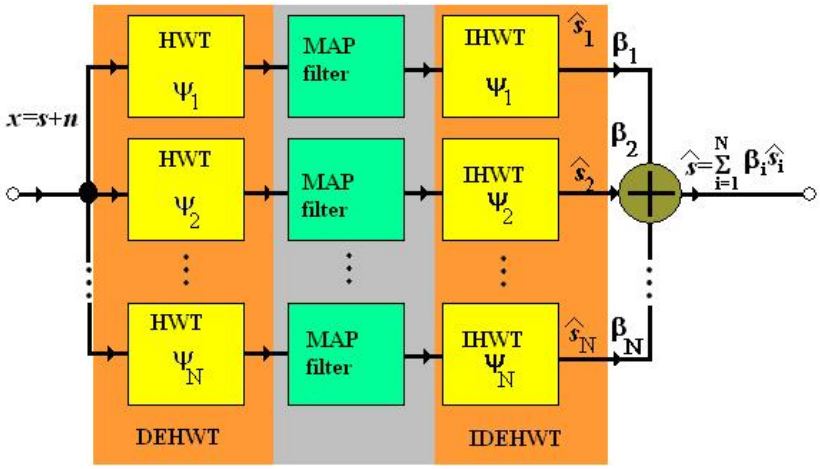


Figure 4.9: The architecture of the denoising system based on the association of the DE HWT with the bishrink filter

Applying the denoising system in figure 4.9, for the particular case of the bishrink filter, to the Lena image perturbed by AWGN with different variances, the results presented in table 4.10 are obtained. The bishrink filter was applied separately to the real and imaginary parts of each HWT.

The results corresponding to the use of the DEHWT are marked with bold fonts. As can be seen analyzing table 4.10, the diversification mechanism brings a denoising gain of around 0.3 dB, compared to the results obtained using only a HWT and 'Daub,20' as mother wavelet. There are other mother wavelets that can provide better denoising results, but they are dependent of the input image (if 'Barbara' test image would have been used, there would have been other results).

Comparing line by line tables 4.7 and 4.10, the superiority of the HWT versus the

σ_n		10	15	20	25	30	35
PSNRi		28.17	24.65	22.14	20.20	18.62	17.28
PSNRo	**Daub,4**	35.00	33.23	31.96	30.94	30.14	29.46
	Daub,6	35.03	33.28	32.02	31.01	30.20	29.51
	Daub,8	35.01	33.27	31.99	30.98	30.17	29.48
	Daub,10	34.97	33.22	31.94	30.92	30.11	29.43
	Daub,12	34.94	33.17	31.89	30.88	30.06	29.38
	Daub,14	34.90	33.13	31.84	30.81	30.00	29.32
	Daub,16	34.86	33.08	31.78	30.76	29.95	29.27
	Daub,18	34.83	33.03	31.73	30.71	29.89	29.21
	Daub,20	34.79	32.99	31.68	30.65	29.84	29.16
	DE HWT ReIm	**35.01**	**33.25**	**31.97**	**30.95**	**30.14**	**29.45**

Table 4.10: The results obtained associating the bishrink filter with the real and imaginary parts of HWT and DEHWT for denoising the image Lena perturbed by AWGN with different variances. The mother wavelets: 'Daub,4' - 'Daub,20' were used.

DWT and the superiority of DEHWT versus DEDWT can be observed especially for low PSNRi.

The magnitudes of different HWTs can be also associated with the bishrink filter using the architecture in figure 4.9. The results presented in the following table are obtained using this denoising strategy to process the image Lena perturbed by AWGN with different variances.

σ_n		10	15	20	25	30	35
PSNRi		28.17	24.65	22.15	20.20	18.62	17.28
PSNRo	**Daub,4**	34.86	33.14	31.89	30.92	30.09	29.40
	Daub,6	34.92	33.17	31.91	30.94	30.11	29.42
	Daub,8	34.92	33.18	31.90	30.91	30.08	29.37
	Daub,10	34.90	33.14	31.86	30.86	30.03	29.34
	Daub,12	34.87	33.11	31.82	30.81	29.98	29.29
	Daub,14	34.83	33.06	31.77	30.76	29.93	29.24
	Daub,16	34.80	33.02	31.72	30.71	29.88	29.19
	Daub,18	34.76	32.98	31.67	30.66	29.84	29.15
	Daub,20	34.72	32.96	31.62	30.61	29.78	29.10
	DE HWT Abs	**34.92**	**33.16**	**31.87**	**30.88**	**30.04**	**29.35**

Table 4.11: Results obtained applying the architecture in figure 4.9 where the magnitudes of HWTs are associated to the bishrink filter to the Lena image perturbed by AWGN with different variances.

Comparing the lines of the tables 4.10 and 4.11 it can be observed that the denoising associations Real and Imaginary parts of HWT - bishrink filter and Magnitudes of HWT - bishrink filter have very similar performances.

Other synthesis mechanisms

The association of the DEDWT with the bishrink filter presented in figure 4.7 and the association of the DEHWT with the bishrink filter presented in figure 4.9 do not make

any difference between the mother wavelets used. The synthesis of the final result from partial results is realized by averaging. These partial results are affected by the sensitivity of the bishrink filter with $\hat{\sigma}_{W_s}$. In the following, a new synthesis mechanism is proposed. It exploits the same diversification mechanism, the computation of the real and imaginary parts or the magnitudes of N HWTs using different mother wavelets, but the fusion of those partial results is realized taking into account their space-frequency localizations. The new synthesis mechanism also solves the problem of the sensitivity of the bishrink filter with $\hat{\sigma}_{W_s}$. It represents a type of NL-means algorithm.

Taking into consideration the space-frequency localization of the DWT [Olt09] and the results from section 2.5.1, the regions of the useful component of the input image with the best space-frequency localization must be treated with a HWT computed using the mother wavelet 'Daub,4' and the regions of the useful component of the input image with the poor space-frequency localization must be treated with a HWT computed using the mother wavelets 'Daub,20'. Consequently, a segmentation of the useful component of the image to be denoised made using as criterion the space-frequency localization could help the synthesis mechanism. The values of the local variances could be used as a measure of space-frequency localization because, as already said in section 4.3.1.5, the local variance of a pixel provides some information about the importance of his space localization and about the frequency content of the region to which the considered pixel belongs.

A first implementation of the new synthesis mechanism is presented in figure 4.10.

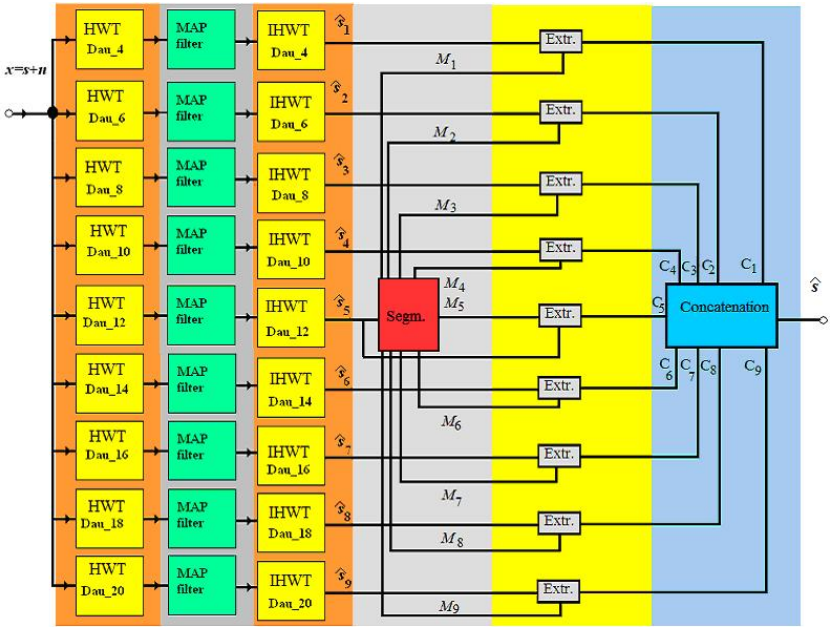


Figure 4.10: A first implementation of the new synthesis mechanism. It can be applied to the association of the real and imaginary parts of the HWT coefficients or of their magnitudes with the bishrink filter

Computing the standard deviation of each pixel of $\hat{s}_5$, a pilot image (similar to the second image in figure 4.6) is obtained. Its pixels are classified in $N = 9$ regions according to their values with the aid of the block 'Segm'. This is equivalent with the image blocks pre-classification proposed in [MS05]. Our segmentation criterion implies the inclusion of

the local standard deviation of the considered pixel to one of nine possible intervals

$$I_k = \left(\epsilon_{k-1} \cdot max\left\{\hat{\sigma}_{\hat{s}_5}\right\}, \epsilon_k \cdot max\left\{\hat{\sigma}_{\hat{s}_5}\right\}\right), \ k = 1, 2, \ldots, 9,$$

where $\epsilon_0 = 0$, $\epsilon_{k-1} < \epsilon_k$ and $\epsilon_9 = 1$. This is the same classification criterion as that proposed in [GSZ06]. The set of coordinates of the pixels belonging to each of these regions will represent one of the $N = 9$ masks denoted by M_k, $k = 1, 2, \ldots, 9$ in figure 4.10. These masks are used to extract from each partial result, $\hat{s}_k$, $k = 1, 2, \ldots, 9$, the corresponding class C_k, $k = 1, 2, \ldots, 9$, of the final result $\hat{s}$. The masks represent the indicator functions of the corresponding set of pixel coordinates. The mask M_1 corresponds to the highest values of the local standard deviation, belonging to the interval I_9 while the mask M_9 corresponds to the interval I_1. The blocs denoted by 'Extr' in figures 4.10 and 4.11, multiply their inputs signals. The classes of the final estimation in figure 4.10 are generated as:

$$C_k = \hat{s}_k \cdot M_k.$$

By their concatenation we obtain the final estimation $\hat{s}$.

This first implementation of the new synthesis mechanism does not take into account the sensitivity $S_{\hat{W}_s^1}^{\hat{\sigma}_{W_s}}$. Next, a second implementation of the new synthesis mechanism, which completes the first implementation by adding some new blocks in the architecture in figure 4.10, which takes into account the sensitivity $S_{\hat{W}_s^1}^{\hat{\sigma}_{W_s}}$ is proposed. The corresponding denoising architecture is presented in figure 4.11.

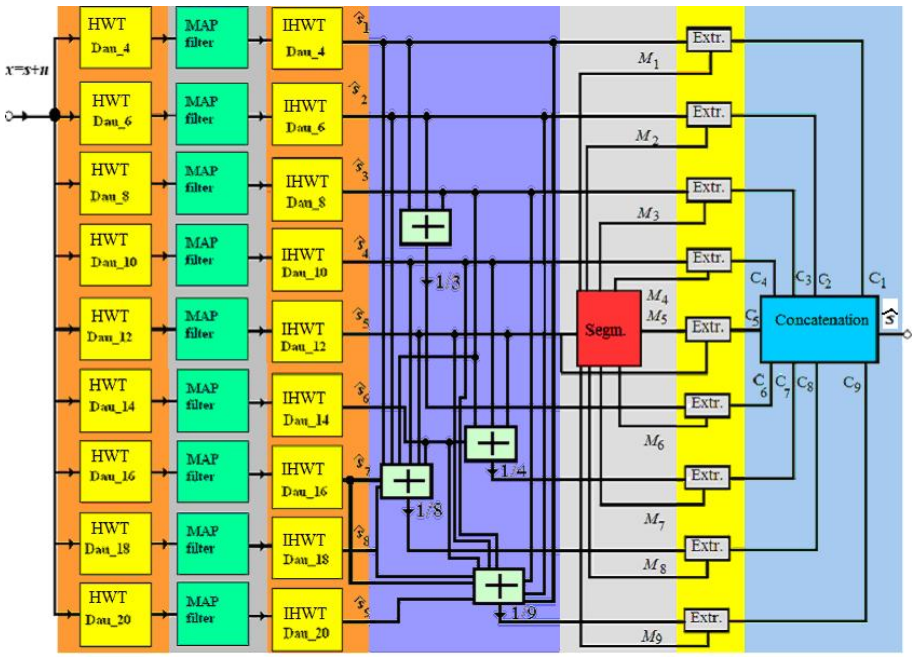


Figure 4.11: The final implementation of the new synthesis mechanism. It can be applied to the association of the real and imaginary parts of the HWT coefficients or of their magnitudes with the bishrink filter

In conformity with equation 4.93, the performance of the bishrink filter degrades with the decreasing of $\hat{\sigma}_{W_s}$. The denoising quality of pixels with slightly different local

variances is different. This difference is higher when the corresponding values of the local variances are smaller. An example can be observed in the first picture of figure 4.12. This picture represents a homogeneous region of the Lena image affected by AWGN with

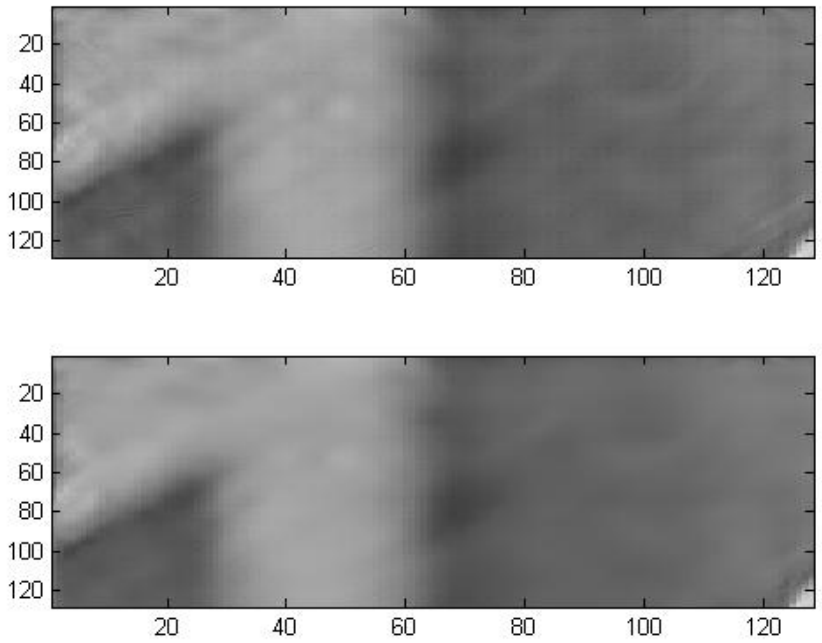


Figure 4.12: A comparison between the results obtained using the association real and imaginary parts of HWT - bishrink (up) and the denoising method proposed in this section (bottom), for the image Lena perturbed by AWGN with $\sigma_n = 35$

$\sigma_n = 35$, denoised with the association bishrink filter - real and imaginary parts of HWT, computed with the mother wavelet 'Daub,20'. Some visible artifacts can be observed. It is necessary to make the denoising more uniform. In the case of the system in figure 4.11, the data is made more uniform with respect to the values of the local variances of the noise-free component of the acquired image, by segmenting the partial result $\hat{s}_5$ according to the values of its local variances, obtaining $N = 9$ masks, like in figure 4.10. Next, some new partial results are computed. To reduce the sensitivity of the bishrink filter with $\hat{\sigma}_{W_s}$, some linear combinations of the partial results $\hat{s}_k$ are computed, obtaining the new partial results (NPR_k):

$$\begin{aligned}
NPR_1 &= \hat{s}_1,\ NPR_2 = \hat{s}_2,\ NPR_3 = \hat{s}_3,\ NPR_4 = \hat{s}_4,\ NPR_5 = \hat{s}_5,\\
NPR_6 &= (\hat{s}_1 + \hat{s}_2 + \hat{s}_3)/3,\\
NPR_7 &= (\hat{s}_1 + \hat{s}_2 + \hat{s}_3 + \hat{s}_4)/4,\\
NPR_8 &= (\hat{s}_1 + \hat{s}_2 + \hat{s}_3 + \hat{s}_4 + \hat{s}_5 + \hat{s}_6 + \hat{s}_7 + \hat{s}_8)/8,\\
NPR_9 &= (\hat{s}_1 + \hat{s}_2 + \hat{s}_3 + \hat{s}_4 + \hat{s}_5 + \hat{s}_6 + \hat{s}_7 + \hat{s}_8 + \hat{s}_9)/9.
\end{aligned}$$

Each new partial result is multiplied with a different mask M_k, obtaining the nine classes

of the final result:

$$C_k = NPR_k \cdot M_{10-k}.$$

In figure 4.13 is presented the fusion system applied for the generation of the last four partial results. All the weights are equal: $\beta_1 = \beta_2 = \ldots = \beta_L$. An averaging block is a

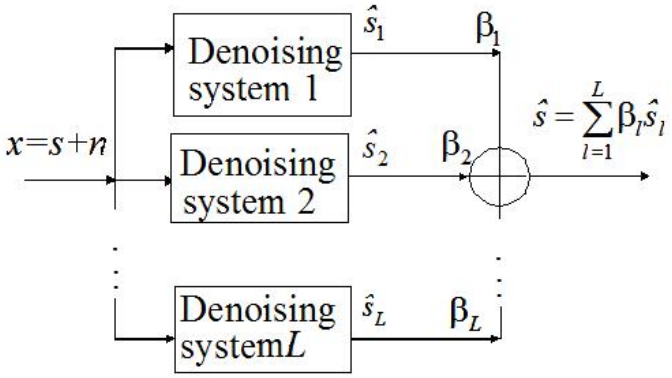


Figure 4.13: The architecture of the fusion system in the interior of one of the classes $C_6 - C_9$ from the system with the architecture in figure 4.11.

linear low-pass filter. Its cut-off frequency is inversely proportional with the number of partial results. The frequency content of a class corresponding to a higher value of local variance is richer than the frequency content of a class that corresponds to a smaller value of local variance. So, for the fusion of a class corresponding to a smaller value of local variance an increased number of partial results are necessary. The fusion procedure uses a different number of new partial results, from class to class. It is based on a NL-means like algorithm.

The amount of noise reduction and the oversmoothing degree, for each of the last four NPR, increases with the increasing of the number of partial results used. The fusion procedure proposed prevents the oversmoothing using a different number of partial results in regions with different local variances of the noise-free component.

Assembling the nine classes $C_1, C_2, \ldots, C_9$, of the system in figure 4.10, by concatenation, the final estimation is obtained. After intensive simulations, the following values for ϵ_k for the intervals I_k for the local standard deviations corresponding to the class C_k, where established: $\epsilon_0 = 0$, $\epsilon_1 = 0.15$, $\epsilon_2 = 0.225$, $\epsilon_3 = 0.25$, $\epsilon_4 = 0.3$, $\epsilon_5 = 0.6$, $\epsilon_6 = 0.7$, $\epsilon_7 = 0.8$, $\epsilon_8 = 0.9$ and $\epsilon_9 = 1$.

The denoising results can be further improved by performing a correction on the homogeneous areas, namely a selection of the pixels belonging to these areas (performed on a local standard deviation basis) followed by a local averaging of these pixel values.

This denoising system was tested on the image Lena perturbed with AWGN with different variances. The results obtained are presented in table 4.12.

The second image in figure 4.12 was obtained with the aid of this new synthesis mechanism and has a better quality than the first image in figure 4.12 (which was obtained using the association of the real and imaginary parts of the HWT coefficients computed with the mother wavelet 'Daub,20' with the bishrink filter) especially in the uniform regions.

σ_n	10	15	20	25	30	35	50	100
PSNRi	28.16	24.65	22.15	20.21	18.62	17.29	14.19	8.16
PSNRo ReIm HWT DE - Bishrink	35.06	33.43	32.19	31.17	30.32	29.67	28.03	24.83
PSNRo Abs HWT DE - Bishrink	35.01	33.34	32.10	31.09	30.28	29.56	27.93	24.78

Table 4.12: Results obtained applying the new synthesis mechanism.

The new synthesis mechanism outperforms the synthesis mechanism which corresponds to the DEHWT from the enhancement of PSNR point of view. The denoising gain is the effect of the new synthesis mechanism with the architecture in figure 4.11. The enhancement of the PSNR is not the only criterion for the appreciation of the quality of a denoising system. Other criteria are proposed in [FNBI09]. One of them refers to the contours treatment. The new synthesis mechanism outperforms the synthesis mechanism which corresponds to the DEHWT from the contours treatment point of view because it does not oversmooth them. The synthesis mechanism which corresponds to the DEHWT oversmooths the contours because it averages the contours of nine partial results. Some of those partial results are obtained using mother wavelets with bad space-frequency localization. In the case of the new synthesis mechanism, the contours belong to a single partial result which is obtained using a mother wavelets with very good space-frequency localization.

In table 4.13 we present the values of the contour errors obtained after the denoising using a single mother wavelet, namely Daub, 6 in association with HWT and bishrink and the new synthesis model (including the homogeneous areas correction). The simulations were performed both using the Real and Imaginary parts of the HWT coefficients and their magnitudes. The countour errors are computed by the sum of the pixels of the

σ_n	10	15	20	25	30	35
ReIm + Daub,6	-153	-530	-866	-1172	-1403	-1649
ReIm + HWT DE	-26	-389	-682	-971	-1175	-1430
Abs + Daub,6	-148	-507	-896	-1229	-1523	-1827
Abs + HWT DE	-30	-389	-739	-1014	-1331	-1590

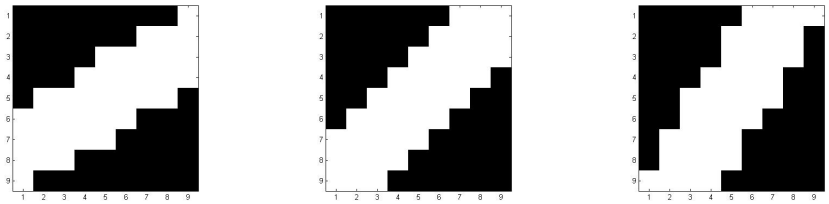
Table 4.13: Contour errors obtained by applying the new synthesis mechanism.

corresponding image of contour errors. A small value of the contour errors indicates a good-quality treatment (the denoising preserves the contours). As can be seen from the results presented in table 4.13, the new synthesis mechanism offers a good treatment of the contours.

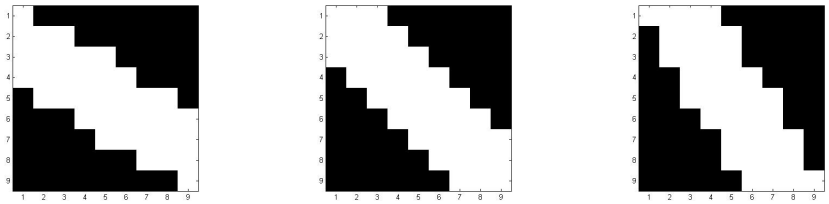
In [FNBI09] is made a comparison between the contour treatment of a denoising method based on the association HWT - adaptive soft-thresholding filter with the contour treatment of a method based on the association of DWT - non-parametric filter, emphasising the superiority of the HWT-based method.

Directional estimation windows

The results already reported were obtained using rectangular estimation windows of size 7x7. The energy clusters in different subbands are mainly distributed along the corresponding preferential directions. For this reason, the estimator using a squared window often leads to downward-biased estimates within and around energy clusters, which is disadvantageous for the preservation of edges and textures in images. In [Shu05] the elliptic directional windows are introduced in conjunction with the 2D DWT to estimate the signal variances in each oriented subband. This idea was generalized in [FNBI09] for the case of the HWT, using constant array elliptic estimation windows with the main axes oriented following the preferential orientations of the subbands: $\pm atan(1/2)$, $\pm\pi/4$ and $\pm atan(2)$. These windows are presented in figure 4.14.



(a) + atan(1/2) (b) + atan(1) (c) + atan(2)



(d) - atan(1/2) (e) - atan(1) (f) - atan(2)

Figure 4.14: Directional elliptic windows

By applying the denoising method relying on the use of Bishrink filter in the context of the real and imaginary parts of HWT having 'Daub,6' and 'Daub,20' as mother wavelets, and of the HWT with enhanced diversity with homogeneous area correction (architecture in fig. 4.11), using the directional estimation windows presented in fig.4.14 we have obtained the PSNR values presented in table 4.14 and for the contours errors the values in table 4.15.

$\sigma_{\mathbf{n}}$	10	15	20	25	30	35	50	100
PSNRi	28.17	24.64	22.15	20.20	18.62	17.29	14.19	8.16
PSNRo Daub,6	35.06	33.31	32.04	31.07	30.25	29.58	27.92	24.77
PSNRo Daub,20	34.82	33.00	31.67	30.69	29.88	29.21	27.57	24.45
PSNRo HWT DE	35.10	33.45	32.19	31.21	30.39	29.72	28.06	24.92

Table 4.14: Results obtained applying the bishrink filter in association with HWT with 'Daub,6' and 'Daub,20', respectively with HWT DE, using directional estimation windows

$\sigma_{\mathbf{n}}$	10	15	20	25	30	35
Daub,6	-141	-539	-819	-1144	-1384	-1618
HWT DE	-32	-384	-640	-950	-1164	-1404

Table 4.15: Contour errors obtained by applying the bishrink filter in association with HWT with Daub,6 and HWT DE, using directional estimation windows

In figure 4.15 is presented the case of input $\sigma_n = 35$. The results obtained using both directional and square windows have been taken into consideration.

(a) Original image

(b) Noisy image ($\sigma_n = 35$)

(c) 'Daub,6', directional windows

(d) HWT DE, directional windows

(e) 'Daub,6', square windows

(f) HWT DE, square windows

Figure 4.15: Simulation results using both directional and square estimation windows

If we compare the results obtained using directional windows with those obtained using square windows from the PSNR and contours error point of view, we notice that they are similar but, from the visual point of view the directional window processing overtakes the other one. Simulations results are presented in [FII10] too.

In table 4.16 we compare the our results with those reported by [Shu05].

$\sigma_{\mathbf{n}}$	10	15	20	25
PSNRi	28.17	24.64	22.15	20.20
PSNRo Daub,6	35.06	33.31	32.04	31.07
PSNRo Shui	34.7	32.8	31.5	30.4

Table 4.16: Comparison HWT - bishrink vs. those reported in [Shu05]

In the experiments already reported the size of the directional windows was kept constant from a decomposition level to next decomposition level. There are references which recommend reducing the size of the directional estimation windows from a decomposition level to the next decomposition level, [Shu05, ZS07, AF01]. Because the simulation results using windows of different sizes did not show substantial improvements, we will not further insist on this aspect.

4.3.1.6 Other statistical models for the wavelet coefficients

The bishrink is one of the best MAP filters. In the following some other MAP filters, with comparable performances will be shortly presented.

MAP filter built on the basis of a bivariate Cauchy statistical model

In [AK05] a similar MAP filter is proposed, but the bivariate a priori pdf of the clean image is that given in 4.75. It is of $S\alpha S$ type and is named Cauchy bivariate distribution. So, the hypotheses of the MAP filter built on the basis of the $S\alpha S$ model are:

$$^{5}p_{W_s}\left(W_s^1, W_s^2\right) = \frac{\gamma}{2\pi\left(\left(W_s^1\right)^2 + \left(W_s^2\right)^2 + \gamma^2\right)^{\frac{3}{2}}} \text{ and } p_{W_n}\left(W_n\right) = \frac{1}{2\pi\sigma_{W_n}^2} \cdot e^{-\frac{\left(w_n^1\right)^2 + \left(w_n^2\right)^2}{2\sigma_{W_n}^2}} \tag{4.94}$$

Substituting these two pdfs in the equation of the MAP filter 4.20, this equation becomes:

$$\begin{aligned}
\hat{W}_s\left(W_x\right) &= \underset{W_s}{argmax}\left\{ln\left(\frac{1}{2\pi\sigma_{W_n}^2}e^{-\frac{(W_x^1-W_s^1)^2+(W_x^2-W_s^2)^2}{2\sigma_{W_n}^2}}\frac{\gamma}{2\pi\left(\left(W_s^1\right)^2+\left(W_s^2\right)^2+\gamma^2\right)^{\frac{3}{2}}}\right)\right\} \\
&= \underset{W_s}{argmax}\left\{ln\frac{\gamma}{4\pi^2\sigma_{W_n}^2} - \frac{\left(W_x^1 - W_s^1\right)^2 + \left(W_x^2 - W_s^2\right)^2}{2\sigma_{W_n}^2} - \right. \\
&\quad \left. -\frac{3}{2}ln\left(\left(W_s^1\right)^2 + \left(W_s^2\right)^2 + \gamma^2\right)\right\}
\end{aligned} \tag{4.95}$$

In the following the argument of the function from the right hand side of the last equation in 4.95 will be denoted by $\Phi\left(W_s^1, W_s^2\right)$. The system of equations to be solved in order to obtain the value that maximize $\Phi\left(W_s^1, W_s^2\right)$ is:

$$\begin{cases} \frac{d\Phi\left(W_s^1, W_s^2\right)}{dW_s^1} = \Phi_1\left(W_s^1\right) = 0; \\ \frac{d\Phi\left(W_s^1, W_s^2\right)}{dW_s^2} = \Phi_2\left(W_s^2\right) = 0. \end{cases}$$

By pursuing with the calculations, we will further get to:

$$\begin{cases} W_s^2 = \frac{W_x^2}{W_x^1} W_s^1, \\ \left(W_s^1\right)^3 - W_x^1 \left(W_s^1\right)^2 + \frac{2\sigma_{W_n}^2 + \gamma^2}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} \cdot W_s^1 - \frac{\gamma^2 W_x^1}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} = 0. \end{cases} \tag{4.96}$$

The second equation in 4.96 is a third order algebraic equation. Making the change of variable $y = W_s^1 - \frac{W_x^1}{3}$ it becomes:

$$\begin{aligned} &\left(y + \tfrac{W_x^1}{3}\right)^3 - W_x^1 \left(y + \tfrac{W_x^1}{3}\right)^2 + \frac{3\sigma_{W_n}^2 + \gamma^2}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} \cdot \left(y + \tfrac{W_x^1}{3}\right) - \frac{\gamma^2 W_x^1}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} = 0 \;\Leftrightarrow \\ \Leftrightarrow\; & y^3 + \left(\frac{\left(W_x^1\right)^2}{3} - \frac{2\left(W_x^1\right)^2}{3} + \frac{3\sigma_{W_n}^2 + \gamma^2}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} \right) y + \frac{\left(W_x^1\right)^3}{27} - \frac{\left(W_x^1\right)^3}{9} + \frac{\left(3\sigma_{W_n}^2 + \gamma^2\right)\frac{W_x^1}{3} - \gamma^2 W_x^1}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} = 0 \;\Leftrightarrow \\ \Leftrightarrow\; & y^3 + \left(\frac{3\sigma_{W_n}^2 + \gamma^2}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} - \frac{\left(W_x^1\right)^2}{3} \right) y - \frac{2\left(W_x^1\right)^3}{27} + \frac{\left(3\sigma_{W_n}^2 + \gamma^2\right)\frac{W_x^1}{3} - \gamma^2 W_x^1}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} = 0. \end{aligned} \tag{4.97}$$

By making the notations:

$$p = \frac{3\sigma_{W_n}^2 + \gamma^2}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} - \frac{\left(W_x^1\right)^2}{3} \text{ and } q = \frac{\sigma_{W_n}^2 W_x^1 - \frac{2}{3}\gamma^2 W_x^1}{1 + \frac{\left(W_x^2\right)^2}{\left(W_x^1\right)^2}} - \frac{2}{27}\left(W_x^1\right)^3, \tag{4.98}$$

the equation 4.97 can be written in the following form:

$$y^3 + py + q = 0. \tag{4.99}$$

Three additional notations are made:

$$\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3;\; A^2 = -\Delta \text{ and } \rho = \left|\sqrt{-\frac{p^3}{25}}\right|. \tag{4.100}$$

The equation 4.97 can have three real solutions (different or a simple and a double solution) or a real solution and two complex conjugated solutions.

First case: three real solutions: one simple and one double:

$$\begin{aligned} &\Delta = 0;\; y_1 = \tfrac{3q}{p} \text{ and } y_2 = y_3 = -\tfrac{1}{2} y_1 \;\Rightarrow \\ &\Rightarrow \hat{W}_{s_1}^1 = \tfrac{3q}{p} + \tfrac{W_x^1}{3} \text{ and } \hat{W}_{s_2}^1 = \hat{W}_{s_3}^1 = -\tfrac{3q}{2p} + \tfrac{W_x^1}{3}. \end{aligned} \tag{4.101}$$

Finally, the right solution is selected by checking which solution, $\hat{W}_{s_1}^1$ or $\hat{W}_{s_2}^1$ verifies the condition:

$$sgn\Phi_1\left(\hat{W}_s^1 - \epsilon\right) > 0 \text{ and } sgn\Phi_1\left(\hat{W}_s^1 + \epsilon\right) < 0 \text{ for any positive, very small } \epsilon \tag{4.102}$$

Second case: a real solution and two complex conjugate solutions:

$$\begin{aligned} &\Delta > 0;\ y_1 = \sqrt[3]{-\tfrac{q}{2} + \sqrt{\Delta}} + \sqrt[3]{-\tfrac{q}{2} - \sqrt{\Delta}} \ \Rightarrow \\ &\Rightarrow \hat{W}^1_{s_1} = \sqrt[3]{-\tfrac{q}{2} + \sqrt{\Delta}} + \sqrt[3]{-\tfrac{q}{2} - \sqrt{\Delta}} + \tfrac{W^1_x}{3}. \end{aligned} \tag{4.103}$$

Third case: three real distinct solutions:

$$\Delta < 0;\ y_1 = 2\,|\sqrt[3]{\rho}|\,cos\frac{\theta}{3};\ y_2 = 2\,|\sqrt[3]{\rho}|\,cos\frac{\theta + 2\pi}{3} \text{ and } y_3 = 2\,|\sqrt[3]{\rho}|\,cos\frac{\theta + 4\pi}{3},$$

where

$$cos\theta = -\frac{q}{2\rho} \text{ and } sin\theta = \frac{A}{\rho}.$$

So:

$$\begin{aligned} \hat{W}^1_{s_1} &= \frac{3q}{p} + \frac{W^1_x}{3}, \\ \hat{W}^1_{s_2} &= -\frac{3q}{2p} + \frac{54}{p}\sqrt{-\Delta} + \frac{W^1_x}{3}, \\ \hat{W}^1_{s_3} &= -\frac{3q}{2p} - \frac{54}{p}\sqrt{-\Delta} + \frac{W^1_x}{3}. \end{aligned} \tag{4.104}$$

The right solution is the one that verifies condition 4.102.

For the computation of the solutions in each of the three cases is necessary to estimate the parameters of the two pdfs in 4.94. The standard deviation of the noise, σ_{W_n} can be estimated using the procedure given in equation 4.9. The estimation of the parameter γ of the pdf of the noiseless image is more complicated. The denoising performance reported in [AK05], obtained associating the DTCWT with the MAP estimator is very good especially for its local form of the MAP estimator.

A marginal MAP filter built on the basis of a Pearson statistical model

A statistical speckle-removal method based on the association of the UDWT with a marginal MAP filter constructed using a Pearson statistical model is proposed in [FBB01]. Both the noise and the noiseless component of the input image are modeled as Pearson random variables:

$$\begin{aligned} p_{W_s}(W_s) &= \left((W_s - a + \gamma)^2 + \delta^2\right)^2 \cdot e^{-\lambda atan\left(\frac{W_s - a + \gamma}{\delta}\right)} \\ p_{W_n}(W_n) &= \left((W_n - a_n + \gamma_n)^2 + \delta_n^2\right)^2 \cdot e^{-\lambda_n atan\left(\frac{W_n - a_n + \gamma_n}{\delta_n}\right)}, \end{aligned} \tag{4.105}$$

where a, γ, δ, a_n, γ_n and δ_n are specific parameters.

Substituting these pdfs in the equation of the MAP filter, 4.20, the latter becomes

$$\begin{aligned}\hat{W}_s(W_x) &= \underset{W_s}{argmax}\left\{ln\left(\left((W_x - W_s - a_n + \gamma_n)^2 + \delta_n^2\right)^2 \cdot e^{-\lambda_n atan\left(\frac{W_x - W_s - a_n + \gamma_n}{\delta_n}\right)}\cdot\right.\right.\\ &\left.\left.\cdot\left((W_s - a + \gamma)^2 + \delta^2\right)^2 \cdot e^{-\lambda atan\left(\frac{W_s - a + \gamma}{\delta}\right)}\right)\right\}\\ &= \underset{W_s}{argmax}\left\{2ln\left((W_x - W_s - a_n + \gamma_n)^2 + \delta_n^2\right) - \lambda_n atan\left(\frac{W_x - W_s - a_n + \gamma_n}{\delta_n}\right) + \right.\\ &\left.+2ln\left((W_s - a + \gamma)^2 + \delta^2\right) - \lambda atan\left(\frac{W_s - a + \gamma}{\delta}\right)\right\}.\end{aligned} \tag{4.106}$$

This equation can be written as, [FBB01]:

$$W_s^3 + c_1 W_s^2 + c_2 W_s + c_3 = 0. \tag{4.107}$$

This is a third order algebraic equation. Making the change of variable $y = W_s + \frac{c_1}{3}$, 4.107 becomes:

$$\begin{aligned}&\left(y - \frac{c_1}{3}\right)^3 + c_1\left(y - \frac{c_1}{3}\right)^2 + c_2\left(y - \frac{c_1}{3}\right) + c_3 = 0 \Leftrightarrow\\ \Leftrightarrow\ & y^3 - c_1 y^2 + \frac{c_1^2}{3}y - \frac{c_1^3}{27} + c_1 y^2 - 2\frac{c_1^2}{3}y + \frac{c_1^3}{9} + c_2 y - \frac{c_1 c_2}{3} + c_3 = 0 \Leftrightarrow\\ \Leftrightarrow\ & y^3 + \left(c_2 - \frac{c_1^2}{3}\right)y + \frac{2c_1^3}{27} - \frac{c_1 c_2}{3} + c_3 = 0.\end{aligned} \tag{4.108}$$

With the denotations:

$$p = c_2 - \frac{c_1^2}{3} \text{ and } q = \frac{2c_1^3}{27} - \frac{c_1 c_2}{3} + c_3 \tag{4.109}$$

the equation 4.108 can be written as:

$$y^3 + py + q = 0. \tag{4.110}$$

The form of the MAP filter equation obtained is the same as that in 4.99 which corresponds to the MAP filter constructed on the basis of the Cauchy model, described in the previous section. As a consequence, the solving procedure already presented can be also applied in the case of the equation 4.110.

The MAP filter obtained is associated in [FBB01] with the UDWT for images denoising purposes. For the computation of the solutions in each of the three cases is necessary to estimate the parameters of the two pdfs in 4.105. This is not a simple task because there are six parameters to be estimated: a, γ, δ, a_n, γ_n and δ_n. One of the merits of [FBB01] is the computation of the first four moments of the two random variables constituted by the noisy and noiseless UDWT coefficients. Using these values, a local estimation of the six parameters a, γ, δ, a_n, γ_n and δ_n can be realized and the MAP filter can be applied pixel wise. The results obtained are improved with the aid of a segmentation similar with that presented in section 4.3.1.5. Pixels belonging to homogeneous regions and to textured regions are treated differently. This treatment strategy does not take into account the type of mother wavelet used.

The denoising results obtained in [FBB01] are impressive, but their visual aspect is quit different, being strongly dependent of the mother wavelet used.

A marginal MAP filter built on the basis of a generalized Gaussian statistical model

In [ABA06] is proposed a statistical speckle reduction method based on the association of the UDWT with a marginal MAP filter constructed using a generalized Gaussian (GG) statistical model. Both the noise and the noiseless component of the input image are modeled with GG random variables:

$$\begin{aligned} p_{W_s}(W_s) &= \left[\frac{\nu \cdot \eta(\nu,\sigma)}{2 \cdot \Gamma(1/\nu)}\right] e^{-(\eta(\nu,\sigma)|W_s|)^{\nu}} \\ p_{W_n}(W_n) &= \left[\frac{\nu_n \cdot \eta(\nu_n,\sigma_n)}{2 \cdot \Gamma(1/\nu_n)}\right] e^{-(\eta(\nu_n,\sigma_n)|W_n|)^{\nu_n}} \end{aligned} \tag{4.111}$$

where σ and σ_n represent the standard deviations of the two pdfs and ν and ν_n represent their shape factors.

The function η is defined as:

$$\eta(a,b) = \frac{1}{b}\left[\frac{\Gamma(3/a)}{\Gamma(1/a)}\right]^{1/2}.$$

Substituting these two pdfs in the equation of the MAP filter 4.20 this equation becomes:

$$\begin{aligned} &\hat{W}_s(W_x) = \underset{W_s}{argmax}\left\{ln\left(\left[\frac{\nu_n \cdot \eta(\nu_n,\sigma_n)}{2 \cdot \Gamma(1/\nu_n)}\right] e^{-(\eta(\nu_n,\sigma_n)|W_x - W_s|)^{\nu_n}} \cdot \right.\right. \\ &\left.\left. \cdot \left[\frac{\nu \cdot \eta(\nu,\sigma)}{2 \cdot \Gamma(1/\nu)}\right] e^{-(\eta(\nu,\sigma)|W_s|)^{\nu}}\right)\right\} \\ &= \underset{W_s}{argmax}\left\{ln\left[\frac{\nu\nu_n\eta(\nu,\sigma)\,\eta(\nu_n,\sigma_n)}{4\Gamma(1/\nu)\,\Gamma(1/\nu_n)}\right] - (\eta(\nu_n,\sigma_n)|W_x - W_s|)^{\nu_n} - (\eta(\nu,\sigma)|W_s|)^{\nu}\right\}. \end{aligned} \tag{4.112}$$

In the following the argument of the function from the right hand side of the last equation will be denoted by $\Phi(W_s)$. The equation that gives the solution to 4.112 is:

$$\frac{d\Phi(W_s)}{dW_s} = \Theta(W_s) = 0,$$

or, more explicitly:

$$\begin{aligned} & \nu_n \cdot \eta(\nu_n,\sigma_n)^{\nu_n} \frac{|W_x - W_s|}{W_x - W_s} \cdot |W_x - W_s|^{\nu_n - 1} - \nu \cdot \eta(\nu,\sigma)^{\nu} \frac{|W_s|}{W_s} \cdot |W_s|^{\nu-1} = 0 \Leftrightarrow \\ \Leftrightarrow\ & \nu_n \cdot \eta(\nu_n,\sigma_n)^{\nu_n} \frac{|W_x - W_s|^{\nu_n}}{W_x - W_s} - \nu \cdot \eta(\nu,\sigma)^{\nu} \frac{|W_s|^{\nu}}{W_s} = 0 \Leftrightarrow \\ \Leftrightarrow\ & \nu_n \cdot \eta(\nu_n,\sigma_n)^{\nu_n} \cdot W_s \cdot |W_x - W_s|^{\nu_n} - \nu \cdot \eta(\nu,\sigma)^{\nu} \cdot (W_x - W_s) \cdot |W_s|^{\nu} = 0 \end{aligned} \tag{4.113}$$

This is not an algebraic equation and it can not be solved analytically. Only numerically solutions can be obtained.

4.4 Summary

In this chapter we have insisted on the denoising of natural images, which are normally affected by additive noise. We have presented both non-parametric and parametric methods, both in the spatial and wavelet domain. After a comparative analysis we can state that the Bayesian methods overtake the non-parametric ones, that denoising in the wavelet domain performs much better than in the spatial domain and that local parameters estimations is better than global estimation. In the wavelet domain, we have associated both classical DWT and the Diversity-Enhanced DWT but also the HWT with MAP filters, and have proved, by simulations, that the latter performs much better than the real transforms.

In order to further increase its performances, we have associated our transform with directional estimation windows and we have tried different methods to increase its diversity. The results obtained are satisfactory.

In the following chapter we will talk about the reduction of speckle noise, which is a multiplicative - type of noise.

Chapter 5

Speckle Reduction

5.1 Introduction

Synthetic Aperture Radar (SAR) images are a particular category of images, comprising both SONAR (originally an acronym for SOund Navigation and Ranging) and RADAR (RAdio Detection and Ranging). In order to capture such an image, one needs a device (sonar or radar) that emits signals (sound pulses in the case of the Sonar or microwave signals for Radar) and then measures the strength and round-trip time of the received signals, signals that are reflected of a distant surface or object.

As a consequence of wave coherence [UKBW86], SAR images are perturbed by a specific type of noise, called speckle noise. Each ground resolution cell is composed of a large number of elementary reflectors backscattering the radar wave in the sensor direction. For a rough surface in comparison with the radar wavelength, these elementary reflectors are present in a number large enough to ensure the statistical independence in phase and amplitude of these backscattered elementary contributions. For this type of target, the speckle is fully developed. Elementary phases are then uniformly distributed random variables. The total component backscattered by the resolution cell is the vectorial sum of these elementary backscattered electrical fields. The energy recorded by the sensor from a resolution cell can be either null or significant according to the constructive or destructive interferences between the elementary contributions. Consequently, the backscattered energy can randomly fluctuate from one resolution cell to another independently of the radar backscattering coefficient. The homogeneous areas within the SAR image present a particular texture called speckle. The ground radar reflectivity proportional to the surface backscattering coefficient cannot then be derived from a single pixel digital number. The latter is only estimated by an average of a pixel set. One way to reduce the radiometric variability due to the speckle is the multi-look processing which consists of taking an average of looks of the same scene produced by a signal bandwidth sub band extraction.

Consequently, we can affirm that the speckle is a signal-dependent noise. Even though there are many speckle scene models ([Tou02]), we chose the one that considers that speckle can be modeled as a multiplicative noise. The expression of the acquired image in this case is:

$$x = s \cdot n, \tag{5.1}$$

where s is the noise-free image or reflectivity and n represents the speckle.

The aim of a speckle reduction algorithm is to reduce the noise level while preserving the image features.

5.1.1 Speckle's Statistics

Throughout time, many researchers have tried to model the speckle noise. Some of the models chosen can be found in [FBB01], [ABA06], [GD06] and so on.

When the speckle is entirely developed, the hypothesis of independence of random processes s and n can be adopted [FBB01]. The speckle that perturbs the SAR images can be modeled as a white noise distributed following a law Gamma, whose parameter is the number of views, L, [FBB01]:

$$p_{N_\Gamma}(n) = \frac{L^L}{\Gamma(L)} \cdot n^{L-1} \cdot e^{-Ln} \text{ for } n \geq 0, \tag{5.2}$$

where Γ represents the Euler's Gamma function:

$$\Gamma(L) = \int_0^\infty t^{L-1} \cdot e^{-t} dt. \tag{5.3}$$

The speckle's mean is

$$\mu_\Gamma = 1, \tag{5.4}$$

and its variance is [IIA06]:

$$\sigma_\Gamma^2 = E\left[N_\Gamma^2\right] - \mu_\Gamma^2 = \frac{L(L+1)}{L^2} - 1 = \frac{1}{L}. \tag{5.5}$$

The Sonar images are a particular case of SAR images, obtained for $L = 1$. In this case, the law Γ becomes a law χ^2.

5.1.2 Speckle Filtering Techniques

Because the noise affecting a SAR image is multiplicative, it is difficult to design effective noise filtering algorithms. The existing strategies rely on transforming the multiplicative noise component into an additive one. These methods can be grouped in two major categories: homomorphic filtering and filtering based on pixel-ratioing.

5.1.2.1 Homomorphic Filtering

Homomorphic filtering relies on the property of the logarithm, namely the logarithm of a product of variables is equal to the sum of the logarithms of each variable:

$$\ln x = \ln s + \ln n. \tag{5.6}$$

An implementation scheme for homomorphic denoising is presented in fig. 5.1. By applying the logarithm, the random variable N_Γ is transformed, resulting a random variable

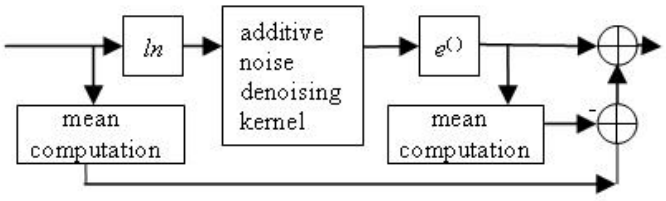


Figure 5.1: The architecture of homomorphic filtering system. The mean correction mechanism and the kernel are highlighted.

denoted $N_{log-\Gamma}$, which describes the speckle noise at the input of the second block in figure 5.1. The probability density function of this new variable is:

$$\begin{aligned} p_{log-\Gamma}(n) &= \left.\frac{p_{N_\Gamma}(n_\Gamma)}{\left|\frac{dn}{dn_\Gamma}\right|}\right|_{n_\Gamma=e^n} = \left.\frac{\frac{L^L}{\Gamma(L)}\cdot n_\Gamma^{L-1}\cdot e^{-Ln_\Gamma}}{\frac{1}{n_\Gamma}}\right|_{n_\Gamma=e^n} = \\ &= \left.\frac{L^L}{\Gamma(L)}\cdot n_\Gamma^L\cdot e^{-Ln_\Gamma}\right|_{n_\Gamma=e^n} = \frac{L^L}{\Gamma(L)}\cdot e^{Ln}\cdot e^{-Le^n}. \end{aligned} \tag{5.7}$$

The value of the mean of the log-Gamma law is given by:

$$\mu_{log-\Gamma} = \sum_{k=1}^{L-1}\frac{1}{k} - \gamma - lnL, \tag{5.8}$$

where γ represents the Euler's number. The variance is:

$$\sigma^2_{log-\Gamma} = \frac{\pi^2}{6} - \sum_{k=1}^{L-1}\frac{1}{k^2}. \tag{5.9}$$

In conclusion, the noise that perturbs the logarithm of the useful part of a SAR image is white and distributed following a log-Gamma law, whose statistical characteristics are given by 5.7, 5.8 and 5.9.

In the case of the SONAR images, after the computation of the logarithm the speckle component will follow a $log-\chi^2$ law, with the pdf:

$$p_{log-\chi^2}(n) = e^{Ln}\cdot e^{-e^n}. \tag{5.10}$$

The value of its mean is :

$$\mu_{log-\chi^2} = -\gamma, \tag{5.11}$$

and the variance is:

$$\sigma^2_{log-\chi^2} = \frac{\pi^2}{6}. \tag{5.12}$$

To obtain the denoising result, the logarithm inversion is performed at the end of the process depicted in figure 5.1. Because the homomorphic treatment could introduce a small bias, the denoising system must contain a mean correction block. The corresponding

block in figure 5.1 computes the mean of the acquired image which is equal to the mean of the noise-free component because the speckle noise has unitary mean (see 5.4). In order to correct the mean of the result, the mean of the image at the output of the block that inverts the logarithm is extracted and the mean of the acquired image is added.

5.1.2.2 Pixel-ratioing - based filtering

Another possibility to transform the multiplicative noise into an additive one is given by the following equation:

$$x = s \cdot n = s + s \cdot (n-1) = s + n', \tag{5.13}$$

where n' is an additive, zero-mean, signal-dependent noise term, proportional to the signal to be estimated. Since s is, in general, non-stationary, the noise n' will be non-stationary as well.

5.1.3 Quality measures

In what concerns the quality measures for the speckle reduction, two are more often encountered: the Equivalent Number of Looks (ENL) and the Signal-to-Mean Squared Error (S/MSE), [GJ97]:

1. The ENL is usually computed using formula 5.14 using pixel values over a homogenous image area:

$$ENL = \left(\frac{\text{mean}}{\text{standard deviation}}\right)^2 = L = constant. \tag{5.14}$$

   Its interpretation would be that the denoised image is equivalent with the result obtained by averaging L different 'snapshots' of the same area.

2. S/MSE corresponds to standard SNR and can be computed with:

$$S/MSE = 10log_{10}\left(\frac{\sum_i^K s\left(i\right)^2}{\sum_i^K \left(\hat{s}\left(i\right) - s\left(i\right)\right)^2}\right), \tag{5.15}$$

   where s is the original image, $\hat{s}$ is the denoised image and K is the number of pixels in the image. An alternative to S/MSE is the PSNR already introduced in the previous chapter.

5.2 Spatial-domain Speckle reduction methods

In this section I will briefly present some classical speckle reduction methods, acting in the spatial domain. In order to compare their performances, the image in figure 5.2, [WD00], is used as input noisy image. It is a synthesis of six images of different nature: optic or radar. The first image (from top-left to bottom-right) is an image acquired by an optical sensor, the second one presents four types of texture generated using different

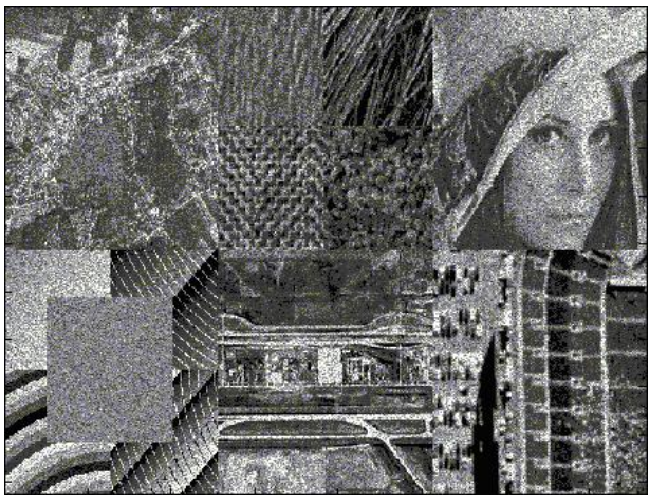


Figure 5.2: Test image

statistical models, the third is obtained by zooming the well known Lena image, the forth is a mosaic of artificial sub-images, the fifth is another optical image and the last one is a SAR image. All the sub-images were perturbed with synthetic 3-look speckle. Tacking the third sub-image and comparing it with the corresponding zone obtained by cropping the clean Lena image, it was established that the mean square error produced by the noise generation method proposed in [WD00] is of 3635.

5.2.1 Frost Filter

Frost filter is a Wiener filter adapted to multiplicative noise, proposed in [FSJ+82]. The parameters of the Frost filter are: the size of the rectangular windows used for the estimation of the local standard deviation of the useful component of the acquired image and the number of looks.

The result of processing the image in figure 5.2 with a Frost filter is presented in figure 5.3.

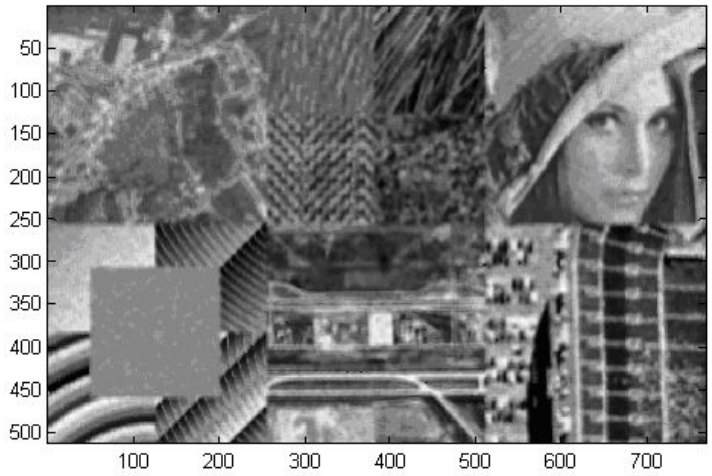


Figure 5.3: The image in figure 5.2 having a number of looks equal to 1 treated with a Frost filter using a rectangular moving window of size 7x7.

5.2.2 Kuan Filter

Kuan considered a multiplicative speckle model and designed a linear filter based on the minimum mean square error (MMSE) criterion, optimal when both the scene and the detected intensities are Gaussian distributed. More about this filter can be read in [KSSC87]. The parameters of the Kuan filter are: the size of the rectangular window used for the estimation of the local standard deviation of the useful component of the acquired image and the number of looks. The result of processing the image in figure 5.2 with a Kuan filter is presented in figure 5.4.

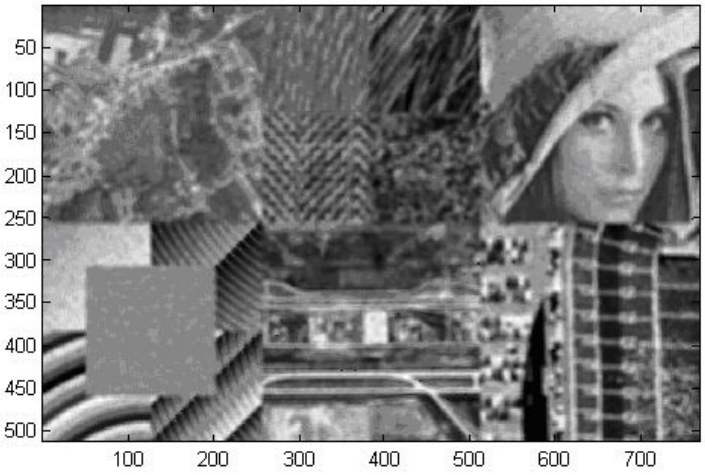


Figure 5.4: The image in figure 5.2 having a number of looks equal to 1 treated with a Kuan filter using a rectangular moving window of size 7x7.

5.2.3 Lee Filter

The Lee filter is a particular case of the Kuan filter due to a linear approximation made for the multiplicative noise model. It was proposed in [Lee81].

The parameters of the Lee filter are: the size of the rectangular windows used for the estimation of the local standard deviation of the useful component of the acquired image and its number of looks.

The result of processing the image in figure 5.2 with a Lee filter is presented in figure 5.5.

An objective comparison of these filters' performances, using the MSE criterion is presented in table 5.1. For the computation of the MSE was considered only the sub-image Lena, from the mosaic in figure 5.2. The parameters of the different filters were selected to minimize the output mean square error.

Noisy	Averager 7-1	Median 7-1	Lee 7-1	Kuan 9-1	Frost 5-1
3635	571.7	569.8	807.5	732.8	566

Table 5.1: Comparison of the performances of some classical speckle reduction systems

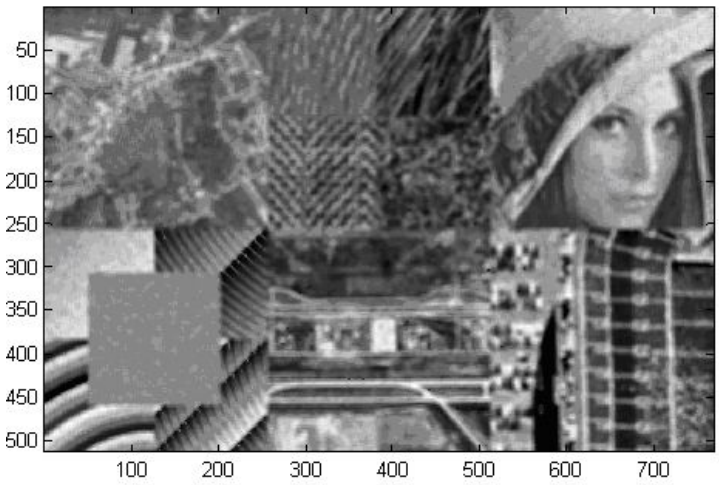


Figure 5.5: The image in figure 5.2 having a number of looks equal to 1 treated with a Lee filter using a rectangular moving window of size 7x7.

5.2.4 Other speckle reduction filters in the spatial domain

In the following, we will briefly analyse the performances of some MAP filters applied in the spatial domain.

5.2.4.1 Zero-order Wiener filter

Applying the filter described in 4.3.1.1 to the logarithm of the image in figure 5.2, the result presented in figure 5.6 was obtained.

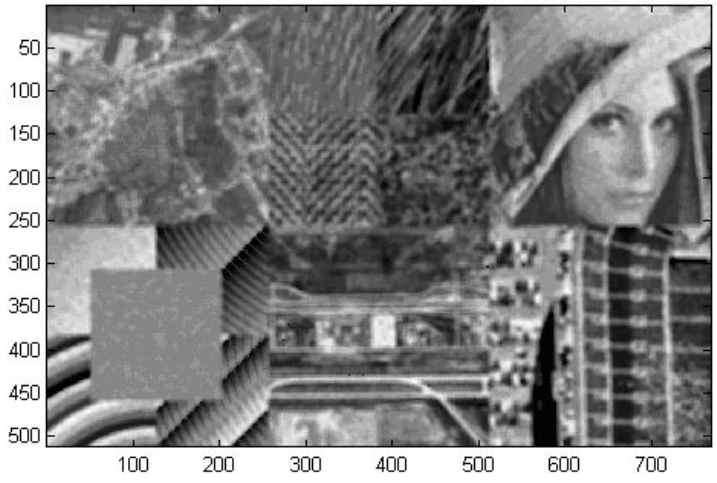


Figure 5.6: The image in figure 5.2 treated with a zero-order Wiener filter.

Comparing the result in figure 5.6 with the results presented in previous figures, it can be observed that the MAP filter performs slightly better than the classical speckle

reduction filters.

5.2.4.2 A MAP filter acting in the spatial domain

As known, the zero-order Wiener type MAP filter is built on the hypothesis that both the noiseless image and the noise are distributed following Gaussian laws. This is not a realistic hypothesis for the noiseless image. In [AKZ06], another MAP filter, built supposing a heavy-tailed Rayleigh model for the logarithm of the noiseless component and a Gamma or Nakagami model for the logarithm of the noise component (depending on whether the image to be filtered is of intensity or amplitude) used in the homomorphic spatial domain filtering is proposed. In order to assess the quality of the filter proposed in [AKZ06], the S/MSE was computed in four different cases, the worst case corresponding to the input image whit the lowest quality (ENL = 1) and the best case corresponding to the image with the highest quality, (ENL = 12).

Method	S/MSE			
	ENL = 1	ENL = 3	ENL = 9	ENL = 12
Median	14.53	15.52	15.85	15.88
Lee	14.69	15.35	16.64	17.28
Proposed in [AKZ06]	15.86	16.83	18.00	18.68

Table 5.2: A comparison of different spatial-domain speckle reduction methods, from the S/MSE point of view.

5.2.4.3 Model-Based Despeckling (MBD)

MBD is a statistical filter introduced in [WD00]. Applying this statistical filter on images from figure 5.7(a) they obtained the result presented in 5.7(b).

(a) Initial image

(b) Denoised image using the MBD method

Figure 5.7: Model-Based Despecking simulation results.

Through visual comparison with the results of the denoising methods acting in the spatial domain already presented, it can be observed that the statistical method proposed in [WD00] is the best.

In [FLA00] is presented a comparison of the best spatial-domain despecklisation methods introduced before 2000. More than 30 different combinations of estimators and neighborhood types were tested. The basic estimators considered were:

- the Kuan filter;
- the Frost filter;
- the improved Frost filter proposed by Touzi;
- the MAP filter, which supposes Gamma distributed texture and speckle;
- the a posteriori mean (APM) filter, also based on the hypothesis of Gamma distributed texture and speckle;
- the T-linear filter, which is an approximation of the APM, with $T = 0.5$ and
- the MBD filter.

Each of these estimators has been associated with different ways of defining the neighborhood where the local statistics are computed:

- fixed size analyzing window without structure detection;
- detection of the biggest homogeneous area within the analyzing window, which includes the central pixel (a zone is considered to be homogeneous if the measured coefficient of variation is below a certain threshold);
- analyzing window with local structure detection based on the generalized likelihood ratio, which constitutes an improvement of the method based on the ratio detector in the general case where the analyzing window is split in two unequally sized parts (asymmetrical binary masks);
- intersection between a fixed size analyzing window and a region defined by a segmentation;
- entire regions defined by a segmentation.

The comparative study in [FLA00] reveals interesting differences between the radiometric estimators employed by different adaptive speckle filters. However, the method used to determine the relevant neighborhood seems to play an even more important role. All adaptive speckle filters perform automatic analysis of the image and may introduce artifacts leading to complicate subsequent automatic analysis or visual interpretation. Direct automatic analysis with robust methods that takes the speckle statistics explicitly into account, or dedicated interactive visualization tools based on non-adaptive transformations of the image (e.g. different spectral weighting functions), can be better than

speckle filtering for many applications. The radar division of the French Space Agency CNES (CNES/DSO/OT/QTIS/SR) intends to continue the comparison of speckle filters on a larger scale. Both a set of SAR images and the corresponding filtered images will be made available on CD-ROM (possibly also over the Internet). Hence different research groups can test their speckle filters on the same images, submit their results, and compare them to the other filtered images. It must be mentioned that the image in figure 5.2 and other images used in this chapter are taken from this CD. Next, I will present the reasons found in [JBFZ03] to select an additive noise denoising kernel based on wavelets.

5.3 Speckle reduction in the wavelet domain

As in the case of the additive noise removal, denoising methods can be applied in the wavelet domain as well. In the following I will continue with the presentation of some of these methods, beginning with a non-parametric one and continuing with parametric filters.

5.3.1 Non-parametric filters

Substituting the MAP filters in figure 4.9 with soft-thresholding filters, we obtain the denoising association DE DWT - soft-thresholding with the architecture presented in figure 5.8, [IMN+06]

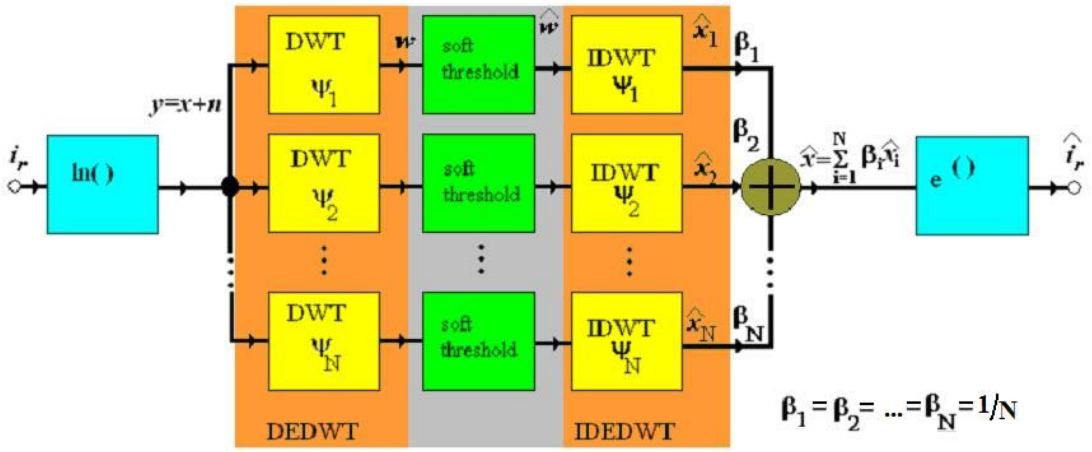


Figure 5.8: A SAR image denoising system based on the association of the DE DWT with the Soft-thresholding filter

For $N = 9$, the result obtained applying the system in figure 5.8 to the image in figure 5.2 is represented in figure 5.9.

Comparing the result in figure 5.9 with the similar results presented in previous figures, one can observe the good treatment of the homogeneous regions (better even than the treatment of homogeneous regions in figure 5.7(b) obtained with the aid of the MBD algorithm). This is the effect of the use of the soft-thresholding filter which is a non-parametric system. The drawbacks of this denoising method lie in the poor treatment of

Figure 5.9: The output of the system in figure 5.8 when at its input is applied the image in figure 5.2.

the contours and textured regions due to the over-smoothing produced by the uniform use of all nine mother wavelets. Another drawback is the reduced directional selectivity of the result caused by the use of the DWTs. Some solutions to these drawbacks are presented in the following.

5.3.2 MAP filters

As presented in chapter 4, MAP filters are a good solution for denoising algorithms, and I will present their performances in speckle reduction purposes.

5.3.2.1 MAP filters associated with 2D UDWT

Pearson model MAP filter

The fist speckle removal method presented is the one based on the association of the 2D UDWT with a marginal MAP filter constructed using a Pearson statistical model presented in section 4.3.1.6 as described in [FBB01]. The multiplicative noise is transformed into additive noise on the basis of the equation 5.13. The application of a Bayesian analysis requires models for the different probability density functions. Consequently, in [FBB01] is proposed an application of the Pearson system of distributions in order to approximate wavelet coefficients' pdf, assuming Gamma distributions for both radar reflectivity and speckle. After a brief review of the usual statistical hypothesis for SAR images and a presentation of the 2D UDWT, in [FBB01] is described the 2D UDWT coefficients' behavior using the second-order moments for a speckled image. The use of the multiplicative model allows a segmentation of the high-frequency images which permits different treatment of regions with different homogeneity degrees: homogeneous, textured and contours. Before applying a MAP criterion, it is also demonstrated in [FBB01] that

the wavelet coefficients' pdf of a Gamma distributed image is well approximated by a Pearson type IV distribution. Therefore, the local Bayesian estimate of the wavelet coefficients of the ground reflectivity is the solution of the third-degree equation solved in section 4.3.1.6. Two types of mother wavelets were considered: orthogonal Daubechies with four vanishing moments (D4) and biorthogonal with five vanishing moments (B5). Results of this method are then compared to the Gamma-MAP filter with edge detection (Touzi's detector). In the following table are presented the experimental results reported in [FBB01].

Denoising Method	Region 1	Region 2	Region 3	Average Value
Proposed in [FBB01] with D4	34.4	49.8	15.0	33.1
Proposed in [FBB01] with B5	42.8	51.5	22.0	38.8
Gamma-MAP with edge detection	48.5	49.7	22.0	40.1

Table 5.3: A comparison of the ENLs of three regions of a SAR image obtained using the denoising method proposed in [FBB01] with ENLs of the same regions obtained using the association of the 2D UDWT with the Gamma-MAP filter which includes an edge detector inside each estimation window.

In the case of the proposed denoising method the size of the windows used for the local statistics estimation is 7x7 on the first 2D UDWT decomposition level and progressively increases with the increasing of the number of decomposition levels.

Generalized Gaussian MAP filter

In the following we will consider the despecklisation method described in section 4.3.1.6. It is based on the use of a marginal MAP filter built on the hypothesis of a generalized Gaussian statistical model described in [ABA06]. The multiplicative noise is transformed into an additive noise using equation 5.13. Starting from the assumption that both components of the input image, the noiseless component and the noise, are distributed following generalized Gaussians (GG) pdfs, a MAP estimator with the equation presented in 4.3.1.6 is derived. The major novelty of the approach presented in [ABA06] is that the parameters of the GG pdf are taken to be space-varying within each wavelet frame. Thus, they may be adjusted to spatial image context, not only to scale and orientation. Since the MAP equation to be solved is a function of the parameters of the assumed pdf model, the variance and shape factor of the GG function are derived from the theoretical moments, which depend on the moments and joint moments of the observed noisy signal and on the statistics of speckle. The solution of the MAP equation yields the MAP estimate of the wavelet coefficients of the noise-free image. The restored SAR image is synthesized from such coefficients. Two types of mother wavelets were considered: orthogonal Daubechies with four vanishing moments (8 taps) (D4) and biorthogonal with 9/7 taps (B9/7). In table 5.4 are presented the experimental results reported in [ABA06].

Gamma Wavelet MAP filter (Γ-WMAP)

Another speckle removal filter acting in the 2D UDWT domain is proposed in [SE04]. In this paper is introduced the homomorphic Γ - WMAP filter, a wavelet-based statistical

Input	Raw	MAP	
		D4	B9/7
1-look	12.1	26.0	26.2
4-look	17.8	29.3	29.6
16-look	23.7	32.9	33.1

Table 5.4: PSNR performances of the proposed despeckling algorithm applied to noisy versions of 'Lena' image.

speckle filter equivalent to the Γ - MAP filter. A logarithmic transformation is performed in order to make the speckle contribution additive and statistically independent of the radar cross section. Further, the use of normal inverse Gaussian (NIG) distribution as a statistical model for the wavelet coefficients of both reflectance image and noise image is proposed.

The stochastic variable X is said to be NIG distributed if it has a pdf of the form:

$$p_X(x,\theta) = \frac{\alpha\delta}{\pi}\frac{e^{p(x)}}{q(x)}K_1[\alpha q(x)] \tag{5.16}$$

where K_1 is the modified Bessel function of the second kind with index 1, $p(x) = \delta\sqrt{\alpha^2-\beta^2} + \beta(x-\mu)$ and $q(x) = \sqrt{(x-\mu)^2+\delta^2}$. The shape of the NIG - density is specified by the four-dimensional vector $\theta = [\alpha, \beta, \delta, \mu]$ which is bounded: $0 \leq \beta < \alpha$, $\delta > 0$ and $-\infty < \mu < \infty$. The NIG distribution is an excellent statistical model for the wavelet coefficients of synthetic aperture radar images, and in [SE04] is presented a method for estimating the parameters. The MAP filter equation (equation 4.20 in previous chapter) must be solved normally and the Γ-WMAP filter is obtained. The performance of this filter is compared in [SE04] with the performance of Γ-MAP filter which are both based on the same statistical assumptions. The homomorphic Γ-WMAP filter is shown to have better performance with regard to smoothing homogeneous regions. It may, in some cases, introduce a small bias, but it is always less than that introduced by the Γ-MAP filter. Further, the speckle removed by the homomorphic Γ-WMAP filter has statistics closer to the theoretical model than the speckle contribution removed with the other filters. The experimental results obtained in [SE04] are presented in table 5.5.

Analyzing table 5.5 it can be observed that the ENL enhancement produced by the homomorphic Γ-WMAP filter is higher than the ENL enhancement produced by the Γ-MAP filter, with a single exception. This exception appears in region A, where the input ENL is smaller than 1 (which indicates that region A is not perfectly homogeneous). So, the application of the MAP filter into the wavelets' domain improves the treatment of homogeneous regions.

A second observation which can be made analyzing table 5.5 refers to the influence of the analysis window's size. This influence is higher in the case of the MAP filter applied into the spatial domain than in the case of the MAP filter applied in the wavelet domain.

Image	ENL			
	Region A	Region B	Region C	Region D
Window size 5x5				
Original	0.9	1	1	1
Γ-WMAP	19.4	50.8	54.8	48.1
Γ-MAP	6.2	10.7	9.4	11.4
Window size 7x7				
Original	0.9	1	1	1
Γ-WMAP	19.6	50.8	54.8	48.1
Γ-MAP	20.2	19.4	18.2	20.2
Window size 9x9				
Original	0.9	1	1	1
Γ-WMAP	19.8	50.8	54.8	48.1
Γ-MAP	13.1	27.6	29.3	30.3

Table 5.5: A comparison of the performance of the homomorphic Γ-WMAP filter (which acts in the 2D UDWT domain) with the performance of the Γ-MAP filter (which acts in the spatial domain) for three analyzing window sizes, in terms of ENL.

5.3.2.2 MAP filters associated with 2D DWT

In [GD06] is presented another version of MAP filter. The multiplicative noise is transformed into additive noise on the basis of the equation 5.13.

This method includes the step of solving the MAP filter equation into a recursive algorithm. For the denoising of SAR images, in this reference is proposed an association of the DWT with a MAP filter. The first level of the Bayesian inference is given by:

$$p\left(\mathbf{w}|\mathbf{y},\theta\right)=\frac{p\left(\mathbf{y}|\mathbf{w},\theta\right)p_{\mathbf{w}}\left(\mathbf{w}|\theta\right)}{p\left(\mathbf{y}|\theta\right)}, \tag{5.17}$$

where $\mathbf{y}$ is the DWT of the noisy image, $\mathbf{w}$ is the DWT of the noise-free component of the considered image and θ represents the model's parameters. The new idea proposed in [GD06] is to improve the estimation of the model's parameters by maximizing the evidence $p\left(\mathbf{y}|\theta\right)$. It is assumed that the image model in the wavelet domain is well approximated with GG distribution, and the speckle-noise can be modeled as non-stationary Gaussian noise with zero mean and with locally varying variance, which is estimated from a given spatial neighborhood, P. As can be seen, from the previous sections, there are a multitude of models for the wavelet coefficients of the speckle noise. The last hypothesis of Gaussianity will be applied in the rest of this document. The prior pdf $p\left(\mathbf{w}|\theta\right)$ is defined in [GD06] as a generalized Gauss-Markov Random Field (GMRF):

$$p\left(w_k|\theta\right)=\frac{\nu\eta\left(\nu,\sigma_w\right)}{2\Gamma\left(\frac{1}{\nu}\right)}\cdot e^{-\left[\eta(\nu,\sigma_w)\left|w_k-\sum_{p\in P}\theta_p\left(w_k^p+w_k^{p'}\right)\right|\right]^{\nu}}, \tag{5.18}$$

where σ_w represents the standard deviation of the approximated image model, $\eta\left(\nu,\sigma_w\right)=\left(\frac{1}{\sigma_w}\right)\sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}$ is the shape factor and Γ represents the Gamma function. The parameters

σ_w and θ define the GMRF in the neighborhood P. The sum $\sum_{p\in P}\theta_p\left(w_k^p + w_k^{p'}\right)$ represents the sum of all distinct cliques of neighboring pixels at a specific sub-band. Using the additive model, $y = w + n'$, the likelihood pdf can be written in the form:

$$p\left(y_k|w_k\right) = \frac{1}{\sqrt{2\pi}\sigma_n}\cdot e^{-\frac{\left(y_k-w_k\right)^2}{2\sigma_n^2}}. \tag{5.19}$$

The MAP estimate is found by solving equation 5.20:

$$\frac{d}{dw_k} lnp\left(w_k|y_k,\theta\right) = 0, \tag{5.20}$$

or:

$$\begin{aligned}
&\frac{d}{dw_k} ln\left(\frac{1}{\sqrt{2\pi}\sigma_n}\cdot e^{-\frac{\left(y_k-w_k\right)^2}{2\sigma_n^2}}\cdot\frac{\nu\eta\left(\nu,\sigma_w\right)}{2\Gamma\left(\frac{1}{\nu}\right)}\cdot e^{-\left[\eta(\nu,\sigma_w)\left|w_k-\sum_{p\in P}\theta_p\left(w_k^p+w_k^{p'}\right)\right|\right]^\nu}\right) = 0 \Leftrightarrow \\
\Leftrightarrow\quad &\frac{d}{dw_k}\left(-\frac{\left(y_k-w_k\right)^2}{2\sigma_n^2} - \eta^\nu\left(\nu,\sigma_w\right)\left|w_k-\sum_{p\in P}\theta_p\left(w_k^p+w_k^{p'}\right)\right|^\nu\right) = 0 \Leftrightarrow \\
\Leftrightarrow\quad &\frac{y_k-w_k}{\sigma_n^2} - \eta^\nu\left(\nu,\sigma_w\right)\cdot\nu\cdot\frac{\left|w_k-\sum_{p\in P}\theta_p\left(w_k^p+w_k^{p'}\right)\right|^\nu}{w_k-\sum_{p\in P}\theta_p\left(w_k^p+w_k^{p'}\right)} = 0.
\end{aligned} \tag{5.21}$$

For $\nu = 1$, the last equation of 5.21 is a first order algebraic equation which can be solved analytically. For $\nu = 2$, the solution of this equation represents the zero order Wiener filter:

$$\hat{w}_k\left(y_k\right) = \frac{\sigma_w^2\cdot y_k + \mu\cdot\sigma_n^2}{\sigma_w^2+\sigma_n^2} \tag{5.22}$$

where

$$\mu = \sum_{p\in P}\theta_p\left(w_k^p + w_k^{p'}\right).$$

For the other values of ν, the equation 5.21 must be solved numerically. The variable ν is incremented with a step of 0.2 in the interval $[0.8, 2]$. To choose an appropriate value for parameter ν, is taken advantage of the second level of Bayesian inference by the maximization of the evidence

$$ev = p\left(\mathbf{y}|\theta\right) = \int p\left(\mathbf{y}|\mathbf{w},\theta\right)p\left(\mathbf{w}|\theta\right).$$

For each value of $\nu = \nu_k$ the evidence $ev\left(\nu_k\right)$ is computed and, the value corresponding to $\underset{k}{argmax}\left\{ev\left(\nu_k\right)\right\}$ is selected. Next, the MAP filter equation is solved and the estimator obtained is applied to the current pixel of the image to be denoised. The experimental result reported in [GD06] can be seen in the figure 5.10.

Analyzing figure 5.10, the better treatment of contours realized by the method which acts in the wavelets domain can be observed. We can also observe that the homogeneous regions are better reproduced in image (b) than in image (c). At the end of [GD06] is written: "All routines were written in ANSI C, and for the selected image, with 256x256

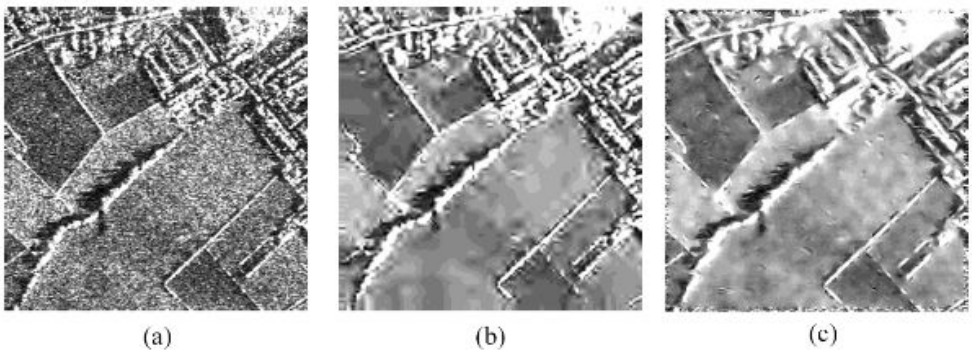


Figure 5.10: A comparison of the results obtained despeckling the image (a) with the algorithm proposed in [GD06] (b) and with the MBD algorithm proposed in [WD00] (c).

pixels as shown in figure, the proposed algorithm for despeckling runs for 7.6 s on a PC with a Pentium 4 processor with a CPU clock 2.6 MHz. The MBD algorithm runs for 41.3 s (on the same computer) using first-order neighborhood and window size 3x3.". So, the computation time required by the despeckling method which acts in the wavelets domain is shorter.

5.3.2.3 MAP filters associated with 2D DTCWT

A possible solution for the construction of the despeckling system in figure 5.1 is to use the additive noise denoising kernel based on the association of the 2D DTCWT with the bishrink filter, described in section 4.3.1.5. One of the papers in which this solution is exploited is [XXM08]. The only difference between the denoising kernel described in 4.3.1.5 and the solution proposed in [XXM08] concerns the computation of the local standard deviation. The estimator in equation 4.85 is replaced in [XXM08] with a Wiener estimator.

Another paper that exploits the association of the 2D DTCWT with a version of the bishrink filter for despecklisation purposes is [IIQ07]. This variant was named 'mixed bishrink filter'. After three iterations of each DWT representing one tree of a DTCWT, the pdf of wavelet coefficients can be considered Gaussian. The mixed bishrink filter acts for the first three iterations of each DWT as a bishrink filter with global estimation of the local variance, for the forth iteration it acts as a local adaptive Wiener filter and for the fifth iteration (the last one) it acts as a hard-thresholding filter with the threshold equal to $3\hat{\sigma}_n$. The bishrink filter with global estimation is a version of the genuine bishrink filter, constructed to increase the precision of the estimation in 4.87. The estimation in 4.87, although based on the correct assumption that W_x^1 and W_x^2 are modeled as zero-mean random variables, is not very precise as these variables' restrictions to the finite neighborhood $N(k)$ are not necessarily zero-mean. The estimation proposed in [IIQ07] uses the estimation of the mean in the neighborhood

$$\hat{\mu}_{W_x} = \frac{1}{M} \sum_{W_x^i \in N(k)} W_x^i,$$

followed by the estimation of the variance:

$$\hat{\sigma}^2_{W_x} = \frac{1}{M} \cdot \sum_{W_x^i \in N(k)} \left(W_x^i - \hat{\mu}_{W_x}\right)^2,$$

and, finally, the relation in 4.87. In the case of the bishrink filter with global estimation of the local variance, the detail wavelet coefficients at the output of the first tree are indexed with re and the ones at the output of the second tree are indexed with im, the standard deviation resulted being the arithmetic mean of the values obtained at each tree:

$$\hat{\sigma} = \frac{_{re}\hat{\sigma} +_{im} \hat{\sigma}}{2}.$$

Using this estimation, the bishrink filter with global estimation of the local variance is applied separately to the real detail wavelet coefficients produced by each of the two trees composing the DTCWT.

The mixed bishrink filter was associated with the 2D DTCWT in [IIM+05]. The experimental result reported in this paper is presented in figure 5.11.

The speckle was entirely suppressed. The good treatment of contours and textured regions can be observed analyzing figure 5.11. The use of the 2D DTCWT, which has an enhanced directional selectivity, allows a good representation of all the details (with any orientation) of the noiseless input image. The input image has a good quality, taking into account its ENL value (which is relatively high), being relatively simple to process. For low quality input images, the despecklisation method must be improved. One way for the improvement of the additive noise denoising kernel is to apply diversification mechanisms obtaining some partial results and then to synthesize these partial results. Such a strategy was applied in [IMI09]. The additive noise denoising kernel proposed in [IMI09] has the architecture in figure 5.12.

The first stage of the algorithm is represented in red. It is composed by four blocks. The first three blocks implement the genuine denoising method based on the use of the bishrink filter with global estimation of local variance.

The first block of the first stage implements a DTCWT and the third one the corresponding inverse transform (IDTCWT). So, a first result $\hat{s}_{2A}$ is obtained. The pilot image is generated by the segmentation of $\hat{s}_{2A}$, done by the block $Segm$. The segmentation is done with regard to the value of local standard deviation of each pixel of the first result. The elements of a class have a value of the local variance, belonging to one of six possible intervals, $I_p = (\alpha_p \hat{\sigma}_{2A_{max}}, \alpha_{p+1}\hat{\sigma}_{2A_{max}})_{1 \le p \le 6}$, where $\alpha_1 = 0$ and $\alpha_7 = 1$. This way the data contained in each class is uniform. The class selector CS_p in figure 5.12 selects the class associated to the interval I_{7-p}. Intensive preliminary tests have led to the following values for the bounds of the intervals I_p: $\alpha_2 = 0.025$, $\alpha_3 = 0.05$, $\alpha_4 = 0.075$, $\alpha_5 = 0.1$ and $\alpha_6 = 0.25$. Recording the coordinates of the pixels belonging to each class, six masks are generated.

The second stage of the denoising system in figure 5.12 is represented in blue. To realize the diversification required in the second stage of the proposed algorithm two types of WT, DT-CWT A and DT-CWT F are computed, obtaining the wavelet coefficients w_A and w_B. The first diversification mechanism refers to the construction of the DT-CWT. Since an image usually consists of several regions of different smoothness, the sparsity of

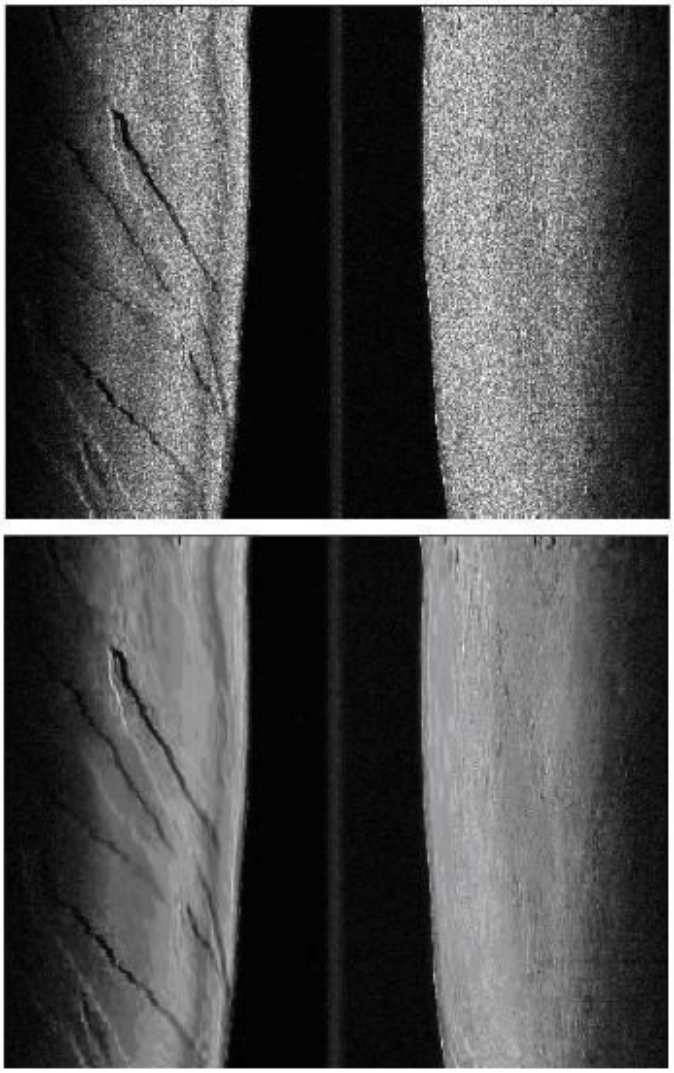


Figure 5.11: The result reported in [IIM+05]. The noisy image was acquired by IFREMER, Brest, France (ENL = 7.34 - up). The denoised image (ENL = 76.64 - down)

its representation in a single wavelet domain is limited. This naturally motivates using multiple wavelet transforms to denoise. This procedure is used for example in [CB04]. Besov balls are convex sets of images whose Besov norms are bounded from above by their radii. Projecting an image onto a Besov ball of proper radius corresponds to a type of wavelet shrinkage for image denoising. By defining Besov balls in multiple wavelet domains and projecting onto their intersection using the projection onto convex sets (POCS) algorithm, an estimate is obtained in [CB04], which effectively combine estimates from multiple wavelet domains. There are two kinds of filters used for the computation of the DT-CWT: for the first decomposition level and for the other levels [Kin01]. The first diversification mechanism is realized through the selection of two types of filters for the first level. The first one is selected from the (9,7)-tap Antonini filters pair and the second one corresponds to the pair of Farras nearly symmetric filters for orthogonal

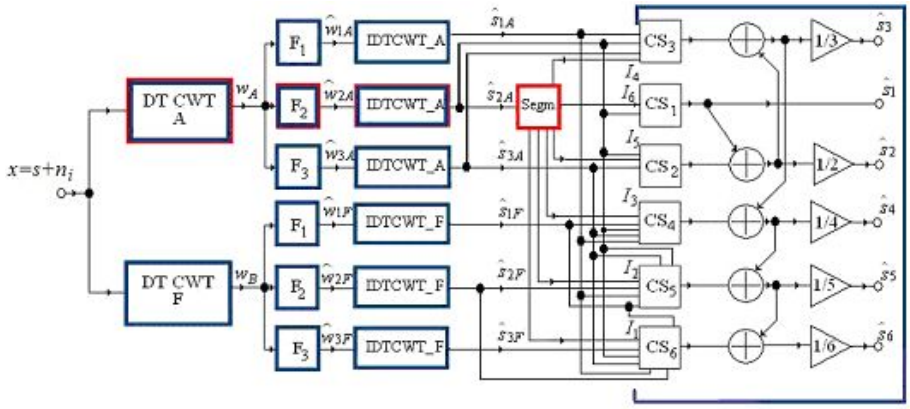


Figure 5.12: The architecture of the additive noise denoising kernel proposed in [IMI09].

2-channel perfect reconstruction filter bank, [AS01]. The idea of diversification by using multiple mother wavelets was also exploited in [IIQ07, FA05], where the bishrink filter was associated with DWT. The same WT was used in [CB04]. The synthesis of the final result was carried out in [IIQ07] by simple averaging and in [CB04, FA05] by variational frameworks.

Next, three versions of bishrink filter: F1-the genuine one, F2-the adaptive bishrink filter with global estimation of the local variance and F3-the mixed bishrink filter, are applied in the field of each DT-CWT. These variants represent another diversification mechanism. Six sequences of estimations of the wavelet coefficients: $\hat{w}_{1A}$, $\hat{w}_{2A}$, $\hat{w}_{3A}$, $\hat{w}_{1F}$, $\hat{w}_{2F}$ and $\hat{w}_{3F}$ are obtained. For each one the inverse WT, IDT-CWT, is computed, obtaining six partial results: $\hat{s}_{1A}$, $\hat{s}_{2A}$, $\hat{s}_{3A}$, $\hat{s}_{1F}$, $\hat{s}_{2F}$ and $\hat{s}_{3F}$. This way, the redundancy was increased because the actual volume of data is six times higher than the initial volume of data. With the aid of the six masks generated at the end of the first stage, the six classes of each partial result are identified. Using the class selectors $CS_1 - CS_6$, the partial results are individually treated. Each mask is used by the corresponding class selector. These systems select the pixels of their input image with the coordinates belonging to the correspondent mask. CS_1 is associated with the class containing the highest values of the local standard deviation and treats the image $\hat{s}_{2A}$. It generates the first class of the final result, $\hat{s}_1$ and contributes to the generation of the classes $\hat{s}_2 \div \hat{s}_6$ of the final result. CS_2 corresponds to the next class of $\hat{s}_{2A}$ and treats the image $\hat{s}_{3A}$, participating to the construction of the classes $\hat{s}_2 \div \hat{s}_6$ of the final result. CS_3 corresponds to the next class of $\hat{s}_{2A}$ and treats the image $\hat{s}_{1A}$. It contributes to the construction of the classes $\hat{s}_3 \div \hat{s}_6$ of the final result and so on. Finally CS_6 is associated to the remaining class of $\hat{s}_{2A}$ and treats the image $\hat{s}_{3F}$. It participates to the construction of the sixth class of the final result, $\hat{s}_6$ (the one that contains the smallest values of the local variance). By NL-averaging (a NL-means like methodology), the six classes of the final result are obtained. The first class of the final result, $\hat{s}_1$, is identical with the first class of the image $\hat{s}_{2A}$ and represents the output of CS_1. The second class of the final result, $\hat{s}_2$, is obtained by averaging the pixels of the outputs of CS_1 and CS_2 and so on. For the last class of the final result, $\hat{s}_6$, containing soft textures and homogeneous zones, all the pixels belonging to the outputs

of CS_1, CS_2, ... and CS_6 are averaged. Assembling these classes by concatenation, the final estimation is obtained.

Some simulations results reported in [IMI09] which refer to synthesized speckle are presented in the following. In this case, the noise is generated following a Rayleigh distribution with unitary mean and is of a multiplicative nature. It is generated computing the square root of a sum of squares of two white Gaussian noises having the same variance. For the Lena image, applying the denoising system in figure 5.1 constructed with the additive noise denoising kernel in figure 5.12, in [IMI09] was obtained the result presented in figure 5.13.

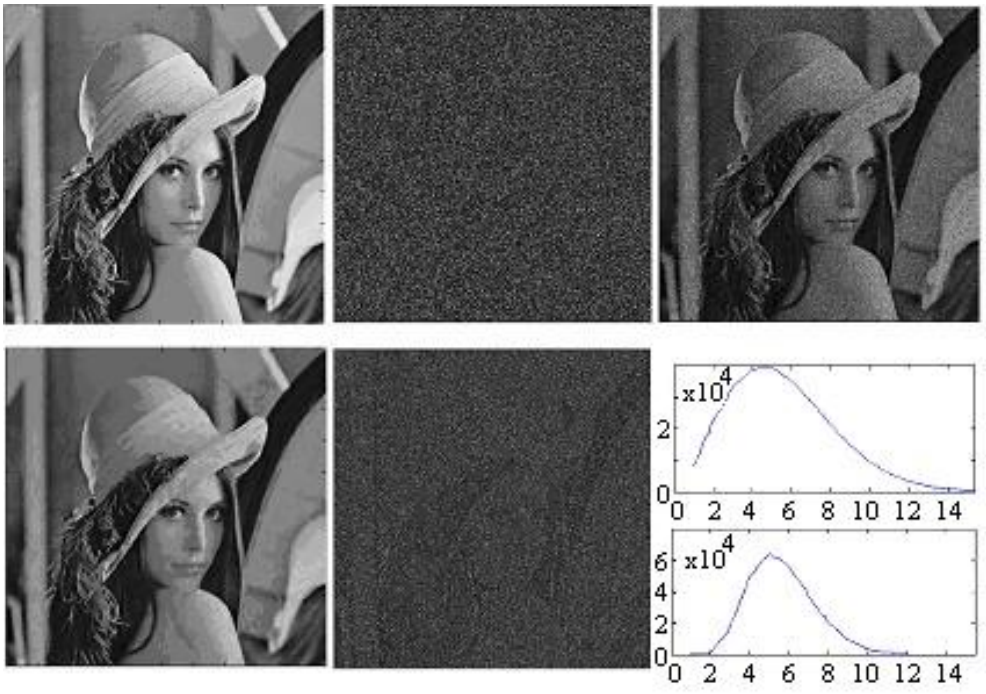


Figure 5.13: Synthesized speckle noise. First line, from left to right: clean image; synthesized speckle; noisy image (PSNR=21.4 dB). Second line, from left to right: denoised image (PSNR=31.4 dB); method noise; histograms of the noise (up) and method noise (bottom).

The PSNR gain performed by the proposed method is in this case of 10 dB. An excellent criterion for the appreciation of the quality of a denoising method conceived for the reduction of the multiplicative noise is based on the computation of the method noise. It represents the ratio of the noisy image by the denoising result, [AKZ06]. The method noise must be identical with the input noise for a perfect denoising method. It can be observed, analyzing figure 5.13, that the input noise (represented in the second picture from the first line) has the same aspect like the method noise (represented in the second picture from the second line). There are some fine differences between the images of the input noise and of the method noise, noticeable especially in the dark regions of the noise-free component of the input image (represented in the first picture of the first line in figure 5.13).

In the last picture from the second line is carried out a comparison between the histograms of the input noise (up) and of the method noise (down), highlighting the statistical differences between these two noises. Although they are distributed following the same type of law (a Rayleigh law), the input noise has a higher variance, meaning that the contrast of the noise-free input image is affected by the proposed denoising method.

It can be observed that the proposed method performs a good treatment of edges and of homogeneous regions. Its drawback is the textures treatment, some of the fine textures of the clean component of the acquired image being erased by the denoising. A better analysis of the visual aspect of the proposed method can be carried out if it is applied to the test image, proposed in [WD00]. In this case the speckle noise is synthesized following the procedure presented in [WD00]. The methods from [FA05] and [WD00] are compared in figure 5.14 with the denoising method proposed in [IMI09].

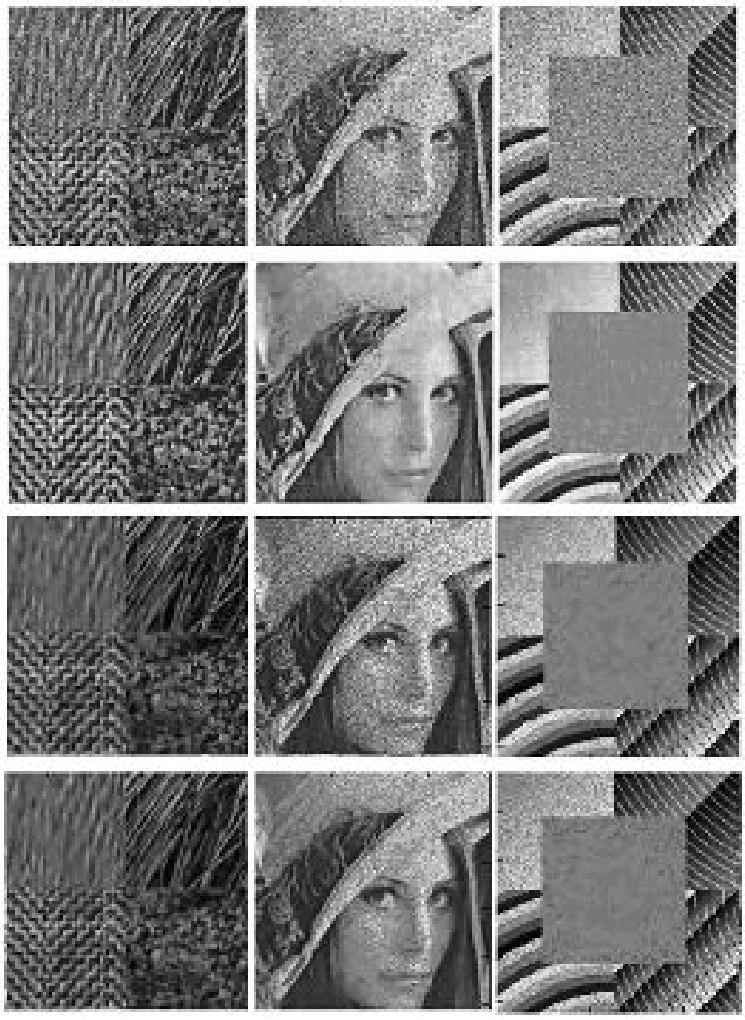


Figure 5.14: From up to bottom: noisy sub-images; results obtained in [WD00]; results obtained applying the method proposed in [FA05]; results of the denoising method proposed in [IMI09].

The better treatment of the homogeneous regions is performed by MBD method in [WD00] but it erases some contours and textures carrying out an over-smooth filtering. The other two methods use the DT-CWT and treat the details better. The method proposed in [FA05], based on the DT-CWT-genuine bishrink filter denoising association does not eliminate all the noise. This effect is easy visible in the homogeneous regions. The denoising method proposed in [IMI09] makes a good treatment of real scenes, completely eliminating the noise and introducing small distortions. The treatment of edges is excellent. The denoising of rough textures is more accurate. Some distortions are visible at the borders of homogeneous zones. In the interior of those regions residual noise can be observed.

Finally, some results concerning real SONAR images are presented in [IMI09]. Figure 5.15 shows the original SONAR image 'Swansea' and the results obtained with the method in [IIQ07] and the method proposed in [IMI09]. The visual analysis of the filtered image proves the correctness of the assumptions made in [IMI09]. Indeed, the result of the proposed method has a better visual aspect, the result in [IIQ07] being slightly over-smoothed. An objective measure of the homogeneity degree of a region of a SAR image is the ENL. The enhancement of the ENL of a denoising method in a homogeneous region is defined by the ratio of the ENLs of the considered region computed before and after the application of the method.

The performance obtained for homogeneous regions through the denoising method proposed in [IMI09] is certificated by the important enhancement of the ENL obtained considering a region of 120x1000 pixels.

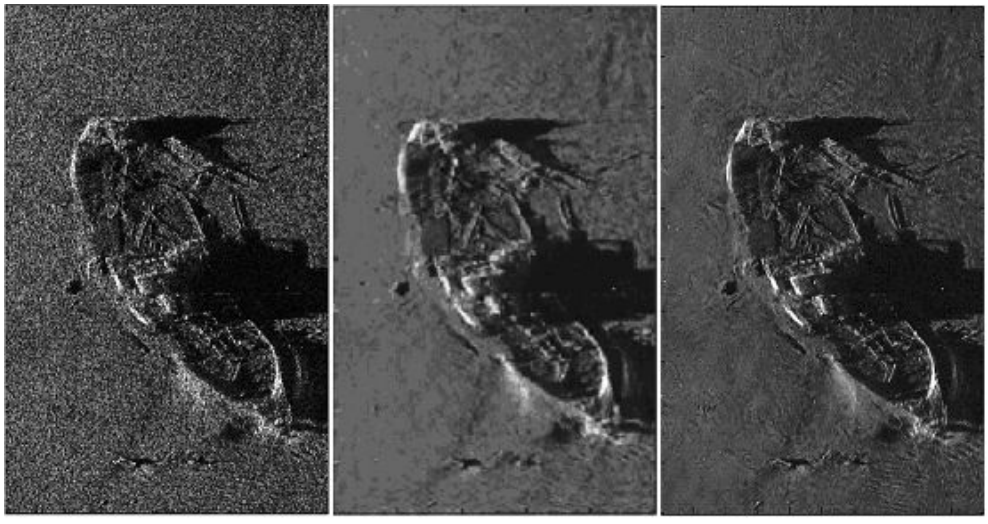


Figure 5.15: Speckle removal for the sea-bed SONAR Swansea image (acquired by GESMA). Left: acquired image (ENL=3.4), Middle: result in [IIQ07] (ENL=106), Right: result of the denoising method proposed in [IMI09] (ENL=101.8).

5.3.2.4 MAP filters associated with HWT

In this section I will present the association of the HWT with different MAP filters applied on images affected by multiplicative noise. Having in mind the properties of natural images depicted in section 4.1.5 and the fact that they are in agreement with the second order statistical analysis of the HWT coefficients, we have selected the HWT for denoising of SAR images, which are a class of natural images. All the methods imply homomorphic approach to speckle filtering. Basically, we have the processing scheme in figure 5.1 and we replace the additive denoising kernel with algorithms already presented in chapter 4 and some improved versions of them. The tests were performed on both synthetic images and SAR images.

The adaptive soft-thresholding filter

First, we have associated HWT with the adaptive soft-thresholding filter presented in section 4.3.1.2. The estimation windows are directional windows. Secondly, we have added a homogeneous-area correction block, in order to achieve an additional smoothing of these areas. This block is basically an averager and, with respect to the value of the standard deviation, uses a different averaging window size, e.g. the smaller the standard deviation, the greater the size of the averaging window.

In table 5.6 we present the values obtained using the adaptive soft-thresholding filter, both the simple version and the one with correction added, applied to the 'Lena' input image, using the same parameters as in [ABA06], namely noise parameters and mother wavelets.

Input	Raw	HWT + Astf		HWT + Astf + cor	
		D4	B9/7	D4	B9/7
1-look	12.1	25.4	25.1	25.6	25.4
4-look	17.8	29.9	29.9	30.0	29.9
16-look	23.7	33.2	32.9	32.9	33.2

Table 5.6: PSNR performances of HWT associated with adaptive soft-thresholding, applied to noisy versions of 'Lena' image

Comparing table 5.6 with table 5.4, we can observe that the results are quite similar, although our approach is less redundant than the one using UDWT.

In figure 5.16 are presented the results on despeckling 'Lena' image.

The noise used in figure 5.16 to disturb the original image has a Rayleigh distribution with unitary expectation. The acquired image was obtained through the multiplication of the original image with the speckle noise. The speckle noise is generated computing the square root of a sum of squares of two white Gaussian noises having same variance. The noiseless component of the input image is shown in figure 5.16(a). The input image can be seen in figure 5.16(b). The result of denoising using the HWT and the adaptive soft thresholding is shown in figure 5.16(c), while the result after the homogeneous-areas correction is presented in 5.16(d). The PSNR enhancement is relatively important, 10.5 dB. The visual aspect of the result in figure 5.16(d) is quite good, but some distortions

(a) Original image

(b) Noisy input image (PSNR = 21.4 dB)

(c) HWT + Astf (PSNR = 32.0 dB)

(d) HWT + Astf + correction (PSNR = 31.94 dB)

Figure 5.16: HWT - Adaptive soft-thresholding denoising results applied on Lena affected by multiplicative noise

are visible. There are some intensity variations in the homogeneous zones, some textures of the hat are lost and some fine contours are over smoothed.

The bishrink filter

In this section I will consider the denoising kernel in figure 5.1 formed by the bishrink filter in the HWT domain.

Properties P1 and P2 (section 4.1.5) clearly show the inability of any wavelet basis to fully decorrelate the covariance matrices of nonstationary self-similar images. Therefore the residual dependencies have to be taken into account. We can take advantage of these correlations when trying to estimate the true image coefficients from noisy observations, since the noise coefficients are decorrelated if the noise is white and stationary. These hypotheses were made for the construction of the bishrink filter. For this reason we have selected the denoising method which associates the HWT with the bishrink filter.

The bishrink filter is largely presented in section 4.3.1.5. In the following I have considered the bishrink filter associated with 8-tap Daubechies and Biorthogonal 9/7, both with and without the supplementary homogeneous area correction and with the HWTDE with the extra correction. Directional elliptic windows were used for the estimation of local parameters, in all of these cases.

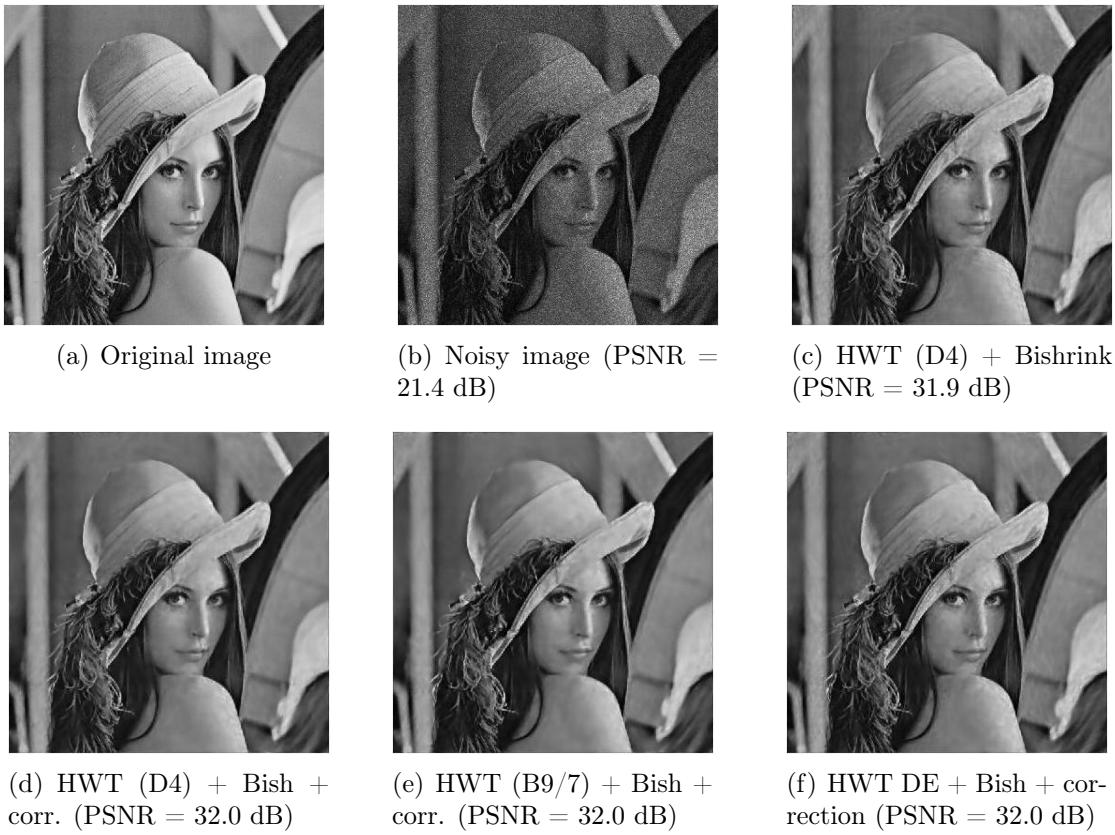
The first test involved the speckle reduction of synthetic noise, applied in a multiplicative form on the 'Lena' input image. From the PSNR point of view, the results obtained are presented in table 5.7.

Input	Raw	HWT + Bish		HWT + Bish + cor		HWTDE + Bish + cor
		D4	B9/7	D4	B9/7	
1-look	12.1	25.7	25.7	25.9	25.9	25.8
4-look	17.8	29.9	30.1	30.2	30.2	30.1
16-look	23.7	33.0	33.2	33.0	33.1	33.2

Table 5.7: PSNR performances of HWT, respectively HWTDE, associated with bishrink, applied to noisy versions of 'Lena' image.

Analyzing table 5.7 we can observe that, from the PSNR point of view, there is a slight increase when the additional correction is performed, and that the additional computation burden introduced by the HWT DE does not necessarily improve the PSNR of the result. Overall, the results are close to the ones presented in [ABA06].

A visual comparison of the results obtained with each of the versions previously mentioned is done in figure 5.17.



(a) Original image

(b) Noisy image (PSNR = 21.4 dB)

(c) HWT (D4) + Bishrink (PSNR = 31.9 dB)

(d) HWT (D4) + Bish + corr. (PSNR = 32.0 dB)

(e) HWT (B9/7) + Bish + corr. (PSNR = 32.0 dB)

(f) HWT DE + Bish + correction (PSNR = 32.0 dB)

Figure 5.17: HWT - Bishrink denoising results applied on Lena affected by multiplicative noise

By comparing figure 5.17(c) with figures 5.17(d) and 5.17(e) we can observe that the correction improves the visual quality of the result. Another observation, deduced by the

analysis of figures 5.17(d), 5.17(e) and 5.17(f) is that the enhanced diversity does not necessarily improve the visual quality of the denoising, thus not justifying the additional computation burden introduced by the use of HWT DE.

In figure 5.18 we present a visual comparison between the results obtained with the association HWT - bishrink and those presented in the literature, associating UDWT and GGPDF-based MAP filter. In table 5.8 are presented the results from the PSNR point of view.

(a) HWT + bishrink

(b) UDWT + GGPDF-based MAP

Figure 5.18: HWT + Bishrink vs. UDWT + GGPDF-based MAP

Input	Raw	HWT + Bish + cor		GGPDF-MAP	
		D4	B9/7	D4	B9/7
1-look	12.1	25.9	25.9	26.0	26.2
4-look	17.8	30.2	30.2	29.3	29.6
16-look	23.7	33.0	33.1	32.9	33.1

Table 5.8: HWT + Bishrink vs. UDWT + GGPDF-based MAP from the PSNR point of view

Another comparison refers to the image in figure 5.2. In figure 5.19 we present the result of the bishrink filter applied in HWT domain (figure 5.19(a)), having B9/7 as mother wavelet and the result of the denosing algorithm presented in section 5.3.1 (figure 5.19(b)). It can be observed that the edges and textures are better preserved in figure 5.19(a) than in 5.19(b).

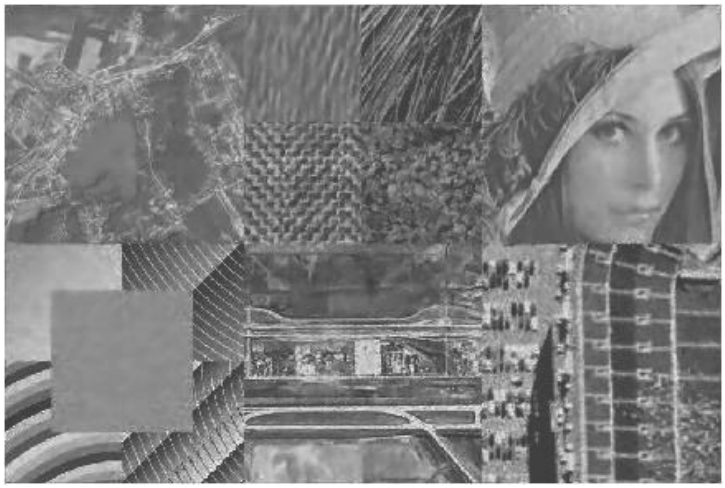
By considering a real SONAR image, namely 'Swansea', the results presented in figure 5.20 are obtained.

In the case of SAR images, the results are presented in 5.21.

Analyzing figures 5.20 and 5.21 we can affirm that our method performs well in the case of SAR images as well, making it suitable for speckle reduction applications.

5.4 Summary

In this chapter I have presented the particular case of SAR images and I have started with an overview of speckle removal techniques both in the spatial domain and in the wavelet

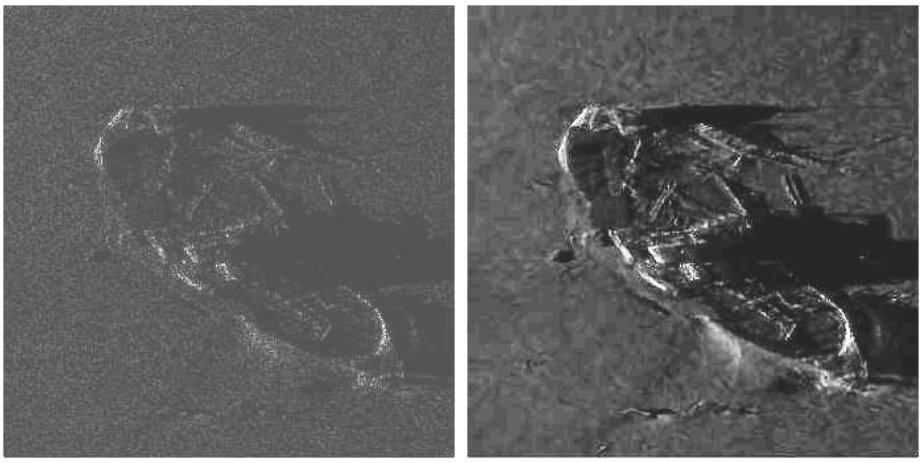


(a) HWT + Bishrink

(b) DEDWT + Soft-thresholding

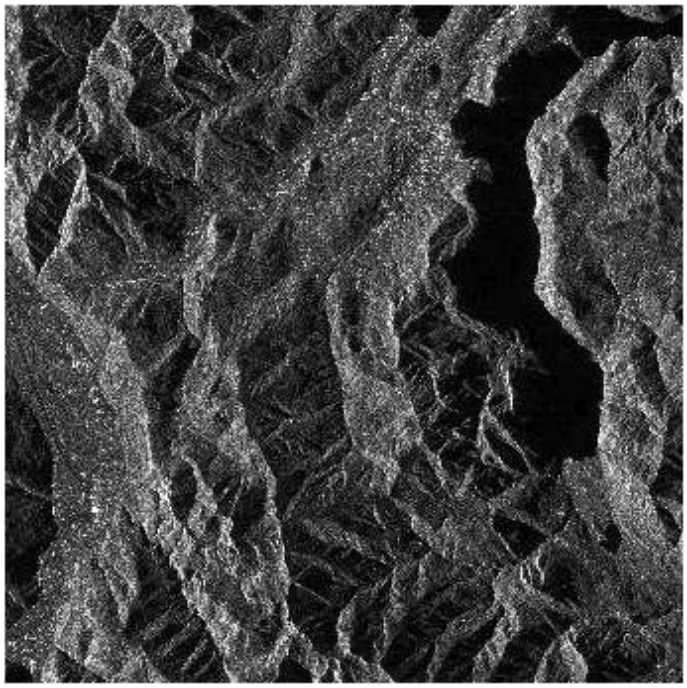
Figure 5.19: HWT - Bishrink denoising results obtained for the test image

domain. In what concerns the methods used in the wavelet domain, they were classified according to the type of the wavelet transform used: 2D DWT, 2D UDWT, 2D DT CWT and HWT. In the case of the HWT, we have used homomorphic filtering techniques and applied the denoising algorithms already discussed in Chapter 4 in the 'additive noise denoising kernel' block, in figure 5.1. The results obtained were satisfactory, especially those in which HWT with Biorthogonal 9/7 as mother wavelet was associated with the bishrink filter, followed by a supplementary correction applied to homogeneous areas.

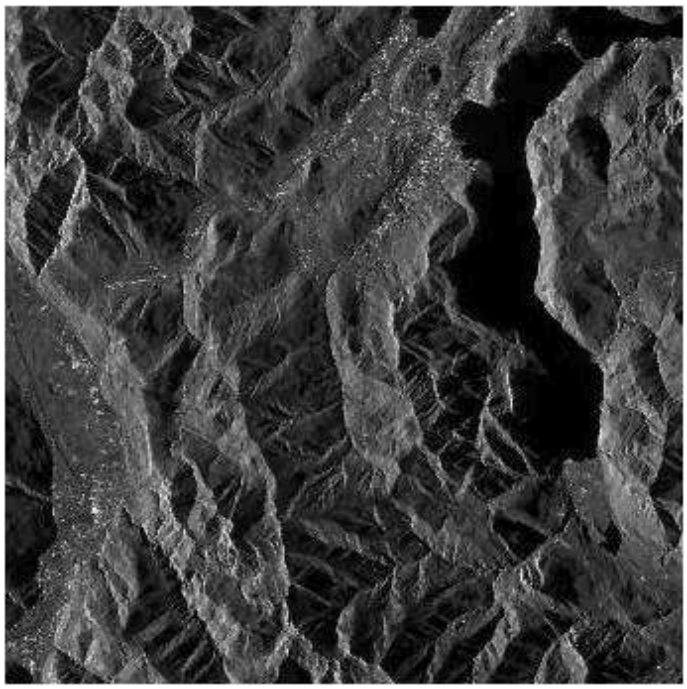


(a) Original image, ENL = 3.32 (b) Denoised image, ENL = 155.04

Figure 5.20: HWT - Bishrink denoising results applied on SONAR image. In this case the ENL is 50 times higher, while in the results presented in 5.15 is only about 30 times higher



(a) Original image



(b) Denoised image

Figure 5.21: Results of HWT - Bishrink denoising applied on SAR image

Chapter 6

Conclusions

6.1 Contributions

In this work, I have proposed:

1. A new one-dimensional wavelet transform called the Analytic Discrete Wavelet Transform (ADWT), introduced to overcome the shift-sensitivity of the Discrete Wavelet Transform (DWT). It is made of two trees, both of them implementing a DWT, one applied to the original signal and the other applied to the Hilbert transform of the original signal. This transform is complex, the complex coefficients having for real part the output of the first tree and for imaginary part the output of the second tree.

2. A new one-dimensional inverse wavelet transform (IADWT). The inverse ADWT is computed by applying 2 IDWT trees, one to the real part of the coefficients, the other to the imaginary part and by **taking the average between the output of the first reconstruction tree and the inverse Hilbert transform of the output of the second reconstruction tree**.

3. The quasi shift-invariance of ADWT was measured in simulations both visually and through the values of the newly introduced measure, namely the **degree of invariance**.

4. The advantage of choosing ADWT over the well-known Dual Tree Complex Wavelet Transform is the possibility to **freely choose the mother wavelet** from the wide range classically associated with the DWT.

5. A new two-dimensional wavelet transform called the HWT, basically representing the generalization in the 2D space of the ADWT. Besides the quasi shift-invariance the HWT inherits from the ADWT, it also has a good directional selectivity, being able to separate different angle orientations.

6. The association of the HWT with the zero-order local Wiener filter for denoising purposes.

7. The association of the HWT with the adaptive soft-thresholding filter, association that outperforms the wiener 0 filter for low signal-to-noise ratios, in denoising applications.

8. I have taken into account the inter-scale dependency of the wavelet coefficients and associated HWT with the bishrink filter in denoising.

9. Taking into account the drawbacks of the bishrink filter, it results that regions with different homogeneity degrees must be treated using different strategies.

   This task can be accomplished through diversification. Doing so, some partial results can be obtained. The final result can be obtained starting from these partial results using a synthesis mechanism. I have tried to propose different synthesis mechanisms and different shapes of the analysis windows in order to increase the quality of the result.

   One of the diversification mechanisms is based on the mother wavelets selection. Such a selection strategy could be based on the time-frequency localization of the mother wavelets.

10. For speckle reduction purposes, I have used a homomorphic filtering, replacing the additive denoising kernel with algorithms using the association HWT - adaptive soft-thresholding and HWT - bishrink, the results obtained being comparable with those in the literature.

6.2 Perspectives

The present work can be continued following a few main directions:

- in what concerns the HWT, it can be improved by finding a better implementation of the Hilbert transform, thus reducing the approximation errors the present implementation introduces;

- a thorough research regarding the criteria to be used in choosing a mother wavelet suitable for the denoising of a particular image.

- associating a wider range of estimators with the HWT, such as the BLS-GSM algorithm and others.

Bibliography

[ABA06] F. Argenti, T. Bianchi, and L. Alparone. Multiresolution MAP Despeckling of SAR Images Based on Locally Adaptive Generalized Gaussian PDF Modeling. *IEEE Transactions on Image Processing*, 15, no. 11: 3385 – 3399, 2006.

[Abr94] P. Abry. *Transformées en ondelettes-Analyses multirésolution et signaux de pression en turbulence.* PhD thesis, Université Claude Bernard, Lyon, France, 1994.

[AF01] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96, no. 455:939 – 955, 2001.

[AK05] A. Achim and E. E. Kuruoglu. Image Denoising Using Bivariate α-Stable Distributions in the Complex Wavelet Domain. *IEEE Signal Processing Letters*, 12 (1): 17 – 20, 2005.

[AKZ06] Alin Achim, Ercan Kuruoglu, and Josiane Zerubia. SAR Image Filtering Based on the Heavy-Tailed Rayleigh Model. *IEEE Transactions on Image Processing*, 15 (9): 2686 – 2693, 2006.

[All77] J. B. Allen. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Transactions on Acoustics, Speech, Signal Processing*, ASSP-25: 235 – 238, 1977.

[ANBI07a] I. Adam, C. Nafornita, J.-M. Boucher, and Al. Isar. A New Implementation of the Hyperanalytic Wavelet Transform. In *Proceedings of IEEE International Symposium ISSCSŠ07, Iasi, Romania*, 2007.

[ANBI07b] Ioana Adam, Corina Nafornita, Jean-Marc Boucher, and Alexandru Isar. A Bayesian Approach of Hyperanalytic Wavelet Transform Based Denoising. In *Proc. IEEE International Conference WISP' 07, Alcala de Henares, Spain*, 2007.

[AOB06] I. Adam, M. Oltean, and M. Bora. A New Quasi Shift Invariant Non-Redundant Complex Wavelet Transform. In *Scientific Bulletin of the ŤPOLITEHNICAŤ University of Timisoara, number dedicated to the Symposium on Electronics and Telecommunications ETC 2006, Seventh Edition, Timisoara*, 2006.

[APM08] Abdourrahmane M. Atto, Dominique Pastor, and Gregoire Mercier. Smooth Sigmoid Wavelet Shrinkage For Non-Parametric Estimation. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Las Vegas, Nevada, USA*, 30 march - 4 april 2008.

[AS01] A. Abdelnour and I. W. Selesnick. Nearly symmetric orthogonal wavelet basis. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001.

[Aus92] Pascal Auscher. Il n'existe pas de bases d'ondelettes régulières dans l'espace de Hardy $H^2(R)$. *C. R. Acad. Sci. Paris*, 315:769–772, 1992.

[BN04] S. Barber and G. P. Nason. Real nonparametric regression using complex wavelets. *J. Roy. Stat. Soc. B*, 66, no. 4: 927 – 939, 2004.

[Bua07] Antoni Buades. *Image and film denoising by non-local means*. PhD thesis, Universitat de les Illes Balears, Dpt. de matematiques i informatica, 2007.

[CB04] Hyeokho Choi and Richard G. Baraniuk. Multiple Wavelet Basis Image Denoising Using Besov Ball Projections. *IEEE Signal Processing Letters*, 11, no. 9: 717 – 720, 2004.

[CD95] R. Coifman and D. Donoho. *Wavelets and Statistics*, chapter Translation invariant de-noising, pages 125–150. Springer-Verlag, 1995.

[Dau88] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure & Appl. Math.*, 41(7):909–996, 1988.

[Dau92] Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

[Dav] C. Davenport. Commutative Hypercomplex Mathematics. available at http://home.comcast.net/ cmdaven/hypercplx.htm.

[DEG+92] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, Ph. Tchamitchian, and B. Torresani. Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies. *IEEE Transactions on Information Theory*, 38, no. 2:644–664, 1992.

[DJ94] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, vol. 81, no. 3: 425 – 455, 1994.

[FA05] Ronan Fablet and Jean-Marie Augustin. Variational Multiwavelet Restoration of Noisy Images. In *Proceedings of IEEE International Conference on Image Processing, ICIP' 2005*, 2005.

[FBB01] S. Foucher, G. B. Bénié, and J.-M. Boucher. Multiscale MAP Filtering of SAR images. *IEEE Transactions on Image Processing*, 10, no.1: 49 – 60, 2001.

[FIB09] Ioana Firoiu, Alexandru Isar, and Jean-Marc Boucher. An Improved Version of the Inverse Hyperanalytic Wavelet Transform. In *Proceedings of IEEE International Symposium SCS' 09, Iasi, Romania*, 2009.

[FIBI09] Ioana Firoiu, Dorina Isar, Jean-Marc Boucher, and Alexandru Isar. Hyperanalytic Wavelet Packets. In *Proc. WISP 2009, 6th IEEE International Symposium on Intelligent Signal Processing, Budapest, Hungary*, 2009.

[FII10] I. Firoiu, Al. Isar, and D. Isar. A Bayesian Approach of Wavelet Based Image Denoising in a Hyperanalytic Multi-Wavelet Context. *WSEAS Transactions on Signal Processing*, 6, Issue 4: 155 – 164, 2010.

[FKE07] Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images. *IEEE Transactions on Image Processing*, 16, no. 5: 1395 – 1411, 2007.

[FLA00] Roger Fjørtoft, Armand Lopès, and Frédéric Adragna. Radiometric and Spatial Aspects of Speckle Filtering. In *Proceedings of IEEE International Conference IGARSSŠ2000, Honolulu, Hawaii*, 2000.

[FNBI09] Ioana Firoiu, Corina Nafornita, Jean-Marc Boucher, and Alexandru Isar. Image Denoising Using a New Implementation of the Hyperanalytic Wavelet Transform. *IEEE Transactions on Instrumentation and Measurements*, 58, Issue 8: 2410 – 2416, 2009.

[FSJ+82] V. S. Frost, J. A. Stiles, A. Josephine, K. S. Shanmugan, and J. C. Holtzman. A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4, no. 2: 157 – 166, 1982.

[FvSB03] Felix C. A. Fernandes, Rutger L. C. van Spaendonck, and C. Sidney Burrus. A New Framework for Complex Wavelet Transforms. *IEEE Transactions on Signal Processing*, 51, no 7:1825–1837, 2003.

[Gab46] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946.

[Gao98] H.Y. Gao. Wavelet shrinkage denoising using the nonnegative garrote. *Journal of Computational and Graphical Statistics*, 7, no. 4: 469 – 488, 1998.

[GD06] D. Gleich and M. Datcu. Gauss-Markov Model for Wavelet-Based SAR Image Despeckling. *IEEE Signal Processing Letters*, 13, no. 6: 365 – 368, 2006.

[GJ97] L. Gagnon and A. Jouan. Speckle Filtering of SAR Images - A Comparative Study Between Complex-Wavelet-Based and Standard Filters. In *SPIE Proc., conference "Wavelet Applications in Signal and Image Processing V", San Diego*, 1997.

[GLG] L. Gagnon, J. M. Lina, and B. Goulard. Sharpening Enhancement of Digitized Mammograms with Complex Symmetric Daubechies Wavelets.

[GM84] A. Grossman and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal*, 15: 723 – 736, 1984.

[GSZ06] Guy Gilboa, Nir Sochen, and Yehoshua Y. Zeevi. Variational Denoising of Partly Textured Images by Spatially Varying Constraints. *IEEE Transactions on Image Processing*, 15, no. 8: 2281 – 2289, 2006.

[Hig84] J. R. Higgins. Bases for the Hilbert space of Paley-Wiener functions. In *Proc. 5 Aachener Kolloqium, Mathematische Methoden in der Signalverarbeitung*, 1984.

[HKMMT89] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. *Wavelets. Time-frequency Methods and Phase Space*, chapter A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform, pages 289–297. Springer-Verlag, Berlin, 1989.

[ICN02] Alexandru Isar, Andrei Cubitchi, and Miranda Nafornita. *Algorithmes et techniques de compression*. Editura ORIZONTURI UNIVERSITARE, Timisoara, 2002.

[IIA06] Alexandru Isar, Dorina Isar, and Ioana Adam. Denoising Sonar Images. In *Proceedings of the Romanian Academy*, volume 7, no. 2 of *A*, pages 1 – 14, 2006.

[IIM+05] A. Isar, D. Isar, S. Moga, J.-M. Augustin, and X. Lurton. Multi-scale MAP Despeckling of SONAR Images. In *Proceedings of IEEE International Conference Oceans'05*, pages 1292 – 1297, 2005.

[IIQ07] A. Isar, D. Isar, and A. Quinquis. Multi-Scale MAP Denoising of SAR and SAS Images. *Sea Technology*, 48 (2): 46 – 48, 2007.

[IMI09] Alexandru Isar, Sorin Moga, and Dorina Isar. A new denoising system for sonar images. *EURASIP Journal on Image and Video Processing*, vol. 2009:14 pages, 2009.

[IMN+06] Al. Isar, S. Moga, C. Nafornita, M. Oltean, and I. Adam. Image Denoising Using Wavelet Transforms With Enhanced Diversity. In *Proceedings of International Conference Communications 2006, Bucharest*, 2006.

[Isa93] Al. Isar. Nouvelles modalités de décomposition multiresolution des signaux à energie finie. In *Quatorzieme Colloque GRETSI*, pages 93–97, 1993.

[JBFZ03] A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia. Natural image modeling using complex wavelets. In *Proc. of SPIE, Wavelets X*, 2003.

[Kin98] N G Kingsbury. The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement. In *Proc. European Signal Processing Conference, EUSIPCO 98, Rhodes*, pages 319–322, 1998.

[Kin99] Nick Kingsbury. Image processing with complex wavelets. *Philosophical Transactions of the Royal Society London A*, 357 no. 1760:2543–2560, 1999.

[Kin00] N. G. Kingsbury. A dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In *Proc. IEEE Conference on Image Processing*, pages 375 – 378, 2000.

[Kin01] N.G. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Journal of Applied and Computational Harmonic Analysis*, 10, no.3:234–253, 2001.

[KM97] N. G. Kingsbury and J. F. A. Magarey. Wavelet transforms in image processing. In *Proc. First Eur. Conf. Signal Anal. Prediction*, pages 23–34, 1997.

[KSSC87] D.T. Kuan, A.A. Sawchuk, T.C. Strand, and P. Chavel. Adaptive restoration of images with speckle. *IEEE Transactions on Acoustic, Speech , Signal Processing*, 35, no. 3: 373 – 383, 1987.

[KV92] J. Kovacevic and M. Vetterli. Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for R. *IEEE Transactions On Information Theory*, 38(2): 533 – 555, 1992.

[Law93] Wayne Lawton. Applications of Complex Valued Wavelet Transforms to Subband Decomposition. *IEEE Transactions on Signal Processing*, VOL. 41. NO. 12:3566–3568, 1993.

[LBU07] F. Luisier, T. Blu, and M. Unser. A New SURE Approach to Image Denoising: Inter-Scale Orthonormal Wavelet Thresholding. *IEEE Transactions on Image Processing*, 16, no. 3: 593 – 606, 2007.

[Lee81] J. S. Lee. Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, 17: 24 – 32, 1981.

[LG95] J.-M. Lina and L. Gagnon. Image enhancement with symmetric Daubechies wavelets. In *Wavelet Applications in Signal and Image Processing III*, 1995.

[LGO+96] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and Jr. R. O.Wells. Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters*, 3, no. 1:10–12, 1996.

[LM95] J.-M. Lina and M. Mayrand. Complex Daubechies Wavelets. *Applied and Computational Harmonic Analysis*, 2:219–229, 1995.

[Mal89] Stephane G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 11, no. 7:674–693, 1989.

[Mal99] Stephane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[MS05] Mona Mahmoudi and Guillermo Sapiro. Fast Image and Video Denoising via Nonlocal Means of Similar Neighborhoods. *IEEE Signal Processing Letters*, 12, no. 12: 839 – 842, 2005.

[Naf08] Corina Nafornita. *Contributii la marcarea transparenta a imaginilor in domeniul transformatei wavelet (Contributions to Image watermarking in the wavelet domain)*. PhD thesis, Universitatea Tehnica Cluj and Universitatea "Politehnica" Timisoara, 2008.

[NII+10] C. Nafornita, I.Firoiu, D. Isar, J.-M. Boucher, and Al. Isar. A Second Order Statistical Analysis of the 2D Discrete Wavelet Transform. In *Proceedings of International Conference Communications 2010, Bucharest*, 2010.

[Olh07] S.C Olhede. Hyperanalytic denoising. *IEEE Transactions on Image Processing*, 16 (6): 1522 – 1537, 2007.

[Olt09] Marius Oltean. *Contributii la optimizarea transmisiei pe canale radio, folosind functii wavelet*. PhD thesis, Universitatea "Politehnica" Timisoara, 2009.

[OM06] S. C. Olhede and Georgios Metikas. The Hyperanalytic Wavelet Transform. Technical report, Imperial College Statistics Section, 2006.

[PP06] A. Pizurica and W. Philips. Estimating the probability of the presence of a signal of interest in multiresolution single and multiband image denoising. *IEEE Transactions on Image Processing*, 15, no. 3: 654 – 665, 2006.

[PSWS03] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using Gaussian scale mixtures in the wavelet domain. *IEEE Transactions on Image Processing*, 12, no. 11: 1338 – 1351, 2003.

[RC92] M.V. Wickerhauser R.R. Coifman, Y. Meyer. *Wavelets and Their Applications*, chapter Wavelet analysis and signal processing, pages 153–178. Jones and Bartlett, Boston, MA, 1992.

[ROF92] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60: 259 – 268, 1992.

[SB86] M.J.T. Smith and T.P. Barnwell. Exact reconstruction for tree-structured subband coders. *IEEE Trans Acoust Speech Signal Processing*, 34: 434 – 441, 1986.

[SBK05] Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury. The Dual-Tree Complex Wavelet Transform. *IEEE Signal Processing Magazine*, 22(6):123–151, 2005.

[SE04] Stian Solbo and Torbjorn Eltoft. Homomorphic Wavelet-Based Statistical Despeckling of SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 42, no. 4: 711 – 721, 2004.

[Sel01] Ivan W. Selesnick. The Characterization and Design of Hilbert Transform Pairs of Wavelet Bases. In *2001 Conference on Information Science and Systems, The Johns Hopkins University*, 2001.

[Sel04] Ivan W. Selesnick. The Double-Density Dual-Tree DWT. *IEEE Transactions on Signal Processing*, 52, No. 5:1304–1314, 2004.

[Sel08] Ivan W. Selesnick. The Estimation of Laplace Random Vector in AWGN and the Generalized Incomplete Gamma Function. *IEEE Transactions on Signal Processing*, 56 (8): 3482 – 3496, 2008.

[SFAH92] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms. *IEEE Transactions on Information Theory*, 38: 587 – 607, 1992.

[Shu03] Panchamkumar D Shukla. Complex wavelet transforms and their applications. Master's thesis, University of Strathclyde, Glasgow, UK, 2003.

[Shu05] P.-L. Shui. Image Denoising Algorithm via Doubly Local Wiener Filtering With Directional Windows in Wavelet Domain. *IEEE Signal Processing Letters*, 12, No. 6: 681 – 684, 2005.

[SI10] M. Salagean and I.Firoiu. Anomaly Detection on Network Traffic Based on Analytical Discrete Wavelet Transform. In *Proceedings of International Conference Communications 2010, Bucharest*, 2010.

[SS02] L. Sendur and I. W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing*, 50, no. 11: 2744 – 2756, 2002.

[Tou02] Ridha Touzi. A Review of Speckle Filtering in the Context of Estimation Theory. *IEEE Transactions on Geoscience and Remote Sensing*, 40, no. 11: 2392 – 2404, 2002.

[UKBW86] F. T. Ulaby, F. Kouyate, B. Brisco, and T. H. L. Williams. Textural information in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24: 235 – 245, 1986.

[VK95] M. Vetterli and J. Kovacevic. *Wavelets and subband coding*. Prentice Hall, Englewood Cliffs, NJ, USA, 1995.

[VU08] Dimitri Van De Ville and Michael Unser. Complex Wavelet Bases, Steerability, and the Marr-Like Pyramid. *IEEE Transactions on Image Processing*, 17(11):2063–2080, 2008.

[WD00] M. Walessa and M. Datcu. Model-Based Despeckling and Information Extraction from SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 38, no. 5: 2258 – 2269, 2000.

[XPZP96] Li-Sheng Tian Xiao-Ping Zhang and Ying-Ning Peng. From the wavelet series to the discrete wavelet transform. *IEEE Transactions On Signal Processing*, 44, No. 1:129–133, 1996.

[XXM08] Shuai Xing, Qing Xu, and Dongyang Ma. Speckle Denoising Based on Bivariate Shrinkage Functions and Dual-Tree Complex Wavelet Transform. ISPRS, 2008.

[You93] Randy K. Young. *Wavelet theory and its applications*. Kluwer Academic Publishers, 1993.

[ZS07] Z.-F. Zhou and P.-L. Shui. Contourlet-based image denoising algorithm using directional windows. *Electronic Letters*, 43, no.2: 92 – 93, 2007.