

Image Denoising Using Wavelets and Spatial Context Modeling

Aleksandra Pizurica

Promotoren: Prof. dr. ir. W. Philips, Prof. dr. ir. M. Achteroy

Proefschrift ingediend tot het behalen van de graad van
Doctor in de Toegepaste Wetenschappen

Vakgroep Telecommunicatie en Informatieverwerking
Voorzitter: Prof. dr. ir. H. Bruneel
Academiejaar 2001-2002



Preface

I would like to thank my promotor, Prof. Wilfried Philips and Prof. Marc Acheroy, for following my research with interest and for giving me constant support. I am thankful to Prof. Philips for his many suggestions and positive criticism during my research. Prof. Acheroy gave me a great encouragement and the opportunity to take part in many international conferences.

I am grateful to Prof. Ignace Lemahieu for expressing interest in my work and for giving me support through all these years of my research.

I also thank to all my TELIN colleagues and ex-colleagues for a friendly atmosphere which makes work pleasant. Annette Nevejans and Alice Verheylesonne gave me often precious administrative help. Here I also want to mention the MEDISIP group and especially the TELIN-MEDISIP meetings which inspire research through constructive discussions.

I am grateful to my parents for their trust, support and love. Finally, I thank my husband Veselin and my daughter Marija for their patience and support that I could always count on.

This research was funded by the Belgian Ministry of defense in the scope of the humanitarian mine detection project HUDEM.

Aleksandra Pizurica
June, 2002.

Principal Symbols

\oplus	: direct sum of two subspaces
\propto	: is proportional to
C	: clique in a neighboring system
∂l	: set of neighboring sites of a site l
\mathbf{f}, f_l	: noise free input image
\mathbf{g}, g_l	: highpass filter in filter bank
\mathbf{h}, h_l	: lowpass filter in filter bank
H_l^0	: hypothesis “signal is absent at position l ”
H_l^1	: hypothesis “signal is present at position l ”
\mathbf{M}	: measure of significance in Bayesian model
n	: number of data points in a discrete finite signal
\mathbf{n}, n_l	: noise in the wavelet domain
$p_Y(y)$: probability density function of random variable Y
$s_{j,l}$: scaling coefficient (level j , position l)
\mathbf{v}, v_l	: noisy input image
\mathbf{w}, w_l	: noisy wavelet coefficients
$w_{j,l}$: noisy wavelet coefficient at level j , position l
z_l	: local spatial activity indicator at position l
x_l	: classification labels (binary)
\mathbf{X}	: vector of classification labels (mask) in Bayesian model
\mathbf{y}, y_l	: noise-free wavelet coefficients
α, α_l	: local Lipschitz exponent
ξ_l	: ratio of conditional pdf's at position l
μ_l	: ratio of prior probabilities of signal presence position l
ρ	: ratio of global prior probabilities of signal presence
ϑ, ϑ_l	: input zero mean noise
\mathbb{R}	: set of real numbers
\mathbb{Z}	: set of integer numbers
$\mathbf{L}^2(\mathbb{R})$: finite energy functions $\int_{-\infty}^{\infty} f(x) ^2 dx < \infty$
g^*	: complex conjugate of g
$\langle f, g \rangle$: inner product for continuous signals $\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g^*(x)dx$ for discrete signals $\langle f, g \rangle = \sum_{l=-\infty}^{\infty} f_l g_l^*$
$\ f\ $: L_2 norm of f ; $\ f\ ^2 = \langle f, f \rangle$
$\lfloor x \rfloor$: integer part of a real number x

Abbreviations

1D	:	one dimensional
2D	:	two dimensional
AWGN	:	Additive white Gaussian noise
ACR	:	Average cone ratio
APR	:	Average point ratio
COI	:	Cone of influence
CWT	:	Continuous wavelet transform
DCT	:	Discrete cosine transform
DCOI	:	Discrete cone of influence
DFT	:	Discrete Fourier transform
DWT	:	Discrete wavelet transform
FN	:	False negatives
FP	:	False positives
GGD	:	Generalized Gaussian distribution
GPR	:	Ground penetrating radar
i.i.d	:	independent identically distributed
IR	:	Infrared
JC	:	Joint conditional
JDE	:	Joint detection estimation
LSAI	:	Local spatial activity indicator
MAD	:	Median absolute deviation
MAP	:	Maximum a posteriori
MLE	:	Maximum likelihood estimation
MMSE	:	Minimum mean squared error
MRA	:	Multiresolution analysis
MRI	:	Magnetic resonance imaging
MRF	:	Markov random field
MSE	:	Mean squared error
PC	:	Principal component
pdf	:	Probability density function
PSNR	:	Peak signal-to-noise ratio
SAR	:	Synthetic aperture radar
SNR	:	Signal-to-noise ratio
STSA	:	Short time spectral amplitude
SURE	:	Stein's unbiased risk estimate
QCF	:	Quadratic cost function

Contents

1	Introduction	1
1.1	Situation and topical outline	1
1.1.1	Organization	2
1.1.2	A summary of contributions and publications	4
2	Wavelets in Image Denoising	5
2.1	Introduction to wavelet representation	5
2.1.1	The wavelet concept and its origins	5
2.1.2	Continuous and dyadic wavelet transforms	7
2.1.3	What makes wavelets useful in signal processing . .	8
2.2	Discrete Wavelet Transform	10
2.2.1	A note on wavelet frames and bases	10
2.2.2	Multiresolution Analysis	12
2.2.3	Wavelets and discrete filters	14
2.2.4	DWT - A fast discrete wavelet transform algorithm	15
2.2.5	DWT in two dimensions	16
2.2.6	Improving the limitations of DWT in denoising . .	17
2.2.7	Non-decimated discrete wavelet transform	18
2.2.8	Choosing a wavelet for image denoising	20
2.2.9	Multiscale edge detection	22
2.2.10	On some extensions of the classical wavelet scheme	23
2.3	Wavelet Domain Image Denoising	24
2.3.1	Noise model and Notation	25
2.3.2	Noise in the wavelet domain	26
2.3.3	Noise variance estimation	27
2.3.4	Performance evaluation in image denoising	27
2.3.5	Ideal coefficient attenuation and ideal selection .	28
2.4	Denoising by wavelet thresholding	29
2.4.1	Hard and soft thresholding	30
2.4.2	Threshold selection	30

2.5	Denoising by singularity detection	32
2.5.1	Lipschitz exponents and the cone of influence	33
2.5.2	Reconstruction from multiscale edges	34
2.5.3	Regularity estimates for coefficient selection	36
2.5.4	Interscale correlations for coefficient selection	38
2.6	Wavelet domain Bayes estimation	39
2.6.1	Bayes risk, MAP and MMSE estimates	39
2.6.2	Prior distributions of image wavelet coefficients . .	41
2.6.3	Simple Bayesian shrinkers	43
2.6.4	Locally adaptive wavelet domain Wiener filters . .	45
2.6.5	MRF and HMT approaches	47
2.7	Summary	49
3	MRF priors in wavelet domain denoising	53
3.1	Introduction	53
3.2	Markov Random Fields	55
3.2.1	Notation and definitions	56
3.2.2	Gibbs distribution	57
3.2.3	Common MRF models	59
3.3	The MAP-MRF approach	60
3.3.1	Joint MAP coefficient estimation	61
3.3.2	MAP Mask estimation	63
3.3.3	Optimization algorithms	64
3.4	Bayesian shrinkage with MRF priors	66
3.4.1	Prior and conditional models	66
3.4.2	Stochastic sampling method of Malfait and Roose	69
3.4.3	A practical example	71
3.5	An alternative estimation approach	73
3.5.1	Motivation and idea	73
3.5.2	Formal description and algorithm	74
3.5.3	A heuristic shrinkage family	75
3.5.4	An application in hydrology	78
3.6	Summary and conclusions	81
4	Statistical Modeling in MRF based wavelet denoising	83
4.1	Introduction	83
4.2	Significance measures	86
4.2.1	A discretized approximation of interscale ratios .	86
4.2.2	Statistical characterization via simulation	89
4.2.3	Performance Evaluation	93

4.2.4	A joint significance measure	98
4.2.5	Implementation details	100
4.3	A New MRF Prior Model	102
4.3.1	A class of anisotropic potentials	102
4.4	Practical algorithm and its performance	103
4.4.1	An upgraded stochastic sampling algorithm	104
4.4.2	The performance of the joint conditional model .	105
4.4.3	The performance of the prior model	107
4.4.4	3-subband versus 2-subband implementation . . .	110
4.4.5	Quantitative results and discussion	112
4.5	Summary and Conclusions	113
5	A Generalized Likelihood Ratio in Denoising	115
5.1	Introduction	115
5.2	Joint detection and estimation (JDE)	117
5.2.1	General principles	117
5.2.2	JDE under statistically independent observations .	120
5.3	A wavelet domain GenLik approach	121
5.3.1	Notation and model assumptions	122
5.3.2	Spatial adaptation	124
5.3.3	Global prior probability ratio	126
5.3.4	Estimation of conditional densities	127
5.3.5	Signal of interest and performance evaluation .	131
5.3.6	Algorithm overview	134
5.3.7	Results and Discussion	134
5.4	An empirical GenLik Approach	148
5.4.1	The main idea and the global concept	149
5.5	A versatile algorithm for various noise types	153
5.5.1	The practical algorithm	155
5.5.2	On medical applications	156
5.5.3	Application to medical ultrasound images	157
5.5.4	The algorithm adapted for medical MRI images .	161
5.5.5	Application in image deblurring	166
5.5.6	Discussion	168
5.6	SAR image despeckling	169
5.6.1	Speckle noise in SAR images	169
5.6.2	Filtering SAR images	170
5.6.3	A new family of wavelet shrinkage functions . . .	174
5.6.4	A practical SAR despeckling algorithm	177
5.6.5	Results and Discussion	180

5.7	Summary and conclusions	187
6	Application to humanitarian demining	189
6.1	Introduction	189
6.2	Humanitarian demining technologies	191
6.2.1	Sensor technologies in humanitarian demining . .	191
6.2.2	A practical research program	195
6.3	Infrared image processing	196
6.3.1	Infrared images of landmines	197
6.3.2	A new nonlinear wavelet denoising technique .	200
6.3.3	Wavelet denoising with spatial priors	203
6.4	Processing of GPR images	204
6.4.1	GPR images of landmines	205
6.4.2	A nonlinear pyramidal filtering technique . . .	208
6.4.3	Application of wavelet denoising with spatial priors	210
6.5	Summary and conclusions	212
7	Conclusions	215
A	Deriving the global prior probability ratio	219
B	MLE parameter estimation	221
B.1	Exponential distribution	222
B.2	Generalized Gamma distribution	223

Chapter 1

Introduction

Owing to its rapidly increasing popularity over last few decades, the wavelet transform has become quite a standard tool in numerous research and application domains. This thesis is about wavelet domain image denoising: we study and develop statistical models and estimators for image wavelet coefficients given their noisy observations. In doing so, we are on a bridge between theory and applications. While merging theory and practice, from time to time we employ heuristics too.

1.1 Situation and topical outline

This thesis has started within the context of a project on humanitarian mine detection. To enhance the quality of images of landmines acquired by different sensors advanced denoising methods are required. Given a great variety of sensor technologies in mine detection, and given that the same technologies appear in other application domains, we decided not to limit our research to any particular type of images.

In general, image denoising imposes a compromise between noise reduction and preserving significant image details. To achieve a good performance in this respect, a denoising algorithm has to adapt to image discontinuities. The wavelet representation naturally facilitates the construction of such spatially adaptive algorithms. It compresses the essential information in a signal into relatively few, large coefficients, which represent image details at different resolution scales.

In this thesis, we further develop some emerging wavelet domain denoising methods. Also, we propose new denoising approaches. Throughout the work, we emphasize the use of spatial context and accurate

statistical modeling. For spatial context modeling, we consider two approaches: (1) encoding pixel interactions in a *Markov Random Field* prior model and (2) a lower complexity local approach, which computes a *local spatial activity indicator* whose marginal densities are characterized. In practical applications, we will often simplify the theory using heuristics, when this leads to algorithms with lower complexity or higher flexibility. Also, we do not propose a universal denoising method, but rather build a library of algorithms suitable for a diversity of applications in close range and remote sensing and in medical imaging.

A common underlying principle in our work is uncertainty as to the signal presence. In other words, in our developments we always start from a viewpoint that a signal (of interest) is not surely known to be present in a given wavelet coefficient. This concept is implicitly present in many wavelet denoising algorithms but is rarely explicitly formulated as such. We devote attention to some interesting questions that arise from this concept: the notion of a “signal of interest” to be reconstructed from a noisy image; confronting a classical estimation concept (signal is present with probability $p = 1$) and a joint detection and estimation concept (signal is present with probability $p < 1$). In this respect, we also point to close relationship between wavelet domain image denoising and seemingly unrelated problems like spectral estimation of speech signals.

The algorithms developed in this thesis fit into a Bayesian framework illustrated in Fig. 1.1. The main novelties and contributions can be grouped in the following categories: selection, statistical characterization and performance evaluation of *significance measures* for wavelet coefficients in the presence of noise; developing new *MRF prior models for spatial clustering* of the coefficients and developing a new class of *low complexity locally adaptive* denoising methods based on joint detection and estimation principles. Practical results of new algorithms are demonstrated in different applications: in infrared imaging, Ground Penetrating Radar (GPR) imaging, Synthetic Aperture Radar (SAR) imaging, medical ultrasound and magnetic resonance imaging (MRI).

1.1.1 Organization

The thesis is organized as follows. In Chapter 2, we first review the background knowledge on wavelet theory. We try to present a concise, but self-contained review emphasizing the most relevant aspects for the topic of this thesis. In the second part of this Chapter, we systematically describe different wavelet denoising approaches and we link them to our

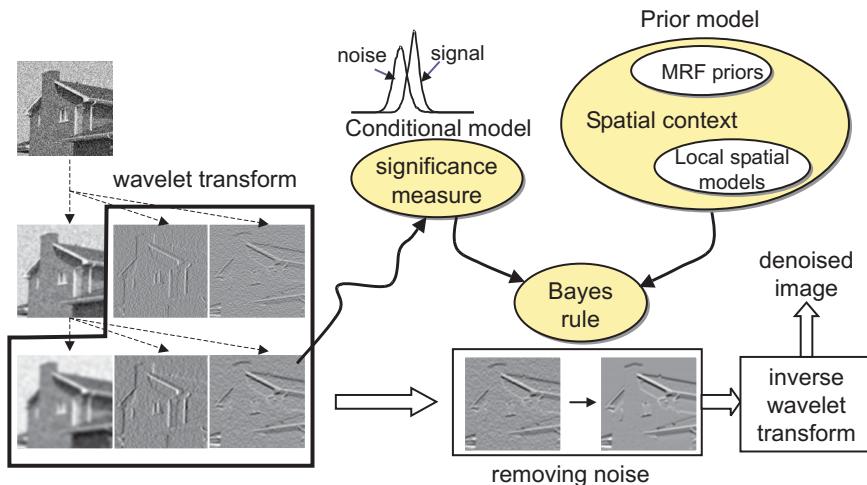


Figure 1.1: A Bayesian wavelet domain denoising framework.

later developments.

In Chapter 3, we explain the use of Markov Random Field models in wavelet domain denoising. The existing wavelet domain approaches from literature use bi-level MRF's as prior models for spatial clustering. With each wavelet coefficient, a given measure of its significance is assigned; the statistical properties of these significance measures are expressed in a conditional probability density model, and are combined with the prior model in a Bayesian framework. We discuss these methods as a special case of a more general framework. An important original contribution is a new MRF based denoising method, which aims at a reduced complexity. We demonstrate its use in hydrology.

Chapter 4, presents our main contributions to wavelet denoising using MRF priors. We statistically characterize different significance measures, for which valid probability density models were not yet available. Then, relying on decision theory, we objectively evaluate the performance of the analyzed significance measures. Accordingly, we propose a new and powerful conditional model. To improve encoding of spatial clustering properties, we develop a new MRF prior model. The results demonstrate the advantage of the new prior and conditional models both in terms of quantitative and visual aspects of image quality.

In Chapter 5, we turn to locally adaptive denoising concepts that do not require MRFs. Instead of modeling the spatial context by MRFs, we model it by much simpler techniques. In this respect, we develop a new class of low complexity locally adaptive methods. Within the new framework, we develop three new practical algorithms: the first one is designed for the suppression of Gaussian noise and achieves denoising results that are among the best state of the art ones. The second algorithm is flexible, designed to allow user interaction and also to adapt to various noise types. The third algorithm is aimed at speckle noise suppression in SAR images. We focus on SAR systems because they are in wide use in remote sensing and have applications in humanitarian mine detection as well.

Chapter 6 is devoted specifically to close range sensors in humanitarian demining applications. We describe a practical mine detection research program Hudem, and illustrate denoising results on infrared and on GPR images of landmines.

1.1.2 A summary of contributions and publications

This research has resulted so far in a publication in an international journal [Pizurica02a], a book chapter [Pizurica01c], an invited talk in a wavelet seminar [Pizurica00e] and 14 papers in Conference Proceedings [Pizurica98a]-[Pizurica01b], all of which as the first author. Also, this research contributed to 7 publications [Duskunovic00a, Duskunovic00b, Duskunovic00c, Rooms01, Stippel00, Verhoest00a, Xue01] through a co-authorship.

Most recent research results, presented in Chapter 5, are submitted to two international journals [Pizurica02b] and [Pizurica02c].

Chapter 2

Wavelets in Image Denoising

This chapter starts with an overview of some basic wavelet concepts. These can be found in many books and papers at many different levels of exposition. Some of the standard books are [Chui92, Daubechies92, Mallat98, Meyer93, Vetterli95]. Introductory papers include [Graps95, Strang94, Vidakovic94], and more technical ones are [Cohen96, Mallat96, Strang95]. The rest of the chapter presents ideas of various wavelet based image denoising methods and reviews the state of the art in this field.

2.1 Introduction to wavelet representation

2.1.1 The wavelet concept and its origins

The central idea to wavelets is to analyze (a signal) according to *scale*. Imagine a function that oscillates like a wave in a limited portion of time or space and vanishes outside of it. The wavelets are such functions: wave-like but localized. One chooses a particular wavelet, stretches it (to meet a given scale) and shifts it, while looking into its correlations with the analyzed signal. This analysis is similar to observing the displayed signal (e.g., printed or shown on the screen) from various distances. The signal correlations with wavelets stretched to large scales reveal gross (“rude”) features, while at small scales fine signal structures are discovered. It is therefore often said that the wavelet analysis is to see both the forest *and* the trees.

In such a scanning through a signal, the scale and the position can vary continuously or in discrete steps. The latter case is of practical interest in this thesis. From an engineering point of view, the discrete wavelet analysis is a two channel digital filter bank (composed of the

lowpass and the highpass filters), iterated on the lowpass output. The lowpass filtering yields an *approximation* of a signal (at a given scale), while the highpass (more precisely, bandpass) filtering yields the *details* that constitute the difference between the two successive approximations. A family of wavelets is then associated with the bandpass, and a family of scaling functions with the lowpass filters. This concept is explained in Section 2.2. At the moment, we address only the general aspects.

The wavelet family is generated from a unique prototype function that is called a *mother wavelet*. Given a real variable x , the function $\psi(x)$ is called a mother wavelet provided that it oscillates, averaging to zero $\int_{-\infty}^{\infty} \psi(x)dx = 0$ and that is well localized (i.e., rapidly decreases to zero when $|x|$ tends to infinity). By convention it is centered around $x = 0$, and has a unit norm $\|\psi(x)\|$. In practice, applications impose additional requirements (Sec.2.2.8), among which, a given number of *vanishing moments* N_v

$$\int_{-\infty}^{\infty} x^k \psi(x)dx = 0, \quad 0 < k \leq N_v - 1. \quad (2.1.1)$$

The mother wavelet $\psi(x)$, generates the other wavelets $\psi_{a,b}(x)$, $a > 0$, $b \in \mathbb{R}$, of the family by change of scale a (i.e., by dilation) and by change of position b (i.e., by translation),

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right), \quad a > 0, \quad b \in \mathbb{R}. \quad (2.1.2)$$

In Fig. 2.1, several wavelets are shown that are obtained from the mother wavelet $\psi(x) = (1 - 2x^2)e^{-x^2}$; this wavelet is the second derivative of a Gaussian function and is called the *Mexican hat*. Its first use was in computer vision, for multiscale edge detection [Witkin83].

The origins of the wavelet analysis can be traced to the 1909 Haar wavelet (that was not called by that name then) and various “atomic decompositions” in the history of mathematics. For a comprehensive review, see [Meyer93, p.13-31]. The current use of the name “wavelet” is due to Grossman’s and Morlet’s work on geophysical signal processing¹, which led to the formalization of the *continuous wavelet transform* [Grossman84]. In the development of wavelets, the ideas from

¹In reflection seismology, the modulated pulses sent underground need to have a short duration at high frequencies in order to separate the returns of fine, closely spaced layers; the Morlet’s idea was thus to send shorter waves at high frequencies simply by scaling a single function, called the “wavelet of continuous shape”

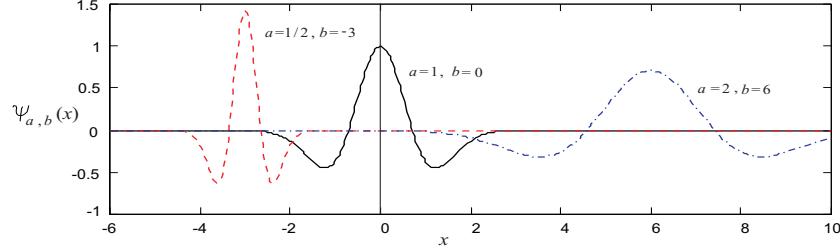


Figure 2.1: A few wavelets obtained from the Mexican hat mother wavelet $\psi(x) = (1 - 2x^2)e^{-x^2}$.

many different fields (including subband coding and computer vision) have merged. Excellent texts on this topic are [Vetterli95, p.1-3] and Daubechies paper “Where do the wavelets come from” [Daubechies96].

2.1.2 Continuous and dyadic wavelet transforms

The continuous wavelet transform (CWT) of a signal $f(x)$ is defined as

$$\mathcal{W}f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x) \psi^* \left(\frac{x-b}{a} \right) dx = \langle f, \psi_{a,b} \rangle, \quad (2.1.3)$$

where $\psi^*(x)$ denotes the complex conjugate of $\psi(x)$. The existence of the inverse transform is guaranteed if $\int_{-\infty}^{\infty} |\hat{\psi}(\omega)|^2 / |\omega| d\omega \triangleq C_\psi < +\infty$, where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(x)$. This is called the *admissibility condition* [Mallat98]. It implies $\hat{\psi}(0) = 0$, and thus $\psi(x)$ can be viewed as an impulse response of a bandpass filter. Obviously, the CWT offers a great degree of freedom in the choice of a wavelet. The inverse transform is defined as

$$f(x) = \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{W}f(a, b) \psi_{a,b}(x) da db / a^2 \right) / C_\psi. \quad (2.1.4)$$

The CWT is highly redundant, and is *shift invariant*. It is extensively used for the characterization of signals [Mallat92a]: the evolution of the CWT magnitude across scales provides information about the *local regularity* of a signal. We return to this aspect in Section 2.5. In those and similar applications of the CWT, for the sake of memory savings, dyadic scales $a = 2^j |_{j \in \mathbb{Z}}$ are commonly used. The corresponding transform $\mathcal{W}f(2^j, b)$ is called the *dyadic wavelet transform*.

In case of images, one can use an arbitrary number $N \geq 1$ of (spatially oriented) wavelets $\psi_{2^j, u, v}^n(x, y) = \psi^n(2^{-j}(x-u, y-v))$

$v)$, $1 \leq n \leq N$, which yield the N -component transform $(\mathcal{W}^1 f(2^j, u, v), \dots, \mathcal{W}^N f(2^j, u, v))$, where

$$\begin{aligned}\mathcal{W}^n f(2^j, u, v) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \psi^{n*} \left(\frac{x-u}{2^j}, \frac{y-v}{2^j} \right) du dv \\ &= \langle f, \psi_{2^j, u, v}^n \rangle\end{aligned}\quad (2.1.5)$$

The conditions, which $\psi_{2^j, u, v}^n(x, y)$, $1 \leq n \leq N$ have to meet for a complete and a stable representation of an image, are presented in [Mallat98, p.160]. The dyadic wavelet transform with spatially oriented wavelets is extensively used in texture processing [Jain91]. In Section 2.2.9, we address another particular type and application of this transform.

2.1.3 What makes wavelets useful in signal processing

In signal processing, the representation of signals plays a fundamental role. David Marr elaborated in [Marr82] on this topic. For example, the Arabic numeral representation permits one to easily notice a power of 10, but more difficult to realize a power of 2. With the binary representation, the situation is reversed. Meyer [Meyer93, p.10], wrote “any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover”.

The Fourier representation reveals the spectral content of a signal, but makes it impossible to recover the particular moment in time (or the particular space coordinates in case of images) where a certain change has occurred. This makes the Fourier representation inadequate when it comes to analyzing *transient* signals. In signal and image processing, concentrating on transients (like, e.g., image discontinuities) is a strategy for selecting the most essential information from often an overwhelming amount of data. In order to facilitate the analysis of transient signals, i.e., to localize both the frequency and the time information in a signal, numerous transforms and bases have been proposed (see e.g., [Mallat98, Vetterli95]). Among those, in signal processing the wavelet and the short time Fourier Transform (STFT) are quite standard. Let us briefly compare the two.

In the STFT transform (which is also called the window Fourier transform or the Gabor transform) the signal is multiplied by a smooth window function (typically Gaussian) and the Fourier integral is applied

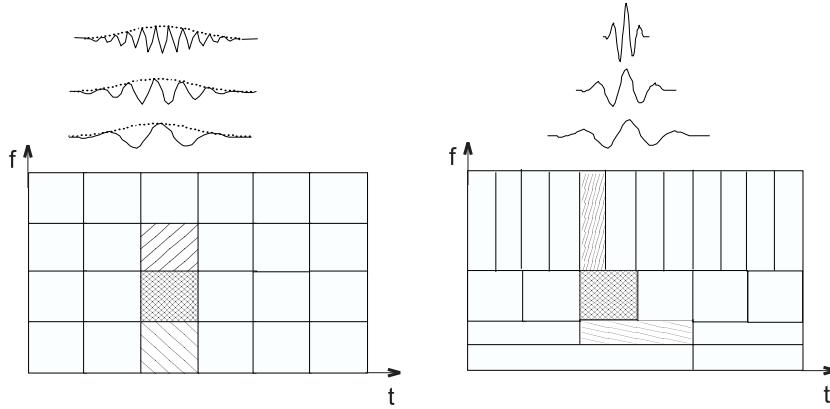


Figure 2.2: Basis functions and the time-frequency tillings of the short time Fourier (a) and the wavelet transform (b) [Vetterli95].

to the windowed signal. For a signal $f(x)$, the STFT is [Mallat98]

$$\mathcal{S}(\tau, \omega) = \int_{-\infty}^{\infty} f(x)g(x - \tau)e^{-j\omega x}dx, \quad (2.1.6)$$

where $g(x)$ is the window function. Note that the basis functions of a STFT expansion are $g(x)$ modulated by a sinusoidal wave and shifted in time; the modulation frequency is changing while the *window remains fixed*. A few of these functions and the corresponding tillings [Nielsen96] of the time-frequency plane are illustrated in Fig. 2.2 (a).

In wavelet analysis, the scale can be interpreted as the inverse of frequency. The corresponding tilling of the time-frequency plane is illustrated in Fig. 2.2 (b). As opposed to STFT, which divides the time-frequency plane into equal blocks, the wavelet transform acts as a microscope [Vetterli95] focusing on smaller time phenomena as the scale decreases. This behaviour permits a *local characterization* of signals, which the Fourier and the window Fourier transform does not.

Other main advantageous properties of the wavelet transform (to be clarified in the following Sections) are:

- *Multiresolution* - a scale invariant representation;
- *Edge detection* - large wavelet coefficients correspond to image edges;
- *Sparsity* - the wavelet transform of natural images tends to be sparse;
- *Fast algorithms* - the complexity of the fast discrete wavelet transform is a linear function of the number of the input samples.

Besides, the wavelet analysis has been also often motivated from

the viewpoint of psychophysical aspects of human vision [Mallat89a]: it corresponds well to the way how we perceive images.

2.2 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is in literature commonly associated with signal expansion into (bi-)orthogonal *wavelet bases*. We shall adopt the same convention in this thesis. Thus, as opposed to the highly redundant CWT, there is no redundancy in the DWT of a signal; the scale is sampled at dyadic steps $a \in \{2^j : j \in \mathbb{Z}\}$, and the position is sampled proportionally to the scale $b \in \{k2^j : (j, k) \in \mathbb{Z}^2\}$.

By no means can a DWT be understood as a simple sampling from a CWT. In the first place, the choice of a wavelet is now far more restrictive: if we are dealing with finite-energy signals $f(x) \in \mathbf{L}^2(\mathbb{R})$, the wavelet $\psi(x)$ has to be chosen such that $\{\psi(2^{-j}(x - 2^j k))\}_{(j,k)\in\mathbb{Z}^2}$ is a basis of $\mathbf{L}^2(\mathbb{R})$. The first such basis was constructed by Alfred Haar in 1909 (Section 2.2.1), and the choice for better ones has culminated in Ingrid Daubechies's work [Daubechies88]. The systematic framework for constructing wavelet bases, known as the *multiresolution analysis* (Section 2.2.2), was mostly developed by Stéphane Mallat [Mallat89b], and has merged the ideas of pyramidal algorithms [Burt83] in computer vision and the filter banks for subband coding [Esteban77]. Books like [Chui92, Daubechies92, Mallat98, Meyer93, Vetterli95] provide a comprehensive treatment of these topics. A particularly comprehensive filter bank point of view is [Strang96].

The orthogonal wavelets are rarely available as closed form expressions, but rather obtained through a computational procedure which uses discrete filters (Section 2.2.3). The link between wavelets and these discrete filters is essential for understanding the Mallat's fast DWT algorithm in Section 2.2.4 and its extension to images in Section 2.2.4. The rest of the Sections address the non-decimated transforms and aspects that are especially important for image denoising.

2.2.1 A note on wavelet frames and bases

We start with a brief reminder of the notion of bases and frames. Recall that a *series expansion* of a signal f from some space S is

$$f = \sum_i c_i \phi_i, \quad (2.2.1)$$

where the elementary ‘‘atoms’’ or building blocks ϕ_i are typically simple waveforms. If the set $\{\phi_i\}_{i \in \mathbb{Z}}$ is *complete* for the space S (meaning that all $f \in S$ can be decomposed as in (2.2.1)), then a *dual* set² $\{\tilde{\phi}_i\}_{i \in \mathbb{Z}}$ exists, such that the expansion coefficients c_i in (2.2.1) are given by inner products $c_i = \langle \tilde{\phi}_i, f \rangle$. A complete and *linearly independent* set $\{\phi_i\}_{i \in \mathbb{Z}}$ is a basis of S ; its dual set is then also a basis $\{\tilde{\phi}_i\}_{i \in \mathbb{Z}}$ of S , and is *biorthogonal* to the primal one: $\langle \phi_i, \tilde{\phi}_j \rangle = \delta(i - j)$, where $\delta(i)$ is the Kronecker Delta. An important special case is when the set $\{\phi_i\}$ constitutes an *orthonormal basis*, where $\langle \phi_i, \phi_j \rangle = \delta(i - j)$ (the dual basis is now equal to the primal one). If the set $\{\phi_i\}$ is complete, but the functions ϕ_i are not linearly independent, the representation is redundant (overcomplete), and is not a basis but a *frame*. We shall address the wavelet frame expansions in Section 2.2.7.

A common wavelet basis of $\mathbf{L}^2(\mathbb{R})$ is a family of functions

$$\left\{ \psi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{x - 2^j k}{2^j}\right) \right\}_{(j,k) \in \mathbb{Z}^2}, \quad (2.2.2)$$

for a suitably constructed $\psi(x)$. For generalizations with other than dyadic scales, which are beyond our scope, see [Daubechies92, ch.10]. Any finite energy signal $f(x)$ can be decomposed in a basis (2.2.2), as

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k} \psi_{j,k}(x), \quad (2.2.3)$$

where $w_{j,k}$ are the *wavelet coefficients*, given by the inner products of $f(x)$ with the dual basis functions $\tilde{\psi}_{j,k}(x)$

$$w_{j,k} = \langle f, \tilde{\psi}_{j,k} \rangle = \int_{-\infty}^{\infty} f(x) \tilde{\psi}_{j,k}^*(x) dx, \quad (2.2.4)$$

The first example of a wavelet basis is the 1909 Haar system, where the wavelet is ‘‘blocky’’:

$$\psi(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1/2, \\ -1, & \text{if } 1/2 \leq x < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (2.2.5)$$

The Haar wavelet has a *compact support* (meaning $\psi(x) = 0$ outside of a bounded interval). Moreover, the corresponding basis is orthonormal. However, these blocky wavelets are far from optimal for

²The dual of a linear space L is by definition the set of all linear forms on L . For more details see [Gao99].

representing smooth functions. There are infinitely many other choices. The systematic way of their construction is indirect and starts from the *scaling functions*, which span the nested approximation spaces. The wavelets span the complementary spaces, which contain the differences between two successive approximations.

2.2.2 Multiresolution Analysis

Multiresolution analysis (MRA) [Mallat89b] results from a sequence of nested approximation spaces $\dots V_3 \subset V_2 \subset V_1 \subset V_0 \dots$. By projecting a signal onto this sequence, a ladder of its approximations is obtained. We use the notation in which the index j refers to the resolution scale 2^j (in [Mallat89b], the indexing is reversed).

For finite energy signals $f(x)$, the V_j are subspaces of $L^2(\mathbb{R})$. By definition [Cohen96], the approximation spaces satisfy the following properties:

- (1) The spaces are embedded $V_{j+1} \subset V_j$.
- (2) The orthogonal projections $P_{V_j} f(x)$ of $f(x)$ onto V_j satisfy $\lim_{j \rightarrow \infty} P_{V_j} f(x) = 0$ and $\lim_{j \rightarrow -\infty} P_{V_j} f(x) = f(x)$.
- (3) The V_j are generated by a *scaling* (father wavelet) function $\varphi(x) \in L^2(\mathbb{R})$, in the sense that, for each fixed j , the family

$$\left\{ \varphi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \varphi\left(\frac{x - 2^j k}{2^j}\right) \right\}_{k \in \mathbb{Z}} \quad (2.2.6)$$

is a stable basis (Riesz basis) of V_j .

From this definition, it follows clearly that $f(x) \in V_{j+1}$ is equivalent to $f(2x) \in V_j$, and that V_j is invariant under translation of 2^j .

The orthogonal projection of a signal $f(x)$ onto V_j , which is its best approximation $f_j(x)$ at the scale 2^j , is:

$$P_{V_j} f(x) \triangleq f_j(x) = \sum_{k=-\infty}^{\infty} s_{j,k} \varphi_{j,k}(x), \quad (2.2.7)$$

where $s_{j,k}$ are the scaling coefficients. The details that constitute the difference between two successive approximations $\Delta f_j(x) = f_{j-1}(x) - f_j(x)$ are contained in the *detail space* W_j , which is a complement of V_j in V_{j+1} :

$$V_j \cap W_j = \{0\} \quad \text{and} \quad V_{j+1} = V_j \oplus W_{j+1}. \quad (2.2.8)$$

The spaces W_j are *differences* between the V_j , and the spaces V_j are *sums* of the W_j . For some $L < J$, $V_L = (\oplus_{L+1}^J W_j) \oplus V_J$, i.e., for a function

in these spaces $f_L(x) = \sum_{j=L+1}^J \Delta f_j(x) + f_J(x)$. This decomposition “telescopes” into the signal at the scale 2^L .

By analogy with the approximation spaces, the detail spaces are built by dilating and shifting the mother wavelet $\psi(x)$, such that the family $\{\psi_{j,k}(x)\}_{k \in \mathbb{Z}} = \{2^{-j/2}\psi(2^{-j}x - k)\}_{k \in \mathbb{Z}}$ is a Riesz basis of W_j . The orthogonal projection of $f(x)$ onto this space is

$$P_{\mathbf{W}_j} f(x) \triangleq \Delta f_j(x) = \sum_{k=-\infty}^{\infty} w_{j,k} \psi_{j,k}(x), \quad (2.2.9)$$

where wavelet coefficients $w_{j,k}$ carry the necessary information to refine the signal approximation.

Desired properties of approximation (e.g., the degree of smoothness) impose a particular choice of $\varphi(x)$, from which the wavelet $\psi(x)$ directly follows. For example, for piece-wise constant approximations, $\varphi(x)$ is a box function: let $\varphi(x) = 1$ for $x \in [0, 1]$ and $\varphi(x) = 0$ elsewhere and $\varphi_{j,k}(x) = \varphi((x - 2^j k)/2^j)$; $f_j(x)$ is then piece-wise constant over intervals $[2^{-j}k, 2^{-j}(k+1)]_{k \in \mathbb{Z}}$. The corresponding wavelet is the Haar wavelet. In this case, $\{\varphi_{j,k}\}_{k \in \mathbb{Z}}$ and $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ are orthogonal bases of V_j and W_j respectively, and $V_j \perp W_j$. Another example of orthogonal wavelet and scaling functions together with their frequency responses is given in Fig. 2.3.

A general MRA is not orthogonal; in a non orthogonal case, the dual functions $\tilde{\varphi}(x)$ and $\tilde{\psi}(x)$ are needed to express the coefficients:

$$s_{j,k} = \langle f, \tilde{\varphi}_{j,k} \rangle, \quad w_{j,k} = \langle f, \tilde{\psi}_{j,k} \rangle. \quad (2.2.10)$$

The dual families $\{\tilde{\varphi}_{j,k}\}_{k \in \mathbb{Z}}$ and $\{\tilde{\psi}_{j,k}\}_{k \in \mathbb{Z}}$ span the spaces \tilde{V}_j and \tilde{W}_j respectively, such that $\tilde{V}_j \perp V_i$ and $\tilde{W}_j \perp W_i$, for $i \neq j$, and $\tilde{V}_j \perp W_j$ and $\tilde{W}_j \perp V_j$ for all j . This biorthogonal setting gives more freedom in designing scaling and wavelet bases [Cohen92]. An illustration of primal and dual scaling and wavelet functions is given in Section 2.2.8.

At this point note that the wavelet series expansion in Eq (2.2.3) can be rewritten as

$$f(x) = \sum_{j=-\infty}^J \sum_{k=-\infty}^{\infty} w_{j,k} \psi_{j,k}(x) + \sum_{k=-\infty}^{\infty} s_{J,k} \varphi_{J,k}(x), \quad (2.2.11)$$

where the expansion of a signal into the scaling basis $\{\varphi_{J,k}\}$ replaces the aggregation of infinitely many details ($J+1 < j < \infty$). According to the notation in Eqs (2.2.10) and (2.2.11), $\tilde{\psi}(x)$ is the *analysis* and $\psi(x)$ is the *synthesis* wavelet.

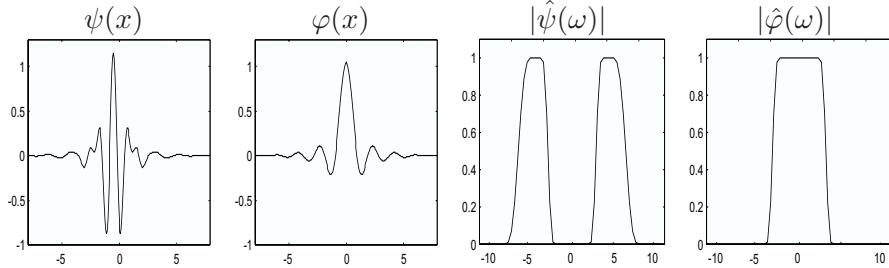


Figure 2.3: Meyer wavelet $\psi(x)$ and scaling function $\varphi(x)$, and amplitudes of their Fourier transforms $|\hat{\psi}(\omega)|$ and $|\hat{\varphi}(\omega)|$.

2.2.3 Wavelets and discrete filters

A direct consequence of the property $V_{j+1} \subset V_j$ of approximation spaces is that the function $2^{-1/2}\varphi(2^{-1}x) \in V_1$ can be decomposed into the basis of V_0 , which is $\{\varphi(x-k)\}_{k \in \mathbb{Z}}$. The same argument holds for the function $2^{-1/2}\psi(2^{-1}x) \in W_1$, since $W_{j+1} \subset V_j$. Formally,

$$\frac{1}{\sqrt{2}}\varphi\left(\frac{x}{2}\right) = \sum_{k \in \mathbb{Z}} h_k \varphi(x-k), \quad (2.2.12)$$

$$\frac{1}{\sqrt{2}}\psi\left(\frac{x}{2}\right) = \sum_{k \in \mathbb{Z}} g_k \varphi(x-k). \quad (2.2.13)$$

Eq (2.2.12) is called the *dilation equation*, *two-scale equation* or the *scaling equation*, while Eq (2.2.13) is referred to as the *wavelet equation*. The sequences \mathbf{h} and \mathbf{g} can be interpreted as discrete filters. In the biorthogonal case, similar relations are defined for the dual scaling and wavelet functions, via the dual $\tilde{\mathbf{h}}$ and the dual $\tilde{\mathbf{g}}$ filters. A case of great importance is when the impulse responses of these filters are finite (FIR): the corresponding wavelet and scaling functions are then of compact support. A necessary condition for the perfect reconstruction [Strang96] (i.e., for the duality of $\varphi(x)$ and $\tilde{\varphi}(x)$ [Cohen96]) is $2 \sum_{n \in \mathbb{Z}} \tilde{h}_n h_{n+2k}^* = \delta_k$. Once \mathbf{h} and $\tilde{\mathbf{h}}$ are specified, the wavelet filters follow as $g_n = (-1)^n \tilde{h}_{1-n}$ and $\tilde{g}_n = (-1)^n h_{1-n}$. The equations (2.2.12) and (2.2.13) are thus the core for the construction of wavelet bases [Daubechies92, Mallat98, Strang96], which is beyond the scope of this thesis, and for the construction of fast discrete wavelet transform algorithms, which we address now.

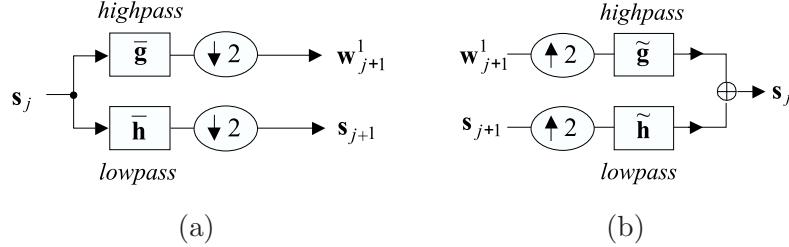


Figure 2.4: A fast orthogonal DWT: a decomposition (a), and a reconstruction (b) step. It is a filter bank algorithm: the lowpass and the highpass filters are followed by the down sampling by 2; at the reconstruction the up sampling by 2 precedes the filtering.

2.2.4 DWT - A fast discrete wavelet transform algorithm

Mallat has introduced a fast, pyramidal filter bank algorithm [Mallat89b] for computing the coefficients of the orthogonal wavelet representation; later it was generalized for the biorthogonal case. This algorithm, is in literature usually referred to as the discrete wavelet transform (DWT). The explanation of the algorithm is simple. One can show³ that the dilation equation (2.2.12), generalizes to $\varphi_{j+1,k} = \sum_{l \in \mathbb{Z}} h_{l-2k} \varphi_{j,l}$, and we have

$$s_{j+1,k} = \langle f, \varphi_{j+1,k} \rangle = \left\langle f, \sum_{l \in \mathbb{Z}} h_{l-2k} \varphi_{j,l} \right\rangle = \sum_{l \in \mathbb{Z}} \bar{h}_{2k-l} s_{j,l}, \quad (2.2.14)$$

where $\bar{h}_k = h_{-k}$ is the mirror filter. The scaling coefficients at the scale 2^{j+1} are thus computed by convolving the scaling coefficients from the previous, finer scale with the filter \bar{h} and downsampling by 2. Similarly, one can show that

$$w_{j+1,k} = \sum_{l \in \mathbb{Z}} \bar{g}_{2k-l} s_{j,l}. \quad (2.2.15)$$

One step of the above decomposition (the forward DWT) is depicted in Fig. 2.4(a). At the reconstruction (the inverse DWT), one has

$$s_{j,k} = \sum_{l \in \mathbb{Z}} h_{k-2l} s_{j+1,l} + \sum_{l \in \mathbb{Z}} g_{k-2l} w_{j+1,l}, \quad (2.2.16)$$

which can be interpreted as up-sampling (by introducing a zero between each two points) followed by filtering and summation of the filtered

³Note that $\varphi_{j+1,k}(x) = \sum_{l \in \mathbb{Z}} \langle \varphi_{j+1,k}, \tilde{\varphi}_{j,l} \rangle \varphi_{j,l}(x)$; the inner product $\langle \varphi_{j+1,k}, \tilde{\varphi}_{j,l} \rangle$ by a simple change of variable becomes $\langle \frac{1}{\sqrt{2}} \varphi(\frac{x}{2}), \tilde{\varphi}(x + 2k - l) \rangle = h_{l-2k}$.

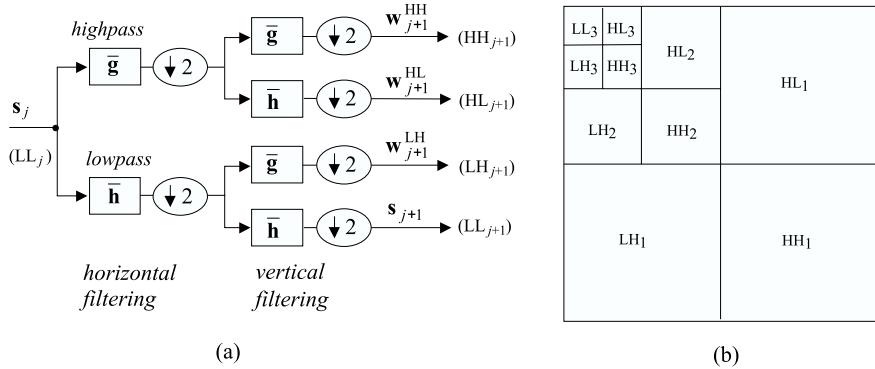


Figure 2.5: Two dimensional DWT. A decomposition step (a) and the usual organization of the subbands (b).

outputs (Fig. 2.4(b)). For an N -sample vector, the algorithm requires $\mathcal{O}(N)$ operations and is faster than the FFT, which has complexity $\mathcal{O}(N \log N)$. Note that in the first step of the DWT decomposition the scaling coefficients are approximated by the input data samples. Specific wavelets, *coiflets*, [Daubechies92, p.258], were designed to make the corresponding error negligible. In image processing, the error is usually neglected for other wavelets as well. The reason is that if the sampling interval is sufficiently small [Wickerhauser94] then physical measurements are good approximations of wavelet scaling coefficients. Another possibility is to pre-filter the samples [Jansen01b, p.23] before computing the wavelet transform.

2.2.5 DWT in two dimensions

The MRA model from Section 2.2.2 can be generalized to any positive dimension $n > 0$. Here we address the conventional *separable* two-dimensional (2D) DWT [Mallat89b]. Non-separable decompositions are described e.g., in [Kovacevic92], and one of those will be briefly addressed in Section 6.4.2. In the separable 2D case, one can show [Mallat89b], that the detail spaces of the (bi)-orthogonal MRA are spanned by the shifts and dilations of the tree “wavelets”: $\Psi^{LH}(x, y) = \varphi(x)\psi(y)$, $\Psi^{HL}(x, y) = \psi(x)\varphi(y)$ and $\Psi^{HH}(x, y) = \psi(x)\psi(y)$. The fast algorithm is a straightforward extension of the one in Section 2.2, where the filter banks are applied successively to the rows and to the columns of an image. A decomposition step is shown in Fig. 2.5(a), and a usual



Figure 2.6: An image (a) and its 2D DWT (b). Black pixels denote large magnitude wavelet coefficients.

representation of the frequency subbands in Fig. 2.5(b).

The DWT of an image yields fairly well decorrelated wavelet coefficients. However, these coefficients are not independent. One can see this in Fig. 2.6; it shows that large-magnitude coefficients tend to occur near each other within subbands, and also at the same relative spatial locations in subbands at adjacent scales and orientations, as noted, e.g., in [Simoncelli99]. Note that the positions of the large wavelet coefficients indicate image edges, i.e., the DWT has an *edge detection* property. (The use of wavelets for a classical edge detection is mentioned in Section 2.2.9). Fig. 2.6 also illustrates the *sparsity* of the DWT of images, which makes it in particular suitable for image coding and compression [Antonini92, Shapiro93]. The 2D DWT is extensively used for image denoising [Banham96], [Donoho92a]-[Donoho95b], [Nason94, Simoncelli96, Vidakovic94, Weyrich98] as well, but there its performance is limited, as the next Section explains.

2.2.6 Improving the limitations of DWT in denoising

A disadvantage of the DWT is that, in contrast to the CWT, this decimated representation is not invariant under translation. The lack of shift invariance makes it unsuitable for pattern recognition [Mallat96] and also limits the performance in denoising [Coifman95]. The latter is perhaps more clear from the viewpoint of the lack of redundancy: the

redundancy of a representation, in general, helps to better estimate a signal from its noisy observation. In this respect, two approaches are common in wavelet based image denoising:

- (1) *Cycle spinning* proposed in [Coifman95]: one averages denoising results of several cyclically shifted image versions and
- (2) Denoising in a *non-decimated* wavelet representation.

There is a slight confusion in literature regarding the two: some authors (e.g., [Fan01, Romberg01]) refer to the first approach as using the redundant discrete wavelet transform (RDWT), while this notion is commonly (and more naturally) associated with the non-decimated transform. Even though the two approaches are sometimes, e.g. in [Chang00a], regarded as “equivalent”, they should not be mixed up. The framework in which one works is quite different. Note that in the first (cycle-spining) case one removes noise from a decimated and thus decorrelated set of coefficients - the i.i.d. models are largely justified and thus the derivation of a MAP or MMSE estimator is facilitated. In the second (non-decimated) case, the realistic statistical modeling of coefficients is far more difficult. There are however other advantages. In the first place, the interscale comparisons between wavelet coefficients yielding the detection of useful image features are largely facilitated. Most of our algorithms, except in Section 5.3, are explicitly designed for the non-decimated transform.

2.2.7 Non-decimated discrete wavelet transform

In an undecimated wavelet transform, a signal is represented with the same number of wavelet coefficients at each scale. These coefficients are samples of the continuous wavelet transform at all integer locations at each dyadic scale $\mathcal{W}f(2^{-j}, k)_{(j,k)\in\mathbb{Z}^2}$. Such a redundant (overcomplete) representation results from decomposing a signal into a family of wavelets $2^{j/2}\psi(2^{-j}(x - k))$, also abbreviated by $\psi_{j,k}(x)$ for notation simplicity. Formally, we again have as in Eq (2.2.3)

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k}(x),$$

but now the wavelets $\psi_{j,k}(x)$ are not linearly independent; they do not constitute a basis but a *frame*. If the frame is *tight*, $\tilde{\psi}(x) = \psi(x)$. With respect to wavelet bases, the choice of a wavelet is less restrictive [Cohen96].

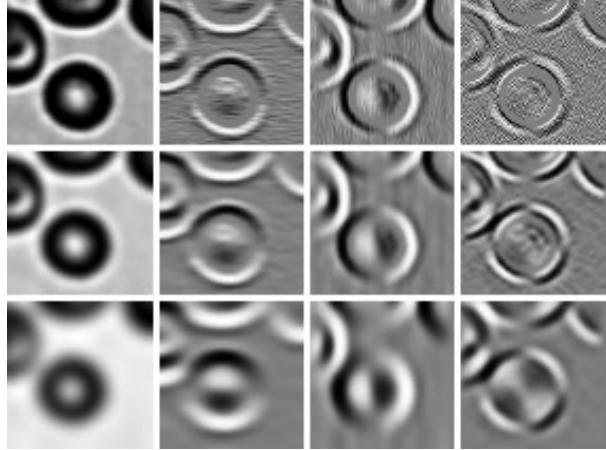


Figure 2.7: An example of the redundant wavelet frame decomposition in three resolution levels. From left to right are represented lowpass images and detail images in LH, HL and HH subbands, respectively.

A non-decimated wavelet transform approaches translation invariance and is therefore also called the *Stationary Wavelet Transform*. It is computed with the *a trous* algorithm [Holschneider89],[Mallat98, p.156]. If we denote by \mathbf{h}^j a filter where $2^j - 1$ zeros are inserted between each two coefficients of the filter \mathbf{h} , then this algorithm is

$$s_{j+1,k} = \sum_{l=-\infty}^{\infty} \bar{h}_{k-l}^j s_{j,l}, \quad w_{j+1,k} = \sum_{l=-\infty}^{\infty} \bar{g}_{k-l}^j s_{j,l} \quad (2.2.17)$$

$$s_{j,k} = \frac{1}{2} \left(\sum_{l=-\infty}^{\infty} \tilde{h}_{k-l}^j s_{j+1,l} + \sum_{l=-\infty}^{\infty} \tilde{g}_{k-l}^j w_{j+1,l} \right). \quad (2.2.18)$$

To understand the inserting zeros, note that (2.2.17) should be a consistent extension of the forward DWT in (2.2.14) and (2.2.15): all the DWT coefficients should reappear in this new transform. To get those coefficients among the redundant set, we have to skip the “extra” ones before applying convolutions. A pictorial explanation is in [Jansen01b, p.28].

Fig.2.7 illustrates the non-decimated wavelet transform of an image. It requires more calculations and calls for bigger memory than the decomposition into wavelet bases. However, it enables a better denoising quality. Also, the nonredundant representation is usually implemented

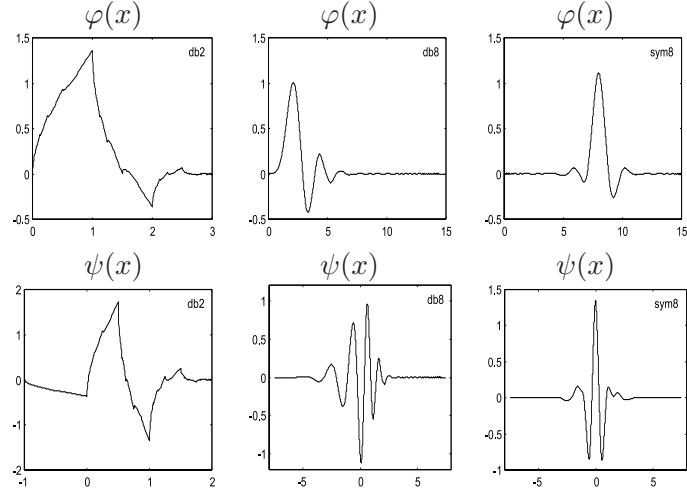


Figure 2.8: Examples of Daubechies dbN_v and Symlet $symN_v$ scaling functions and wavelets.

for discrete signals or images whose size is a power of two, because the number of coefficients is halved in subsequent resolution levels. In contrast, the non-decimated transform is equally implemented for arbitrary input sizes.

2.2.8 Choosing a wavelet for image denoising

Important questions are which wavelet(s) to choose for image denoising and why. Firstly, denoising in general is facilitated in a sparse representation (i.e., one with relatively few non-negligible coefficients). Secondly, when choosing a wavelet for imaging applications, its influence on the visual quality should be taken into account as well.

One goal is thus to produce as many as possible wavelet coefficients $\langle f, \tilde{\psi}_{j,k} \rangle$ that are close to zero. Apart from the regularity of the analyzed signal, this depends on the number of *vanishing moments* \tilde{N}_v (see Eq (2.1.1)) and on the *support size* K of the *analysis wavelet* $\tilde{\psi}(x)$: \tilde{N}_v should be as large as possible and K as small as possible. $\langle f, \tilde{\psi}_{j,k} \rangle$ is large only if a signal discontinuity is located within the support of $\tilde{\psi}(x)_{j,k}$. Moreover, any polynomial component of $f(x)$ of a degree less than \tilde{N}_v lies in the complementary scaling space [Strang89], yielding thus zero wavelet coefficients.

For the visual quality of images, the *regularity* and the *symmetry*

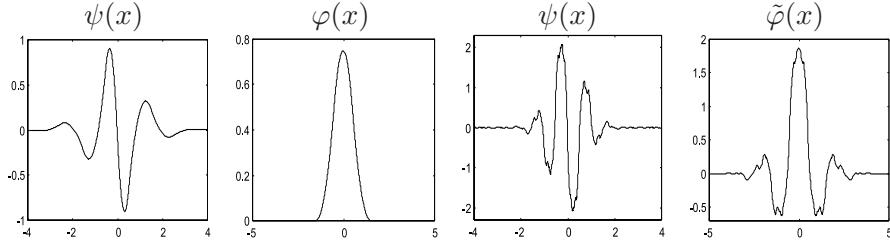


Figure 2.9: An example of spline biorthogonal wavelets and scaling functions of compact support of [Cohen92], with $N_v = 7$, $\tilde{N}_v = 3$.

try of the *synthesis* wavelet $\psi(x)$ are important. For simplicity, recall the wavelet series expansion $f(x) = \sum_j \sum_k w_{j,k} \psi_{j,k}(x)$. When reconstructing a signal from its (thresholded or quantized) wavelet coefficients, an error ϵ added to a coefficient $w_{j,k}$ will add the wavelet component $\epsilon \psi_{j,k}(x)$ to the reconstructed signal. If $\psi(x)$ is smooth then this error is smooth as well, and if we work with images it is less visible. The regularity of a wavelet usually increases with the number of its vanishing moments, even though this is not guaranteed in general [Mallat98, p.244]. The preference of symmetrical wavelets is due to the fact that our visual system is more tolerant of symmetric errors than asymmetric ones [Daubechies92, p.254].

Having in mind the above requirements, now we address some of the wavelets at our disposal. In the orthogonal case, it is difficult to achieve a large number of vanishing moments and a small support size at the same time. The theoretical limit is $K = 2N_v - 1$ and is achieved in the Daubechies wavelets [Daubechies88], usually denoted as db N_v . The shortest member of this group db1 is in fact the Haar wavelet. Two others, db2 and db8, are illustrated in Fig. 2.8. Note the lack of smoothness in the shorter one and the asymmetry of both. Except the Haar wavelet, compactly supported orthogonal wavelets cannot be symmetrical. This comes from the properties of filter banks, see e.g. [Strang96]. The least asymmetrical compactly supported orthogonal wavelets, also constructed by Daubechies, are the *symlets* sym N_v (see [Daubechies92]); an example, sym8, is illustrated in Fig. 2.8. With biorthogonal wavelets [Cohen92], illustrated in Fig. 2.9, the desired properties are easier to meet. Firstly, the compact support does not contradict the symmetry. $\psi(x)$ and $\tilde{\psi}(x)$ have equal support size K , but in general a different number of vanishing moments and a different regularity (one can “com-

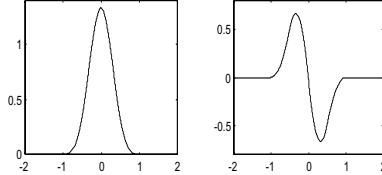


Figure 2.10: The cubic spline smoothing function $\theta(x)$ (left) and its first derivative, the quadratic spline wavelet [Mallat92b] $\psi(x)$ (right).

promise” the properties of the analysis and the synthesis wavelet).

The non-decimated wavelet transform, gives even more freedom in choosing a wavelet (see e.g., [Mallat98, p.153]). The number of vanishing moments is less important, since the sparsity is not the main argument now. In this framework, for image denoising, the *quadratic spline wavelet* [Mallat92b] (Fig. 2.10), is often used (e.g., [Xu94],[Malfait97]) because it is short and smooth. The next Section gives additional arguments. We shall use this wavelet extensively. Finally, since in image denoising literature, the db8 and the sym8 are among the most standard ones for reporting results, we shall also use these wavelets for ease of comparison of our results with the related state of the art methods.

2.2.9 Multiscale edge detection

A specific type of the dyadic wavelet transform, introduced in [Mallat92b] acts as a multiscale extension of the Canny [Canny86] edge detector. The Canny algorithm detects points of sharp variation in an image $f(x, y)$ by calculating the modulus of its gradient vector

$$\vec{\nabla} \mathbf{f} = [\partial f / \partial x \quad \partial f / \partial y]^T. \quad (2.2.19)$$

A point (x_0, y_0) is defined as an edge if the modulus of $\vec{\nabla} \mathbf{f}(x, y)$ is locally maximum at (x_0, y_0) in the direction parallel to $\vec{\nabla} \mathbf{f}(x_0, y_0)$. The multiscale version [Mallat92b] of this edge detector uses two wavelets that are the partial derivatives of a *smoothing function* $\theta(x, y)$

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial(x)} \quad \text{and} \quad \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial(y)}. \quad (2.2.20)$$

The dyadic wavelet transform $\mathcal{W}f(2^j, u, v)$ now consists of the two components $(\mathcal{W}^1 f(2^j, u, v), \mathcal{W}^2 f(2^j, u, v))$, where $\mathcal{W}^n f(2^j, u, v) =$

$\langle f, \psi_{2^j, u, v}^n \rangle$ as in Eq (2.1.5). The local maxima of the modulus

$$\mathcal{M}f(2^j, u, v) = \sqrt{\mathcal{W}^1 f(2^j, u, v) + \mathcal{W}^2 f(2^j, u, v)}, \quad (2.2.21)$$

in the direction of the angle $\mathcal{A}f(2^j, u, v)$, defined by

$$\tan(\mathcal{A}f(2^j, u, v)) = \mathcal{W}^2 f(2^j, u, v) / \mathcal{W}^1 f(2^j, u, v), \quad (2.2.22)$$

are exactly the Canny's edges of the smoothed image $(f * \theta_{2^j})(x, y)$, where $\theta_{2^j}(x, y) = (1/2^j)\theta(x/2^j, y/2^j)$. From these multiscale edges only, the algorithm of [Mallat92b] computes an image approximation that is visually identical to the original one. An important application is in image denoising [Mallat92a].

A particular choice, leading to a fast implementation, is when the smoothing function is separable. The two components of the wavelet transform are then obtained by convolving only the rows and only the columns of the image with a 1D wavelet $\psi(x)$, respectively. In [Mallat92b], the quadratic spline wavelet from Fig. 2.10 was proposed for this purpose: it has a short support and is continuously differentiable, and its dual is also a spline. This wavelet is the first derivative of the cubic spline $\theta(x)$ in Fig. 2.10 (which results from three convolutions of the box function defined on $[0, 1]$ with itself).

In practice, the above transform is implemented, using a 2D version of the *à trous* algorithm from Sec. 2.2.7, yielding the discretized wavelet coefficients $w_{j,m,n}^l = \mathcal{W}^l f(2^j, m, n)$ and the corresponding scaling coefficients $s_{j,m,n}$ (see Fig. 2.11(a)). It is thus a specific 2D non-decimated wavelet transform, with two instead of the classical three orientation subbands. Note that the reconstruction now involves *two* bandpass filters. An example illustrating the lowpass and the detail images is shown in Fig. 2.11(b).

In Chapter 4, we use both the 2-subband and the classical 3-subband non-decimated wavelet transform with the quadratic spline wavelet. The filter coefficients for this wavelet are tabulated in [Mallat92b] and [Mallat98, p.156].

2.2.10 On some extensions of the classical wavelet scheme

Numerous extensions of the “classical wavelet scheme” exist, which will not be addressed in this thesis. A nice overview is in [Sweldons96]. Here we briefly mention the *wavelet packets* [Coifman92, Wickerhauser94]

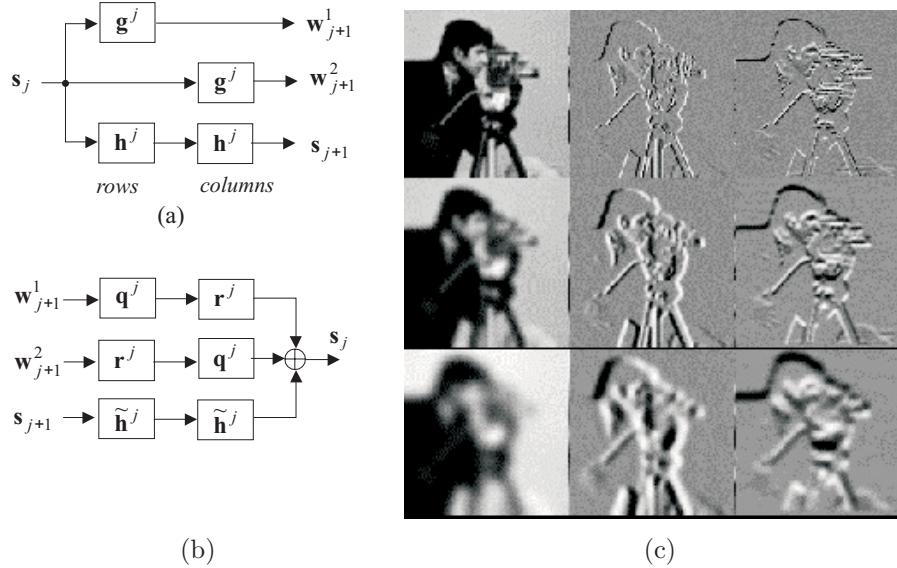


Figure 2.11: The non-decimated wavelet transform with two orientation subbands. A decomposition (a) and a reconstruction (b) step, and a decomposition example (c).

which yield more flexible and signal-adapted [Ramchandran96] representations at the cost of slightly more complex algorithms; *local trigonometric bases* [HessNielsen96]; *multiwavelets* [Geronimo94, Strang95], where instead of only one, several mother wavelets are used in order to combine their useful properties; *second generation wavelets*, using, e.g., the Sweldons' lifting scheme [Sweldons94, Sweldons95], where the idea of translation and dilation is abandoned, and the wavelet construction is adapted to irregular samples, weights, and manifolds [Schroder95, Schroder96].

Recent trends like *ridgelets* [Candes99], *curvelets* [Candes00] and *bandelets* [Pennec00] appear as competitors to wavelets in image processing.

2.3 Wavelet Domain Image Denoising

This Section serves as an introduction to the rest of the chapter, where use of wavelets in image denoising is reviewed.

In denoising there is always a trade-off between noise suppression and preserving actual image discontinuities. To remove noise without excessive smoothing of important details, a denoising algorithm needs

to be *spatially adaptive*. The wavelet representation, due to its sparsity, edge detection and multiresolution properties, naturally facilitates such spatially adaptive noise filtering. A common procedure is: (1) Compute the DWT or non-decimated wavelet transform; (2) Remove noise from the wavelet coefficients and (3) Reconstruct the denoised image. The scaling coefficients are usually kept unchanged, unless in certain cases of signal dependent noise (Section 5.5.4).

2.3.1 Noise model and Notation

We denote discrete images as vectors $\mathbf{f} = [f_1, \dots, f_n]$, where the index l refers to the spatial position (like in a raster scanning). Most noise reduction methods to be reviewed in this Chapter, start from the following additive model of a discrete image \mathbf{f} and noise $\boldsymbol{\vartheta}$

$$\mathbf{v} = \mathbf{f} + \boldsymbol{\vartheta}. \quad (2.3.1)$$

The vector \mathbf{v} is the input image. The noise $\boldsymbol{\vartheta}$ is a vector of random variables, while the unknown \mathbf{f} is a deterministic signal. Some descriptions (Section 2.6) start from a fully stochastic model, considering \mathbf{f} as well to be a specific realization of a random vector. One usually assumes that the noise has zero mean ($E(\boldsymbol{\vartheta}) = \mathbf{0}$), so that the covariance matrix is

$$Q = E[(\boldsymbol{\vartheta} - E(\boldsymbol{\vartheta}))(\boldsymbol{\vartheta} - E(\boldsymbol{\vartheta}))^T] = E(\boldsymbol{\vartheta}\boldsymbol{\vartheta}^T). \quad (2.3.2)$$

On its diagonal are the variances $\sigma_l^2 = E(\vartheta_l^2)$. If the covariance matrix is diagonal, i.e., if $E(\vartheta_l, \vartheta_k) = 0$ for $l \neq k$, the noise is uncorrelated and is called *white*. If all ϑ_l follow the same distribution, they are said to be *identically distributed*. This implies $\sigma_l^2 = \sigma^2$, for all $l = 1, \dots, n$.

An important noise type is Gaussian with the probability density

$$p_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(Q)}} e^{-\frac{1}{2}\boldsymbol{\vartheta}^T Q^{-1} \boldsymbol{\vartheta}}. \quad (2.3.3)$$

If Gaussian noise variables are uncorrelated, they are also statistically *independent* $p_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) = \prod_l p_{\vartheta_l}(\vartheta_l)$. The reverse implication (independent variables are uncorrelated) holds for all densities. A common assumption is that the noise variables are independent, identically distributed (i.i.d.). Most of the methods in this chapter are specifically designed for the case of additive white Gaussian noise, which is often abbreviated as AWGN.

2.3.2 Noise in the wavelet domain

In the wavelet domain, the most essential information in a signal is compressed into relatively few, large coefficients, which coincide with the areas of major spatial activity (edges, corners, peaks, ...) in the image. On the other hand, noise is spread over all coefficients, and at typical noise levels (that are of practical importance) the important coefficients can be well recognized. Now we describe the assumed noise model in the wavelet domain formally.

Due to linearity of the wavelet transform, the additive model (2.3.1) remains additive in the transform domain as well:

$$\mathbf{w} = \mathcal{W}_d \mathbf{v} + \mathbf{n}, \quad (2.3.4)$$

where $\mathbf{w} = \mathcal{W}_d \mathbf{v}$ are the observed wavelet coefficients, $\mathbf{y} = \mathcal{W}_d \mathbf{f}$ are the noise-free coefficients, $\mathbf{n} = \mathcal{W}_d \boldsymbol{\vartheta}$ is additive noise, and \mathcal{W}_d is an operator that yields the discretized wavelet coefficients.

An *orthogonal* wavelet transform maps the white noise in the input image into a white noise in the wavelet domain. Under such an orthogonal transform, i.i.d. noise with a variance σ^2 remains i.i.d. with the same variance σ^2 .

The situation is slightly more complicated for *bi-orthogonal* and *non-decimated* transforms. In general, one can show that if the input noise $\boldsymbol{\vartheta}_i$ is i.i.d., then in the wavelet domain the noise variance depends only on the resolution level and on the subband orientation. In other words, in each detail image \mathbf{w}_j^d noise has a constant variance $(\sigma_{n,j}^d)^2$, where

$$(\sigma_{n,j}^d)^2 = S_j^d \sigma^2, \quad (2.3.5)$$

and S_j^d is a function of the coefficients of the particular lowpass \mathbf{h} and the highpass \mathbf{g} filters used in the decomposition. Recall the scheme from Fig. 2.5. To produce \mathbf{w}_j^{LH} and \mathbf{w}_j^{HL} subbands, the lowpass filter \mathbf{h} is applied for both horizontal and vertical directions applied $2j - 1$ times in total, and the highpass filter \mathbf{g} only once (either horizontally or vertically). The subband \mathbf{w}_j^{HH} , results from in total $2(j - 1)$ lowpass filtering and 2 highpass filtering. Starting from this, and deriving the second-order cumulants of the wavelet coefficients, one can show [Foucher01] that

$$S_j^{LH,HL} = \left(\sum_k g_k^2 \right) \left(\sum_l h_l^2 \right)^{2j-1}, \quad S_j^{HH} = \left(\sum_k g_k^2 \right)^2 \left(\sum_l h_l^2 \right)^{2(j-1)}, \quad (2.3.6)$$

which holds for both decimated and non-decimated case. In the same way, for the specific decomposition with two orientation subbands $d \in \{x, y\}$ from Section 2.2.9, one can derive

$$S_j^x = S_j^y = \left(\sum_k g_k^2 \right) \left(\sum_l h_l^2 \right)^{2(j-1)}. \quad (2.3.7)$$

The above equations show that for common AWGN case, one can easily compute the noise variance in the wavelet domain, provided that the input noise variance σ^2 is known.

2.3.3 Noise variance estimation

In some applications of image denoising, the value of the input noise variance σ^2 is known, or can be measured based on information other than the corrupted data. If this is not the case, one has to estimate it from the input data, eliminating the influence of the actual signal. Wavelet based methods commonly use the highest frequency subband of the decomposition for this purpose. In the DWT of an image, the HH_1 subband contains mainly noise. A robust estimate $\hat{\sigma}$ is obtained with a median measurement, which is highly insensitive to isolated outliers of potentially high amplitudes. In [Donoho92a], it was proposed

$$\hat{\sigma} = \text{Median}(|\mathbf{w}_1^{HH}|)/0.6745. \quad (2.3.8)$$

The motivation is: if $\{u_n\}$ are N independent Gaussian random variables of zero mean and variance σ^2 , then $E(\text{Median}(|u_n|)_{0 \leq n < N}) \approx 0.6745\sigma$. One often denotes $\text{Median}(|\mathbf{w}|) = MAD(\mathbf{w})$, where MAD stands for *Median Absolute Deviation*.

The estimate in Eq (2.3.8) is commonly used in image denoising [Chang00a, Coifman95] and we use it in this thesis as well. Other ways of estimating σ in the wavelet domain include [Jansen99].

2.3.4 Performance evaluation in image denoising

The ultimate objective of image denoising is, of course, to produce an estimate $\hat{\mathbf{f}}$ of the unknown noise-free image \mathbf{f} , which approximates it best, under given evaluation criteria. Like in any estimation problem, an important objective goal is to minimize the error of the result as compared to the unknown, uncorrupted data. In this respect, a common

criterion is minimizing the *mean squared error* (MSE)

$$\text{MSE} = \frac{1}{N} \|\mathbf{f} - \hat{\mathbf{f}}\|^2 = \frac{1}{N} \sum_{i=1}^N (f_i - \hat{f}_i)^2. \quad (2.3.9)$$

One can express the *signal to noise ratio* (SNR) in terms of the mean squared error as

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{f}\|^2}{\|\mathbf{f} - \hat{\mathbf{f}}\|^2} = 10 \log_{10} \frac{\|\mathbf{f}\|^2/N}{\text{MSE}}, \quad (2.3.10)$$

where SNR is in dB. In image processing, another common performance measure is the *peak signal to noise ratio* (PSNR), which is for grey scale images defined in dB as

$$\text{PSNR} = 10 \log_{10} \frac{255^2/N}{\text{MSE}}. \quad (2.3.11)$$

The above, objective performance measures, treat an image simply as a matrix of numbers. As such, they do not reflect exactly the human perception images, i.e., their *visual quality*, which is also important. Our visual system is, for example, more tolerable to a certain amount of noise than to a reduced sharpness. Moreover, the visual quality [Barten99] is highly subjective, and it is difficult to express it in objective numbers. In this respect, there exist certain objective criteria, for expressing the degree of edge preservation (e.g., [Sattar97]). Our visual system is also highly intolerant to various artifacts, like "blips" and "bumps" [Donoho95a] in the reconstructed image. The importance of avoiding those artifacts is not only "cosmetic"; in certain applications (like astronomy, or medicine) such artifacts may give rise to a wrong data interpretation. The conclusion is that validation of image processing techniques requires both quantitative performance measures and visual inspection of results.

2.3.5 Ideal coefficient attenuation and ideal selection

Having in mind the nature of the image denoising problem, a joint estimation (of all coefficients) seems as a natural strategy. However, all the wavelet domain denoising methods to be examined in this Chapter, as well as those that we develop in the later ones, derive an estimator for each wavelet coefficient separately. The reason is simplicity and the speed of algorithms. Therefore, the wavelet coefficients will be often

modeled as independent or conditionally independent random variables, to decompose a joint estimation problem into independent ones. It is convenient at this point to examine what an ideal coefficient estimate is. In this respect, following [Mallat98], we consider a *nonlinear* estimate

$$\hat{y} = \theta(y)w \quad (2.3.12)$$

of the noise-free wavelet coefficient y given the noisy coefficient w . For AWGN noise with variance σ_n^2 , one can show that the expected squared error is $E\{(y - w\theta(y))^2\} = y^2(1 - \theta(y))^2 + \sigma_n^2\theta(y)^2$. The estimator $\theta_{ideal}(y)$ that minimizes this error is the ideal attenuator

$$\theta_{ideal}(y) = \frac{y^2}{y^2 + \sigma_n^2}, \quad (2.3.13)$$

which reminds of the Wiener filter [Papoulis84], but is in essence basically different from it (the Wiener filter performs filtering with constant values, which depend on the covariance eigenvalues). If we restrict $\theta(y) \in \{0, 1\}$, resulting in a simpler problem, which is the *coefficient selection*, we conclude from Eq (2.3.13) that the ideal selection is

$$\theta_{ideal/sel.} = \begin{cases} 1, & \text{if } |y| > \sigma_n, \\ 0, & \text{if } |y| \leq \sigma_n. \end{cases} \quad (2.3.14)$$

Even though this is an unrealistic procedure, since we do not know the noise-free values of y , this ideal selection principle is important for our later developments (Chapters 4 and 5) in specifying a "signal of interest".

2.4 Denoising by wavelet thresholding

Wavelet thresholding is a popular approach for denoising due to its simplicity. In its most basic form, this technique operates in the orthogonal wavelet domain, where each coefficient is *thresholded* by comparing against a threshold; if the coefficient is smaller than the threshold it is set to zero, otherwise, it is kept or modified. One of the first reports about this approach was by Weaver et al [Weaver92]. A systematic theory was developed mainly by Donoho and Johnstone [Donoho92a]-[Donoho95b]. They have shown that various wavelet thresholding schemes for denoising have near optimal properties in the minimax sense and perform well in simulation studies of one dimensional curve estimation. An extensive review of wavelet thresholding in image processing is in [Jansen01b].

2.4.1 Hard and soft thresholding

Two standard thresholding policies are: *hard-thresholding*, (“keep or kill”), and *soft-thresholding* (“shrink or kill”). In both cases, the coefficients that are below a certain threshold are set to zero. In hard-thresholding, the remaining coefficients are left unchanged

$$T^{hard}(w) = \begin{cases} 0, & \text{if } |w| \leq T, \\ w, & \text{if } |w| > T. \end{cases} \quad (2.4.1)$$

In soft thresholding, the magnitudes of the coefficients above threshold are reduced by an amount equal to the value of the threshold

$$T^{soft}(w) = \begin{cases} 0, & \text{if } |w| \leq T, \\ \text{sgn}(w)(|w| - T), & \text{if } |w| > T. \end{cases} \quad (2.4.2)$$

We can say that in both cases each wavelet coefficient is multiplied by a given *shrinkage factor*, which is a function of the magnitude of the coefficient (Fig. 2.12).

In soft thresholding, the estimates are biased: large coefficients are always reduced in magnitude; therefore, the mathematical expectations of their estimates differ from the observed values. The reconstructed image is often oversmoothed. On the other side, a disadvantage of the hard thresholding is its abrupt discontinuity: estimates have a larger variance and may be highly sensitive to small changes in the data. In practice, especially when the noise level is high, hard thresholding yields abrupt artifacts in the reconstructed image. Due to this, in image processing applications the soft thresholding is usually preferred over the hard one (see, e.g., [Chang00a]).

Various thresholding policies have been proposed, which are a compromise between the classical hard and soft ones, like, e.g., the “hyperbola” function [Vidakovic94] $T^{hyper}(w) = \text{sgn}(w)\sqrt{w^2 - T^2}$ for $|w| > T$, and $T^{hyper}(w) = 0$ otherwise; it attenuates large coefficients less than soft thresholding, and is continuous. Other shrinkage functions, that are less *ad hoc* result from Bayes rules (Section 2.6).

2.4.2 Threshold selection

In wavelet thresholding, the choice of the threshold is a central question, and a number of publications have been devoted to it. Here we briefly review some of the ideas and well known threshold selection rules.

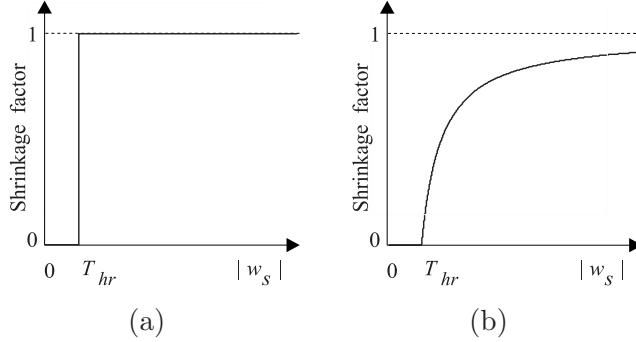


Figure 2.12: Shrinkage factors that multiply the wavelet coefficients in (a) hard-thresholding and (b) soft-thresholding.

Most methods for estimating the threshold assume AWGN noise and an orthogonal wavelet transform. Among those, well known is the *universal threshold* of Donoho and Johnstone [Donoho92a]

$$T_{univ} = \hat{\sigma}_n \sqrt{2\log(n)}, \quad (2.4.3)$$

where $\hat{\sigma}_n$ is the estimate of the standard deviation of additive white noise and n is the total number of the wavelet coefficients in a given detail image. The rationale behind this threshold is to remove all the coefficients that are smaller than the expected maximum of i.i.d. normal noise: if $\{u_i\}$ is a sequence of n i.i.d. random variables with normal distribution $\mathcal{N}(0, 1)$, then the maximum $\max_i\{|u_i|\}$ is smaller than $\sqrt{2\log(n)}$ with a probability approaching one when n tends to infinity. Moreover, the probability that $\max_i\{|u_i|\}$ exceeds $\sqrt{2\log(n)}$ by a value t is smaller than $e^{-t^2/2}$ [Donoho92a, Vidakovic94]. At different resolution scales, the threshold (2.4.3) differs only in the constant factor that is related to the number of the coefficients in a given subband.

Other thresholds that are estimated in an adaptive way for each level were proposed, e.g., in [Donoho95b, Hilton97, Jansen97, Nason94, Weyrich98]. Among those, well known is the *SURE* threshold of [Donoho95b], derived from minimizing the Stein's unbiased risk estimate [Stein81] when soft-thresholding is used. Nason [Nason94] proposed a threshold selection based on a *cross-validation* procedure, which is further extended in [Jansen97, Weyrich98] and applied to correlated noise. Other methods, like [Chang00b, Ruggeri99], derive the optimum threshold by minimizing the mean squared error in a thresholded signal under an assumed prior distribution of the wavelet coefficients. Hilton's *data*

analytic threshold [Hilton97] takes into account the spatial clustering properties of wavelet coefficients. However, this threshold as well as all the others mentioned above is *spatially uniform*, i.e., of the constant value for the whole detail image.

It is obvious that spatially uniform thresholding is not the best thing one can do. Instead of applying a constant threshold to all the coefficients (in a given subband) it would be better to decide for each coefficient separately what is better: keeping or killing (a nice discussion is in [Jansen01b, p.102]). It was shown in Sec. 2.3.5 that the mean squared error would be minimized by selecting the coefficients the signal component of which is above noise standard deviation and removing the others. A *spatially varying* threshold selection can better approach this unrealistic “dream”. In this respect, spatially adaptive thresholding with *context modeling* of wavelet coefficients [Chang98, Chang00a] is a state-of-the art approach for image denoising. Briefly, this approach applies a soft thresholding with the threshold equal to σ_n^2/σ_X , where σ_n is the noise standard deviation and σ_X is the standard deviation of the signal; to estimate σ_X at a given position, the coefficients with “similar context” are clustered; actually the context variable in [Chang00a] is a weighted average of the coefficient magnitudes in a moving window. This method appears as a reference method in Table 5.1. Other approaches, which rely on the decay of individual coefficients across scales will be reviewed in the next Section.

2.5 Denoising by singularity detection

In real world signals and images the actual transitions are usually less abrupt, or “softer” than those produced by noise. In mathematical language, this means that the *local regularity* [Mallat98, p.165] of actual signal transitions and noise is different. This fact can be effectively exploited for denoising using the wavelet transform. The basic, underlying concept is the following: the evolution (i.e., the rate of increase or decrease) of the amplitude of the wavelet transform through resolution scales at a particular spatial position is directly related to the local regularity of the signal at that position: actual edges produce large coefficients across many scales, i.e., they show a *persistence* across scales, while noise dies out swiftly as the scale increases. This is illustrated in Fig. 2.13. In this Section, we briefly review several representative denoising approaches all of which employ a *non-decimated* wavelet representation.

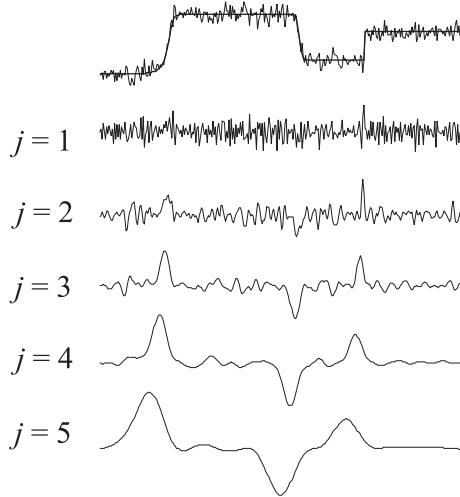


Figure 2.13: A noisy signal and its wavelet transform at five successive resolution scales.

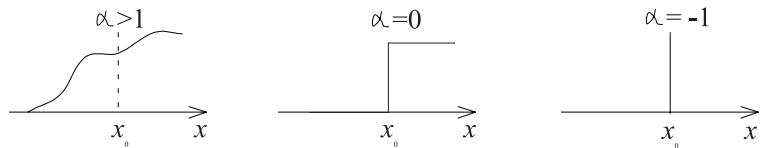


Figure 2.14: An illustration of the pointwise Lipschitz exponents.

2.5.1 Lipschitz exponents and the cone of influence

Uniform and pointwise regularity of signals is characterized by *Lipschitz* exponents that are in the mathematical literature also called *Hölder* exponents. For an extensive elaboration, see e.g., [Daoudi98], [Daubechies92, p.289–311], [Jaffard91], [LevyVehel01] [Mallat98, p.165–184], [Mallat92a]. Lipschitz exponents yield upper bounds on the error in polynomial approximations of functions, in the following way: let $P_{x_0}(x)$ be the Taylor expansion of $f(x)$ in the neighborhood of x_0 : $P_{x_0}(x) = \sum_{k=0}^{m-1} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$. Now we can define the Lipschitz exponent as follows:

Definition 2.5.1 *A function $f(x)$ is called (uniformly) Lipschitz $\alpha \geq 0$ over an interval $[a, b]$ if for all $x \in [a, b]$ there exist a polynomial $P_{x_0}(x)$*

of degree $m = \lfloor \alpha \rfloor$, and there exist a constant $C > 0$ such that

$$\forall x \in [a, b], \quad |f(x) - P_{x_0}(x)| \leq C|x - x_0|^\alpha. \quad (2.5.1)$$

The Lipschitz regularity of $f(x)$ at x_0 over $[a, b]$ is the supremum of the set of all numbers α for which $f(x)$ is "Lipschitz α ".

If $\alpha < 1$ at x_0 , then $f(x)$ is *singular* (not differentiable) at x_0 , and α characterizes the singularity type. If $0 \leq \alpha < 1$, then $P_{x_0}(x) = f(x_0)$ and the condition (2.5.1) becomes: $|f(x) - f(x_0)| \leq C|x - x_0|^\alpha$.

If a function is continuously differentiable at some point, its Lipschitz regularity at that point is positive. Exponents $\alpha \leq 0$ are determined through the primitives of functions, as follows: if $f(x)$ is Lipschitz α at a certain point, then its primitive function is Lipschitz $\alpha + 1$ at the same point. In this way, one can deduce that a bounded discontinuity (step function) has Lipschitz regularity 0, and a Dirac pulse has Lipschitz regularity -1, as it is illustrated in Fig. 2.14.

Following [Jaffard91] and [Mallat98, p.174] one can show that if a function $f(x)$ is Lipschitz α at a point x_0 , and if the wavelet function $\psi(x)$ has N_v vanishing moments with $N_v \geq \alpha$, then

$$|\mathcal{W}f(s, x)| \leq A s^{\alpha+\frac{1}{2}}, \quad \text{for } |x - x_0| \leq Ks, \quad (2.5.2)$$

where K is the support size of the mother wavelet $\psi(x)$.

Eq (2.5.2) describes the evolution of the amplitude of the wavelet transform through scales, inside the *cone of influence* (COI) of the point x_0 . The COI of x_0 is the region in the scale-space plane (s, u) , for which x_0 is included in the support of $\psi_{s,u}(x)$. In other words it consists of the points (s, u) for which the value of $\mathcal{W}f(s, u)$ is influenced by $f(x_0)$. Property (2.5.2) tells us that if the function $f(x)$ has Lipschitz regularity $\alpha > -0.5$ at a certain point, then the amplitude of the wavelet transform increases through scales within the COI of that point. In the oposite case, for $\alpha < -0.5$, the amplitude of the wavelet transform decreases through scales. This is illustrated in Fig. 2.15.

2.5.2 Reconstruction from multiscale edges

An often cited method of Mallat and Hwang [Mallat92a] reconstructs a denoised image from the detected *multiscale edges*. This method uses the wavelet transform from Section 2.2.9, the modulus maxima of which correspond to Canny's edges [Canny86] at different resolution scales. Image

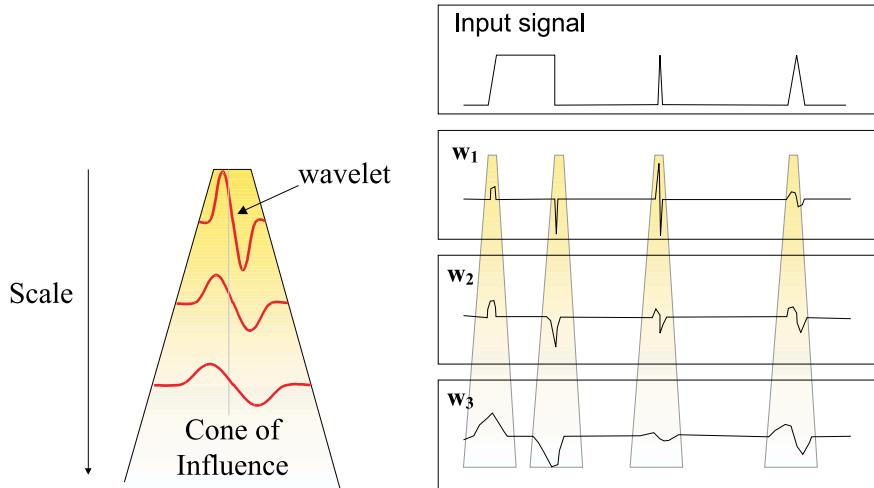


Figure 2.15: Cone of influence (COI) (left) and evolution of the wavelet transform through scales (right).

reconstruction from its multiscale edges is in itself an interesting problem. Marr [Marr82] was among the first ones to formulate solutions for this problem. Despite the non-completeness of dyadic wavelet maxima [Meyer93, p.104–109], the algorithm of Mallat and Zhong [Mallat92b] yields a close image approximation that is visually identical to the original one.

In the presence of noise, the multiscale edges are detected using the estimates of the Lipschitz exponents α , obtained through Eq (2.5.2). In particular, in [Mallat92a], the wavelet transform modulus maxima (WTMM) in the COI of each point x_0 are linked across scales, to produce the *maximum line*. The Lipschitz regularity α at x_0 is then estimated as the maximum slope of $\log_2 |\mathcal{W}f(2^j, x)|$ with respect to j along the maximum line. If the estimated $\alpha < 0$ the singularity at x_0 is supposed to be noise, and all the WTMM along the corresponding maximum line are removed. In the case of images, the method of [Mallat92a] performs additional chaining at each scale and uses prior knowledge about geometrical properties of image edges to refine the edge selection: important image structures often generate longer contours, while the short ones are likely to be created by noise. A denoised image is reconstructed from the remaining modulus maxima only. The method achieves a high performance in terms of MSE; we use it as a reference method in Table

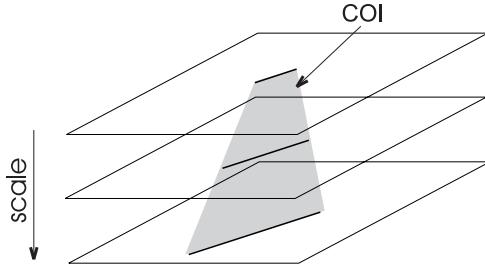


Figure 2.16: Directional cone of influence in [Hsung99]. It is support of wavelets in different scales with direction indicated by the wavelet transform angle at a given point.

4.1. The visual quality of images is satisfactory, apart from the fact that all the textures disappear.

The reconstruction step [Mallat92b] is a computationally demanding iterative procedure, the core of which are the projections onto convex sets (POCS) [Youla82]. Several techniques, like [Cetin94], have been proposed to improve the convergence of the reconstruction process; other reconstruction algorithms were studied in [Cvetkovic95]. An alternative is to avoid the WTMM representation, and to incorporate the knowledge about the evolution of the wavelet transform across scales in a classical thresholding procedure, as the next Section shows.

2.5.3 Regularity estimates for coefficient selection

One can use the rate of increase (decrease) of the wavelet transform through scales, as an (additional) criterion for selecting “important” coefficients from which a denoised image is reconstructed. Computationally, this approach is much simpler than the reconstruction from the modulus maxima only; on the other hand, with respect to simple thresholding that operates on coefficient magnitudes only, a much better performance is achieved with often only a slight increase in complexity. The selection of significant coefficients can be based on *interscale ratios* [Hsung99, Malfait97] or *interscale products* [Xu94] of the coefficients. Both of these approaches play an important role in the later developments in this thesis, so we describe them now in some detail.

Looking for a simpler approach to estimate the local Lipschitz regularity, Hsung *et al* [Hsung99] define the integral over the COI of a point

x_0

$$\mathcal{N}f(s, x_0) = \int_{|x-x_0| \leq Ks} |\mathcal{W}f(s, x)| dx, \quad (2.5.3)$$

that they denote as the wavelet transform modulus sum (WTMS), and show that $\mathcal{N}f(s, x_0) \leq A's^{\alpha+1}$ where A' is a constant. According to this result, one can estimate the exponents α from the upper bound of the slope of WTMS, instead of tracing the irregularly located wavelet maxima across scales. For a practical application in denoising, the wavelet coefficient $W(2^j, x_0)$ is selected as being useful if

$$\frac{\mathcal{N}f(2^{j+1}, x_0)}{\mathcal{N}f(2^j, x_0)} \cong 2^{\alpha+1} \geq 2, \quad (2.5.4)$$

because for useful signal transitions it is likely to have that $\alpha > 0$. In [Hsung99], the authors further observe that some “small irregular signals” can falsely fulfill the above criterion due to overlapping of COI’s and errors in estimation of $\mathcal{N}f(2^{j+1}, x_0)$; they thus introduce an additional criterion for useful coefficients, which is the interscale difference

$$\mathcal{N}f(2^{j+1}, x_0) - \mathcal{N}f(2^j, x_0) > T, \quad (2.5.5)$$

where the threshold T is a tuning parameter.

In the 2D case, the method of [Hsung99] uses the wavelet transform with two orientation subbands from Section 2.2.9. The WTMS $\mathcal{N}f(2^{j+1}, x_0, y_0)$ is now the integral of the wavelet transform modulus $\mathcal{M}f(2^j, x, y)$ from (2.2.21) over the *directional cone of influence* (DCOI)

$$\{(x, y) : (x - x_0)^2 + (y - y_0)^2 \leq K(2^j)^2, \frac{y - y_0}{x - x_0} = \tan(\mathcal{A}f(2^j, x_0, y_0))\},$$

where $\mathcal{A}f(2^j, x, y)$ is the wavelet transform angle (2.2.22). The DCOI is just a COI in a particular direction on a 2D-plane, as illustrated in Fig. 2.16. The line integral over this DCOI is implemented by linear interpolation since not all wavelet coefficients lie on the direction indicated by $\mathcal{A}f(2^j, x_0, y_0)$. Based on the 2D equivalents to the criteria (2.5.4) and (2.5.5), the wavelet coefficients in the two detail images $W^1(2^j, x_0, y_0)$ and $W^2(2^j, x_0, y_0)$ are *both* either selected or rejected for reconstruction. Obviously, this approach applies a hard coefficient classification, based on rough estimates of α , and employs heuristics in motivating and choosing a tuning parameter T . The results in [Hsung99] demonstrate its advantage over the classical wavelet thresholding, for different types of noise, in terms of MSE and the visual quality of images.

Remark. We note that the forced selection or rejection of both coefficients $W^1(2^j, x_0, y_0)$ and $W^2(2^j, x_0, y_0)$ is not optimal: for example in case of a vertical or horizontal edge only one of these two bears the signal and the other one is pure noise. In Chapter 4, we compute the WTMS in each detail image separately; the DCOI is then either horizontal or vertical and an additional benefit is that no interpolation is needed. More importantly, in Section 4.2.1, we generalize the interscale-ratio formulations and investigate their statistical properties in Section 4.2.2.

2.5.4 Interscale correlations for coefficient selection

Now we turn to the use of *interscale products*, instead of ratios, for the coefficient selection. Rosenfeld demonstrated in [Rosenfeld70] that a direct multiplication of the subband decompositions of an image yields an efficient localization of important edges. Xu *et al* [Xu94] developed a related noise filtration technique using a non-decimated wavelet transform. They use “correlation” (actually, products) among wavelet coefficients at adjacent scales in order to detect the edge coefficients. The detected edge coefficients are left unmodified while all the others are set to zero; a denoised signal or image is then reconstructed by applying the inverse transform. In the 1D case, the correlation mentioned above is

$$\mathcal{C}_L(2^j, x) = \prod_{r=0}^{L-1} \mathcal{W}f(2^{j+r}, x). \quad (2.5.6)$$

Since the wavelet coefficients of actual edges propagate well across scales, while noise dies out swiftly with increasing scale, the above correlation enhances major edges. Instead of choosing one threshold, the method of [Xu94] from each subband $\{\mathcal{W}f(2^j, k)\}_{1 \leq k \leq n}$ extracts gradually more and more “edge” coefficients, until only noise remains; in particular, the procedure is

1. Rescale the power of $\{\mathcal{C}_L(2^j, k)\}_{1 \leq k \leq n}$ to that of $\{\mathcal{W}f(2^j, k)\}_{1 \leq k \leq n}$
2. Identify an edge at position k if $|\mathcal{C}_L(2^j, k)| > |\mathcal{W}f(2^j, k)|$
3. At the detected edge positions, set to zero $\mathcal{C}_L(2^j, k)$ and $\mathcal{W}f(2^j, k)$; store the edge positions
4. Repeat the whole procedure until the power of unextracted data points in $\{\mathcal{W}f(2^j, k)\}_{1 \leq k \leq n}$ is equal to some reference noise level.

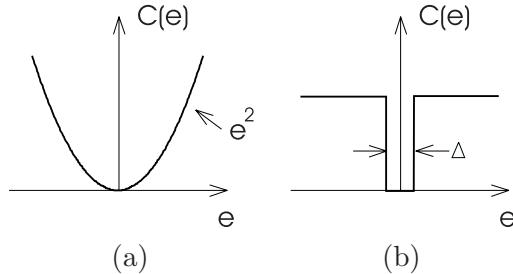


Figure 2.17: Typical cost functions: (a) quadratic and (b) uniform (delta).

Experimentally, $L = 2$ was found optimal. In the 2D case, a non-decimated transform with 3 orientation subbands was used in [Xu94] and the above procedure was applied to each orientation separately. The method is robust and has been proved efficient for various types of natural noise, e.g., in medical [Xu94] and satellite [Foucher96] images.

Remark. *In our experiments, the above iterative correlation procedure was often found unstable: a small deviation in the estimated noise level may cause a large difference in the detected edges. A similar problem was observed in [Foucher96], and therefore the stopping criterion was modified: the procedure stops when the number of detected edge coefficients reaches a predefined percentage of the total number of image pixels. We develop in Section 5.4.1 another, non-iterative correlation approach.*

2.6 Wavelet domain Bayes estimation

Bayesian approaches to wavelet shrinkage are less ad-hoc than earlier proposals and were shown to be effective [Abramovich98, Clyde98, Vidakovic98]. In general, Bayes rules are shrinkers and their shape in many cases has a desirable property: it can heavily shrink small arguments and only slightly shrink large arguments. The resulting actions on wavelet coefficients can be very close to thresholding.

2.6.1 Bayes risk, MAP and MMSE estimates

Consider a classical estimation problem: we want to estimate an unknown signal s from its noisy observation v . The probability density governing the observation process is the conditional density $p_{V|S}(v|s)$.

The unknown signal s is treated as a realization of a random vector S . The Bayes estimate \hat{s} is the estimate that minimizes the *Bayes risk* \mathcal{R} , which is the expected value of a cost $C(s, \hat{s})$

$$\mathcal{R} \triangleq E\{C(s, \hat{s})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(s, \hat{s}) p_{S,V}(s, v) ds dv. \quad (2.6.1)$$

In a given problem one chooses a cost function to accomplish two objectives [VanTrees68]: first, the cost function should measure user satisfaction adequately. The second objective is to assign one that results in a tractable problem. One usually starts by assuming that the cost depends only on the error of the estimate $s_\epsilon = \hat{s} - s$; the cost function $C(s_\epsilon)$ is then reduced to a function of a single variable. Since the joint density can be rewritten as $p_{S,V}(s, v) = p_V(v)p_{S|V}(s|v)$, the risk becomes

$$\mathcal{R} = \int_{-\infty}^{\infty} p_V(v) dv \int_{-\infty}^{\infty} C(s - \hat{s}) p_{S|V}(s|v) ds. \quad (2.6.2)$$

The estimate that minimizes this risk follows from differentiating the inner integral and setting it zero: $\frac{d}{ds} \int_{-\infty}^{\infty} C(s - \hat{s}) p_{S|V}(s|v) ds = 0$. To proceed, the cost function needs to be specified. Two typical cost functions are illustrated in Fig. 2.17.

In Fig. 2.17(a), the cost function is the square of the error $C(s_\epsilon) = s_\epsilon^2$, and is commonly referred to as the squared error [VanTrees68] or the *quadratic cost* function [Middleton68]. It accentuates the effects of large errors. The corresponding estimate is called the *minimum mean squared error* (MMSE) estimate \hat{s}_{ms} . One can easily show that it is the *conditional mean*

$$\hat{s}_{ms} = \int_{-\infty}^{\infty} s p_{S|V}(s|v) ds. \quad (2.6.3)$$

The *uniform cost* function in Fig. 2.17(b) assigns zero cost to all errors less than $\pm\Delta/2$. In other words, an error s_ϵ less than $\Delta/2$ in magnitude is as good as no error. If $|s_\epsilon| > \Delta/2$ the cost has a uniform value 1. Of particular interest is the case where Δ is arbitrarily small but nonzero number. One can show [VanTrees68, p.57] that the risk is in this case minimized by choosing the value s at which the posterior density $p_{S|V}(s|v)$ has its maximum. Hence the name the *maximum a posteriori* (MAP) estimate

$$\hat{s}_{map} = \arg \max_s p_{S|V}(s|v). \quad (2.6.4)$$

We may wish to rewrite the above expression such as to separate the role of the observation v and of the a priori knowledge. Since $p_{S|V}(s|v) =$

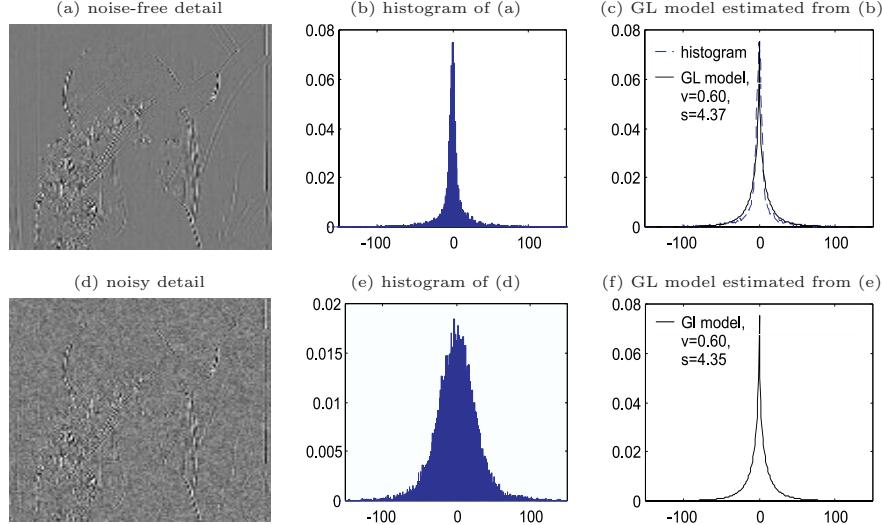


Figure 2.18: An illustration of the generalized Laplacian (GL) prior for the wavelet coefficients, and the estimation of its parameters.

$p_{V|S}(v|s)p_S(s)/p_V(v)$, and observing that $p_V(v)$ is not a function of s , the expression 2.6.4 can be rewritten as

$$\hat{s}_{map} = \arg \max_s p_{V|S}(v|s)p_S(s). \quad (2.6.5)$$

The limiting case of the MAP estimate, in which the prior is not available is the *maximum likelihood* (ML) estimate: $\hat{s}_{ml} = \arg \max_s p_{V|S}(v|s)$.

In many cases of interest (see, e.g., [VanTrees68, p.59], the MAP and the MMSE estimates coincide. For a large class of cost functions the optimum estimate is the conditional mean whenever the a posteriori density is symmetric about the conditional mean.

2.6.2 Prior distributions of image wavelet coefficients

For natural noise-free images, the histograms of wavelet coefficients are typically sharply peaked at zero. Mallat [Mallat89b] proposed to model the marginal prior distribution of image wavelet coefficients $p_Y(y)$ as a generalized Laplacian distribution

$$p_Y(y) = \frac{\nu}{2s\Gamma(\frac{1}{\nu})} \exp(-|y/s|^{\nu}), \quad s, \nu > 0 \quad (2.6.6)$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the Gamma function. The above model is often used in image processing applications, e.g., [Antonini92, Chang00a, Chang00b, Moulin99, Simoncelli96, Yoo99], where some authors, e.g., [Chang00a, Moulin99] call it the generalized Gaussian distribution (GGD). The model parameters s and ν can be computed from the histogram of noise-free wavelet coefficients. Specifically, if σ_y^2 is the sample variance and κ_y the kurtosis of the noise-free histogram, then following [Simoncelli96]

$$\sigma_y^2 = \frac{s^2 \Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})}, \quad \kappa_y = \frac{\Gamma(\frac{1}{\nu}) \Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})}. \quad (2.6.7)$$

In practice, of course, the histogram of noise-free wavelet coefficients is not available and the model parameters need to be estimated from the noisy coefficients w . In case of additive white Gaussian noise (AWGN), and for the above prior model, it is a simple task. The variance σ_w^2 and the fourth moment $m_{4,w}$ of the generalized Laplacian signal corrupted by AWGN are [Simoncelli96]

$$\sigma_w^2 = \sigma_n^2 + \frac{s^2 \Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})}, \quad m_{4,w} = 3\sigma_n^4 + \frac{6\sigma_n^2 s^2 \Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})} + \frac{s^4 \Gamma(\frac{5}{\nu})}{\Gamma(\frac{1}{\nu})}, \quad (2.6.8)$$

where σ_n^2 is the noise variance. For natural images, the shape parameter ν is typically $\nu \in [0, 1]$. Fig. 2.18 illustrates the generalized Laplacian prior, and the estimation of its parameters.

In practice, the Laplacian prior, where $\nu = 1$, is often used due to simplicity. The scale parameter is then simply computed as $s = [0.5(\sigma_w^2 - \sigma_n^2)]^{1/2}$. We shall use the Laplacian and the generalized Laplacian prior model in this thesis. Other long-tailed distributions of wavelet coefficients have been proposed for specific types of images, like, e.g., the Pearson distributions for SAR images in [Foucher01], and the α -stable distributions for medical ultrasound in [Achim01].

Common marginal prior models for wavelet coefficients are also *Gaussian mixture* models $p_Y(y) = \gamma \mathcal{N}(0, \sigma_{y,1}^2) + (1 - \gamma) \mathcal{N}(0, \sigma_{y,2}^2)$, where $0 \leq \gamma \leq 1$ is a parameter, which can be a constant in a given subband [Chipman97] or estimated for each coefficient [Crouse98, Fan01, Romberg01]. A number of methods, e.g., [Mihcak99, Strela00, Portilla01], to be reviewed in Section 2.6.4, use *Gaussian scale mixture* models (GSM) [Andrews74], where each coefficient is modeled as the product of two independent random variables: $y = \sqrt{z}u$, where z is a positive scalar, and u is an element of a Gaussian random field.

Joint histograms of image wavelet coefficients in a given subband were investigated and modeled in [Simoncelli99]. Another way of modeling the spatial interactions amongst wavelet coefficients is the use of Markov Random Field (MRF) prior models for spatial clustering [Malfait97].

Remark. *In Chapter 4, we develop a new MRF prior model for spatial clustering, and a new joint inter- and intrascale statistical model for the wavelet coefficients. Further on, in Chapter 5, we investigate and develop analytical and empirical prior models for averaged magnitudes and other functions of wavelet coefficients in a local surrounding.*

2.6.3 Simple Bayesian shrinkers

A large class of Bayesian wavelet domain filtering techniques assume an orthogonal wavelet transform and approximate the wavelet coefficients as mutually independent. In this case, optimum Bayes estimates act as simple “shrinkers” of the coefficients. Such Bayesian shrinkage rules are often similar to soft-thresholding, but are less ad-hoc and usually outperform the classical thresholding in terms of the mean squared error [Chipman97, Vidakovic98]. If the wavelet coefficients are assumed to be i.i.d., their MAP estimates are

$$\hat{y}_{map} = \arg \max p_{W|Y}(w|y)P(Y = y), \quad (2.6.9)$$

where we have omitted the indices, since the same procedure is applied to each coefficient in a given subband. The relationship between such i.i.d. MAP estimators and wavelet thresholding was studied in [Moulin99]. There it was shown that the soft-thresholding method is equivalent to the MAP estimation assuming a Laplacian prior on the wavelet coefficients, with standard deviation equal to $\sigma_y = \sigma_n^2 T_{univ}^{-1} \sqrt{2}$, where σ_n^2 is the noise variance and T_{univ} is the universal threshold.

Under the minimum mean square error (MMSE) criterion, the optimum estimate is the conditional mean

$$\hat{y}_{ms} = \int_{-\infty}^{\infty} y p_{Y|W}(y|w) dy = \frac{\int_{-\infty}^{\infty} y p_{W|Y}(w|y) p_Y(y) dy}{\int_{-\infty}^{\infty} p_{W|Y}(w|y) p_Y(y) dy}. \quad (2.6.10)$$

These estimates were studied by Vidakovic [Vidakovic94], for general one dimensional estimation problems; for the prior $p_Y(y)$, a flat-tailed distribution was used to produce the shrinkage rule similar to thresholding functions. Simoncelli and Adelson [Simoncelli96] apply and discuss

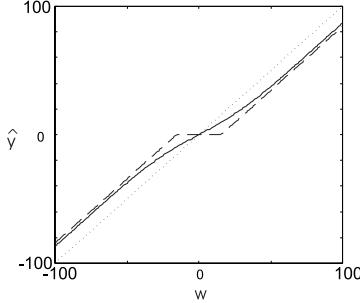


Figure 2.19: A MMSE estimator for the Laplacian prior (solid) in comparison with the soft thresholding (dashed).

the above MMSE estimator in the context of image denoising, assuming the generalized Laplacian prior from Eq (2.6.6). They consider the AWGN case, under which Eq (2.6.10) becomes

$$\hat{y}_{ms} = \frac{\int_{-\infty}^{\infty} y p_N(w-y) p_Y(y) dy}{\int_{-\infty}^{\infty} p_N(w-y) p_Y(y) dy}, \quad (2.6.11)$$

where $p_N(n)$ is the normal distribution $\mathcal{N}(0, \sigma_n^2)$. For the Laplacian prior $p_Y(y)$ the above estimator has no closed form solution and is computed numerically. Its shape is illustrated in Fig. 2.19: small coefficients are highly attenuated and the large ones slightly, which is known as the “coring” operation. In [Simoncelli96] this operation was implemented in a particular, oriented multiresolution representation, known as the steerable pyramid and was reported to outperform classical Wiener filtering. A related approach, but with the orthogonal 2D DWT was applied to medical ultrasound in [Achim01].

Another important class of Bayesian shrinkers are derived from priors on $p_Y(y)$ that are mixtures of Gaussians. Chipman *et al* [Chipman97] start with the model $p_Y(y) \propto \gamma \mathcal{N}(0, (c\tau)^2) + (1-\gamma) \mathcal{N}(0, \tau^2)$, where γ is a realization of a binary random variable Γ , with $P(\Gamma = 1) = \pi$, and the parameters π , c and τ depend on the resolution level, but are constant within a given subband. The MMSE estimate has an explicit form

$$\hat{y}_{ms} = \left(P(\Gamma = 1|w) \frac{(c\tau)^2}{\sigma_n^2 + (c\tau)^2} + P(\Gamma = 0|w) \frac{\tau^2}{\sigma_n^2 + \tau^2} \right) w, \quad (2.6.12)$$

where

$$P(\Gamma = 1|w) = \frac{\pi p_{W|\Gamma}(w|1)}{(1-\pi)p_{W|\Gamma}(w|0) + \pi p_{W|\Gamma}(w|1)}, \quad (2.6.13)$$

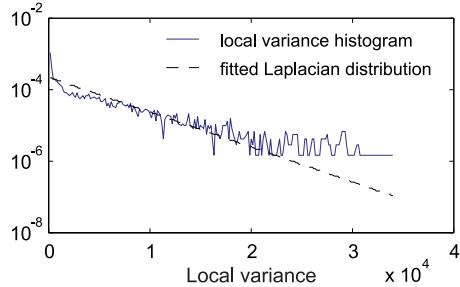


Figure 2.20: Local variance modeling [Mihcak99]. Histogram of the ML estimates of local variances of the wavelet coefficients calculated over 5x5 windows, and the fitted Laplacian distribution.

and $p_{W|\Gamma}(w|1) \propto \mathcal{N}(0, \sigma_n^2 + (c\tau)^2)$ and $p_{W|\Gamma}(w|0) \propto \mathcal{N}(0, \sigma_n^2 + \tau^2)$. Since the parameters π , c and τ can be different at different resolution levels, this method is *level-wise adaptive*. Related approaches are, e.g., [Clyde96, Clyde98] and a nice overview is given in [Vidakovic98].

Remark. Starting from a different point of view, we derive in Chapter 5 a related estimator, which is both level-wise and spatially adaptive, and which is not restricted to priors that are mixtures of Gaussians.

2.6.4 Locally adaptive wavelet domain Wiener filters

Locally adaptive wavelet domain Wiener filtering schemes, like [Li00, Mihcak99, Strela00, Portilla01] result from the MMSE criterion when local Gaussian scale mixture (GSM) models are used as priors $p_Y(y)$: each coefficient is modeled as a product of a Gaussian random variable and a multiplier derived from its local surrounding. We briefly outline the main idea.

Recall that a Wiener filter [Papoulis84] is known as a linear estimator that minimizes the mean squared error. The MMSE estimator in Eq (2.6.11) for the case where $p_Y(y) = \mathcal{N}(0, \sigma_y^2)$ has a well known closed-form solution $\hat{y}_{ms} = \sigma_y^2 w / (\sigma_y^2 + \sigma_n^2)$, which is a simple linear rescaling of the measurement w . When applied to the coefficients of a Fourier transform, this estimator corresponds to the Wiener filter. When applied to subbands of a wavelet transform, the solution is an approximation to the Wiener filter, in which the power spectral density information is averaged over each of the subbands [Simoncelli96].

As we have seen in Section 2.6.2, in the case of images, the Gaussian

prior on $p_Y(y)$ is not realistic. However, Mihcak *et al* [Mihcak99], note that the histogram of the wavelet coefficients, in a given subband, *scaled* by their local standard deviations approaches well the Gaussian distribution. They model the wavelet coefficients as conditionally independent zero mean Gaussian random variables, *given* their local variances. Under this model, the MMSE estimator is a locally adaptive Wiener filter $\hat{y}_l = \sigma_{y,l}^2 w / (\sigma_{y,l}^2 + \sigma_n^2)$, which is practically implemented as

$$\hat{y}_l = \frac{\hat{\sigma}_{y,l}^2}{\hat{\sigma}_{y,l}^2 + \sigma_n^2} w_l, \quad (2.6.14)$$

where $\hat{\sigma}_{y,l}$ is an estimate of $\sigma_{y,l}$. This estimate is formed based on a local surrounding N_l , which was chosen as a square window. A usual, maximum likelihood (ML) estimate of the variance is then

$$\hat{\sigma}_{y,l}^2 = \arg \max_{\sigma_y^2 > 0} \prod_{i \in N_l} P(w_i | \sigma_y^2) = \max \left(0, \frac{1}{n} \sum_{i \in N_l} w_i^2 - \sigma_n^2 \right), \quad (2.6.15)$$

since $P(w_j | \sigma_y^2)$ was modeled as $\mathcal{N}(0, \sigma_y^2 + \sigma_n^2)$, and n is the total number of the coefficients in a subband. The main novelty of the approach [Mihcak99] is the introduction of a prior on the variances: by observing the histograms of the ML estimates from Eq (2.6.15) (see Fig. 2.20), they propose an exponential prior $p(\sigma_y) = \lambda e^{-\lambda \sigma_y^2}$, and accordingly derive the MAP estimate, $\hat{\sigma}_{y,l}^2 = \arg \max_{\sigma_y^2 > 0} \prod_{i \in N_l} P(w_i | \sigma_y^2) p(\sigma_y^2)$, as

$$\hat{\sigma}_{y,l}^2 = \max \left(0, \frac{n}{4\lambda} \left(-1 + \sqrt{1 + \frac{8\lambda}{n^2} \sum_{i \in N_l} w_i^2} \right) - \sigma_n^2 \right), \quad (2.6.16)$$

where the parameter λ is computed for each subband, by fitting the exponential prior to the histogram of $\hat{\sigma}_{y,l}^2$ from Eq (2.6.16). The above approach was in [Mihcak99] abbreviated by *LAWMAP*, which stands for locally adaptive window-based denoising using MAP.

Related approaches, [Strela00, Portilla01] also use Gaussian scale mixture prior models for $p_Y(y)$ and Wiener filtering. The latter includes the adjacent scales into the neighborhood N_l and uses a log-normal prior for the multiplier, with a more involved procedure for estimating its parameters. Another related method [Li00], applies different filtering in supposedly smooth and “edgy” regions: the locally adaptive Wiener filtering using Eqs. (2.6.14) and (2.6.15) is applied in smooth areas, and a directional filtering along the edges. Some of the above methods

appear as reference in Table 5.1 for comparison with our method from Section 5.3.

Remark. *The methods examined in this Section are typical representatives of the locally adaptive MMSE estimation, where the signal of interest is assumed to be present with probability $p = 1$: all the wavelet coefficients, including those with the smallest magnitudes, carry a signal to be estimated. in Section 5.3.7, we compare this approach with a new locally adaptive MMSE method of similar complexity, where each coefficient is assumed to carry a useful signal with $p < 1$.*

2.6.5 MRF and HMT approaches

The Hidden Markov Tree (HMT) modeling framework is well studied and quite often used in recent wavelet denoising literature, e.g., [Fan00, Fan01, Crouse98, Romberg00, Romberg01]. Use of Markov Random Field (MRF) models for spatial clustering of the wavelet coefficients [Malfit97, Jansen99, Jansen01a] has been considerably less studied and we devote to this subject the following two Chapters. In this Section, without going into much detail, we briefly outline the main ideas, conceptual differences and similarities between the HMT and the MRF based approaches.

The HMT modeling framework is naturally related to a *decimated* wavelet representation. To improve the results of denoising, related approaches usually use the cycle spinning method (Section 2.2.6). In a decimated wavelet transform one can represent the “parent - child” relationships among the coefficients at two subsequent resolution levels on a *quadtree* structure. Due to downsampling, each coefficient at the scale 2^j corresponds to four coefficients at the next finer scale 2^{j-1} (recall Fig. 2.6). In a sense, “predecessors” of the HMT wavelet methods are multiscale stochastic processes on quadtrees studied in [Basseville92, Banham96, Luettgen93], where the wavelet coefficients are modeled using Markov relationships of the type “parent - child” on a quadtree (see Fig. 2.21(a)). In the HMT approaches (see Fig. 2.21(b)) similar relationships are established among the *hidden state variables* rather than among the coefficients themselves.

The wavelet domain HMT model of Crouse *et al* [Crouse98] forms a basis for other recent methods from this branch. In this approach, the pdf of the wavelet coefficients is a Gaussian mixture model, related to that of [Chipman97] from Section 2.6.3, but spatially varying and

capturing the interscale dependencies. With each wavelet coefficient w_l a hidden state variable $S_l \in \{S, L\}$ is associated, where S stands for a small and L for a large coefficient. In a simplified version (the so-called independent mixture model) where the vertical links in the quadtree Fig. 2.21(b)) are omitted, the marginal pdf of a noise free coefficient is a mixture of two normal densities with zero mean and variances $\sigma_{S,l}^2$ and $\sigma_{L,l}^2$: $p_Y(y) = p_l^S \mathcal{N}(0, \sigma_{S,l}^2) + p_l^L \mathcal{N}(0, \sigma_{L,l}^2)$, where p_l^S denotes the probability that $S_l = S$, and $p_l^S = 1 - p_l^L$. In the HMT model, in addition, each parent-child state-to-state link has a corresponding state transition matrix

$$A_l = \begin{bmatrix} p_l^{S \rightarrow S} & p_l^{S \rightarrow L} \\ p_l^{L \rightarrow S} & p_l^{L \rightarrow L} \end{bmatrix} \quad (2.6.17)$$

with $p_l^{S \rightarrow L} = 1 - p_l^{S \rightarrow S}$ and $p_l^{L \rightarrow S} = 1 - p_l^{L \rightarrow L}$. The parameters $p_l^{S \rightarrow S}$ ($p_l^{L \rightarrow L}$) can be read as “the probability that the wavelet coefficient w_l is small (large) given that its parent is small (large).” These are often called the *persistency probabilities*, while $p_l^{S \rightarrow L}$ and $p_l^{L \rightarrow S}$ are called the *novelty probabilities* [Romberg01], for they express the probability that the state values will change from one scale to the next. The HMT model is specified in terms of (1) the mixture variances $\sigma_{S,l}^2$ and $\sigma_{L,l}^2$; (2) the state transition matrices A_l and (3) the probability of a large state at the root node. Grouping these parameters into a vector Θ , the HMT model provides a parametric model for the joint pdf $p_{Y|\Theta}(y|\Theta)$ in each of the three orientation subbands. To compute the parameters one uses “upward-downward” algorithms through the tree, and model training procedures; these algorithms, detailed in [Crouse98], are out of our scope. Once the parameters are estimated, the wavelet coefficients are estimated as

$$\hat{y}_l = E(y_l|w_l, \Theta) = \sum_{q \in \{S, L\}} P(S_l = q|w_l, \Theta) \frac{\sigma_{q,l}^2}{\sigma_n^2 + \sigma_{q,l}^2} w_l \quad (2.6.18)$$

where σ_n is the noise standard deviation. Commonly mentioned problems are: (1) a large number of unknown parameters; it implies simplifications in practice (usually, a parameter invariance within the scale) (2) convergence (in estimating the parameters) can be relatively slow [Romberg99] and (3) lack of spatial adaptation - the links in the quadtree from Fig. 2.21(b) do not capture the *intrascalar* dependencies. In this respect, a local contextual HMT model of Fan *et al* [Fan01] is an improvement: an additional hidden state is attached to each coefficient; this additional hidden variable is a function of the surrounding wavelet

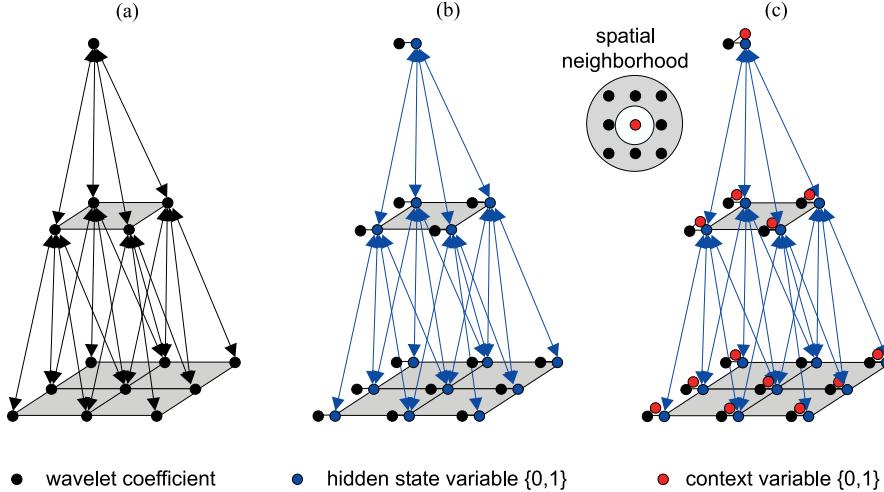


Figure 2.21: A schematic representation of (a) the multiscale stochastic process on a quadtree used in [Basseville92]; (b) the HMT model of [Crouse98] and (c) the Local contextual hidden Markov model of [Fan01].

coefficients, as illustrated in Fig. 2.21(c). We note that the actual “interactive communication” between the state variables is still only in the vertical direction and not within the scale.

Precisely the opposite type of interactions are present in a MRF based approach of [Malfait97]. There, a *significance measure* for each coefficient is computed taking into account interscale dependencies, while the interactive communication among the state variables goes horizontally, i.e., within the scale. Fig. 2.22 depicts this concept schematically.

Remark. In this thesis, we do not treat further the HMT approaches. Some of the methods listed above appear as reference methods in Table 5.1. A detailed description of the MRF based approach and our own contributions to this method follow in the next two Chapters.

2.7 Summary

In this Chapter, we reviewed briefly the principles of wavelet theory, the common concepts in wavelet based image denoising and the representatives of the state-of-the-art methods in this fields.

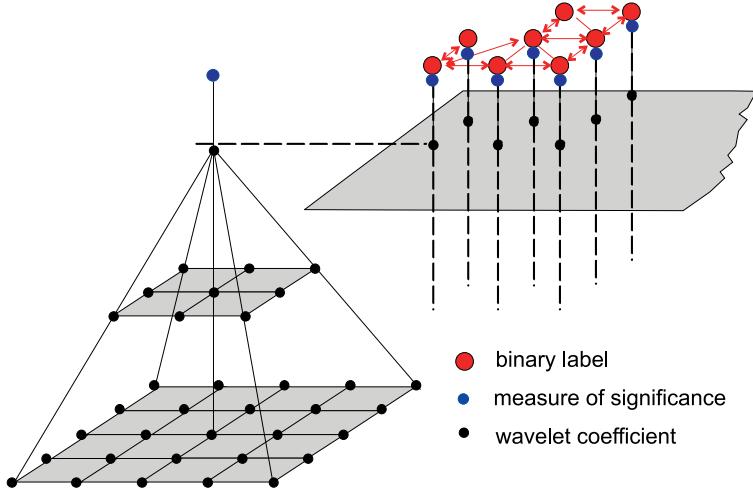


Figure 2.22: A schematic representation of the interactions among the attached (hidden) variables in a MRF based approach.

While the continuous wavelet transform is extensively used in functional analysis, the discrete wavelet transform is commonly used in signal and image processing. In this thesis, we focus therefore on the discrete wavelet transform. The main properties of this representation, which make it attractive for image denoising are: *locality*, the *multiresolution* property, *sparsity* and the *edge detection* property.

The redundant discrete wavelet representation offers a better denoising performance as compared to the orthogonal one. In this respect, either a *non decimated* transform is used or a redundant representation is achieved by implementing denoising in an orthogonal transform within a *cycle-spinning* method. Each of the two approaches has certain advantages. A non decimated transform facilitates interscale comparisons among the coefficients; in the orthogonal case, statistical modeling of the wavelet coefficients is usually simpler. In this thesis, we use almost exclusively a non decimated transform.

In imaging applications, one usually gives preference to *symmetrical* wavelets over non symmetrical ones, since the human visual system is more tolerant of symmetrical errors. In denoising, a higher SNR is usually achieved by using wavelets with a small *support size* and with a large number of *vanishing moments*; these requirements are often conflicting. Common wavelets for reporting results in image denoising literature are

the Daubechies wavelets and symlets; in methods involving regularity analysis and multiscale edge detection, the quadratic spline wavelet is commonly used.

Due to the sparsity of the wavelet transform, additive white Gaussian noise is often well suppressed by a simple *thresholding* of the wavelet coefficients, with a spatially uniform threshold. Moreover, such classical wavelet thresholding was shown to meet various optimality criteria (in the asymptotic sense). In imaging practice, the application of a uniform threshold usually yields unpleasant artefacts and/or oversmoothed results.

More advanced wavelet based denoising methods make use of interscale and intrascale dependencies among wavelet coefficients. In some approaches (including some of the state of the art methods) the inter and intrascale dependencies are exploited to “mimick” an optimum coefficient selection; the procedure is similar to standard hard or soft thresholding, with the significant difference that the threshold is *spatially varying*. Other approaches derive the MAP or MMSE estimators, relying on the *prior distribution* of the wavelet coefficients. The most popular prior distributions are the generalized Laplacian (or the generalized Gaussian distribution GGD) and mixtures of (usually two) Gaussian distributions.

While the generalized Laplacian distribution models the coefficient histogram better than the mixture of two Gaussians, we observe that the latter approach has an advantage of taking into account a probability (or uncertainty) that a particular coefficient is small (non significant) or large (significant). We find this concept in particular interesting, and develop it further in the following Chapters. The existing state-of-the art approaches, which start from such a reasoning, usually employ the Hidden Markov Tree (HMT) or the Markov Random Field (MRF) models. The HMT approach has recently become quite popular and a number of publications are devoted to this subject. In contrast, the MRF approach, proposed in [Malfait97] is less well studied and leaves much room for research and further developments.

Chapter 3

MRF priors in wavelet domain denoising

Markov Random Field (MRF) models play an important role in this thesis. This Chapter will provide the necessary theoretic background. Representative wavelet denoising approaches using MRF models will be discussed as an instance in a more general framework, where other solutions will be also proposed. A basis for our later developments will be formed.

3.1 Introduction

The use of spatial context is of great importance in image processing and pattern recognition. The intensity, i.e., the brightness level of a pixel in an image is highly statistically dependent on the intensities of surrounding pixels, unless the image is simply random noise. Using knowledge about the image structure one can recover missing information or correct erroneous data. The image structure is also reflected in the detail images of its wavelet decomposition. Spatial dependencies among wavelet coefficients are especially evident in a non-decimated wavelet representation, as in Fig. 2.7.

An example in Fig. 3.1 illustrates a noise-free wavelet detail image and its perturbation due to noise. For typical images, *spatial continuity* of pixel intensities and also of the coefficients in a wavelet detail image is *a priori* probable. Indeed, as illustrated in the figure, large coefficients are likely to form *spatially connected* clusters. One can encode the prior knowledge about the spatial dependencies among coefficient

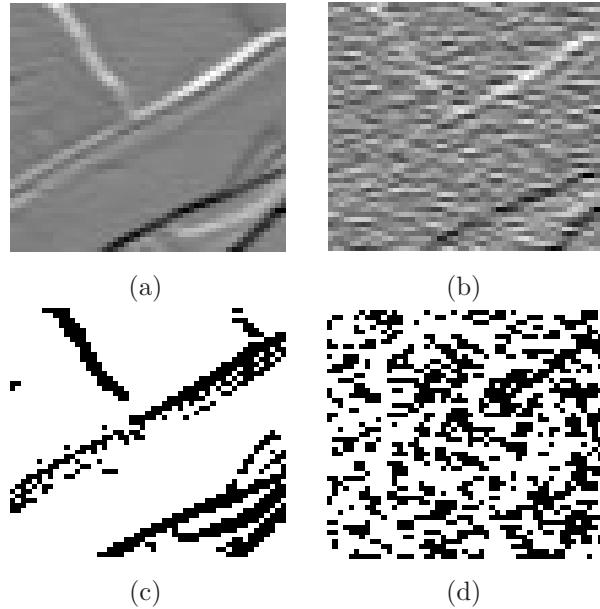


Figure 3.1: (a) A noise-free detail image and (b) its perturbation due to noise. (c) and (d): the results of applying a uniform threshold to the above coefficient magnitudes; a noise-free detail typically yields spatially connected clusters.

values or about their spatial clustering using appropriate image models. The Markov Random Field (MRF) model is one such model.

MRFs are widely used in image processing, especially for texture modeling and classification [Abend65, Chellappa85, Cross83], for image restoration [Geman84, Gidas89], and for segmentation of noisy and textured images [Bouman91, Comer99, Derin87, Krishnamachari97]. The classical concepts of single-resolution image restoration using MRFs and the maximum a posteriori (MAP) estimation, known as the MAP-MRF approach, were established in [Geman84]. Their extension to a multi-grid framework [Gidas89], has even emphasized their advantages. The application of these ideas in the wavelet domain is an active research field.

This Chapter consists of four parts. Section 3.2 reviews the basic concepts of MRF's and some common MRF models. Next, in Section 3.3, we address the wavelet domain MAP-MRF approach. Two different directions are discussed: (1) a joint MAP estimation of the wavelet coefficients in a given detail image and (2) the MAP estimation of masks, which indicate the positions of significant coefficients. In the first ap-

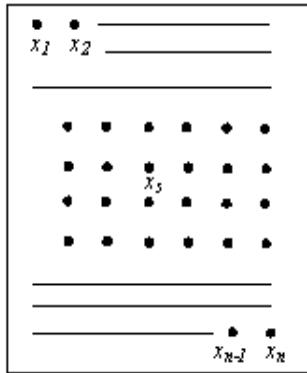


Figure 3.2: A configuration (realization) of an image field sampled on a rectangular lattice.

proach, the prior knowledge gives preference to detail structures like the ones in Fig. 3.1(a) over those in Fig. 3.1(b). In the second approach, preference is given to connected clusters like in Fig. 3.1(c) over those in Fig. 3.1(d). The difference is that only the location of large coefficients are taken into account, but not their values. Due to practical reasons, related with the complexity of the optimization techniques, in the remainder we focus on the second approach. In Section 3.4, related wavelet shrinkage approaches using bi-level MRF priors and *marginal* posterior estimators are discussed and a basis for the latter developments is formed. An alternative new method is proposed in Section 3.5, and its practical applications on images with natural noise are demonstrated. The results presented here are mainly aimed to illustrate the applications of MRF based denoising; we do not present a performance analysis of the algorithms (such analysis appears in the next Chapter).

3.2 Markov Random Fields

Many problems in image processing, such as image restoration, segmentation or edge detection fit into a general image *labeling* framework, where a given label is assigned to each pixel. In image restoration, the label that is assigned to a pixel is the “true” gray value of its intensity. In image segmentation the label determines the category of a pixel, i.e., its belonging to a specific class. In edge detection the labels are binary: if a pixel belongs to an edge it is assigned one label and in the

opposite case the other. The values of the wavelet coefficients, in this general framework, can also be interpreted as labels assigned to the corresponding pixels in detail images. In the language of random image models the statistical dependencies among pixel labels are called *spatial interactions*.

Markov random field models provide a convenient way of modeling *local* spatial interactions, i.e., they describe statistical dependence of a pixel label on labels in its local spatial surrounding. First, we introduce the notation.

3.2.1 Notation and definitions

Let $L = \{1, \dots, n\}$ be a finite index set - the set of *sites* on a regular rectangular lattice. The elements of L correspond to points at which an image is sampled, i.e., to the location of image pixels. A family of random variables $\mathbf{X} = \{X_1, \dots, X_n\}$ defined on the set L is called the *image field*. The notation $\mathbf{X} = \mathbf{x}$ will be used to abbreviate the joint event $(X_1 = x_1, \dots, X_n = x_n)$. The vector $\mathbf{x} = \{x_1, \dots, x_n\}$ is a *configuration* of \mathbf{X} , corresponding to a given realization of the image field. The space of all possible configurations of \mathbf{X} will be denoted by \mathcal{X} . A subscript in the notation of a vector will be used to indicate that only some variables are present in the vector. For example, $\mathbf{X}_{L \setminus l} = \{X_k : k \in L \setminus \{l\}\}$.

A *random field* is a family of random variables $\mathbf{X} = \{X_1, \dots, X_n\}$ such that all its possible configurations have strictly positive probability. A specific class of random field models called *Markov* random fields furthermore requires that the label of each pixel is influenced only by pixels that are its *neighbors*. Not necessarily, but usually these neighbors are the pixels that are surrounding the current one. Formally, the neighboring relation is defined as follows.

Definition 3.2.1 *A collection $\partial = \{\partial l : l \in L\}$ of subsets of L is called a **neighborhood system**, if the neighborhoods ∂l associated with the sites l satisfy:*

1. $l \notin \partial l$
2. $l \in \partial k$ if and only if $k \in \partial l$.

*The sites $l \in \partial k$ are called **neighbors** of k .*

Two examples of neighborhoods are shown in Fig. 3.3. The neighborhood in Fig. 3.3(a) is called the four-point or the *first order* neighborhood of the center pixel, and the one in Fig. 3.3(b) is called the

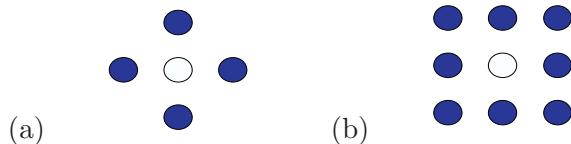


Figure 3.3: (a) First order and (b) second order neighborhood.

eight-point or the *second order* neighborhood of that pixel. For practical reasons, these two neighborhood systems are the most frequently used in image processing. The formal definition of a Markov random field is:

Definition 3.2.2 The random field \mathbf{X} is a Markov field with respect to the neighborhood system ∂ , if for all $\mathbf{x} \in \mathcal{X}$

$$P(X_l = x_l | \mathbf{X}_{L \setminus l} = \mathbf{x}_{L \setminus l}) = P(X_l = x_l | \mathbf{X}_{\partial l} = \mathbf{x}_{\partial l}). \quad (3.2.1)$$

According to this definition, the probability of a pixel label x_l , given all other labels in the image, reduces to a function of neighboring labels only. By choosing an arbitrarily large neighborhood, the MRF model can be applied to every image. MRF models that are used in image processing are often *homogeneous* (i.e., strictly stationary), meaning that the distribution $P(X_l = x_l | \mathbf{X}_{\partial l} = \mathbf{x}_{\partial l})$ is the same for all pixels l .

3.2.2 Gibbs distribution

Gibbs random fields (GRF) [Malyshev91] are used in statistical mechanics as probability models for the fluctuations of large physical systems around their equilibrium state. While MRFs characterize local aspects, GRFs explicitly express the joint probability of the system variables, providing thus a model for *global context*. The link between Markov and Gibbs random fields provides an excellent framework for specifying the global context via local spatial interactions. A Gibbs random field is generally a random field \mathbf{X} for which the configurations \mathbf{x} obey a Gibbs distribution

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} e^{-H(\mathbf{x})/T}, \quad (3.2.2)$$

where $H(\mathbf{x})$ is called the *energy function* and Z and T are constants. Z is the normalizing constant $Z = \sum_{\mathbf{x} \in \mathcal{X}} e^{-H(\mathbf{x})/T}$ and is called the *partition function*. T is the *temperature* and it actually controls the “peaking”

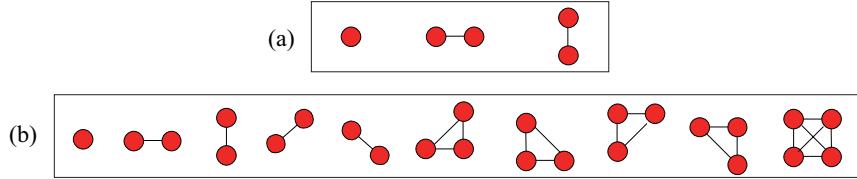


Figure 3.4: (a) The clique types for (a) the first order and (b) the second order neighborhood.

in the probability density: if T is smaller there is a bigger difference between joint probabilities of different configurations \mathbf{x} and peaks in the density that correspond to configurations with lowest energy are sharper (that is “easy to find”).

The relationship between Gibbs and Markov random fields is established through the notion of cliques (Fig. 3.4). A *clique* is a set of sites in which any two different elements are neighbors. The set of all cliques will be denoted by \mathcal{C} . The *clique potential* $V_C(\mathbf{x})$ is a function of *only* those variables x_l for which $l \in C$. If the energy $H(\mathbf{x})$ can be decomposed as a sum of clique potentials then the corresponding Gibbs random field is called a neighbor Gibbs field, and is equivalent to a Markov random field. This is stated in the often cited Hammersley-Clifford theorem:

Theorem 3.2.1 *A random field is a Markov field for the neighborhood system ∂ if and only if it is a neighbor Gibbs field for ∂ .*

Different proofs of this theorem can be found in, e.g., [Li95, Winkler95]. Its practical value is that it provides a simple way of specifying the joint probability of MRFs:

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{1}{T} \sum_{C \in \mathcal{C}} V_C(\mathbf{x})\right). \quad (3.2.3)$$

The equation (3.2.3) characterizes the *global* context (joint distribution of all pixels in an image) in terms of *local* spatial characteristics that are expressed through clique potentials. One chooses the appropriate clique potential functions to give preference to certain local spatial interactions. In this way, the *prior* knowledge about an image is encoded. For example, the isolated abrupt changes of pixel intensities are usually produced by noise. This prior knowledge can be encoded by choosing such a clique potential function that the lowest potential

corresponds to the case when all pixel labels in a clique are equal. In this way we give a preference to spatial continuity over pixels that differ strongly from their neighbors.

3.2.3 Common MRF models

For an extensive review of different MRF models see e.g., [Li95, Winkler95]. The particularly simple *Ising* model is often used as a starting point for studying the more complex ones. This model was proposed by the German physicist E. Ising in 1925 for the purpose of modeling the behaviour of ferromagnets. We present it here in terms of images. The energy function in the Ising model takes the form

$$H(\mathbf{x}) = \alpha \sum_l x_l + \beta \sum_{\{k,l\} \in \mathcal{C}} x_k x_l, \quad (3.2.4)$$

where α and β are constants and the label values are binary $x_l = \pm 1$. The neighbors $\{k, l\} \in \mathcal{C}$ are the horizontally and vertically adjacent pixels. The first term in (3.2.4) is the *cost* that is paid for the occurrence of a given label in the image irrespective of the values of its neighbors: depending on the sign of α one of the labels -1 or +1 will contribute to the higher energy, being thus “more expensive” and less probable. Therefore, by choosing α we encode the prior preference of a given label. If both labels are equally probable a priori, we choose $\alpha = 0$. The second term in (3.2.4) corresponds to the interaction of neighboring labels. If $\beta < 0$ then equal neighboring labels contribute to the lower energy, and therefore $x_k = x_l$ is more probable than $x_k \neq x_l$. For $\beta > 0$ the opposite is true. The Ising model is *homogeneous* and is also *isotropic* (i.e., rotationally invariant). Fig. 3.5 illustrates three samples of this random field, for $\alpha = 0$ and different values of β . We have sampled these configurations using the Metropolis algorithm (Fig. 3.7) at the unit temperature.

For practical reasons, image processing methods mostly use MRF models with cliques consisting of two sites (*pair-site* cliques). When the prior probabilities of different labels are known (which is less frequently the case in practice), single-site cliques are additionally used. In general, the energy of such MRF models is

$$H(\mathbf{x}) = \sum_{C \in \mathcal{C}} V_C(\mathbf{x}) = \sum_l V_1(x_l) + \sum_{\{k,l\} \in \mathcal{C}} V_2(x_k, x_l), \quad (3.2.5)$$

where $V_1(x_l)$ denotes the potential function for single-site cliques, and $V_2(x_k, x_l)$ is the potential function for pair-site cliques.

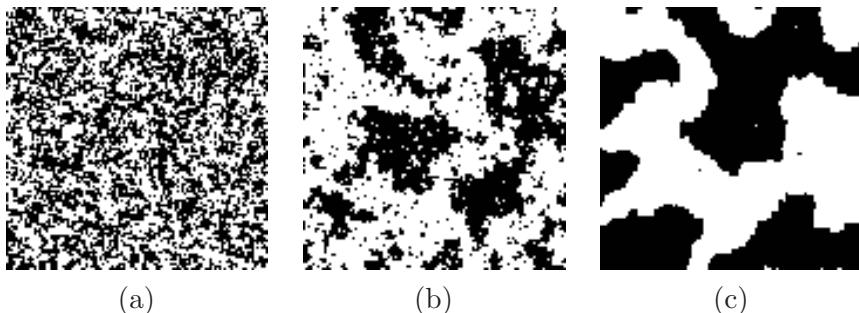


Figure 3.5: Different configurations of the Ising model. In all cases $\alpha = 0$ and (a) $\beta = -0.25$, (b) $\beta = -0.45$ and (c) $\beta = -0.9$.

Commonly used are *automodels* [Besag74], where $V_1(x_l) = x_l g_l(x_l)$ and $V_2(x_k, x_l) = \beta_{k,l} x_k x_l$; $g_l(x)$ are arbitrary functions and the constants $\beta_{k,l}$ reflect the strength of pair-site interactions. If the constants $\beta_{k,l}$ are different for cliques of different orientations then preference is given to spatial clusters in specific directions; such models appear in texture analysis [Cross83].

Important classes of auto models are *auto-logistic* models [Li95], *auto-binomial* models [Cross83] and *auto-normal* models, which are also called Gaussian Markov Random Field (GMRF) models [Chellappa83, Chellappa85, Kashiap83, Khotanzad99]. GMRF models are often used for image segmentation [Bouman91, Comer99, Krishnamachari97].

In the *multi-level logistic* (MLL) model [Derin87, Geman84], for cliques containing two or more sites the potentials are

$$V_C(\mathbf{x}) = \begin{cases} -\gamma_c, & \text{if all } x_l, l \in C \text{ are equal,} \\ +\gamma_c, & \text{otherwise.} \end{cases} \quad (3.2.6)$$

where γ_c is the potential for the type c cliques. In [Geman84], this model was called the *generalized Ising* model and was applied to image restoration.

3.3 The MAP-MRF approach

Now we address the MAP approach to wavelet based image denoising, applying a MRF prior model. The additive noise model from Eq (2.3.4) is assumed.

3.3.1 Joint MAP coefficient estimation

This Section is more theoretical, and serves as a part of a more general framework, into which the later MRF-based algorithms fit also. Recall the MAP estimation assuming independent wavelet coefficients from Section 2.6.3; its natural extension with MRF priors is the classical method of Geman [Geman84], but applied to the detail images of the wavelet decomposition. The objective thus is to find the joint MAP estimate

$$\hat{\mathbf{y}} = \arg \max_{\mathbf{y}} p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y})P(\mathbf{Y} = \mathbf{y}), \quad (3.3.1)$$

where \mathbf{y} is the noise free detail image, \mathbf{w} is the observed detail image and the joint probability $P(\mathbf{Y} = \mathbf{y})$ is a Gibbs distribution (3.2.3). If the noise in the wavelet coefficients is spatially noncorrelated, one has $p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y}) = \prod_l p_{W_l|Y_l}(w_l|y_l)$. Furthermore, one can present an arbitrary conditional distribution in the exponential form $p(w|y) = A \exp(-V(w|y))$ by choosing $V(w|y) = \ln(A) - \ln(p(w|y))$, where A is a given constant. Then the conditional probability model becomes

$$p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y}) = A \exp\left(-\sum_l V(w_l|y_l)\right). \quad (3.3.2)$$

If the joint probability $P(\mathbf{Y} = \mathbf{y})$ is the Gibbs distribution with the partition function Z , the temperature T and the energy function $H(\mathbf{y}) = \sum_{C \in \mathcal{C}} V_C(\mathbf{y})$, the posterior probability is also a Gibbs distribution

$$P(\mathbf{Y} = \mathbf{y}|\mathbf{W} = \mathbf{w}) = \frac{1}{\zeta} \exp\left(-\frac{H(\mathbf{y}|\mathbf{w})}{\tau}\right), \quad (3.3.3)$$

where the partition function is $\zeta = A/(ZP(\mathbf{W} = \mathbf{w}))$, which is independent of \mathbf{y} . The temperature is $\tau = T$ and the *posterior energy*, for the conditional model in (3.3.2) is of the form

$$H(\mathbf{y}|\mathbf{w}) = \sum_l V(w_l|y_l) + \lambda \sum_{C \in \mathcal{C}} V_C(\mathbf{y}). \quad (3.3.4)$$

The constant λ represents the influence of the prior knowledge that is encoded in the model. Thus, the MAP solution (3.3.1) is the global minimum of the posterior energy (3.3.4). In [Geman84], the posterior probability is described as an imaginary physical system whose lowest energy state is exactly the MAP estimate of the noise-free image given the noisy data. For white Gaussian noise with zero mean and variance

σ^2 , the expression (3.3.4) becomes

$$H(\mathbf{y}|\mathbf{w}) = \sum_l (w_l - y_l)^2 + \frac{2\sigma^2}{T} \sum_{C \in \mathcal{C}} V_C(\mathbf{y}). \quad (3.3.5)$$

A practical way of computing the MAP solution $\hat{\mathbf{y}}$ by minimizing the posterior energy $H(\mathbf{y}|\mathbf{w})$ is addressed in Section 3.3.3. Such a solution should yield better results than one assuming independent wavelet coefficients, if the clique potentials are correctly defined.

The specification of clique potentials is a key problem: they should penalize the abrupt changes due to noise, but not the discontinuities due to edges in the noise free image. This problem is in general important in image denoising using MRF priors. In classical, single-resolution techniques some approaches involve “line processes”, which are updated in parallel with the Markov random field [Figueiredo97, Geman84]. Many different “discontinuity adaptive” (DA) single resolution methods [Li95] have been proposed. Generally speaking, DA potentials are nonlinear functions of the absolute difference between neighboring pixel intensities, which tend to turn off the smoothing when this difference is large.

The construction of discontinuity adaptive potentials in the wavelet domain is an interesting problem, which is not well studied yet. For example, the energy of the wavelet coefficients in local surroundings can be used to indicate the presence of an edge, like in the HMT approach of [Fan01]. Such a local measurement can be introduced as an additional variable in the clique potentials. Similarly, the measurement derived from several resolution scales, at a given spatial position, may be used to control the clique potential. Here, we introduce one rough idea:

$$H(\mathbf{y}|\mathbf{w}) = f(\alpha_l) \sum_l (w_l - y_l)^2 + f(\alpha_k) \sum_{\{k,l\} \in \mathcal{C}} (|y_l| - |y_k|)^2, \quad (3.3.6)$$

where $f(\alpha_l)$ is a monotonic increasing function of the local Lipschitz exponent α_l at position l , and such that $0 < f(\alpha_l) < 1$ for $\alpha_l < 0$ and $f(\alpha_l) > 1$ for $\alpha_l > 0$. In this way, the first term in Eq (3.3.6), which expresses “closeness to the data” gives more weight to the measurements w_l that are likely to represent a useful signal; the second term in Eq (3.3.6), which takes care of spatial continuity, gives more weight to neighbors that are likely to represent actual edges. We are currently investigating the construction of practical algorithms based on this idea.

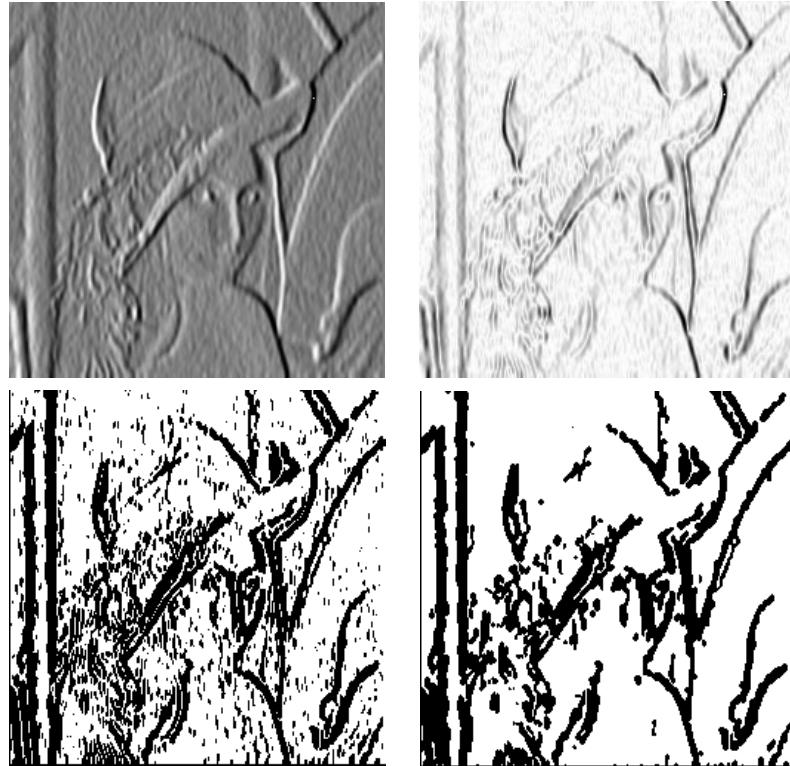


Figure 3.6: MAP-MRF mask estimation example. Top: a noisy detail and its significance map. Bottom: initial mask and its MAP estimate.

3.3.2 MAP Mask estimation

As an alternative to the previously analyzed approach, consider now a simpler one: we search for the estimate of a binary *mask* which indicates the positions of important coefficients in a detail image. This concept is among the principal ones in this thesis.

The mask \mathbf{x} is a set $\{x_1, \dots, x_n\}$ of binary labels: $x_l = -1$ if w_l is dominated by noise and $x_l = 1$ if w_l is dominated by a useful signal. The MRF prior model $P(\mathbf{X} = \mathbf{x})$ encodes now the prior knowledge about the “geometry” or spatial clustering properties of detail images. Following Malfait’s idea from [Mal97], we associate with each wavelet coefficient a measure of its *significance* m_l , which is computed from the observed wavelet coefficients. For example, one can define m_l as the coefficient magnitude, an estimate of the local Lipschitz exponent, or the amount of

interscale correlation at the position l . We call the set $m = \{m_1, \dots, m_n\}$ the *significance map*. To find the MAP mask estimate

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x})P(\mathbf{X} = \mathbf{x}) \quad (3.3.7)$$

we shall start from an initial mask, that is obtained by applying a threshold to \mathbf{m} . The labels of this mask are then iteratively updated, using a given optimization algorithm. The concept is illustrated with an example in Fig. 3.6, and the details of its implementation are explained in the next Section. Note that if $\{m_1, \dots, m_n\}$ are conditionally independent given \mathbf{x} , and if we write $p_{M_l|X_l}(m_l|x_l) = A \exp(-V(m_l|x_l))$, the posterior energy takes the same form as in Eq (3.3.4), i.e.,

$$H(\mathbf{x}|\mathbf{m}) = \sum_l V(m_l|x_l) + \lambda \sum_C V_C(\mathbf{x}), \quad (3.3.8)$$

where λ is a constant. One can use the estimated MAP mask estimate $\hat{\mathbf{x}}$ in different ways to modify the wavelet coefficients. The simplest approach is a coefficient selection: $\hat{y}_l = w_l$ if $x_l = 1$, and $\hat{y}_l = 0$ if $x_l = -1$. It already offers better results than the classical thresholding of coefficient magnitudes. Even better results are obtained when the hard binary decision is replaced by a soft modification. In this respect, we propose one approach in Section 3.5. An alternative is Malfait's method (Section 3.4.2), which does not actually use the MAP mask estimates, but a related marginal estimation concept.

3.3.3 Optimization algorithms

For both problems that were addressed in this Section, the MAP estimation of the detail images and the MAP estimation of masks, we need to minimize the posterior energies of the same form. The significant difference is that in the second case we operate with binary labels, which is usually an easier computational problem. If the values of the wavelet coefficients are quantized, the same combinatorial minimization techniques can be used for both problems.

The minimization of the energy function by deterministic gradient descent algorithms is likely to end up in a local minimum. To overcome this problem, random search methods are often used, such as the *Metropolis algorithm* [Li95] and the *Gibbs sampler* [Geman84]. These algorithms allow occasional increases of the energy, in order to get out of a local minimum. Random search starts from some initial estimate and

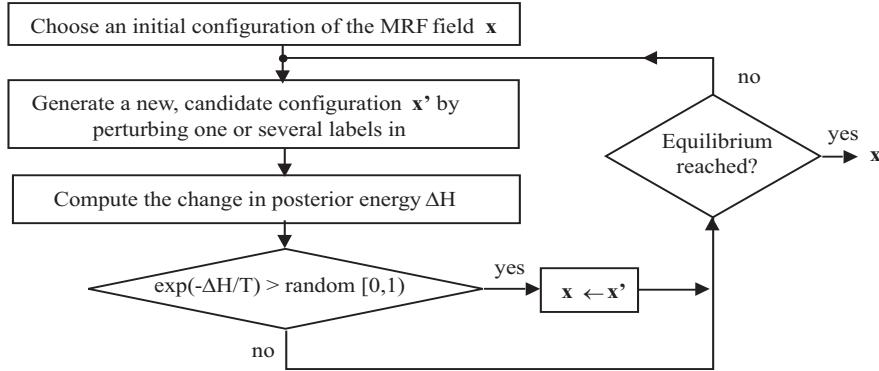


Figure 3.7: The Metropolis algorithm at the temperature T .

a chain of configurations leading to the optimum is generated by randomly perturbing labels. The convergence is faster if the initial estimate is good.

The Metropolis algorithm is depicted in Fig. 3.7. During the random search process, a new “candidate” configuration is obtained by a random perturbation of the previous configuration; subsequently, it is accepted or not, depending on the change in the posterior energy ΔH . If $\Delta H \leq 0$ the change is always accepted and if $\Delta H > 0$ the change is accepted with probability p . Practically, a random number with uniform distribution on $[0,1)$ is generated and compared with $\exp(-\Delta H/T)$. When all the labels are updated, one iteration is completed. This method does not guarantee the global minimum solution, but a low energy configuration is found with a large probability. In general, a sufficient number of generated sample configurations to reach the equilibrium is 50 times the size of the image [Li95]. For masks that we consider, 10 iterations usually suffice if the initial estimate is good. For a more theoretic background on the above and other samplers, see [Winkler95].

In image processing applications, the *iterated conditional modes* (ICM) technique [Besag86] is often used. It searches for a solution close to the MAP one, by maximizing the marginal conditional probabilities sequentially. In terms of our mask estimation problem, the ICM algorithm sequentially updates each label x_l by maximizing $P(X_l = x_l | \mathbf{M} = \mathbf{m}, \mathbf{X}_{L \setminus l} = \mathbf{x}_{L \setminus l})$. Two assumptions are made in calculating these probabilities: the conditional independence $p_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x}) = \prod_l p_{M_l|X_l}(m_l|x_l)$ and the Markovianity of \mathbf{X} . From the two assumptions and the Bayes

theorem, it follows that [Li95, p.189]

$$\begin{aligned} P(X_l = x_l | \mathbf{M} = \mathbf{m}, \mathbf{X}_{L \setminus l} = \mathbf{x}_{L \setminus l}) &= \\ &= C p_{M_l|X_l}(m_l|x_l) P(X_l = x_l | \mathbf{X}_{\partial l} = \mathbf{x}_{\partial l}), \end{aligned} \quad (3.3.9)$$

where C does not depend on x_k . For the serial updating case, convergence is guaranteed and is rapid (5 to 6 raster scans of an image are sufficient). In our applications, we found this method very efficient (Section 3.5).

Other optimization techniques, like random and deterministic *annealing* algorithms, are discussed in [Li95]. Annealing algorithms tend to overcome the problem of getting trapped in local minima by decreasing the value of a given parameter (e.g., temperature in the Gibbs distribution) during the iterative minimization. For the MAP estimation of detail images in the continuous case, like in Eq (3.3.6), a deterministic annealing algorithm *graduated non-convexity* (GNC) [Blake89] is attractive.

3.4 Bayesian shrinkage with MRF priors

The core of this Section is a stochastic sampling wavelet shrinkage method, first proposed by Malfait and Roose [Malfait97], and further motivated and developed in [Jansen99, Jansen01a, Jansen01b]. It uses the same prior and conditional model formulations as in Section 3.3.2, and we start from this overlap first.

3.4.1 Prior and conditional models

Recall the MAP-MRF mask estimation problem from Eq (3.3.7). It requires the MRF prior model $P(\mathbf{X} = \mathbf{x})$ for the masks \mathbf{x} and the conditional model $p_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x})$ for a chosen significance map \mathbf{m} . These models appear also in the Bayesian shrinkage algorithm of this Section.

As illustrated in Section 3.2.3, there are a number of possibilities in designing the MRF prior model. In making a specific choice, the complexity of realization is an important thing to bear in mind. In [Malfait97], the isotropic auto-logistic MRF model, with the second order neighborhood N_l , was used. In this case, and for the label set

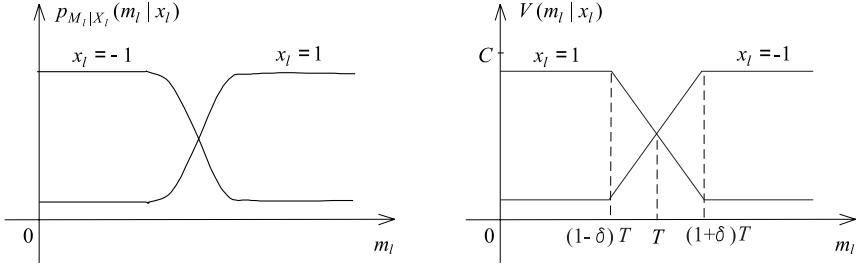


Figure 3.8: The heuristic conditional model of Malfait and Roose [Malfait97], where m_l is a rough estimate of the local Lipschitz exponent. Left: conceptual. Right: practical, where $p_{M_l|X_l} = A \exp(-V(m_l|x_l))$.

$\{-1, 1\}$, the joint probability of the MRF becomes¹

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp\left(-\sum_{l \in L} V_{N_l}(\mathbf{x})\right), \quad V_{N_l}(\mathbf{x}) = -\gamma \sum_{k \in N_l} x_l x_k, \quad (3.4.1)$$

where γ is a positive constant. In Section 4.3, we shall introduce a new, *anisotropic* model, which adapts to the presence of edge-components in a given neighborhood. Another prior model, which uses a different reasoning, and attempts to compute an average degree of isolation of the "1" labels was recently proposed in [Jansen01a].

Now we come to the second important issue: the specification of the conditional model. Regardless of the specific choice for the significance measure m_l , the starting assumption [Malfait95, Malfait97, Jansen99, Jansen01a, Jansen01b] is the conditional independence

$$p_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x}) = \prod_l p_{M_l|X_l}(m_l|x_l). \quad (3.4.2)$$

Jansen [Jansen01a] motivates it from the viewpoint that the conditional model describes *local* significance (while the spatial correlations are included in the prior model). A similar argument was given in [Malfait97]. However, one should note, that in the non-decimate wavelet transform domain, where these methods operate, the conditional independence does not actually hold (the wavelet functions are not linearly independent). We shall accept the model 3.4.2 as an approximation, which greatly facilitates the computation and which has been proved to yield

¹In [Malfait97], the set of labels was $\{0, 1\}$, and this expression was represented in a less compact way, but it is only a matter of notation.

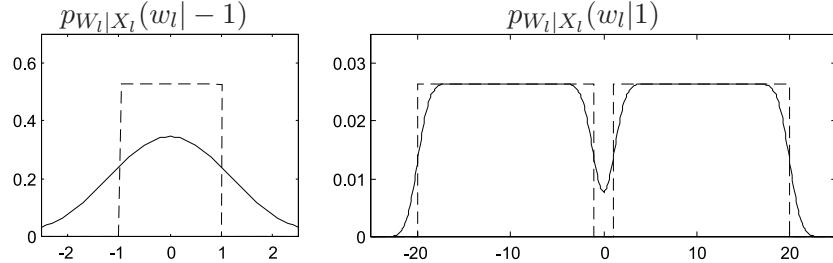


Figure 3.9: The conditional model of [Jansen01b]. Solid lines represent the conditional densities of the noisy coefficients, and dashed lines represent the assumed underlying distributions of noise-free coefficients.

good results in practice. It still remains to specify the marginal conditional densities $p_{M_l|X_l}(m_l|x_l)$ for the chosen m_l . We shall focus on this topic in Chapter 4. At the moment, we proceed with the representatives from the related literature.

Malfait and Roose [Malfait95, Malfait97] incorporate the local regularity information into the conditional model. The significance measure m_l is chosen as the interscale ratio of coefficient magnitudes, averaged over a certain number d of resolution scales:

$$m_l = \frac{1}{d} \sum_{j=1}^d \left| \frac{w_{j+1,l}}{w_{j,l}} \right| \simeq 2^\alpha, \quad (3.4.3)$$

yielding a rough estimate of the local Lipschitz exponent α at spatial position l . Reasoning that for $x_l = 1$ larger m_l are more likely than for $x_l = -1$, a heuristic model from Fig. 3.8 was proposed. Practically, the conditional probabilities were implemented as $p_{M_l|X_l}(m_l|x_l) = A \exp(-V(m_l|x_l))$, where the functional forms of the potentials $V(m_l|x_l)$ consist of two constant parts and a linear transition around a threshold T . A semi-automatic procedure was proposed in [Malfait97] to compute the threshold T , but still there are other unknown parameters of this model, which may complicate its practical use.

Looking for less parametrized and more realistic conditional models, Jansen and Bultheel [Jansen01a, Jansen01b] chose a simpler significance measure: $m_l = |w_l|$. Their model for the conditional densities $p_{W_l|X_l}(w_l|x_l)$ is shown in Fig. 3.9, and the explanation follows. First, referring to the ideal coefficient selection procedure from Section 2.3.5, the authors treat *noise-free* coefficients y_l as significant if

$|y_l| > \sigma_n$, where σ_n is the noise standard deviation. They further simplify: $p_{Y_l|X_l}(y_l|1)$ is uniform on $[-\sigma_n, \sigma_n]$ and $p_{Y_l|X_l}(y|1)$ is uniform on $[-\mu, \sigma_n] \cup [\sigma_n, \mu]$, where μ is a parameter, which is estimated from the observed coefficients. Convolution of these densities with the noise distribution $p_N(n) = \mathcal{N}(0, \sigma_n)$ yields the corresponding densities of noisy coefficients $p_{W_l|X_l}(w_l|x_l)$ from Fig. 3.9. The corresponding model for the coefficient magnitudes is straightforward. Equivalently, one can derive the conditional models for more realistic prior distributions $p_Y(y)$, like the generalized Laplacian, as we do in Chapter 4.

3.4.2 Stochastic sampling method of Malfait and Roose

Malfait and Roose proposed in [Malfait97] the following probabilistic wavelet shrinkage scheme: each wavelet coefficient is multiplied with the marginal probability that it is noise-free, given the computed significance map:

$$\hat{y}_l = P(X_l = 1|\mathbf{M} = \mathbf{m})w_l. \quad (3.4.4)$$

The motivation was: the shrinkage factor $P(X_l = 1|\mathbf{M} = \mathbf{m})$ is always between zero and one, and it suppresses more those coefficients that are less likely to represent a useful signal. In [Jansen01b, p.154], this rule is further motivated as a “posterior expected action”, where “action” is the hard coefficient classification according to labels X_l .

The exact computation of the marginal probability in (3.4.4) requires the summation of the posterior probabilities $P(\mathbf{X} = \mathbf{x}|\mathbf{M} = \mathbf{m})$ of all possible configurations \mathbf{x} for which $x_l = 1$:

$$P(X_l = 1|\mathbf{M} = \mathbf{m}) = \sum_{\mathbf{x} \in \mathcal{X}} f_l(\mathbf{x})P(\mathbf{X} = \mathbf{x}|\mathbf{M} = \mathbf{m}),$$

where $f_l(\mathbf{x}) = \begin{cases} 0, & x_l = -1, \\ 1, & x_l = +1. \end{cases} \quad (3.4.5)$

Since this is an intractable task, one typically alleviates it by using a relatively small, but “representative” subset of all possible configurations. Such a representative subset is obtained via an importance-type sampling: the probability that a given mask is sampled should be proportional to its posterior probability. An estimate of $P(X_l = 1|\mathbf{M} = \mathbf{m})$ is then obtained by computing the fractional number of all sampled masks for which $x_l = 1$. This sampling is typically realized using the Metropolis and the Gibbs samplers that were mentioned in Section 3.3.3, in the context of the MAP estimation. Note that now

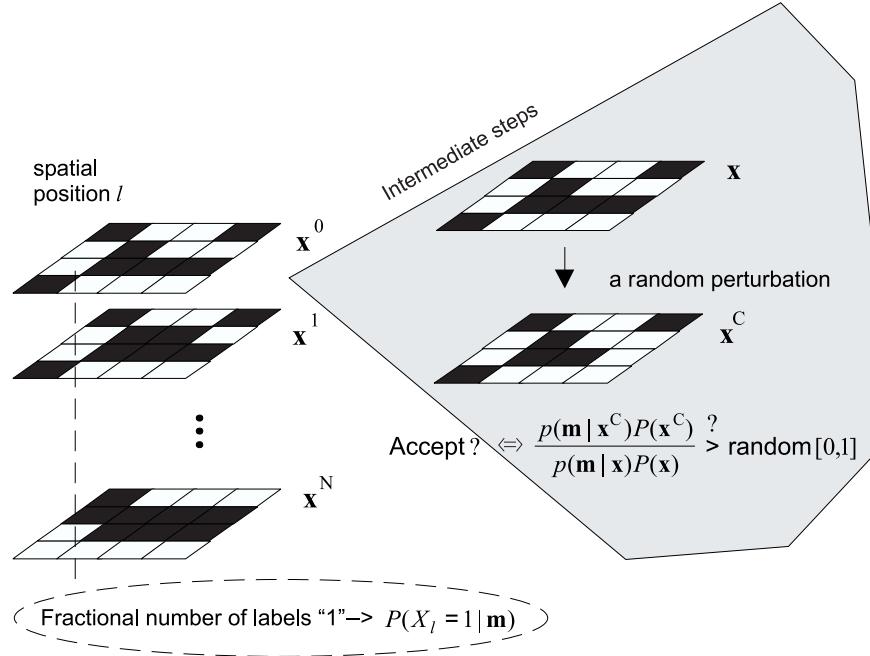


Figure 3.10: Estimation of marginal probabilities using the Metropolis sampler.

the MAP mask estimate is not the objective (or at least not a direct one), but the same random search algorithms are employed in the role of the *marginal* posterior estimators. In particular, all the methods in [Malfait95, Malfait97, Jansen99, Jansen01a, Jansen01b] use the Metropolis sampler. Recall that from each configuration \mathbf{x} , it generates a new, “candidate” configuration \mathbf{x}^c by switching the binary label at a randomly chosen position l . The decision about accepting the change is based on the change in the posterior energy ΔH , or equivalently, the decision is based on the ratio $p = \exp(-\Delta H/T)$ of the posterior probabilities of the new and the old configurations

$$p = \frac{P(\mathbf{x}^c | \mathbf{m})}{P(\mathbf{x} | \mathbf{m})} = \frac{p_{\mathbf{M}|\mathbf{X}}(\mathbf{m} | \mathbf{x}^c)P(\mathbf{x}^c)}{p_{\mathbf{M}|\mathbf{X}}(\mathbf{m} | \mathbf{x})P(\mathbf{x})}. \quad (3.4.6)$$

If $p > 1$ the local change is accepted, and if $p < 1$, the change is accepted with probability p .

Practically, using the prior model from Eq (3.4.1), and the conditional model from Fig. 3.8, the method of [Malfait97] runs the stochastic

sampling procedure with the posterior probability

$$P(\mathbf{x}|\mathbf{m}) \propto \exp\left(-\left(\alpha \sum_l V(m_l|x_l) + \beta \sum_l V_{N_l}(x)\right)\right), \quad (3.4.7)$$

where the two parameters α and β are experimentally optimized. Note that for other conditional models, like in Fig. 3.9, the representation from Eq (3.4.7), is also possible, but may unnecessarily complicate the representation. In general, $P(\mathbf{x}|\mathbf{m}) \propto \exp(\sum_{c \in C} V_c(\mathbf{x})) \prod_l p_{M_l|X_l}(m_l|x_l)$.

We illustrate the whole procedure in Fig. 3.10. The number of iterations depends on the required accuracy of the marginal probability estimates. This and other aspects of the sampling procedure, like the use of several short runs as opposed to one long run are discussed in detail in [Malfait95]. In [Jansen99], one run with 10 sampled masks was found sufficient in practice, provided that the initial mask is well chosen.

3.4.3 A practical example

Now we illustrate a simplified version of the stochastic sampling procedure of Malfait and Roose on a real noisy image. The idea is to give the reader a feeling of how the whole procedure works, on a simple-to-implement example. Let $m_l = |w_l|$, and assume a heuristic conditional model from Fig. 3.8, implemented with piece-wise linear potentials $V(m_l|x_l)$ having two constant parts and a linear transition of the width $2\delta T$ around the threshold T . The prior model $P(\mathbf{X} = \mathbf{x})$, is given in Eq (3.4.1). For the two models, we derive the posterior probability ratio p from Eq (3.4.6) as

$$p = \exp(-\Delta V_{cond} - \Delta V_{prior}), \quad (3.4.8)$$

where for the label switching $x_l^C = -x_l$

$$\Delta V_{cond} = \begin{cases} -\alpha x_l, & \text{if } m_l/T < (1 - \delta), \\ \alpha x_l, & \text{if } m_l/T > (1 + \delta), \\ \alpha x_l(m_l/T - 1)/\delta, & \text{otherwise.} \end{cases} \quad (3.4.9)$$

and

$$\Delta V_{prior} = \beta x_l \sum_{k \in N_l} x_k. \quad (3.4.10)$$

Obviously, the value of $\alpha > 0$ equals the difference of the two constant parts in the piece-wise linear model for $V(m_l|x_l)$, and $\beta = 2\gamma > 0$, where

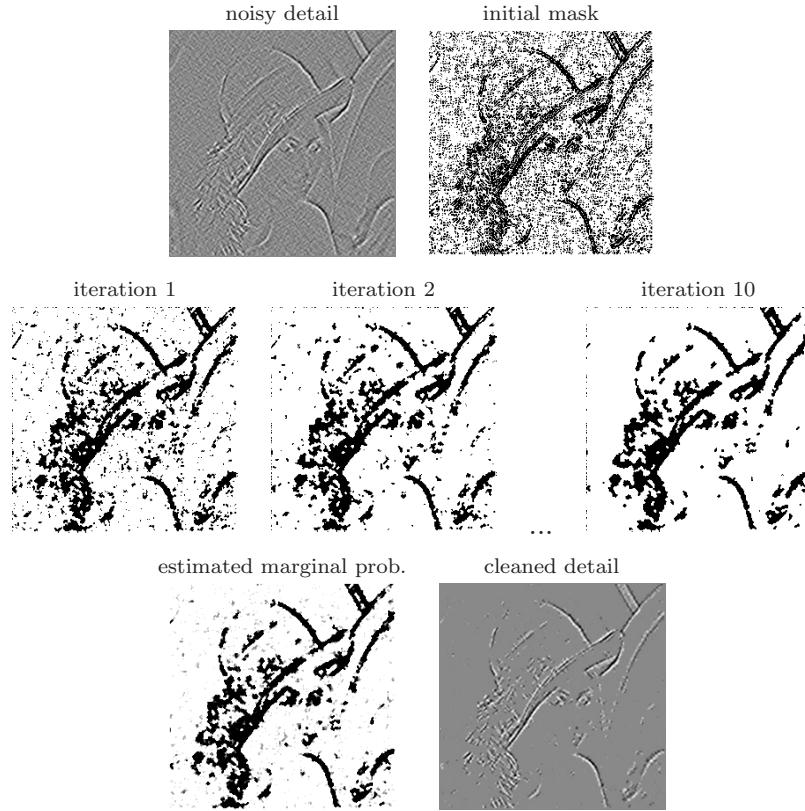


Figure 3.11: A stochastic sampling procedure, applied to the noise removal from the HH2 subband of the noisy image from Fig. 3.12.

$\pm\gamma$ is the potential of pair-wise cliques in the model (3.4.1). Practically, α and β are two constants, which reflect the relative importance given to the observed data, and to the prior model, respectively.

We choose $T = \hat{\sigma}$, where $\hat{\sigma}$ is the estimate of the noise in a given detail image computed as in Eq. (2.3.8). We have used the same threshold to form the initial mask in Fig. 3.12. After some experiments, we set $\alpha = 10$, $\beta = 8$, $\delta = 1$, and apply 10 iterations with the Metropolis sampler. The shrinkage factors $P(X_l = 1 | \mathbf{M} = \mathbf{m})$, for $l = \{1, \dots, n\}$ take in this case 11 different values and are visualized in Fig. 3.11 as a grey scale image. The same procedure was applied to all the detail images. The final denoising result is shown in Fig. 3.12 in comparison with a classical soft-thresholding in an orthogonal wavelet representation. This simple



Figure 3.12: Left: 256x256 lena image with additive white Gaussian noise of zero mean and standard deviation 20. Center: the result of the stochastic sampling algorithm, with the heuristic conditional model, PSNR=28.0dB. Right: the result of the hard thresholding in the orthogonal transform, with Donoho’s universal threshold; PSNR=25.6dB.

example demonstrates already a potential of the Bayesian shrinkage with MRF priors. Further improvements will result from more realistic conditional and from better MRF prior models, that we propose in the next Chapter.

3.5 An alternative estimation approach

As an alternative to the stochastic sampling procedure, we now introduce a simple and fast edge adaptive denoising method. The specific form of the shrinkage factor that we derive here will also inspire a broader new class of denoising techniques, in Chapter 5. Our main intention here is a *reduced complexity* of the MRF based denoising approach.

3.5.1 Motivation and idea

In certain applications, like SAR image processing, one deals with extremely large matrices of pixels. In such cases, stochastic sampling approaches may take rather long time, which motivates our search for simpler solutions.

Our reasoning behind this approach is practical: using a fast algorithm, like ICM (see Section 3.3.3), one can find a solution close to the MAP mask in a few raster scans. In our experiments a good mask is found already after one iteration; after two or at the most three iterations, the solution is stable. Naturally, the idea comes of using this mask directly for an edge-adaptive coefficient suppression (as opposed

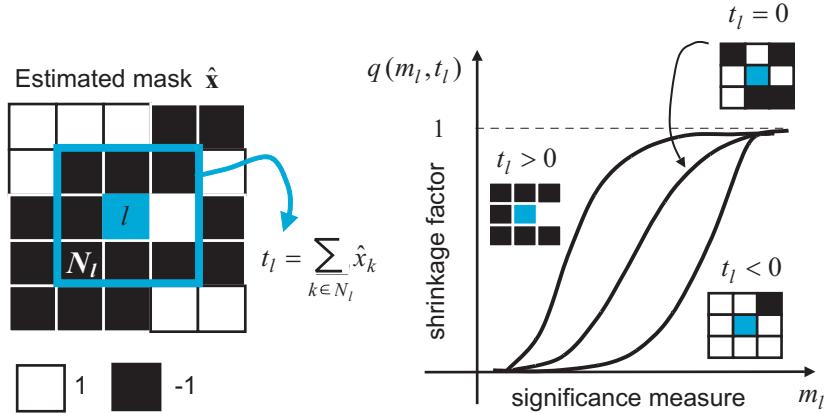


Figure 3.13: The principle of edge-adaptive wavelet shrinkage using masks.

to using a number of intermediate masks in a stochastic sampling process). We propose one possibility in this respect, where the coefficients are estimated as:

$$\hat{y}_l = q_l w_l, \quad q_l = q(m_l, t_l), \quad (3.5.1)$$

where $q(m_l, t_l)$ is a shrinkage factor, m_l the value of chosen significance measure and $t_l = t(\hat{\mathbf{x}}_{\partial l})$ a *spatial activity indicator*, computed from the mask. Fig. 3.13 illustrates this concept: we have a family of shrinkage characteristics $q(m_l)$, and the value of t_l dictates the choice of the particular curve. The local spatial context information, which is in this case derived from the binary mask, thus refines a wavelet shrinkage estimator.

3.5.2 Formal description and algorithm

We are interested in a closed form expression for the shrinkage factor $q(m_l, t_l)$ in Eq (3.5.1), which leads to a fast computation. Inspired by the idea of the ICM algorithm (Sec. 3.3.3), we propose

$$q_l = P(X_l = 1 | \mathbf{M} = \mathbf{m}, \mathbf{X}_{L \setminus l} = \hat{\mathbf{x}}_{L \setminus l}), \quad (3.5.2)$$

which has a clear interpretation and which can be easily computed. In particular, applying Eq (3.3.9), and observing that

$$P(X_l = -1 | \mathbf{m}, \hat{\mathbf{x}}_{L \setminus l}) + P(X_l = 1 | \mathbf{m}, \hat{\mathbf{x}}_{L \setminus l}) = 1, \quad (3.5.3)$$

we easily derive

$$q(m_l, t_l) = \frac{\xi_l \mu_l}{1 + \xi_l \mu_l}, \quad (3.5.4)$$

where ξ_l is the likelihood ratio and μ_l is the ratio of prior probabilities:

$$\xi_l = \frac{p_{M_l|X_k}(m_l|1)}{p_{M_l|X_l}(m_l|-1)} \quad \text{and} \quad \mu_l = \frac{P(X_l = 1|\hat{\mathbf{x}}_{\partial l})}{P(X_l = -1|\hat{\mathbf{x}}_{\partial l})}. \quad (3.5.5)$$

The particular form of the likelihood and prior ratios in Eq (3.5.5) results from the assumed statistical models. For the isotropic auto-logistic MRF model with pairwise cliques only, one can show that

$$\mu_l = \exp(\gamma t_l), \quad t_l = \sum_{k \in \partial l} \hat{x}_k \quad (3.5.6)$$

which is the case depicted in Fig. 3.13. For the auto-binomial models from [Cross83], t_l is a weighted sum of neighboring labels.

The new algorithm is:

Compute the non-decimated wavelet transform

For each orientation and for the resolution scales 2^j , $1 \leq j \leq 3$

- Run the ICM algorithm:

 - Initialize the binary mask \mathbf{x}

 - Repeat N times ($N=1$ or 2)

 - Update the labels x_l sequentially: for $l = 1 \dots N$

 - Compute the likelihood ratio ξ_l and the prior ratio μ_l from Eq (3.5.5); compute $r_l = \xi_l \mu_l$.

 - if $r_l > 1$ set $x_l = 1$; otherwise: $x_l = -1$

- For $1 \leq l \leq n$:

 - Recompute $r_l = \xi_l \mu_l$

 - Estimate the coefficients: $\hat{y}_l = [r_l/(1+r_l)]w_l$

Apply the inverse wavelet transform.

3.5.3 A heuristic shrinkage family

Here we illustrate the approach proposed above with simplified conditional models $p_{M_l|X_l}(m_l|x_l)$. This method has led to some nice practical results in humanitarian demining (Sections 6.3.3 and 6.4.3) and in hydrology (Section 3.5.4), and is thus worth mentioning. In particular, we

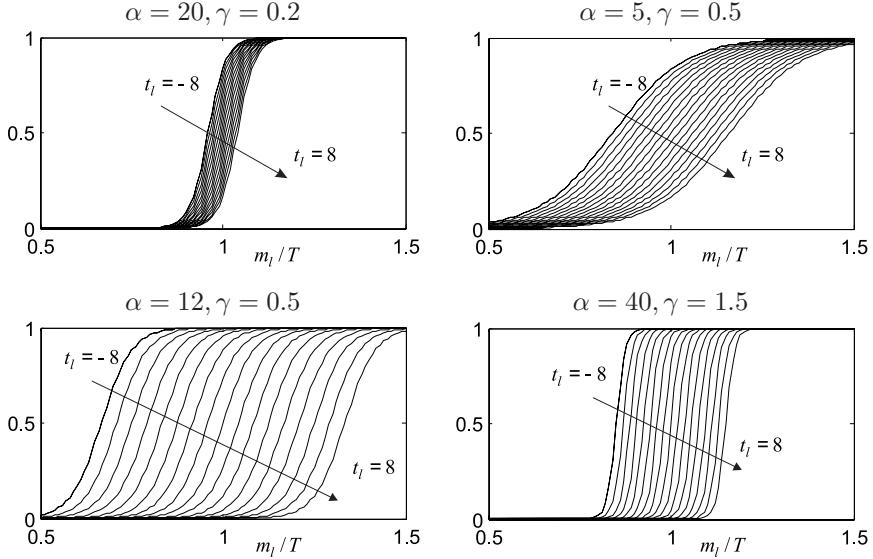


Figure 3.14: The proposed heuristic shrinkage factor $q_l = q(m_l, t_l)$ versus normalized measure m_l/T .

shall assume Malfait's conditional model from Fig. 3.8, for which the likelihood ratio ξ_l becomes

$$\xi_l = \begin{cases} \exp(-\alpha) & m_l/T < (1 - \delta), \\ \exp(\alpha(m_l/T - 1)/\delta) & (1 - \delta) \leq m_l/T \leq (1 + \delta), \\ \exp(\alpha) & m_l/T > (1 + \delta). \end{cases} \quad (3.5.7)$$

Several examples of the resulting shrinkage functions q_l from Eq (3.5.4) are shown in Fig. 3.14. It can be seen that if the value of the significance measure is equal to the chosen threshold ($m_l/T = 1$) and if the neighborhood ∂l contains the same number of edge and non-edge labels ($t_l = 0$), then the shrinkage factor $q(m_l, t_l)$ equals 0.5. Depending on its neighborhood, a coefficient can be significantly reduced even if m_l is highly above the threshold. On the other hand, if the neighborhood indicates an edge, the coefficient is less suppressed even if the value of m_l is small. The greater γ is, the greater is the influence of the neighboring labels.

Since the underlying model is heuristic, the parameters need to be experimentally optimized for a given application. We do not analyze in

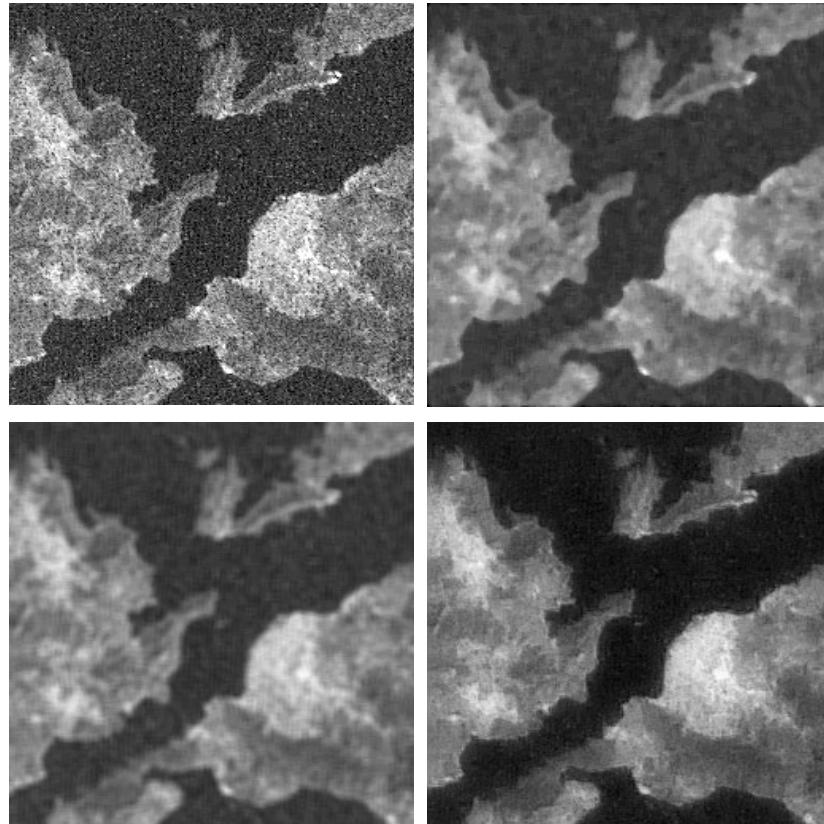


Figure 3.15: Top left: original SAR image. Top right: median filtered. Bottom left: Gamma MAP filtered. Bottom right: the result of the *HeurShrink* method.

detail the influence of these parameters on the SNR for specific noise types. Instead, we shall introduce more realistic conditional models for such cases in the following Chapters. The main advantage of the heuristic shrinkage functions of this Section, apart from their simplicity, is their adaptability to various problems through a user interaction. In practice, we used $\delta = 0.5$ and $10 \leq \alpha/\gamma \leq 50$.

Fig. 3.15 illustrates an application in remote sensing. Images are highly corrupted by speckle noise [Goodman76]. Later, in Section 5.6, we shall address this type of noise in more detail. The speckle suppression is important for further image processing tasks, like segmentation. Here we

compare visually the performance of the new method with two standard filters for speckle noise: median filter and the Gamma MAP [Lopes90] filter. The Gamma MAP filter is considered among the best state of the art filters for speckle noise suppression [Foucher01, ShiFung94]. In this case, however, its performance appears to be rather poor and not better than that of the simple median filtering. The new method clearly outperforms the two other filters: sharpness and image details are much better preserved.

3.5.4 An application in hydrology

The method described above has found an interesting application in hydrology. In particular, Verhoest [Verhoest00b] has confirmed that the proposed noise reduction method facilitates the extraction of the soil moisture information from SAR images, and that it outperforms other filters that are commonly used in this application. Here we briefly introduce the problem and demonstrate the results.

An important problem in hydrologic science is the prediction of the effect of rain on the outflow of a river. In order to predict extreme events, like floods, modeling of the spatial redistribution of soil moisture after rainfall is needed. In developing such models frequent measurements of the surface moisture are required. For studies covering large scales (which go even beyond 1000 km) remote sensing is used.

Due to its high penetration power, Synthetic Aperture Radar (SAR) is very attractive for this purpose. In [Verhoest98], a technique was proposed to separate the soil moisture information from the other physical factors that dominate the radar backscattering, such as topography and land cover. Basically, this technique applies the principal component (PC) analysis of a time series of eight European Remote Sensing (ERS) SAR images. The second principal component (PC2) is supposed to contain most of the soil moisture information. The resulting image is highly corrupted with noise, which is partially related to the speckle observed within the SAR images themselves. In order to reduce the noise and to obtain an image that is more appropriate for hydrologic modeling schemes, we have applied the method from the previous Section.

In our experiments, the PC2 image of the Zwalm catchment was used [Verhoest00b]. The Zwalm catchment is located approximately 20 km South of Ghent in Belgium; its total area is 114 km²; the topography is characterized by hills and mild slopes, with a maximum elevation difference of 150 m. The drainage map, shown in Fig. 3.16(a) was derived

Table 3.1: Classification results for the original PC2 and the denoised image from Fig. 3.16 in comparison with the corresponding result of the FKA method [VanMeirvenne00] applied to the same original PC2 image. The user's accuracy A_u and the producer's accuracy A_p are calculated according to [Lillesand94].

ACCURACY	WET AREAS		DRY AREAS	
	A_u	A_p	A_u	A_p
Original PC2 image	70.5%	66.0%	47.8%	53.0%
FKA method [VanMeirvenne00]	73.1%	58.5%	46.1%	62.3%
New method	77.1%	73.2%	58.1%	63.1%

from the digital soil map of the catchment. This drainage map classifies different soils into well-drained and poorly-drained classes: b and A are well drained soils, while e, f, g and h correspond to the poorly-drained classes. The c, D and d classes have an intermediate drainage capacity. A visual comparison between second principal component PC2 (Fig. 3.16(b)) and the drainage map (Fig. 3.16(a)) already shows that high PC2 values indeed correspond to the poorly-drained soils, while low PC2 values correspond to the well-drained uphill soils. The visual similarity with the drainage map has improved remarkably after applying the proposed filter (see Fig. 3.16(c)).

To provide an objective comparison between these figures, Verhoest [Verhoest00b] calculated the user's and the producer's accuracies defined in [Lillesand94]:

- *The user's accuracy A_u* - the probability that a pixel classified into a given category actually belongs to that category;
- *The producer's accuracy A_p* - the probability that a pixel from a given category is classified correctly.

The results are presented in Table 3.1. The Table also lists the classification results of a standard filtering technique in hydrology: factorial kriging analysis (FKA) [VanMeirvenne00]. The results demonstrate that the new wavelet domain filter offers a better performance with respect to the FKA filter: the classification results in all areas have improved; in wet areas, the producers' accuracy has increased for 14.7%; in dry areas, the users' accuracy has increased for 12%. The achieved improvement in the delineation of the spatial distribution of the wet and dry areas was found significant in [Verhoest00b].

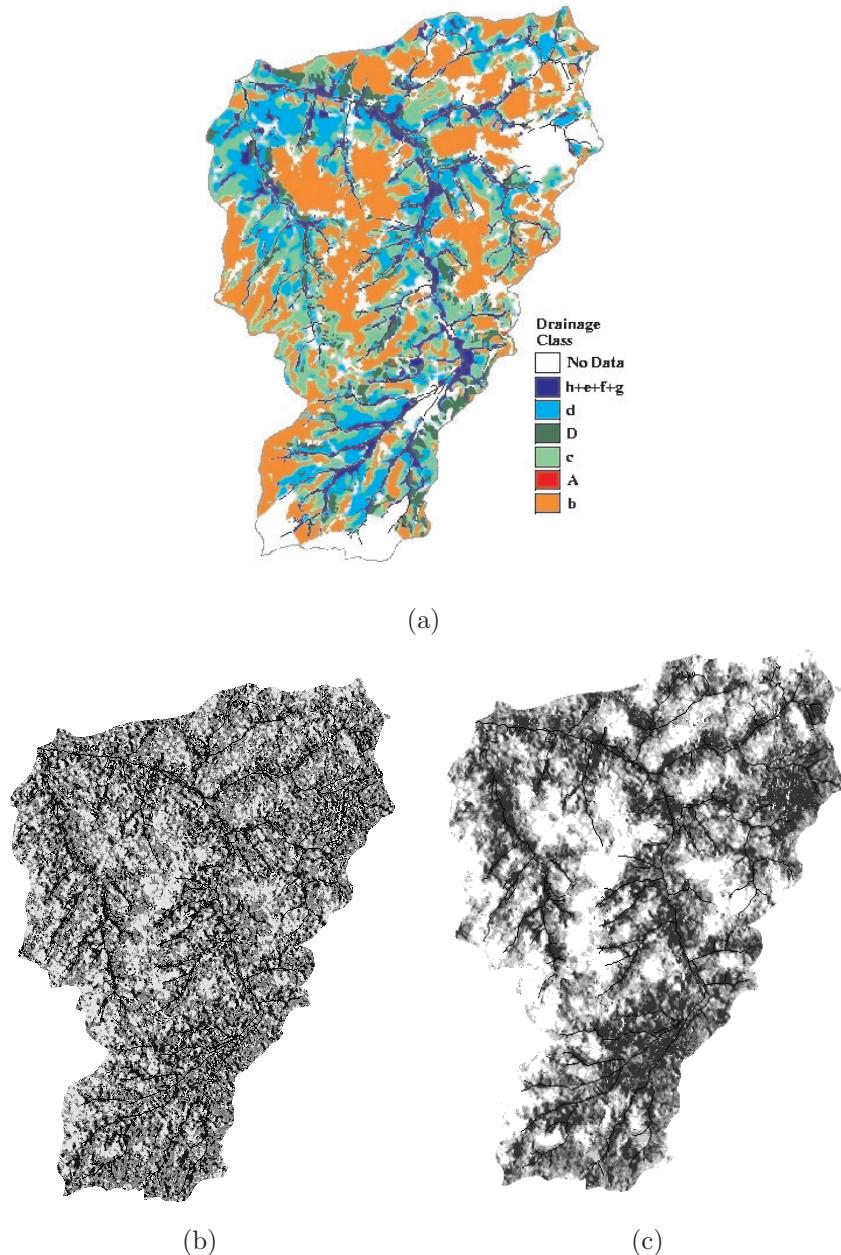


Figure 3.16: (a) Drainage map of the Zwalm catchment (ground truth) (b) Second principal component (PC2) image obtained from a time series of ERS-1/ERS-2 radar images, after [Verhoest98]. (c) The result of the proposed method applied to the PC2 image. The river network is draped on top of the three images.

3.6 Summary and conclusions

Markov random field models provide a convenient way of modeling *local* spatial interactions, as well as the *global* image context in terms of its local characteristics. We reviewed some common MRF models and explained some optimization methods, like the Metropolis sampler and the iterated conditional modes (ICM), that are commonly used in MRF based approaches.

In Section 3.4, we summarized the existing wavelet domain methods that employ MRFs. Basically, these methods use bi-level MRFs as prior models for spatial clustering and employ marginal posterior estimation using the Metropolis sampler. In order to “demystify” the formalism of these methods, we presented a pictorial illustration (Fig. 3.10) and a practical, easy-to-implement example (Section 3.4.3).

The main contribution of this Chapter is that we described a more general framework of the MRF based wavelet denoising, where the existing methods appear as one possibility. In this context, we discussed possibilities for the joint MAP estimation of the wavelet coefficients with MRF priors, which is a direct extension of the classical single-resolution results. Construction of discontinuity adaptive (DA) potentials in the wavelet domain seems a challenging topic. In Section 3.3.1, we outlined one idea in this respect. Regarding the use of bi-level MRFs, we proposed a new algorithm in Section 3.5. The main idea of this approach is to compute a *local spatial activity indicator* from the MAP mask estimate. The value of this indicator dictates the choice of a particular curve in a family of shrinkage characteristics; in this way, local spatial context information refines a wavelet shrinkage estimator. The main advantage of the new algorithm is a reduced complexity: instead of a stochastic sampling procedure, a suboptimal but fast ICM method can be used. This is interesting in cases where large images need to be processed. We demonstrated the usefulness of the new algorithm in one such application (Section 3.5.4).

In the next Chapter, our main goal is to improve the noise suppression performance, and therefore we shall use a more powerful stochastic sampling approach. In Chapter 5, we return to the idea of our new algorithm from 3.5 and develop it further and beyond the scope of MRFs.

Some parts of this Chapter appeared in [Pizurica01c]. The algorithm of Section 3.5 with various conditional models was published in [Pizurica99c, Pizurica00c, Pizurica00d], and its applications in [Pizurica00a, Pizurica00b, Verhoest00a].

Chapter 4

Statistical Modeling in MRF based wavelet denoising

This Chapter concentrates on advanced concepts in Bayesian wavelet shrinkage with MRF priors. The necessary background was already formed. Now we are aiming at an improved noise suppression performance through statistical modeling. The contents of the Chapter are original and present our own contributions in this field. The main novelties with respect to related approaches are: (1) the interscale-ratios of wavelet coefficients are statistically characterized, and different local criteria for distinguishing useful coefficients from noise are evaluated; (2) a joint conditional model is introduced, and (3) a novel anisotropic Markov Random Field prior model is proposed. The results demonstrate an improved denoising performance over related existing approaches.

4.1 Introduction

The previous Chapter has established the necessary background theory for wavelet domain denoising using prior models for spatial clustering. Fig. 4.1 summarizes pictorially the basic concept of this approach. Now we concentrate on the following issues:

Significance measures - The main novelties are:

a statistical characterization of the interscale-ratios of the wavelet coefficients; an evaluation of different local criteria for distinguishing useful coefficients from noise; a joint conditional model, which combines coefficient magnitudes and their evolution across scales.

We start from different interscale-ratio formulations, [Malfait97], [Hsung99], and unify them in a general framework. Instead of using heuristic models for the ratios, we determine empirically their conditional probability densities in regions containing pure noise, and in edge regions, respectively. Using the empirical densities and employing a statistical decision theory, we compare the performance of different significance measures. Such a performance evaluation clearly motivates a joint significance measure, which relies on both coefficient magnitudes and on their evolution across scales. The resulting joint conditional model offers a superior denoising performance with respect to earlier ones that use only interscale ratios [Malfait97], or only coefficient magnitudes [Jansen01a].

A new Markov Random Field prior model for spatial clustering

We propose an original anisotropic MRF prior model. As compared to the isotropic model from [Malfait97], our model, which is slightly more complex, preserves image details significantly better. We use the same formal description as in Eq (4.1.1), so the joint label distribution is still given by

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp\left(-\sum_{l \in L} V_{N_l}(\mathbf{x})\right), \quad (4.1.1)$$

but with a new form of $V_{N_l}(\mathbf{x})$, which adapts to the presence of edge segments in a given neighborhood.

New prior and conditional models are combined into an upgraded stochastic sampling procedure. We adopt the estimation approach from Section 3.4.2: $\hat{y}_l = P(X_l = 1 | \mathbf{M} = \mathbf{m}) w_l$, and run the stochastic sampling procedure, which was depicted in Fig. 3.10. We shall rewrite the ratio p of prior probabilities from Eq (3.4.6) simply as

$$p = \frac{p_{M_l|X_l}(m_l|x_l^c)}{p_{M_l|X_l}(m_l|x_l)} \exp\left(V_{N_l}(\mathbf{x}) - V_{N_l}(\mathbf{x}^c)\right). \quad (4.1.2)$$

We use a non-decimated wavelet transform with the quadratic spline wavelet from Fig. 2.10, and consider both the decomposition with two orientation subbands from [Mallat92b] (Section 2.2.9) and the classical decomposition with three orientation subbands. The two-subband decomposition is often used for edge detection, local regularity characterization and denoising [Mallat92a, Hsung99]. We find it attractive from the computational point of view: e.g., the stochastic sampling runs

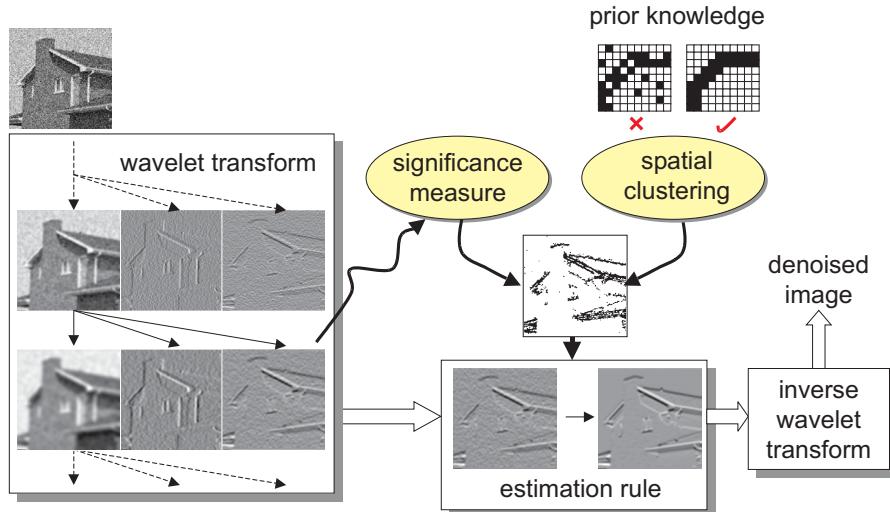


Figure 4.1: A MRF based approach that we now further develop.

on 2 instead on 3 detail images per scale and is thus 50% faster. It is therefore interesting to compare the noise suppression of the method for the two and three subband case. We present the results for both. For the illustrations throughout the Chapter, the 2-subband decomposition was used.

The Chapter is organized as follows. In Section 4.2, we first propose a common and slightly more general formulation for different previously used interscale ratios. Then we present a simulation method for characterizing statistical properties of these and possibly other significance measures. Further on, we conduct a performance evaluation. The results motivate clearly a joint significance measure, which relies on both the coefficient magnitudes and their evolution across scales. We therefore propose the corresponding joint conditional model, and describe it both conceptually and in terms of a practical application. In Section 4.3, we propose a new MRF prior model. In Section 4.4, we describe the upgraded stochastic sampling algorithm, which uses the proposed prior and the conditional models. Then we evaluate the influence of those models on the noise suppression performance and on the visual quality of images. Further on, we compare the performance of the method (in terms of SNR and visually) in cases when two and when three orientation subbands are used. We then summarize and discuss the quantitative results

of the method in comparison with related approaches from literature. In Section 4.5, we summarize the main conclusions.

4.2 Significance measures

For optimal performance of the MRF based denoising approach from Fig. 4.1, the choice of the significance measure m_l and the characterization of its conditional densities $p_{M_l|X_l}(m_l|x_l)$ are important. Earlier models proposed in literature [Malfit97, Jansen01a] were explained in Section 3.4.1 (see Fig. 3.8 and Fig. 3.9 and the related text). In summary, m_l was chosen either as an interscale ratio, with a heuristic and parametrized conditional model, or m_l was chosen as the coefficient magnitude, in which case more realistic probability models were considered. To our knowledge, there has been no attempt to quantitatively compare the performance of these two and possibly other significance measures m_l . Also, for interscale ratios real densities were not examined.

We note that two different formulations of interscale ratios have been used in recent literature, albeit with the same intention: to estimate roughly the local regularity of an image and accordingly to make a distinction between useful edges and noise. One formulation [Malfit97], was given in Eq (3.4.3), and the other one [Hsung99], was given in Eq (2.5.4). In the first case, the ratio $|w_{j+1,l}/w_{j,l}|$ was averaged over a given number of scales, yielding the estimate of 2^α , where α is the local Lipschitz exponent; in the second case, averaging through scales was not considered, but the cone of influence was taken into account, and the estimator of $2^{\alpha+1}$ was proposed.

4.2.1 A discretized approximation of interscale ratios

We formulate the two mentioned interscale ratios from Eqs (3.4.3) and (2.5.4) in a unifying and slightly more general way. Let us define for each spatial position l the coefficient $\alpha_{n \rightarrow k, l}$ which determines the average rate of increase of the magnitudes of the wavelet coefficients between any two dyadic scales 2^n and 2^k , where $n, k \in \mathbb{Z}$ and $k \geq n + 1$

$$\alpha_{n \rightarrow k, l} \triangleq \log_2 \left(\frac{1}{k-n} \sum_{j=n}^{k-1} \frac{|w_{j+1,l}|}{|w_{j,l}|} \right). \quad (4.2.1)$$

The logarithm in Eq. (4.2.1) is used in order to make $\alpha_{n \rightarrow k, l}$ behave as a rough estimate of the local Lipschitz exponent α . This quantity describes the evolution of the individual wavelet coefficients at the spatial

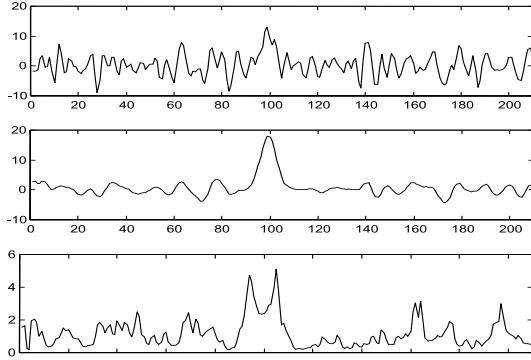


Figure 4.2: An illustration of the “rabbit ears” effect, associated with interscale ratios. Top to bottom: details at two successive scales and the corresponding interscale ratio.

position l . Similarly we define a second quantity $\beta_{n \rightarrow k, l}$, which describes the “collective” evolution of the wavelet coefficients inside a cone of influence centered at the spatial position l . We denote by $C(j, l)$ the discrete set of wavelet coefficients at the resolution scale 2^j , which belong to the cone of influence of the spatial position l , and we define $\beta_{n \rightarrow k, l}$ as

$$\beta_{n \rightarrow k, l} \triangleq \log_2 \left(\frac{1}{k - n} \sum_{j=n}^{k-1} \frac{|I_{j+1, l}|}{|I_{j, l}|} \right), \quad I_{j, l} \triangleq \sum_{m \in C(j, l)} |w_{j, m}|, \quad (4.2.2)$$

which is an estimate of $\alpha + 1$. We call the quantity $\alpha_{n \rightarrow k, l}$ the *average point ratio* (APR) and the quantity $\beta_{n \rightarrow k, l}$ the *average cone ratio* (ACR).

In the case of images, $C(j, l)$ refers to a directional cone of influence (DCOI), similar to the one in Fig. 2.16. Note that in contrast to the formulation from Section 2.5.3 (see also the remark there), we operate on each detail orientation separately; the direction of the cone of influence is now simply perpendicular to the subband orientation.

Fig. 4.2 illustrates an interscale ratio: at positions of useful edges characteristic pulses with “rabbit ears” occur. Due to this, the coefficients adjacent to edges are falsely detected as significant. In order to avoid false broadening of the estimated regions with significant edges, one should avoid averaging of these ratios over a large number of scales.

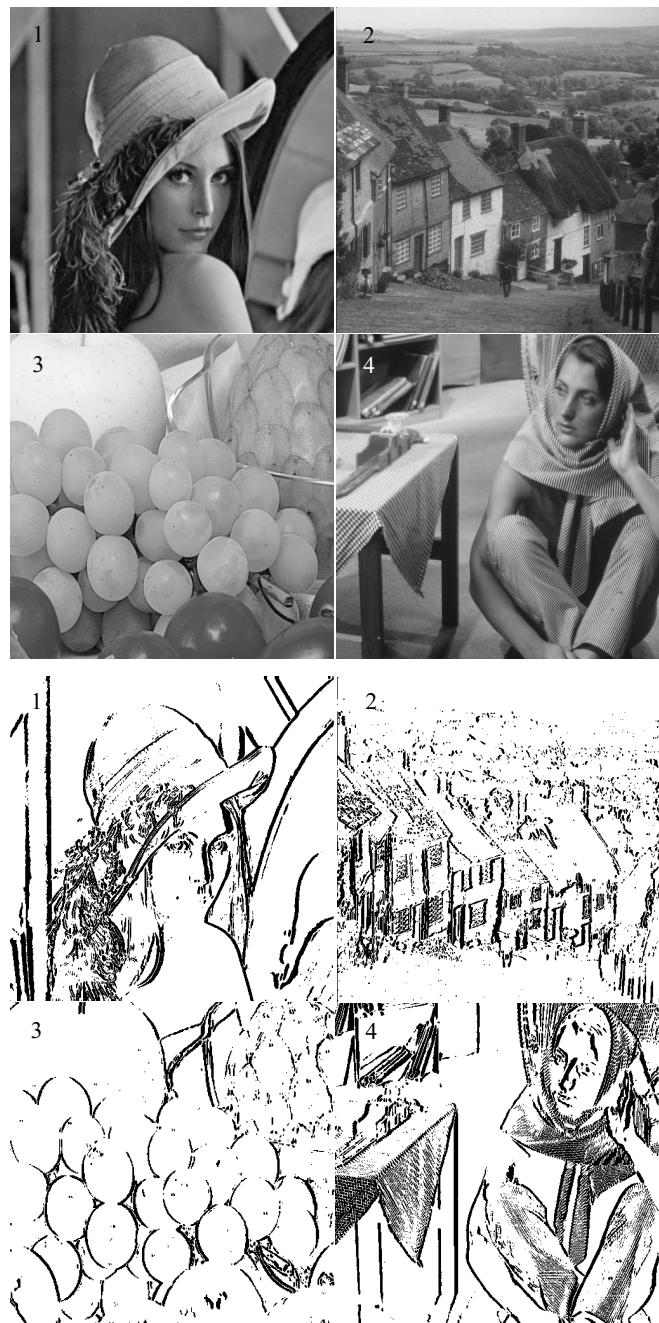


Figure 4.3: Top: reference images: 1 - “Lena”, 2 - “Goldhill”, 3 - “Fruits”, 4 - “Barbara”. Bottom: reference edge positions for vertical orientation of details at resolution scale 2^2 .

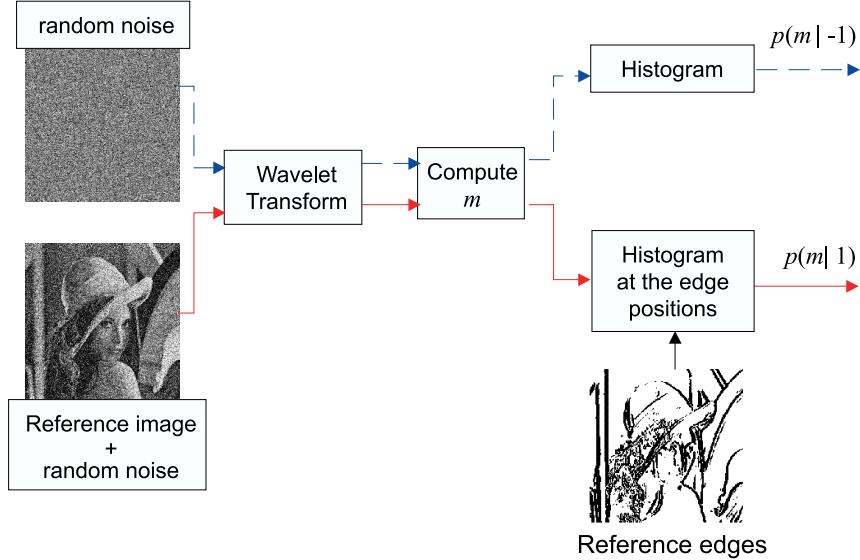


Figure 4.4: The simulation setup in our experiments.

4.2.2 Statistical characterization via simulation

Now we introduce a simulation method to determine the empirical conditional densities $p_{M_l|X_l}(m_l|x_l)$ of any chosen significance measure m_l and for an arbitrary type of noise. Practically we restrict ourselves to additive white Gaussian noise. The benefit of knowing the empirical densities is twofold. Firstly, the performance of different significance measures can be objectively compared relying on the statistical estimation theory [VanTrees68]. Secondly, the empirical densities can be employed in an actual denoising procedure instead of the heuristic models, which should lead to a more accurate estimation.

To determine $p_{M_l|X_l}(m_l|1)$ one needs a statistical model for the actual significant image discontinuities, i.e., the “ground truth” edges. To achieve realistic results, we extract these discontinuities from various natural noise-free images, like those in Fig. 4.3. Edge quality measures are discussed, e.g., in [Abdou79, Bryant79, Kitchen81, Strickland93]. In our approach, for a given reference image, at each resolution scale and orientation the reference edge positions are obtained by thresholding the magnitudes of noise-free wavelet coefficients. For characterizing densities of interscale ratios, the choice of these thresholds is not critical. In particular, we choose a threshold for each resolution scale that is equal

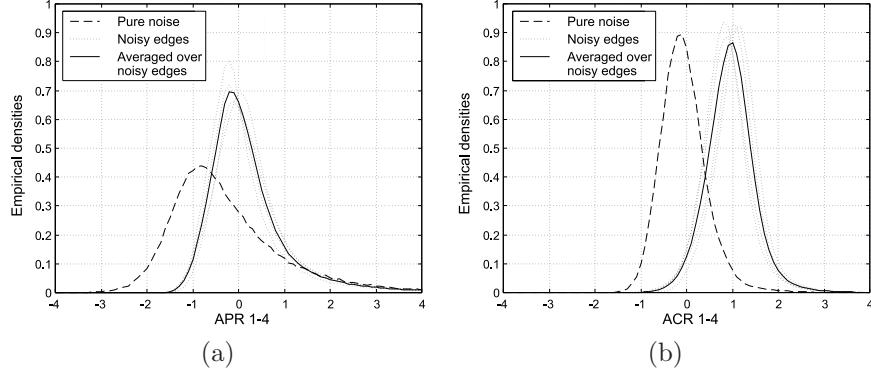


Figure 4.5: Conditional densities of (a) APR and (b) ACR, computed from scales $2^1 - 2^4$. Different reference edges from Fig. 4.3 are used. The standard deviation of added noise is $\sigma = 25$.

to the magnitude of a wavelet coefficient produced by an ideal step edge of amplitude A , where A equals to $1/16$ of the dynamic range of the image.

The complete simulation procedure is depicted in Fig. 4.4. To estimate $p_{M_l|X_l}(m_l|1)$, we use an image which consists of pure noise. We compute the discrete wavelet transform of this image and find the histogram of the given significance measure m_l . To estimate $p_{M_l|X_l}(m_l|1)$, we apply a similar procedure, except that the noise is added to the reference image and m_l is computed only in the reference positions of significant coefficients. The whole procedure is iteratively repeated, adding every time random noise with the same variance to the noise-free image. Therefore, we refer to this method as simulation.

Conditional densities of APR and ACR

The conditional densities $p_{A_l|X_l}(\alpha_{n \rightarrow k,l}|1)$ and $p_{B_l|X_l}(\beta_{n \rightarrow k,l}|1)$, of APR and ACR, respectively, given pure noise do not depend on the noise variance: the increase of noise affects the magnitudes of the wavelet coefficients at all scales equally and the statistical distribution of their averaged ratios does not change. These conditional densities, shown as dashed curves in Fig. 4.5(a) and Fig. 4.5(b), peak in the vicinity of -1 and 0, respectively. The empirical densities $p_{A_l|X_l}(\alpha_{n \rightarrow k,l}|1)$ and $p_{B_l|X_l}(\beta_{n \rightarrow k,l}|1)$, for different reference edges from Fig. 4.3 are also illustrated in Fig. 4.5(a) and Fig. 4.5(b), and peak in the vicinity of 0 and 1, respectively. We have verified that the choice of a threshold that

specifies the reference edges in Fig. 4.3 is not critical for characterizing these densities.

It should be noticed that the overlap between $p_{B_l|X_l}(\beta_{n \rightarrow k,l} - 1)$ and $p_{B_l|X_l}(\beta_{n \rightarrow k,l}|1)$ is smaller than the overlap between $p_{A_l|X_l}(\alpha_{n \rightarrow k,l} - 1)$ and $p_{A_l|X_l}(\alpha_{n \rightarrow k,l}|1)$. It suggests that ACR provides a better separation between noise and useful edges than APR. We examine this further in Section 4.2.3. The robustness of interscale ratio statistics with respect to noise level is illustrated in the left part of Fig. 4.6. One can see that the overlap between conditional densities of ACR given noise and given edges does not change much as the noise variance increases.

Conditional densities of the coefficient magnitudes

The right part of Fig. 4.6 illustrates densities $p_{M_l|X_l}(m_l|x_l)$, for the case where m_l is the magnitude of the wavelet coefficient, which we denote by ω_l . In case of white Gaussian noise, there is no need to simulate conditional densities of pure noise; obviously one has

$$p_{\Omega_l|X_l}(\omega_l|1) = \frac{2}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{\omega_l^2}{2\sigma_n^2}\right), \quad \omega_l > 0, \quad (4.2.3)$$

where σ_n is the standard deviation of the noise in the given detail image. For $p_{\Omega_l|X_l}(\omega_l|1)$, close approximations can be computed directly from the noisy image itself (Section 4.2.4). For comparative performance analysis with respect to interscale ratios, we use empirical densities $p_{\Omega_l|X_l}(\omega_l|1)$.

It is interesting to compare how the overlap between the conditional densities of ω_l in Fig. 4.6 evolves with increasing of noise level, with respect to that of ACR that is shown in the same figure. A simple visual inspection suggests that for relatively small standard deviations of noise the coefficient magnitude provides a better separation between noise and useful signal, whereas the opposite is true for high noise levels. This is further investigated in Section 4.2.3. It is not shown here, but one can easily extrapolate, that if a different threshold was chosen for defining the edges of interest in Fig. 4.3, the densities $p_{\Omega_l|X_l}(\omega_l|1)$ are peaked at correspondingly greater or smaller values ω_l , but important is that their qualitative behaviour does not change.

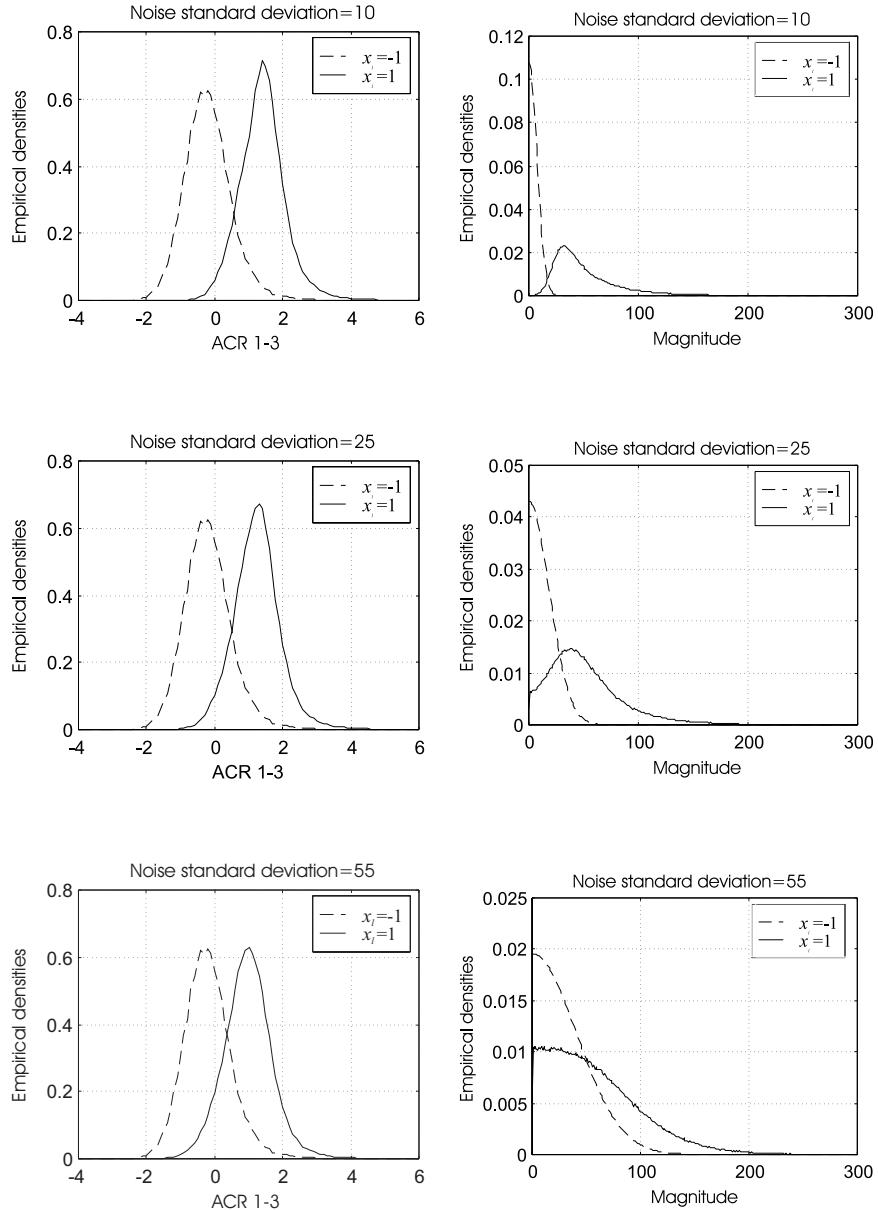


Figure 4.6: Comparison between empirical densities of ACR computed from scales $2^1 - 2^3$ (left) and empirical densities of coefficient magnitudes at scale 2^2 (right), for reference edges 1 from Fig. 4.3 and for different noise levels.

4.2.3 Performance Evaluation

By performance of a given significance measure we mean its ability to distinguish between noise and useful signal without making use of the prior model. This is equivalent to analyzing the performance under the Bayes labeling [Duda73, VanTrees68] with $P(X_l = -1) = P(X_l = 1) = 0.5$. We shall compute the receiver operating characteristics (ROC) for a given standard deviation σ of added noise, and the error probabilities depending on σ . The ROC curves plotted in this thesis show the fraction of the labels $x_l = 1$ that are wrongly classified as $x_l = -1$ (false negatives, FN), versus the fraction of the labels $x_l = -1$ that are wrongly classified as $x_l = 1$ (false positives, FP). A customary definition is slightly different¹. For $P(X_l = -1) = P(X_l = 1)$, FN and FP are given by

$$FN = \int_{-\infty}^T p_{M_l|X_l}(m_l|1) dm_l \quad \text{and} \quad FP = \int_T^{\infty} p_{M_l|X_l}(m_l|1) dm_l. \quad (4.2.4)$$

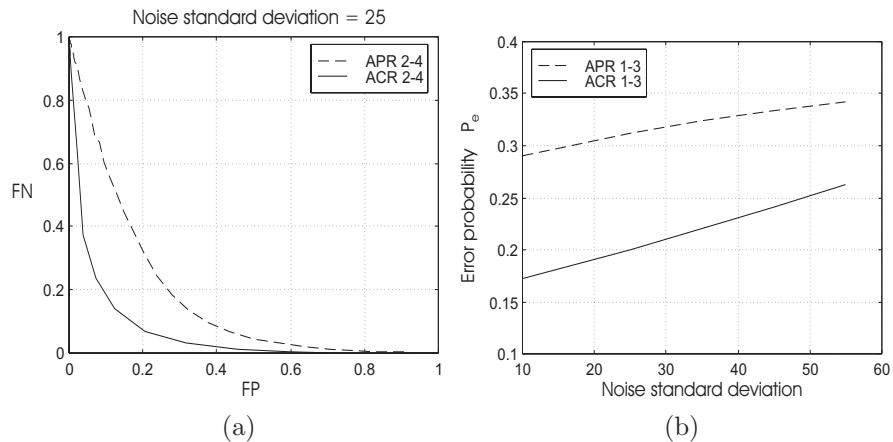
It can be shown that the optimal decision threshold T_{opt} (for which the total probability of wrongly classified labels is minimum) is the one for which $p_{M_l|X_l}(T_{opt}|1) = p_{M_l|X_l}(T_{opt}| -1)$. For the case $T = T_{opt}$, we shall compute the error probability $P_e = 0.5(FN + FP)$.

Recall that we have derived the empirical densities $p_{M_l|X_l}(m_l|1)$ from realistic edges corrupted by noise, and $p_{M_l|X_l}(m_l| -1)$ from pure noise. Therefore, FN, FP and P_e that we consider refer to the ability to distinguish realistic noisy edges from *pure noise*. We call this the ideal performance (Section 4.2.3). The actual performance for different test images, where instead of pure noise one deals with a combination of *noise and image texture* is addressed later in this Section.

Ideal performance

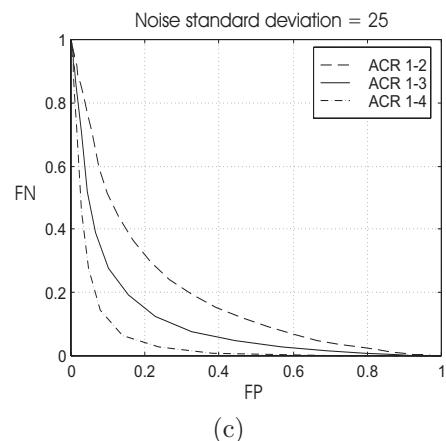
Using the empirical densities derived from the test images in Fig. 4.3, we compare first the two interscale ratios: APR and ACR. When discussing Fig. 4.5, we have already remarked that ACR should perform better. The computed ROC in Fig. 4.7(a) and error probabilities as a function of σ in Fig. 4.7(b) confirm this observation. Fig. 4.7(c) illustrates that the expected performance of interscale ratios improve, when they are averaged over a greater number of resolution scales.

¹By a classical definition [VanTrees68] the ROC plots true positives TP=1-FN versus false positives FP, i.e., it plots the probability of detection versus the probability of a false alarm.



(a)

(b)



(c)

Figure 4.7: (a) ROC for APR and for ACR, calculated from the scales $2^2 - 2^4$.
 (b) Error probabilities for optimal decision thresholds versus standard deviation of noise. (c) ROC for ACR averaged over different scales.

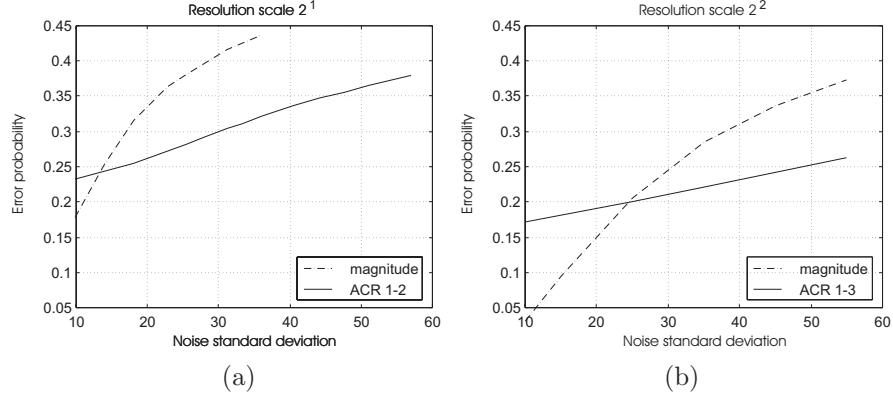


Figure 4.8: A comparative performance of ACR and coefficient magnitudes at the scales (a) 2^1 and (b) 2^2 .

Let us now compare the performance of ACR (which was the better one among the two considered interscale-ratios) and the magnitudes of the wavelet coefficients. Fig. 4.8(a) illustrates that at the finest resolution scale 2^1 , ACR offers a better performance than the coefficient magnitude even for small noise levels. However, this is not the case at coarser resolution scales. At the scale 2^2 , as shown in Fig. 4.8(b), ACR offers a better performance only for relatively high standard deviations σ of noise.

Actual performance

The above presented ideal performance comparison between different significance measures relies on the empirical densities $p_{M_l|X_l}(m_l|x_l)$ and refers to the distinction between noisy edges and pure noise. To investigate the performance in practice, where a non-edge region not only contains noise but also (partly) texture, we apply a decision threshold directly to the computed significance map \mathbf{m} , and actually count the number of misclassified labels, with respect to the assumed “ground truth” edges. A reference edge label at the position l is denoted by x_l^R , and the one that results from thresholding m_l is denoted by x_l^T . The empirical FP and FN, are then

$$\begin{aligned} FP_e &= \#\{l | x_l^R = -1, x_l^T = 1\} / \#\{l | x_l^R = -1\}, \\ FN_e &= \#\{l | x_l^R = 1, x_l^T = -1\} / \#\{l | x_l^R = 1\}, \end{aligned} \quad (4.2.5)$$

where $\#S$ denotes the cardinality of the set S . Now, FN_e and FP_e , refer to the separation between realistic edges on the one hand and a combination of noise and image texture on the other hand. For the coefficient magnitudes, the ideal and real distribution hardly differ. For the interscale ratios, the real FN_e and FP_e are greater than those that were derived from the estimated densities, and the difference is image-dependent.

In Fig. 4.9, we display the empirical error probabilities computed as $(FN_e + FP_e)/2$ for a range of decision thresholds. One can see that the main conclusions regarding the comparison between the three significance measures remain valid on these empirical curves. Also, it can be seen that for interscale ratios, a well chosen decision threshold provides nearly optimal performance for different images and for different noise levels, as was expected according to the empirical densities.

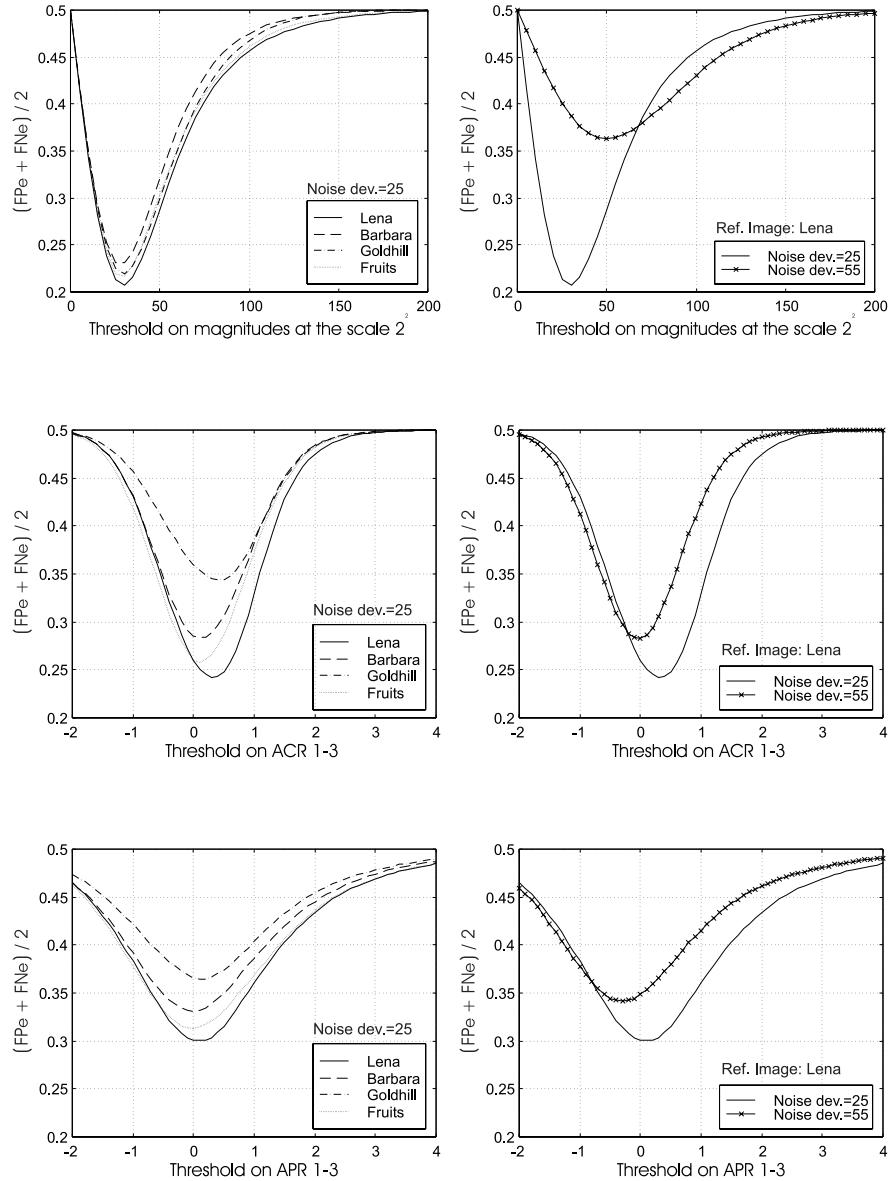


Figure 4.9: Empirical error probabilities computed as $(FN_e + FP_e)/2$. Left: for different images and the same noise deviation. Right: for the same image and different noise deviations.

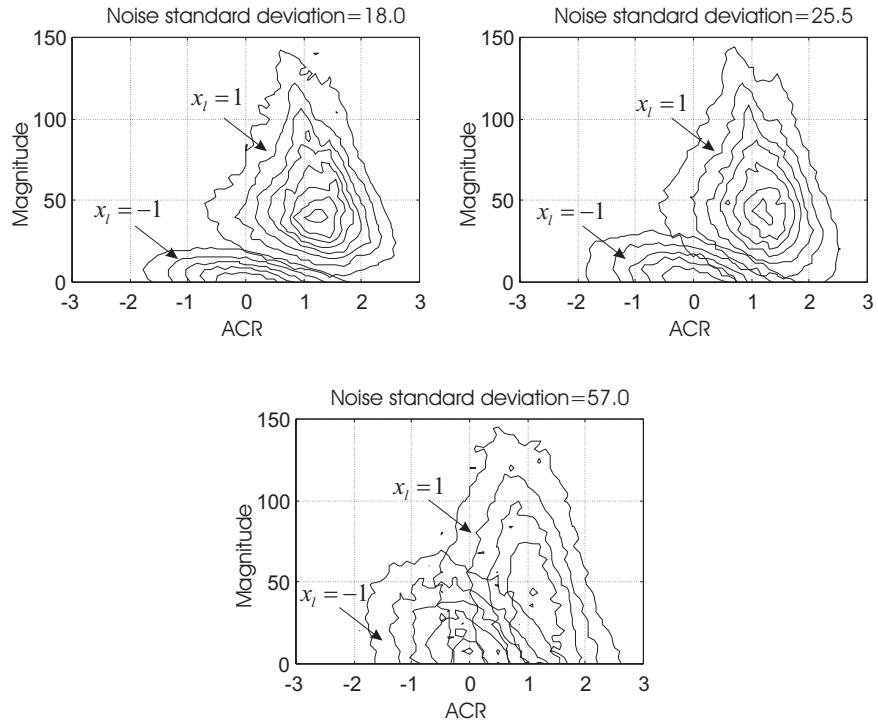


Figure 4.10: Empirical joint densities at the scale 2^2 , for different standard deviations of input noise.

4.2.4 A joint significance measure

On the basis of the previous analysis, we now propose a joint significance measure m_l , which relies on both the coefficient magnitude ω_l and on the interscale ratio at the corresponding spatial position. It was shown that ACR offers a better performance than APR. Therefore, for wavelet coefficients at resolution scale 2^j we compute ACR $\beta_{1 \rightarrow j+1, l}$. It was the best choice according to our experiments: averaging ratios over scales coarser than 2^{j+1} suppresses background noise better, but increases the number of falsely selected coefficients in regions adjacent to image edges. To simplify notations, in the remainder we use only the location index l . The proposed significance measure is thus $m_l = (\omega_l, \beta_l)$, where ω_l and β_l are computed from the observed wavelet coefficients, and are realizations of random variables Ω_l and B_l , respectively.

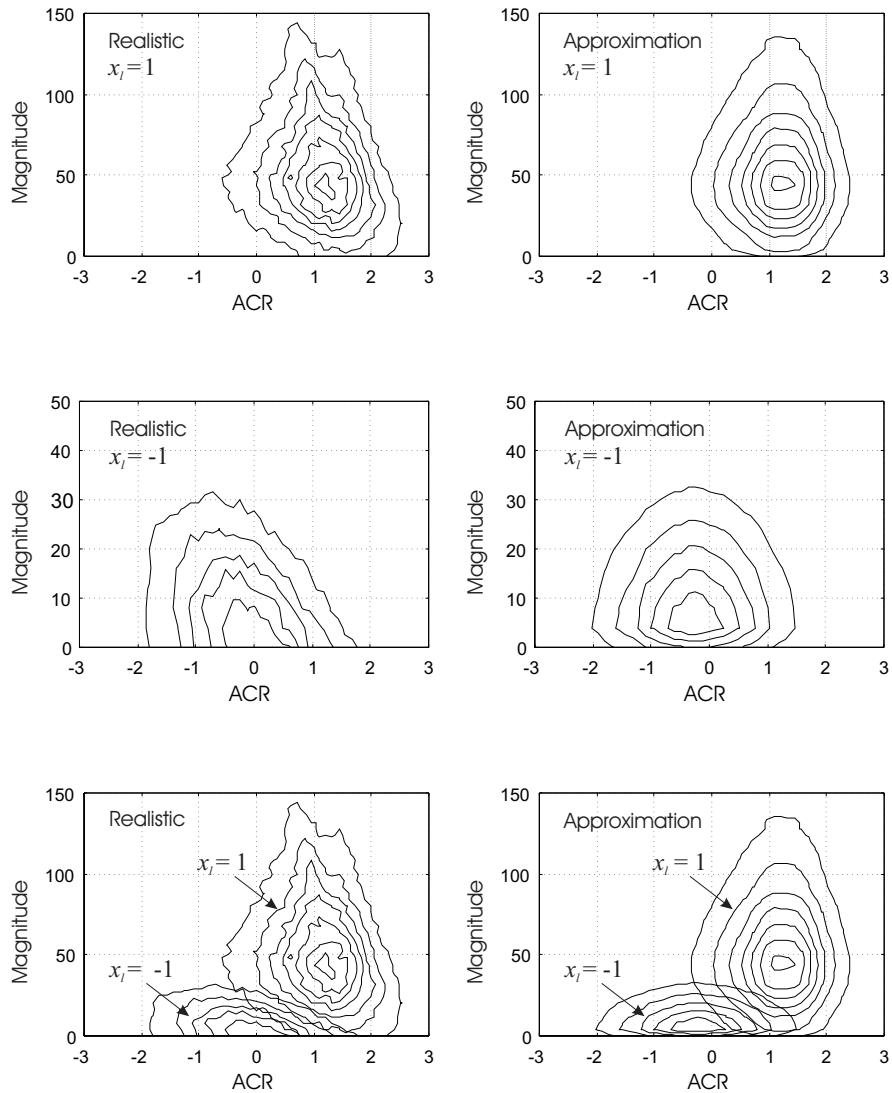


Figure 4.11: Joint densities for ACR and coefficient magnitudes at the scale 2^2 , for $\sigma = 25.5$. Left: Empirical joint densities. Right: Products of empirical one dimensional densities.

We still need to specify conditional densities $p_{M_l|X_l}(m_l|x_l)$ for the joint significance measure. Using the simulation method from Section 4.2.2, we obtain its empirical conditional densities, examples of which are given in Fig. 4.10. From a practical point of view, it is convenient to assume that β_l and ω_l are conditionally independent given x_l ,

$$p_{M_l|X_l}(m_l|x_l) = p_{\Omega_l|X_l}(\omega_l|x_l)p_{B_l|X_l}(\beta_l|x_l). \quad (4.2.6)$$

To investigate the validity of this assumption, we compare the actual joint conditional densities, $p_{M_l|X_l}(\omega_l, \beta_l|x_l)$, with products of the corresponding one-dimensional densities adapted for the two-dimensional case. Fig 4.11 shows the contour plots of the empirical joint densities in comparison with the product of two independent densities from Eq (4.2.6). These diagrams suggest that the assumption about the conditional independence is not true, but is acceptable as an approximation.

With the proposed conditional probability model (4.2.6), the parameter p in Eq(4.1.2) becomes

$$p = \left(\frac{p_{\Omega_l|X_l}(\omega_l|x_l^c)}{p_{\Omega_l|X_l}(\omega_l|x_l)} \right)^{\lambda_1} \left(\frac{p_{B_l|X_l}(\beta_l|x_l^c)}{p_{B_l|X_l}(\beta_l|x_l)} \right)^{\lambda_2} e^{\lambda_3(V_{N_l}(\mathbf{x}) - V_{N_l}(\mathbf{x}^c))}, \quad (4.2.7)$$

where $\lambda_1 = \lambda_2 = 1$ and λ_3 is a positive constant. In practice, it may be useful to allow other values of λ_1 and λ_2 in order to change the relative importance given to ω_l and β_l . For example, in the special case where $\lambda_1 = 0$, this model reduces roughly to the one in [Malfit97], except for the fact that there APR was used instead of ACR, and that another, heuristic density model was assumed. For $\lambda_2 = 0$ the joint conditional model basically reduces to the one of [Jansen99] and [Jansen01a], again with a slight difference that here more realistic densities of coefficient magnitudes will be actually implemented. The next Section explains the details regarding the implementation of those densities in our practical algorithm. The influence of different values λ_1 , λ_2 and λ_3 on denoising results is demonstrated in Section 4.4.2.

4.2.5 Implementation details

In an actual denoising procedure, for interscale ratios, the densities of which are robust with respect to noise level, we exploit the empirical results. We have verified that using empirical ratios of ACR densities $p_{B_l|X_l}(\beta_l|1)/p_{B_l|X_l}(\beta_l|-1)$ derived from different reference edges and for

different noise variance, does not affect denoising performance noticeably. Therefore, in our practical algorithm, one ratio

$$p_{B_l|X_l}(\beta_l|1)/p_{B_l|X_l}(\beta_l|-1)$$

was fitted from the empirical densities, stored and used for all the results presented in this thesis.

For $p_{\Omega_l|X_l}(\omega_l|-1)$, storing the empirical curves obviously makes no sense, since one can easily compute their close approximation directly from the observed image. Since white Gaussian noise is assumed, $p_{\Omega_l|X_l}(\omega_l|-1) = 2p_N(\omega_l)$, $\omega_l \geq 0$, where $p_N(n) = \mathcal{N}(0, \sigma_n)$, and σ_n is the noise standard deviation in the given subband. In case where the standard deviation of input noise is not known, a robust estimate from Eq (2.3.8) is used. To find $p_{\Omega_l|X_l}(\omega_l|1)$, we use a related model to that of Jansen *et al* from Section 3.4.1; the difference is that instead of approximating the underlying pdf of noise free coefficients $p_Y(y)$ by a uniform distribution, we now use a realistic, generalized Laplacian model $p_{Y_l}(y_l) \propto \exp(-y_l/s)^\nu$. A global approach is thus: the density of the significant noise-free wavelet coefficients $p_{Y_l|X_l}(y_l|1)$ is equal to zero for $|y_l| < \sigma_n$, and proportional to $\exp(-y_l/s)^\nu$ for $|y_l| > \sigma_n$. Since $w_l = y_l + n_l$, where y_l and n_l are statistically independent, the density of w_l is given by convolution [Papoulis84] $p_{W_l|X_l}(w_l|1) = p_{Y_l|X_l}(y_l|1) * p_N(n)$, and it is straightforward to show that $p_{\Omega_l|X_l}(\omega_l|1) = 2p_{W_l|X_l}(\omega_l|1)$, $\omega_l \geq 0$.

The parameters s and ν of the generalized Laplacian model $p_Y(y)$ are computed according to Eq(2.6.8) from the noisy coefficients' histogram, using its standard deviation σ_w^2 , and its fourth moment $m_{4,w}$. One can apply different numerical procedures to solve the above equations. Our implementation is summarized below. We first derive from Eq(2.6.8) the following two expressions

$$\kappa_y = \frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})} = \frac{m_4 + 3\sigma_n^4 - 6\sigma_n^2\sigma_w^2}{(\sigma_w^2 - \sigma_n^2)^2}, \quad (4.2.8)$$

and

$$s = \left((\sigma_w^2 - \sigma_n^2) \frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})} \right)^{\frac{1}{2}}. \quad (4.2.9)$$

From Eq (4.2.8), we solve the parameter ν numerically (which is easy since $\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})/\Gamma^2(\frac{3}{\nu})$ is a monotonic decreasing function of ν). From the estimated shape parameter ν , the scale parameter s follows directly, using Eq (4.2.9).

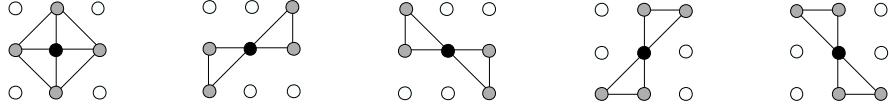


Figure 4.12: The proposed sub-neighborhoods.

This completes the specification of the proposed conditional model, both conceptually and in terms of its practical computation. Now we turn to the specification of the prior model.

4.3 A New MRF Prior Model

The simplest isotropic MRF model is not well suited to encode prior knowledge about the spatial clustering of the wavelet coefficients. Significant coefficients form relatively narrow clusters with predominantly horizontal, vertical or diagonal directions, depending on the orientation subband. At coarser scales, the isotropic model with a small neighborhood still performs well, because the neighborhoods are small compared to the width of the edges. However, the problem arises at the finest scales, where details are very thin lines.

The idea behind our model is the following: for each spatial position l , we define a given number of oriented sub-neighborhoods, which contain possible edge segments centered at the position l . The binary label value $x_l = 1$ (edge) should be assigned a high probability if *any* of the oriented sub-neighborhoods indicates the existence of an edge element in a certain direction. On the contrary, the binary label value $x_l = -1$ (non-edge) should be assigned a high probability only if *no one* of the oriented neighborhoods indicates the existence of such an edge element.

4.3.1 A class of anisotropic potentials

One, but not the only possible, realization of the above explained idea follows. We choose the sub-neighborhoods $N_{l,i}$, $1 \leq i \leq 5$ shown in Fig. 4.12; each $N_{l,i}$ contains four neighbors of the central pixel l . We use only pair-wise cliques $\{k, l\}$, with the potential function $V_2(k, l) = -\gamma x_k x_l$, where γ is a positive constant. The potential $V_{N_{l,i}}(\mathbf{x})$ of the sub-neighborhood $N_{l,i}$, is equal to the sum of all pair-wise potentials

$$V_{N_{l,i}}(\mathbf{x}) = \sum_{k \in N_{l,i}} V_2(k, l) = -\gamma x_l \sum_{k \in N_{l,i}} x_k. \quad (4.3.1)$$

The potential associated with the *complete* neighborhood N_l of the position l will be denoted by $V_{N_l}(\mathbf{x})$. To determine this potential we follow the idea that was introduced at the beginning of this Section. An edge-label $x_l = 1$ should be assigned a high probability if *any* of $N_{l,i}$ indicates the presence of an edge element. Therefore,

$$V_{N_l}(\mathbf{x}|x_l = 1) = \min_i \{V_{N_{l,i}}(\mathbf{x})\} = -\gamma \max_i \left\{ \sum_{k \in N_{l,i}} x_k \right\}. \quad (4.3.2)$$

A non-edge label $x_l = -1$ should be assigned a high probability if *none* of $N_{l,i}$ indicates the presence of an edge element. This is accomplished if we choose

$$V_{N_l}(\mathbf{x}|x_l = -1) = \max_i \{V_{N_{l,i}}(\mathbf{x})\} = \gamma \max_i \left\{ \sum_{k \in N_{l,i}} x_k \right\}, \quad (4.3.3)$$

which is exactly the opposite with respect to $V_{N_l}(\mathbf{x}|x_l = 1)$. Both of the above two potentials can be represented with the same expression, if we keep the label x_l as a variable:

$$V_{N_l}(\mathbf{x}) = -\gamma x_l \max_i \left\{ \sum_{k \in N_{l,i}} x_k \right\}. \quad (4.3.4)$$

This completes the specification of the new prior model. The expression for the joint probability is of the form (3.4.1), but with the new neighborhood potential $V_{N_l}(\mathbf{x})$, given in (4.3.4).

One should note that the specification of a MRF model in terms of a neighborhood potential is not common. We did so in order to enable a clear comparison of the new prior model with the isotropic one from Eq (3.4.1) and to give a clear insight in the actual pixel interactions that it involves. An alternative representation of the same model is to treat the expression in Eq (4.3.4) as the potential of 3x3 *cliques*, which formally implies the 5x5 neighborhood, where the potentials of all other possible cliques are set to zero. Due to the specific potential function, the label of each pixel is influenced only by labels of its immediately adjacent pixels, which are its only actual neighbors.

4.4 Practical algorithm and its performance

Here we summarize the practical implementation of the proposed algorithm and demonstrate the influence of the proposed conditional and prior models on the noise suppression performance and on the visual quality of the images.

4.4.1 An upgraded stochastic sampling algorithm

The main advantage of the proposed method with respect to earlier related approaches is that it combines both the coefficient magnitudes and their evolution across scales, through a joint conditional (JC) model. We denote this algorithm as the MRF-JC, and summarize it as:

Compute the discrete wavelet transform without down-sampling

For each orientation and for resolution scales 2^j , $1 \leq j \leq 3$

- For all coefficients w_l , $1 \leq l \leq n$ compute significance measures $m_l = (\omega_l, \beta_l)$
- Run the Metropolis algorithm:

Initialize binary mask \mathbf{x} and set $N_{local} \leftarrow 0$ and $S_l \leftarrow 0$,
for $1 \leq l \leq n$

Repeat N_{global} times

Repeat until all the positions l have been visited

Choose l at random and set

$x_l^c \leftarrow -x_l$ and $N_{local} \leftarrow N_{local} + 1$

Compute p from Eq (4.2.7) and generate a random number u from $U[0,1]$

if $p > u$ set $x_l \leftarrow x_l^c$

If $x_l = 1$ set $S_l \leftarrow S_l + 1$

- For $1 \leq l \leq n$: estimate the wavelet coefficients as

$$\hat{y}_l = w_l S_l / N_{local}$$

Applying the inverse wavelet transform.

As *initial masks*, we use their maximum likelihood estimates, that are obtained by applying thresholds T_{MAG} and T_{ACR} to the coefficient magnitudes and to ACR, respectively. The thresholds are found from the intersections of the corresponding densities: $p_{\Omega_l|X_l}(T_{MAG}|1) = p_{\Omega_l|X_l}(T_{MAG}|1)$ and $p_{B_l|X_l}(T_{ACR}|1) = p_{B_l|X_l}(T_{ACR}|1)$. Actually, we compute only T_{MAG} for each detail image, and use $T_{ACR} = 0.5$.

In our experiments, increasing the number of iterations N_{global} above 10 did not contribute to an improved noise suppression performance. However, we have observed (as was noted in [Malfait95]) that better results are obtained when several short runs of the Metropolis sampler are applied instead of the one long run.

The computation of p in Eq (4.2.7), involves three parameters λ_1 , λ_2 and λ_3 . As it is motivated in the next Section (see Fig. 4.13), we use in practice $\lambda_1 : \lambda_2 : \lambda_3 = 1.5 : 0.75 : 1$.

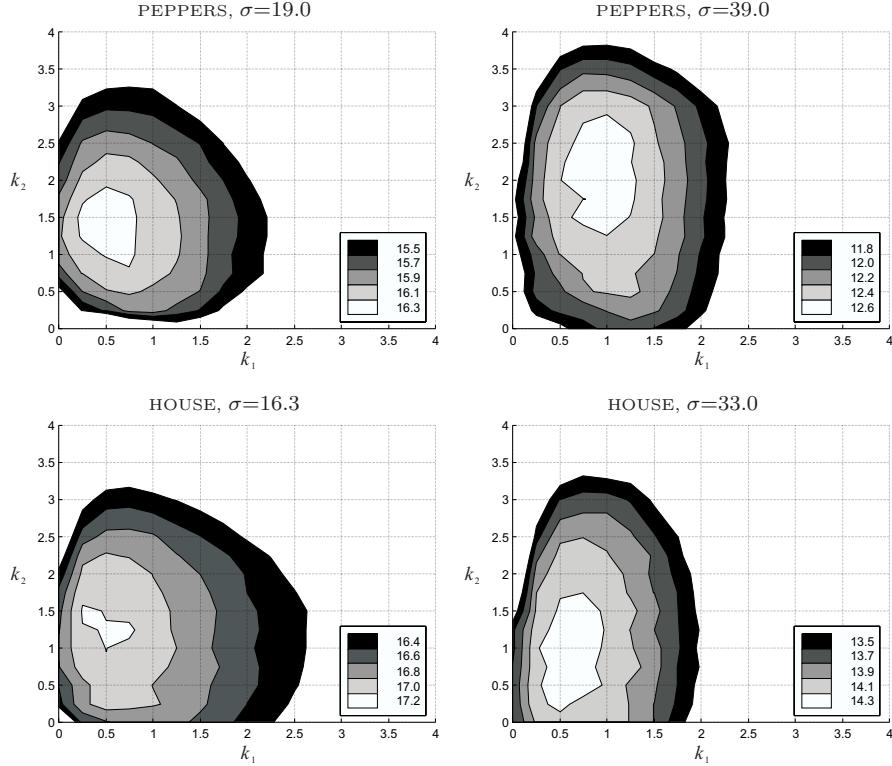


Figure 4.13: Resulting SNR corresponding to different values of the parameters $k_1 = \lambda_1/\lambda_3$ and $k_2 = \lambda_2/\lambda_3$, for different test images and different standard deviations σ of added noise.

4.4.2 The performance of the joint conditional model

The proposed method relies on the magnitudes of wavelet coefficients, their interscale dependencies and on prior knowledge about the spatial clustering. The relative importance given to these three sources is expressed through the parameters λ_1 , λ_2 and λ_3 , respectively, in Eq (4.2.7). To quantify the relative influence of the two significance measures with respect to each other, and with respect to the prior model at the same time, we set $\lambda_3 = \lambda$, $\lambda_1 = k_1\lambda$, and $\lambda_2 = k_2\lambda$. The choice of λ is then equivalent to choosing a different “temperature” [Li95] of the Metropolis sampler, which is not critical in this method. Therefore, we fix it to a constant value (in particular, $\lambda = 10$) and concentrate on the optimum choice of the relative influences $k_1 = \lambda_1/\lambda_3$ and $k_2 = \lambda_2/\lambda_3$.

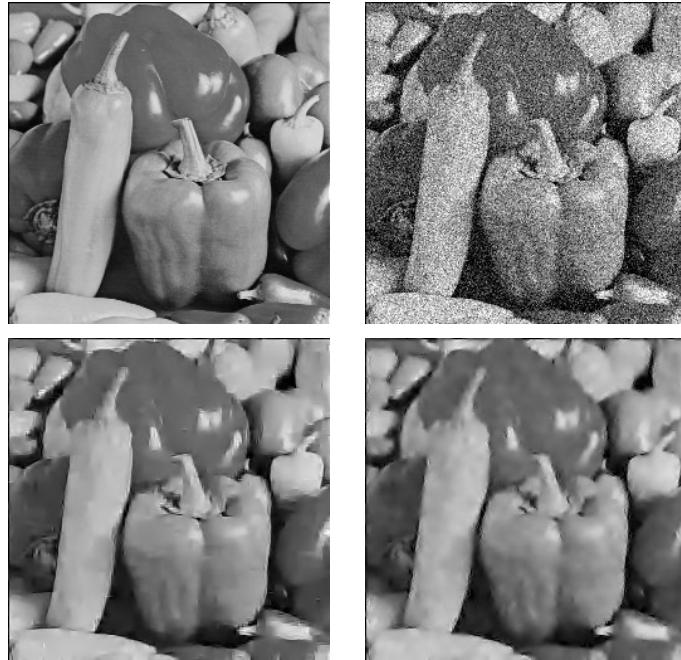


Figure 4.14: Influence of the proposed joint conditional model on the visual quality of results. Top left: noise-free image. Top right: noisy image SNR=3dB. Bottom left: the result of the proposed method using the joint conditional model. Bottom right: the corresponding result for $\lambda_1 = 0$.

The performance will be illustrated on two representative images: the peppers and the house images from Fig. 4.14 and Fig. 4.15, with two different levels of artificial noise. The size of both images is 256x256 pixels. The first one is rich with slow intensity variations and natural texture, while the second one is mainly characterized by sharp edges and flat background.

To each noisy image, we apply the algorithm from Section 4.4.1 for 16 values of k_1 and k_2 in the range 0 to 4 (i.e., 256 denoising procedures) and compute each time the resulting SNR. The contour plots of the resulting SNR values for four images are shown in Fig. 4.13. These contour plots show that the performance of the method is stable for a wide range of k_1 and k_2 . The SNR plots confirm that better results are always achieved by combining both significance measures instead of using either of them alone; we also notice that the benefit from a joint conditional model is image-dependent. If we examine the centers of

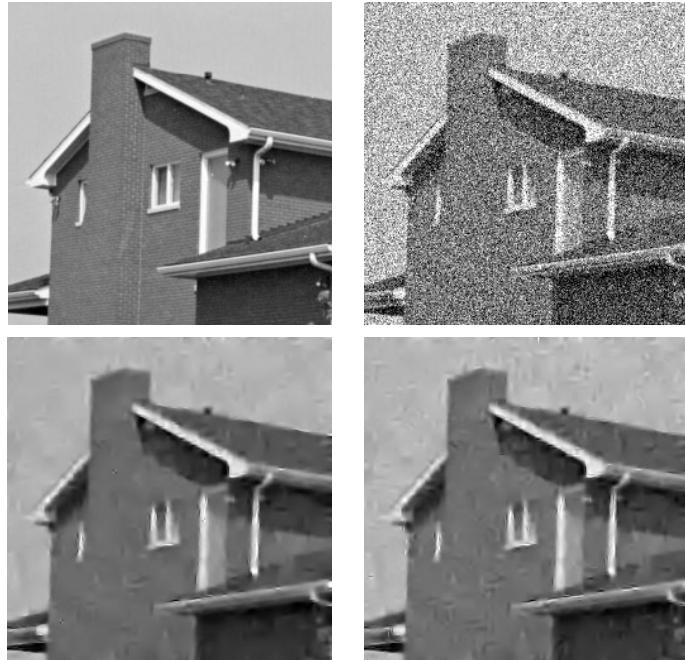


Figure 4.15: Influence of the proposed joint conditional model on the visual quality of results. Top left: noise-free image. Top right: noisy image SNR=0dB. Bottom left: the result of the proposed method using the joint conditional model. Bottom right: the corresponding result for $\lambda_2 = 0$.

the regions with maximum SNR, then it indeed appears that for higher noise variance the optimal proportion of k_2/k_1 (i.e., λ_2/λ_1) is bigger, i.e., that the use of interscale statistics is more important. The parameters $k_1 = \lambda_1/\lambda_3 = 1.5$ and $k_2 = \lambda_2/\lambda_3 = 0.75$ are near to optimal for all analyzed cases, so we use these parameter values for all subsequent results. The visual quality of results in Fig. 4.14 and Fig. 4.15 confirms the advantage of using the joint conditional model over special cases, where $\lambda_1 = 0$ or $\lambda_2 = 0$.

4.4.3 The performance of the prior model

Here we demonstrate the performance of the proposed MRF model with respect to the isotropic one from Eq(3.4.1). Earlier, in Section 3.4.3, it was demonstrated that the isotropic MRF model performs rather well on coarser resolution scales (see Fig. 3.11). Now we use the same im-

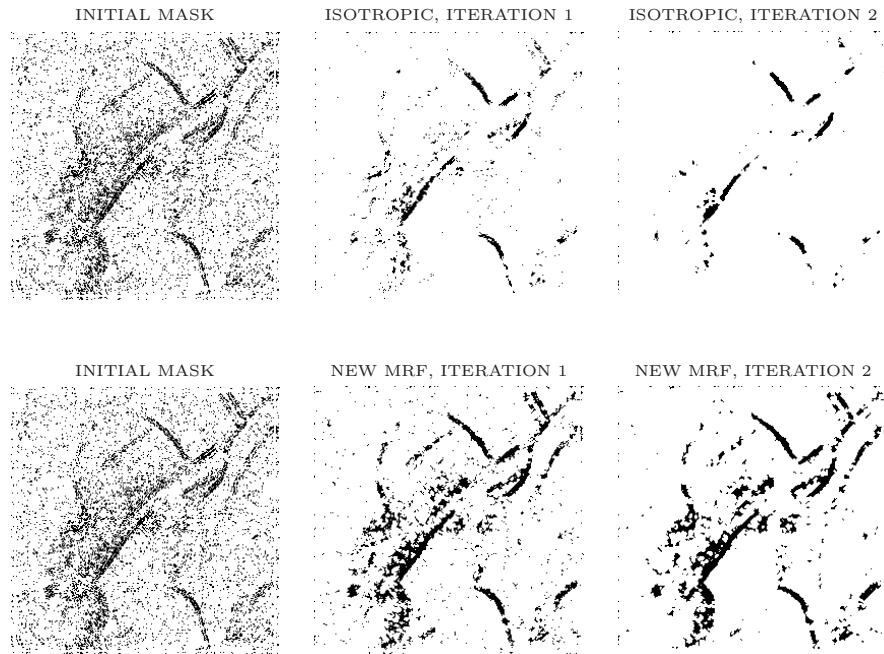


Figure 4.16: A comparison between the isotropic and the new MRF model. Several iterations of the Metropolis sampler are shown. Initial mask (top) and the conditional model is the same in both cases.

age for illustration, as there, but choose the finest resolution subband. Fig. 4.16 compares the performance of the isotropic model and the new, anisotropic one, by showing several iterations of the Metropolis sampler, starting from the same initial mask. In both cases, the same (above described) conditional model is used and the same relative importance is given to the prior model. Another example is given in Fig. 4.17, which corresponds to vertical details at the resolution scale 2^1 , for the house image with input SNR=9dB. The figure shows the initial mask and the results after 10 iterations of the Metropolis sampler, with the isotropic and with the new prior model. One can see that the isotropic model removes background noise, but it also removes useful edges. The new model, in contrast to this, preserves the useful edges well.

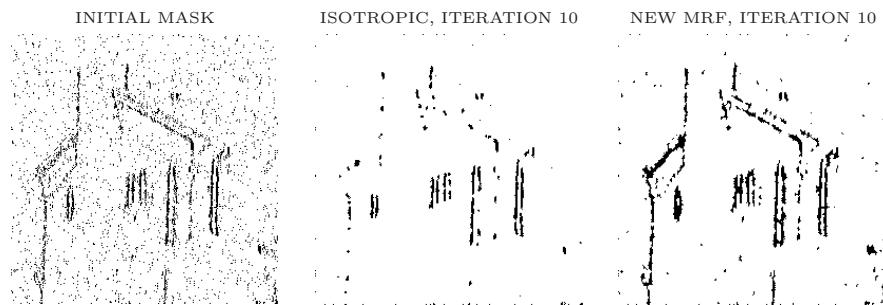


Figure 4.17: Masks at the scale 2^1 , for SNR=9dB. Left: initial. Middle: using the isotropic MRF model. Right: using the new, anisotropic MRF model.

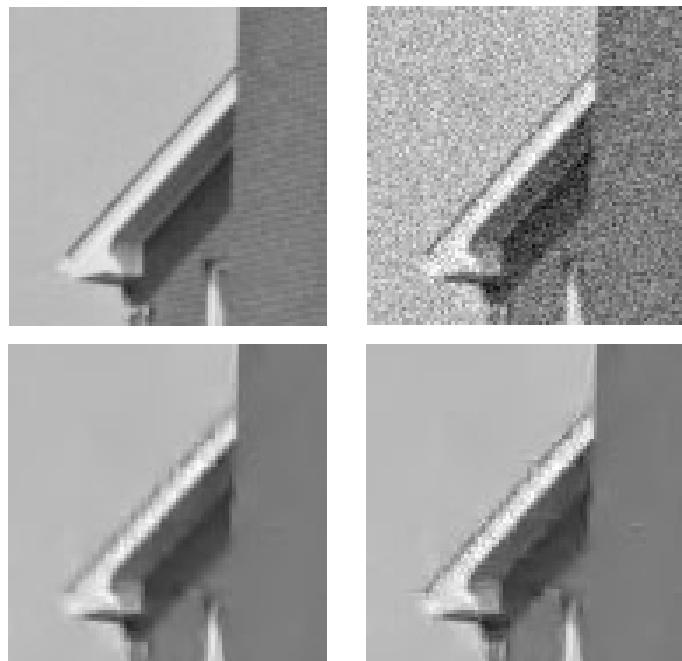


Figure 4.18: Comparing the performance of the isotropic and the anisotropic MRF prior model on a part of the house image. Noise-free image (top left), noisy image $SNR = 9dB$ (top right) and the results of the 3-subband version of the proposed method, using isotropic MRF model (bottom left) and using the new MRF model (bottom right).

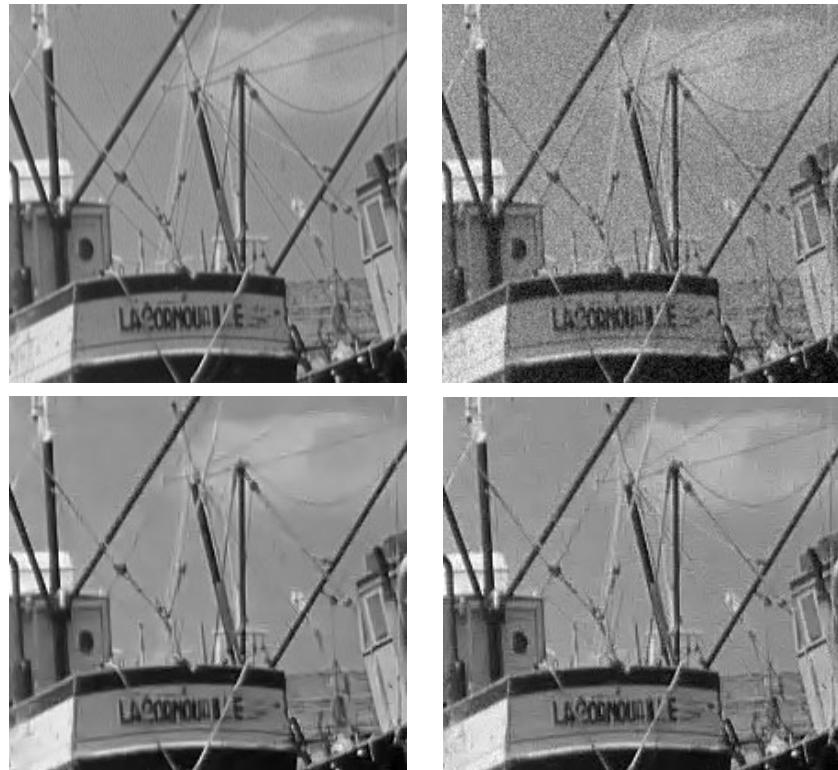


Figure 4.19: Influence of the new MRF model on the visual quality of results. Top: noise-free image (left) and noisy image (right). Bottom: the results using isotropic MRF model (left) and using the new MRF model (right).

The influence of the proposed model on the visual quality of denoised images can be noticed in fine details. In Fig. 4.18, one can see that the roof edge is better reconstructed. Another example is the boat image from Fig. 4.19; in this image the thin lines are considerably better reconstructed from noise when the new MRF model is used.

4.4.4 3-subband versus 2-subband implementation

We have implemented the proposed method with a 3-subband decomposition as well, using the quadratic spline wavelet from [Mallat92a] like in [Malfait97]. For all results, we used the same parameters $\lambda_1 : \lambda_2 : \lambda_3 = 1.5 : 0.75 : 1$ as in the two-subband case.

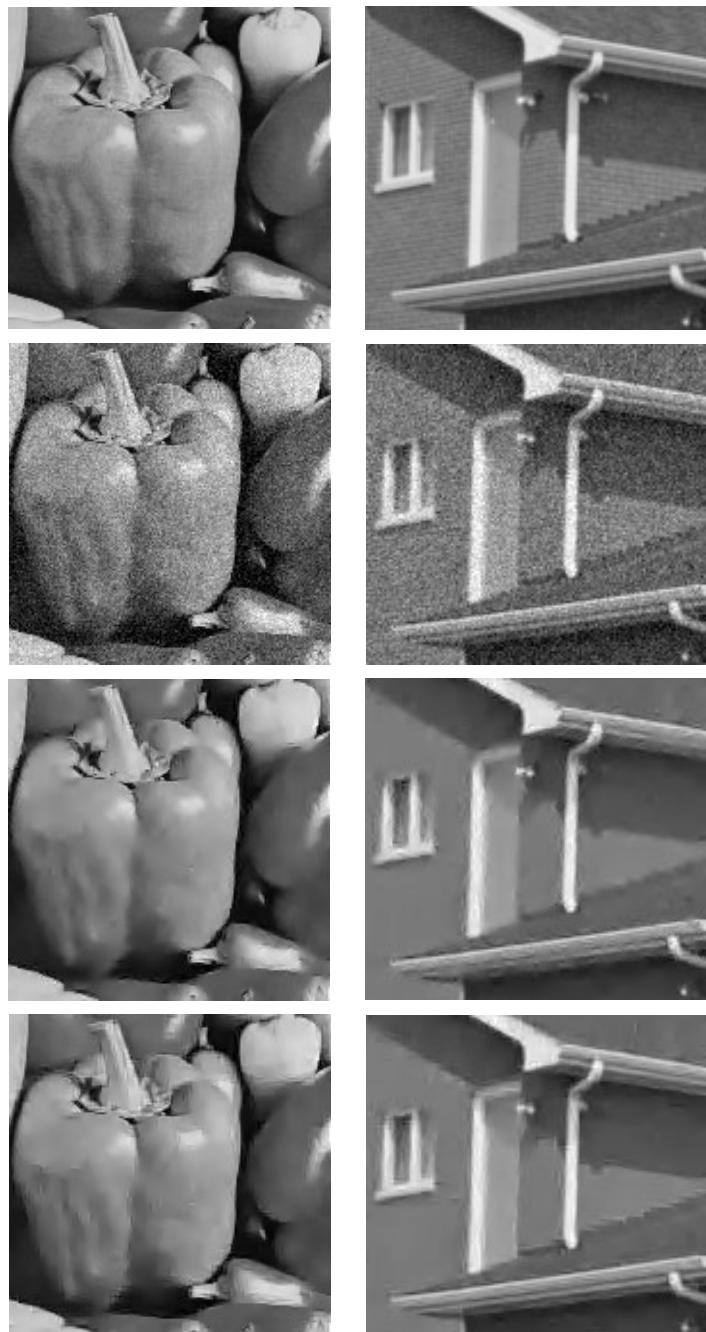


Figure 4.20: From top to bottom: enlarged parts of noise-free images, noisy images $SNR = 9dB$, and the results of the proposed method using 2 and using 3 orientation subbands.

Table 4.1: Comparison of quantitative results, expressed in $SNR[dB]$, for different methods. Notations 2-s. and 3-s. refer to the wavelet decompositions with 2, and with 3 orientation subbands, respectively.

Input	PEPPERS				HOUSE			
	9	6	3	0	9	6	3	0
MRF-JC, 2-s.	16.3	14.6	13.0	11.3	17.2	15.6	14.4	13.2
MRF-JC, 3-s.	16.6	15.0	13.4	11.6	17.9	16.4	14.9	13.4
[Malfait97], 3-s.	15.0	13.7	12.4	11.0	18.0	16.4	14.9	13.3
[Mallat92a], 2-s.	14.6	13.7	13.5	11.0	16.4	15.6	14.2	12.2
Spatially adaptive Wiener	15.4	13.5	11.7	9.7	15.4	13.7	11.8	10.0

With respect to the 2-subband case, a certain image-dependent improvement is achieved at the expense of computation time. In particular, in our experiments, use of three orientation subbands instead of two, lead to the improvements of the SNR ranging from 0.2dB up to 0.8dB. The difference in visual quality can be judged from Fig. 4.20. In most cases there is no striking difference in the visual appearance of the results. One can see the difference in small details: the roof lines on the house image from Fig. 4.20 are much better reconstructed when the three subband decomposition is used. We have measured the total execution time for 256x256 images, on a computer with a Pentium III processor. For the implementations with two and with three orientation subbands, the corresponding execution times are 45s and 30s, respectively.

4.4.5 Quantitative results and discussion

Quantitative results for the two test images are summarized in Table 4.1. The table lists the signal to noise ratio values $SNR[dB]$. To compute the corresponding $PSNR[dB]$, 13.6 should be added to all values for the peppers image, and 14.9 should be added to all values for the house image.

The first method, that is used as a basis for comparison, is the one of [Malfait97]. It uses 3 orientation subbands and is similar to a special case of our method, with $\lambda_1 = 0$. However, note that [Malfait97] uses APR with a heuristic conditional model, the parameters of which need to be estimated for each detail image at each scale and that it applies the isotropic MRF prior model. For the peppers image, our method achieves a significant improvement, which is bigger for higher input SNR. For ex-

ample, for input SNR=9dB the new method is 1.6dB better, and for input SNR=0dB it is 0.6dB better. For the house image, the results of the new method are almost the same as those in [Malfait97]. One can also see from the contour plots in Fig. 4.13 that for the house image setting $k_1 = \lambda_1/\lambda_3 = 0$ does not incur such a big penalty on the resulting SNR as it does for the peppers image. The computation time is approximately the same for both methods, since the Metropolis sampling procedure accounts for the majority of the computing time. The differences in prior and conditional models are insignificant in this respect: both prior models practically involve pair-wise pixel interactions only and converge equally fast.

Table 4.1 also lists the results from [Mallat92a]. The method of [Mallat92a], previously explained in Section 2.5.2, is considered one of the best for detecting multiscale edges from a noisy image. However, it makes all textures disappear, as was noted by the authors themselves. Our method is also not specifically designed to deal with textures, but since it applies a probabilistic shrinkage rule to all the wavelet coefficients (instead of reconstructing the image from its edges only), the natural texture is better preserved.

We have also included the results of Matlab's spatially adaptive Wiener filter, which was derived in [Lee80]. Its window size was optimized for each input image to produce the maximum SNR. For both test images these sizes were 3x3 for input SNR 9dB, 5x5 for input SNR 6 and 3dB and 7x7 for input SNR 0dB.

The performance of the new method on several other images is illustrated later, in Chapter 5.

4.5 Summary and Conclusions

In this Chapter, we addressed several issues to improve Bayesian image denoising using prior models for spatial clustering. First we investigated different significance measures and conducted their statistical characterization. Then we examined performance evaluation of those significance measures, relying on the decision theory. Motivated by the results of this performance evaluation, we proposed a joint conditional model, which combines both the coefficient magnitudes and their evolution across scales. For the new conditional model, we proposed a simple practical realization and motivated it by simulations. In order to preserve image details better, we developed a new MRF prior model. The advantages of the new prior and conditional models in terms of

noise suppression performance were demonstrated on different images. These developments generalize and improve previous methods from this branch.

Some aspects that were analyzed here may be useful for other denoising schemes as well: the realistic conditional densities of interscale ratios obtained via simulations, and objective criteria for evaluating noise suppression performance of different significance measures.

This work is published in [Pizurica02a].

Chapter 5

A Generalized Likelihood Ratio in Denoising

This Chapter introduces a new class of locally adaptive, low-complexity wavelet based denoising methods. The underlying principle is joint signal detection and estimation, which is now explicitly formulated. Apart from Section 5.2, where the general theoretic background is reviewed, the contents of the Chapter are original and present our own contributions. Most important among those is the estimation of the probability of signal presence in a given coefficient from the global coefficient histogram and from a local indicator of the spatial activity in the image. Three practical algorithms are developed: a fully analytic and automatic one for the Gaussian noise, and two heuristic extensions: a versatile algorithm for various noise types, and an algorithm for speckle suppression in SAR images.

5.1 Introduction

In several previously reviewed and in the newly developed algorithms in this thesis, the *uncertainty of signal presence* in a given coefficient was *implicitly* used as a starting assumption. In this context, e.g., the approach of Chipman *et al* [Chipman97] from Section 2.6.3 uses a *level-wise* adaptive Gaussian mixture model for the pdf of noise-free coefficients; we can say, in this case the uncertainty of signal presence is assumed to be equal in each coefficient from a given subband. More sophisticated approaches of this kind, which are locally adaptive as well, were using either hidden Markov trees (HMT) models or the Markov Random Field

(MRF) models. Now we propose an alternative approach, which uses a different concept and which is of low-complexity but compares well with the best state of the art methods.

We shall also recognize a great body of the related theory [Middleton68], and noise-suppression algorithms in a much broader framework than wavelet based image denoising. In particular, there are numerous examples in telecommunication and data processing systems where the signal is not surely known to be present in a given point of the observation space. Radar and speech transmission are among the most obvious ones. In such problems it is often most natural to join, i.e., to couple, the detection and the estimation of the signal. The statistical theory of joint detection and estimation [Middleton68] provides a consistent approach in this respect. The application of this theory, and in particular the application of rather simple estimators that employ the *generalized likelihood ratio*, has made a great breakthrough in speech processing [McAulay80, Ephraim84, Cohen00]; it has also proved the advantages for image denoising in the Discrete Fourier Transform (DFT) and the Discrete Cosine Transform (DCT) domains [Aach96a, Aach96b]. To our knowledge, wavelet domain counterparts of such simple, but powerful algorithms have received little if any attention so far.

The approach that we propose here fits in the above general framework, but was initially inspired by a different reasoning. Our starting point was an edge-adaptive algorithm that we have introduced in Section 3.5.2. There, in the scope of the MRF approach we have proposed to estimate the wavelet coefficients as $\hat{y}_l = P(X_l = 1 | \mathbf{m}, \hat{\mathbf{x}}_{\partial l}) w_l$, where \mathbf{m} was the significance map and $\hat{\mathbf{x}}$ the estimated mask. Starting from that expression we have further derived in Eq (3.5.4)

$$\hat{y}_l = \frac{\xi_l \mu_l}{1 + \xi_l \mu_l} w_l, \quad (5.1.1)$$

where ξ_l is the likelihood ratio, and μ_l is the ratio of prior probabilities of the edge and non-edge labels, conditioned on $\hat{\mathbf{x}}_{\partial l}$. Practically, we had $\mu_l = \exp(t_l)$, where t_l is a local spatial activity indicator (L.S.A.I.), expressed as a weighted sum of the labels $\{\hat{x}_k : k \in \partial l\}$. From here on we can observe several steps that lead us to the algorithms of this Chapter:

- One can use masks to estimate the densities $p_{M_l|X_l}(m_l|x_l)$ directly from the observed image, by finding the corresponding histograms of m_l (whatever m_l is)

- Instead of conditioning the ratio of prior probabilities μ_l on a discrete L.S.A.I. that is derived from the binary labels, one can use a continuous L.S.A.I. that is derived from the observed neighboring coefficients $\{w_k : k \in \partial l\}$ themselves
- The use of MRF's and the related optimization algorithms is not necessary for estimating masks in this approach. An alternative, simple and robust classification of coefficient labels will be introduced
- Masks are not necessary at all, in case where a continuous L.S.A.I. is used: we shall express all the parameters in terms of the observed wavelet coefficients themselves

The first three points lead to various heuristic solutions, and flexible algorithms which use rough estimates of masks to empirically estimate the required parameters and densities. The last one leads to a fully analytic and automatic approach, which fits in a theoretic, joint detection and estimation framework.

We shall present the contents of the Chapter in exactly the opposite direction from the above flow of mind, in order to start from theory and end with applications. First, we introduce briefly the background of joint detection and estimation principles, in Section 5.2. The proposed fully analytic wavelet domain denoising method for Gaussian noise, is described and discussed in Section 5.3. A general concept, which underlines various empirical extensions, leading to flexible algorithms for various noise types, is described in Section 5.4. In the above, empirical framework, two practical algorithms are described, in Sections 5.5 and 5.6. The first one is versatile and applicable to various noise types. Its application is demonstrated on medical ultrasound and MRI images and in image deblurring. The algorithm of 5.6 is aimed for speckle suppression in SAR images. Summary and conclusions are in Section 5.7.

5.2 Joint detection and estimation (JDE)

5.2.1 General principles

In their famous work [Middleton68], Middleton and Esposito have established a systematic theory of joint signal detection and estimation. This theory formalizes optimum (in the Bayesian sense) estimators, which act on the data when there is uncertainty as to the presence of a signal ($p(H_1 < 1)$). We read in [Middleton68]: "... the occasion frequently

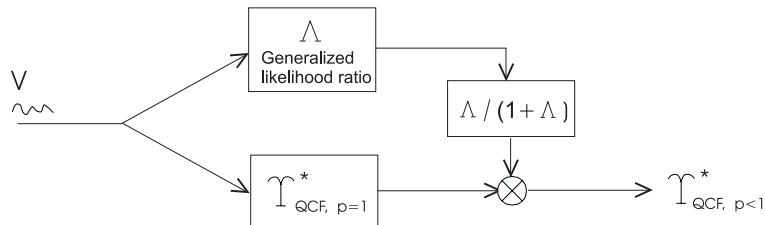


Figure 5.1: Bayes estimation with QCF of signal waveform or its parameters when $p < 1$, after Middleton and Esposito [Middleton68].

arises where the joint or simultaneous detection and extraction of signals is essential. Here optimal processing often requires the mutual coupling of detection and extraction for most effective results, so that each operation may significantly influence the other, as well as overall expected performance. Moreover, signal estimation now takes place under the condition $p(H_1) < 1$: the signal is *not* surely known to be present. Examples involving joint detection and extraction [$p(H_1) < 1$] occur in all circumstances where properties of the signal source or transmission medium (amplitude, phase, range, velocity, waveform structure, etc.), as well as signal presence or absence or signal classification, are jointly desired: in radar, communications, adaptive systems, sequential and multi-decision processes generally, and in filtering and prediction, to mention some of the more obvious and important areas.”

The exposition in [Middleton68] is remarkably general: it addresses arbitrary signal vectors, corrupted in an arbitrary fashion and a general form of the Bayes risk, as well as its various specific forms. Among the principal results is the one depicted in Fig. 5.1. It relates the Bayes estimate $\Upsilon_{QCF,p<1}^*$ under the quadratic cost function (QCF) (i.e., the MMSE estimate) and under the uncertainty of signal presence to the corresponding estimate $\Upsilon_{QCF,p=1}^*$ where this uncertainty is not encountered. The link between the two estimates is a scalar factor: $\Upsilon_{QCF,p<1}^* = \Lambda / (1 + \Lambda) \Upsilon_{QCF,p=1}^*$, where Λ is the *generalized likelihood ratio*. As we shall later develop a related wavelet-domain estimator under specific constraints and model assumptions, it is useful at this point to see its general meaning.

We shall use different notations than in [Middleton68] to make them consistent with the ones used in this thesis before. Suppose the following situation: the input of a reception system is either a signal $\mathbf{S}(\Theta)$ corrupted (in an arbitrary fashion) by a noise process \mathbf{N} (state H^1), or

the noise process alone (state H^0). In the notation $\mathbf{S}(\Theta)$, Θ are the parameters of the waveform \mathbf{S} (e.g., its amplitude and phase). At the output of the reception system, either the estimate of the waveform \mathbf{S} or the estimates of (some of) its parameters Θ are delivered. Let \mathbf{V} denote the input data (that are formed as described above). The spaces of all possible realizations of the signal (\mathbf{S} or Θ) and data \mathbf{V} are denoted by Ω and Γ , respectively. Let us further denote

The pdf of the data \mathbf{V} under the hypothesis H^1 (i.e., when a signal $\mathbf{s} \in \Omega$ is present):

$$p_{\mathbf{V}}(\mathbf{v}|H^1) = \int_{\Omega} p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{s})p_{\mathbf{S}}(\mathbf{s})d\mathbf{s} = \langle p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{s}) \rangle_{\mathbf{s}}. \quad (5.2.1)$$

The pdf of the data \mathbf{V} under the hypothesis H^0 (i.e., when the signal is absent, or equivalently, when $\mathbf{s} = \mathbf{0}$):

$$p_{\mathbf{V}}(\mathbf{v}|H^0) = p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{0}) = p_{\mathbf{N}}(\mathbf{v}). \quad (5.2.2)$$

The risk associated with the estimation of \mathbf{S} can now be written as

$$\mathcal{R} = \int_{\Gamma} \left\{ q p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{0})C(\mathbf{0}, \hat{\mathbf{s}}) + p \int_{\Omega} p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{s})C(\mathbf{s}, \hat{\mathbf{s}})p_{\mathbf{S}}(\mathbf{s})d\mathbf{s} \right\} d\mathbf{v},$$

where $p = P(H_1)$ denotes the probability of signal presence and $q = 1-p$. For the quadratic cost function $C(\mathbf{s}, \hat{\mathbf{s}}) = |\mathbf{s} - \hat{\mathbf{s}}|^2$, it can be shown that the above risk is minimized for $\hat{\mathbf{s}} = \hat{\mathbf{s}}_{QCF, p < 1}$, where

$$\hat{\mathbf{s}}_{QCF, p < 1} = E(\mathbf{s}|\mathbf{v}) = p \int_{\Omega} \mathbf{s}p(\mathbf{v}, \mathbf{s})d\mathbf{s} / \left\{ q p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{0}) + p \int_{\Omega} p_{\mathbf{V}, \mathbf{S}}(\mathbf{v}, \mathbf{s})d\mathbf{s} \right\}.$$

Observing that the classical QCF (i.e., the MMSE) estimate, where $p = 1$ is $\hat{\mathbf{s}}_{QCF, p=1} = \int_{\Omega} \mathbf{s}p_{\mathbf{V}, \mathbf{S}}(\mathbf{v}, \mathbf{s})d\mathbf{s} / \int_{\Omega} p_{\mathbf{V}, \mathbf{S}}(\mathbf{v}, \mathbf{s})d\mathbf{s}$, and denoting

$$\Lambda = \frac{p}{q} \frac{\langle p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{s}) \rangle_{\mathbf{s}}}{p_{\mathbf{V}|\mathbf{S}}(\mathbf{v}|\mathbf{0})} = \frac{p}{q} \frac{p_{\mathbf{V}}(\mathbf{v}|H^1)}{p_{\mathbf{V}}(\mathbf{v}|H^0)} \quad (5.2.3)$$

it follows clearly that

$$\hat{\mathbf{s}}_{QCF, p < 1} = \frac{\Lambda}{1 + \Lambda} \hat{\mathbf{s}}_{QCF, p=1}, \quad (5.2.4)$$

as it is depicted in Fig. 5.1. Equivalently, one can show that the above relationship holds between the corresponding estimates of the waveform parameters Θ , which can be e.g., the phase or the amplitude, like in the next Section.

5.2.2 JDE under statistically independent observations

The above approach simplifies considerably in the case where statistically independent data samples are assumed. Precisely such a treatment has been used for decades in speech processing [McAulay80, Ephraim84] and has led to the state-of-the art methods for speech enhancement [Cohen00]. The subject of these methods is the estimation of the amplitude of the short time Fourier transform of speech signals, which is often denoted as the short time spectral amplitude (STSA) estimation. It is interesting to recognize a close link between the STSA estimation of speech signals and seemingly unrelated wavelet domain image denoising. In this respect, we shall first summarize the relevant ideas of the above mentioned, speech-oriented methods.

Let us start from the abstract signal reception problem from the previous Section. The samples of the data \mathbf{V} are now assumed to be statistically independent. The input measurement is a random variable $V_l = S_l + N_l$, where N_l are i.i.d. random variables. Let $S_l = A_l \exp(j\Theta_l)$ and suppose that we are interested in estimating the amplitude A_l only. The amplitude of the input noisy signal will be denoted by B_l . The minimum mean squared estimate of A_l under the signal presence uncertainty, as treated in [McAulay80, Ephraim84] is

$$\hat{A}_l = E(A_l|B_l, H_l^1)P(H_l^1|V_l) + E(A_l|B_l, H_l^0)P(H_l^0|V_l). \quad (5.2.5)$$

Since $E(A_l|V_l, H_l^0)$ is the expected value of A_l given an observation V_l and the fact that signal is absent, it has to be zero. The above equation thus reduces to

$$\hat{A}_l = E(A_l|B_l, H_l^1)P(H_l^1|B_l). \quad (5.2.6)$$

Further on, the Bayes rule gives

$$P(H_l^1|B_l) = \frac{p_B(b_l|H_l^1)P(H_l^1)}{p_B(b_l|H_l^1)P(H_l^1) + p_B(b_l|H_l^0)P(H_l^0)} \quad (5.2.7)$$

and one can rewrite Eq(5.2.6) as

$$\hat{A}_l = \frac{\Lambda_l}{1 + \Lambda_l} E(A_l|B_l, H_l^1), \quad (5.2.8)$$

where

$$\Lambda_l = \mu_l \frac{p_B(b_l|H^1)}{p_B(b_l|H^0)}, \quad (5.2.9)$$

and $\mu_l = P(H_l^1)/P(H_l^0)$. Since speech signals contain large portions of silence, signal absence in the noisy observations is frequent. This absence of signal implies its absence in the spectral components as well. In [Ephraim84], it is underlined that it is also possible that the signal *is present*, but appears with *insignificant energy* in some spectral components, which are randomly determined. In the latter case, we immediately see the analogy with the wavelet representation of a noisy image: in some wavelet coefficients, the actual signal appears with insignificant energy.

In most of the approaches mentioned above and in other related ones the probability of signal presence in a given spectral component $p_l = P(H_l^1)$, was given either a fixed value or it has been treated as a free parameter that is experimentally optimized. In [Cohen00], a heuristic formula which exploits the strong correlation of speech presence in the neighboring frequency bins was proposed.

A related problem was also addressed in [Aach96a, Aach96b], where the above formalism was applied to *image denoising* in the Discrete Fourier Transform (DFT) and in the Discrete Cosine Transform (DCT) domains. In these methods, lacking the prior knowledge to estimate $E(A_l|B_l, H_l^1)$ the authors propose a simple approximation

$$E(A_l|B_l, H_l^1) \cong B_l, \quad (5.2.10)$$

and use it with equality in practice. Further on, μ_l has been seen as a free parameter in [Aach96a], and in [Aach96b] a certain heuristic formula was proposed to estimate it. Specifically, the authors take into account that the DFT of an image block concentrates the coefficient energy along the line that is perpendicular to the dominant spatial orientation in that image block; a larger μ_l was thus assigned to those coefficients, which are in the 2D spatial frequency domain located closer to the line of the dominant energy concentration. The actual expression is rather complicated and is empirically motivated.

5.3 A wavelet domain GenLik approach

From this Section on, we present our own work, where the above joint detection and estimation principles are employed in wavelet based image denoising. Apart from the common underlying principle, which is the uncertainty about the signal presence, we are *not* just re-applying the existing theory to a new problem. We shall propose an original approach for estimating the probability of signal presence in a given coefficient,

which is less ad hoc with respect to what has been done in other transform domains from the previous Section. Since the proposed method employs the generalized likelihood ratio, we have denoted it as *GenLik*.

5.3.1 Notation and model assumptions

We start from the following noise model in the wavelet domain

$$w_l = y_l + n_l, \quad (5.3.1)$$

where y_l is an unknown noise-free wavelet coefficient and n_l are independent, identically distributed (i.i.d.) Gaussian random variables. The analysis that follows can be easily extended to other noise distributions, provided that the i.i.d. assumption holds. The method will be practically implemented both with the orthogonal and the non-decimated wavelet transforms. In the first case, the above assumed model is realistic. To make it realistic in the second (non-decimated) case as well, one can separate the coefficients of a given subband at the scale 2^j into 2^{2j} sets containing non-correlated coefficients, as it was done in [Chang00a]. We use the i.i.d. model as a simplification without the coefficient separation.

In a wavelet decomposition of a noisy image there is rarely actual signal absence in any coefficient. However, in numerous coefficients the signal component appears with insignificant energy. The latter case we characterize as the absence of the signal “of interest”. The two hypotheses H_l^0 and H_l^1 now have the following explicit meaning

$$H_l^0 : \text{“signal of } interest \text{ is absent”} \Leftrightarrow |y_l| < T$$

$$H_l^1 : \text{“signal of } interest \text{ is present”} \Leftrightarrow |y_l| \geq T$$

where T is a threshold that defines what our signal of interest is. Its precise specification will be addressed later.

The wavelet coefficients representing the signal of interest in a given subband are assumed to be independent identically distributed (i.i.d.) random variables with the probability density (pdf) $p(w_l|H_l^1)$. Similarly, the coefficients in the same subband, corresponding to the absence of the signal of interest, are assumed to be i.i.d. random variables with the pdf $p(w_l|H_l^0)$. With these model assumptions, the MMSE estimate of y_l is the following conditional mean,

$$\hat{y}_l = E(y_l|w_l) = E(y_l|w_l, H_l^1)P(H_l^1|w_l) + E(y_l|w_l, H_l^0)P(H_l^0|w_l). \quad (5.3.2)$$

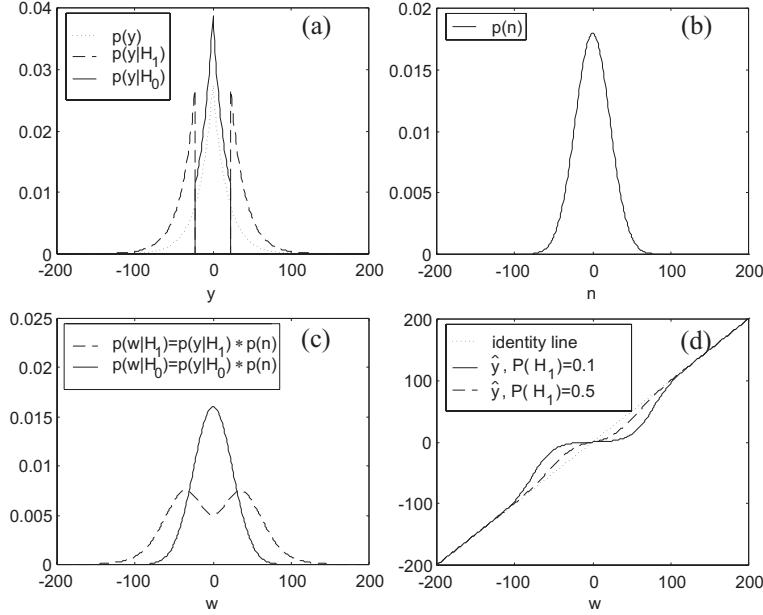


Figure 5.2: An illustration of the involved pdf models (a)-(c) for additive white Gaussian noise and the estimator (d), where the probability of signal presence is given as a parameter.

Since H_l^0 refers to the absence of the signal (of interest) it is reasonable to take $E(y_l|w_l, H_l^0) = 0$, like in all the related approaches from Section 5.2.2. For practical reasons, following Eq (5.2.10) here we also simplify $E(y_l|w_l, H_l^1) = w_l$. It is a subject for further research to see whether the exact expressions would yield a considerable gain in practice. By applying the Bayes rule to $P(H_l^1|w_l)$ like in Eq (5.2.7), our final estimator becomes

$$\hat{y}_l = \frac{\Lambda_l}{1 + \Lambda_l} w_l, \quad \text{with} \quad \Lambda_l = \mu_l \frac{p(w_l|H_l^1)}{p(w_l|H_l^0)}, \quad (5.3.3)$$

where for latter convenience we write, like in [Aach96b]:

$$\mu_l = \frac{P(H_l^1|\mathcal{P})}{P(H_l^0|\mathcal{P})}, \quad (5.3.4)$$

where \mathcal{P} denotes abstract *prior knowledge* that is involved in estimating the probability of $H_l^{0,1}$.

To summarize, our wavelet domain estimator, given by Eqs (5.3.3) and (5.3.4) has the same general form as the ones that have been used in

other transformation domains, and also other types of signals as listed in Section 5.2.2. What follows now is essentially different from all those approaches.

5.3.2 Spatial adaptation

The critical point in specifying the estimator from Eq (5.3.3) is the ratio of prior probabilities μ_l . The crucial question is thus how do we estimate the probability of the presence of the signal of interest in a given coefficient w_l . We shall propose an original approach in this respect, where μ_l is derived from both the *global* coefficient histogram in a given subband *and* from the *local* spatial context of each coefficient. The idea itself is rather clear. We first explain our reasoning behind it.

Clearly, if the probability of signal presence is *local*, in the sense that it is specifically related to the spatial position l , then it has to be a function of w_l and/or its surrounding coefficients. To simplify the formalism, let z_l be an arbitrary *indicator of the local spatial activity*, which is computed from the set $\{w_k : k \in \partial l\}$. The probability of the hypothesis H^1 at spatial position l is thus a function of z_l ; in other words this local probability is specified *given* z_l and we may denote it as $P(H_l^1|z_l)$. On the other hand, the unconditional probability $P(H_l^1)$, which is not a function of any local measurement related to l , does not actually depend on the spatial position l , but is global and may be denoted as $P(H^1)$. To summarize, from this point on we denote

$$P(H_l^1|z_l) \text{ local probability of signal presence}$$

$$P(H_l^1) = P(H^1) \text{ global probability of signal presence}$$

Following the above reasoning, we now replace μ_l in Eq (5.3.4) by

$$\mu_l = \frac{P(H_l^1|z_l)}{P(H_l^0|z_l)}, \quad (5.3.5)$$

and applying Bayes rule rewrite it as

$$\mu_l = \rho \frac{p(z_l|H_l^1)}{p(z_l|H_l^0)}, \quad (5.3.6)$$

where ρ is the ratio of unconditional, i.e., global prior probabilities

$$\rho = \frac{P(H^1)}{P(H^0)}. \quad (5.3.7)$$

Note that the ratio of prior probabilities μ_l in Eq (5.3.6) through ρ depends on global statistical properties of the wavelet coefficients in a given subband and through $p(z_l|H_l^{0,1})$ it depends also on a local measurement and its statistical properties. Moreover, the derived expression has a clear interpretation and what is of particular interest in practice, it can be easily computed as we show next. We shall estimate both ρ and $p(z_l|H_l^{0,1})$ directly from the input noisy image, as it will be explained in Sections 5.3.3 and 5.3.4.

In order to compute both ρ and the conditional densities $p(z_l|H_l^{0,1})$ and $p(w_l|H_l^{0,1})$, which completely specify the estimator from Eq (5.3.3), we shall need the underlying distribution of noise-free wavelet coefficients $p(y)$. We assume a model for $p(y)$ and compute its parameters directly from the input noisy image. We consider two candidates for $p(y)$: the generalized Laplacian (GL) prior

$$p_Y(y) = \frac{\nu}{2s\Gamma(\frac{1}{\nu})} \exp(-|y/s|^\nu), \quad (5.3.8)$$

and the Laplacian prior (LP)

$$p_Y(y) = \frac{1}{2s} \exp(-|y/s|). \quad (5.3.9)$$

Recall from Sections 2.6.2 and 4.2.5 that in case of AWGN the parameters of these densities are easily computed from the observed noisy image. For the latter convenience, let us repeat here briefly: if σ_w is the standard deviation of the noisy coefficients histogram, and $m_{4,w}$ the fourth moment, then the parameter ν from Eq (5.3.8) is found by solving

$$\frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})} = \frac{m_4 + 3\sigma_n^4 - 6\sigma_n^2\sigma_w^2}{(\sigma_w^2 - \sigma_n^2)^2}. \quad (5.3.10)$$

which is easily solved numerically, and the scale parameter s follows directly as

$$s = \left((\sigma_w^2 - \sigma_n^2) \frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})} \right)^{\frac{1}{2}}. \quad (5.3.11)$$

For the simpler, Laplacian prior from 5.3.9, one has

$$s = [0.5(\sigma_w^2 - \sigma_n^2)]^{1/2}. \quad (5.3.12)$$

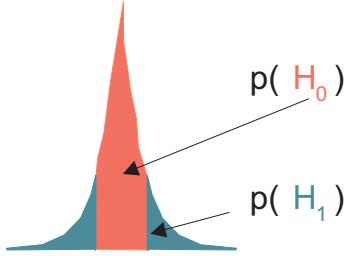


Figure 5.3: An illustration of the proposed global probabilities of the presence and absence of a signal of interest.

5.3.3 Global prior probability ratio

Now we turn to the actual estimation of the ratio $\rho = P(H^1)/P(H^0)$. We reason that $P(H^1)$ is the fraction of the wavelet coefficients in a given subband, which represent the signal of interest, i.e.,

$$P(H^1) = \int_{-\infty}^{-T} p_Y(y)dy + \int_T^{\infty} p_Y(y)dy, \quad (5.3.13)$$

as it is depicted in Fig. 5.3. It leads to

$$\rho = \frac{1 - \int_{-T}^T p(y)dy}{\int_{-T}^T p(y)dy}. \quad (5.3.14)$$

For the generalized Laplacian prior we derive (see the Appendix A):

$$\rho = \frac{1 - \Gamma_{inc}\left(\left(\frac{T}{s}\right)^{\nu}, \frac{1}{\nu}\right)}{\Gamma_{inc}\left(\left(\frac{T}{s}\right)^{\nu}, \frac{1}{\nu}\right)}, \quad (5.3.15)$$

where

$$\Gamma_{inc}(x, a) = \frac{1}{a} \int_0^x t^{a-1} e^{-t} dt \quad (5.3.16)$$

is the *incomplete gamma function*. The parameters ν and s are computed directly from the noisy coefficient histogram using Eqs (5.3.10) and (5.3.11).

For the Laplacian prior $p(y)$ from Eq (5.3.9), ρ has a simple form (see the Appendix A):

$$\rho = \frac{\exp\left(\frac{T}{s}\right)}{1 - \exp\left(\frac{T}{s}\right)}, \quad (5.3.17)$$

and the parameter s is computed from the noisy coefficient histogram using Eq (5.3.12).

5.3.4 Estimation of conditional densities

The specification of the generalized likelihood estimator further requires the conditional densities $p(w_l|H_l^{0,1})$, and $p(z_l|H_l^{0,1})$, where w_l is the observed coefficient, and z_l is the given local spatial activity indicator. Actually, only the specification of $p(z_l|H_l^{0,1})$ needs to be explained; the derivation of $p(w_l|H_l^{0,1})$ is already clear from the earlier exposition in Section 4.2.5. Nevertheless, since a different estimation framework, and a different notation is used in this Chapter, with respect to that in Chapter 4, we proceed with a complete treatment to avoid any confusion.

Since we consider here additive white Gaussian noise (AWGN), the pdf of noisy wavelet coefficients under the hypothesis H^0 is

$$p(w_l|H_l^0) = p(y_l|H_l^0) * p_N(n), \quad (5.3.18)$$

where $p_N(n) = \mathcal{N}(0, \sigma_n)$ is the normal distribution with zero mean and standard deviation σ_n . Equivalently, the density of noisy wavelet coefficients under the hypothesis H^1 is

$$p(w_l|H_l^1) = p(y_l|H_l^1) * p(n), \quad (5.3.19)$$

where $p(y_l|H_l^1)$ is the pdf of the noise-free coefficients representing the signal of interest: $p(y_l|H_l^1) = 0$ for $|y_l| < T$, and $p(y_l|H_l^1) \propto p(y_l)$ for $|y_l| \geq T$.

For the sake of tractability, we use the locally averaged magnitude of the wavelet coefficients as the indicator of local spatial activity,

$$z_l = \frac{1}{N} \sum_{k \in \partial l} \omega_k, \quad \text{where } \omega_k = |w_k| \quad (5.3.20)$$

and ∂l is the square window $W \times W$ centered at position l , and $N = W^2$. Due to intrinsically image properties the wavelet coefficients in a relatively small local window are likely to be identically distributed, i.e., according to either $p(w_l|H_l^1)$ or $p(w_l|H_l^0)$. With our model assumptions, the wavelet coefficients are also assumed to be locally i.i.d. given H_l^0 or H_l^1 . Therefore, $p(z_l|H_l^1)$ is simply computed via $N = W^2$ convolutions of $p(\omega_l|H_l^1)$, and $p(z_l|H_l^0)$ is computed via N convolutions of $p(\omega_l|H_l^0)$:

$$\begin{aligned} p(Nz_l|H_l^1) &= \underbrace{p(\omega_l|H_l^1) * p(\omega_l|H_l^1) * \dots * p(\omega_l|H_l^1)}_N, \\ p(Nz_l|H_l^0) &= \underbrace{p(\omega_l|H_l^0) * p(\omega_l|H_l^0) * \dots * p(\omega_l|H_l^0)}_N, \end{aligned} \quad (5.3.21)$$

where $p(\omega_l|H_l^{0,1}) = 2p(w_l|H_l^{0,1})$, $\omega_l > 0$. The validity of this model is illustrated in Fig. 5.4. The positions of the coefficients representing the signal of interest in a given subband $S_{H_1} = \{l : |y_l| > \sigma_n\}$, shown in this figure, were obtained from noise-free wavelet coefficients. The right-hand side of the figure shows the histograms of $\{z_l : l \in S_{H_1}\}$ in comparison with the model from Eq 5.3.21, where $p(w_l|H_l^1)$ was estimated from the noisy image. On the basis of these and similar results on other images, we have concluded that a simple model 5.3.21 is reasonable for a relatively small window size.

In the experiment described above, we have related the threshold T that specifies the signal of interest to the noise level, like in Chapter 4. In the next Section, we investigate whether in the proposed method such a specification of the value of T is indeed optimal in the sense of the mean squared error.

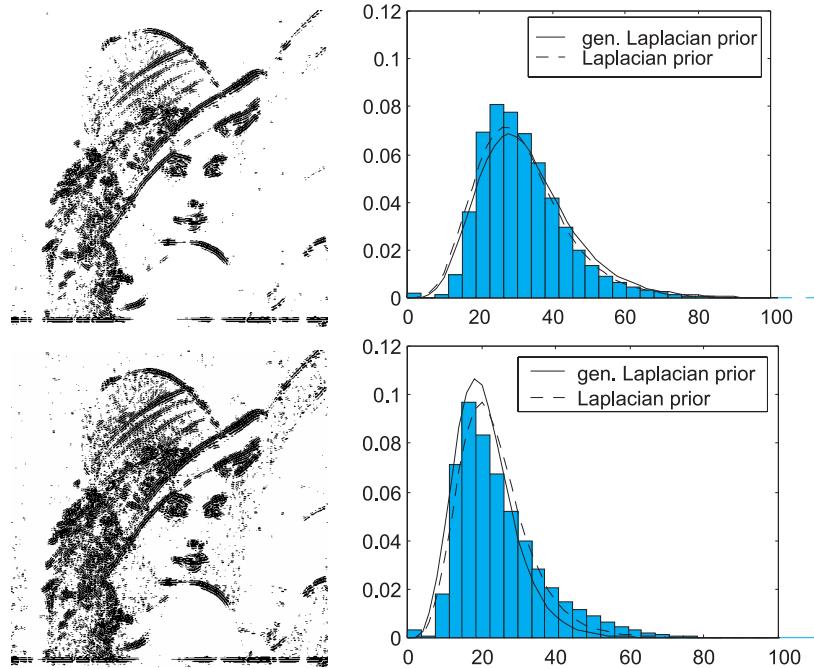


Figure 5.4: Left: reference positions for the signal of interest in the LH subband at the scale 2^2 for the Lena image, for two noise standard deviations: 30 (top) and 16 (bottom). Right: the corresponding histograms of z_l compared to the analytical models $p(z_l|H_l^1)$ estimated from the noisy image under the Laplacian (dashed line) and the generalized Laplacian (solid line) priors $p(y)$. The window size is 7x7.

Apart from the specification of the threshold that specifies the signal of interest, we have now completed the description of the key points in the new method. A general form of the estimator is presented in Fig. 5.2. This estimator tends to leave untouched large magnitude coefficients, while the others are suppressed to a degree that depends on the global and on the local image context. In contrast to this, a classical MMSE estimator from Section 2.6.3, depicted in Fig. 2.19, applies the same shrinkage action to each coefficient (no matter of the spatial context) and introduces a bias on large magnitude coefficients.

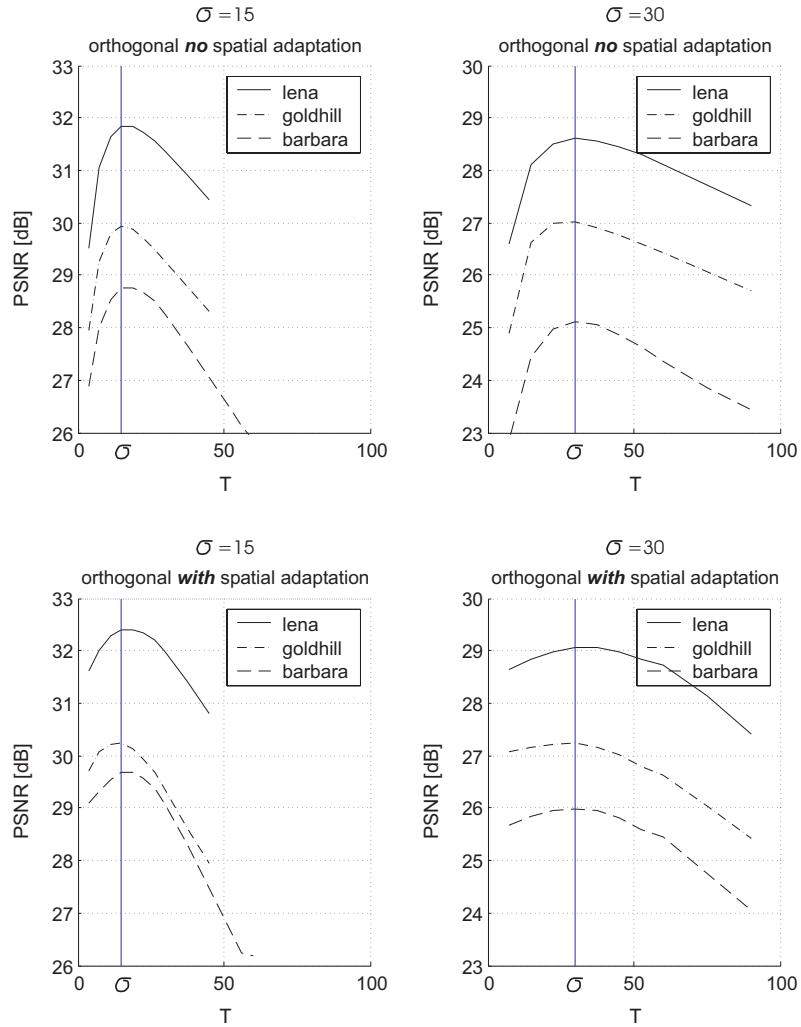


Figure 5.5: Influence of the threshold that specifies the signal of interest, on the denoising performance, when an *orthogonal transform* (with the wavelet *dB8*) is used. Top: a simplified version of the method, without spatial adaptation. Bottom: full method. The standard deviation of the added noise σ is denoted on each diagram.

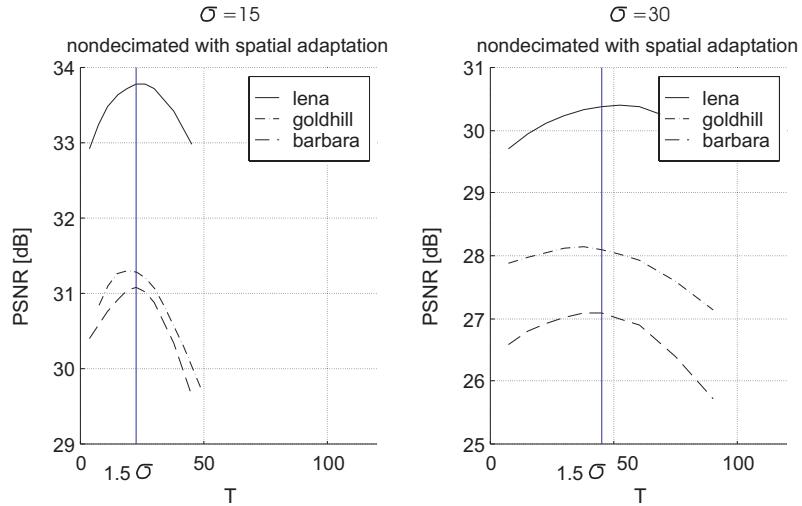


Figure 5.6: Influence of the threshold that specifies the signal of interest, on the denoising performance, when a *non-decimated transform* (with the wavelet *dB8*) is used. The standard deviation of added noise σ is denoted on each diagram.

5.3.5 Signal of interest and performance evaluation

Now we come to a crucial question, which is what actually our signal of interest is, i.e., how do we practically specify the threshold T , which appears in the previous equations.

Our reasoning is purely objective. Noise has a precisely known distribution (Gaussian) and the signal of interest needs to be specified such that the overall performance of the method is optimized in terms of the mean squared error. One should expect that in this respect it is most reasonable to relate the threshold T to the standard deviation σ_n of noise. One reason is intuitive: the lower the noise level, the more image features can be actually reconstructed from it. Another reason is theoretical: recall the optimal coefficient selection principle from Section 2.3.5. In the related, MRF-based denoising approach, the conditional densities of coefficient magnitudes were drawn from this reasoning as well.

We have performed extensive simulations to investigate the optimal specification of the threshold T in the current estimation framework. Apart from the main one, other interesting conclusions will be drawn

from those results as well: the benefit from using a local context in estimating μ_l will be quantified, as well as the benefit from using a non-decimated representation with respect to the orthogonal one. In all cases, we shall use the more accurate, generalized Laplacian prior, and the three representative images: "Lena", "Barbara" and "goldhill" from Fig. 4.3. Now we take one step at a time.

First, we investigate the performance of the method, with respect to T , when an *orthogonal* wavelet transform is used. In particular, the Daubechies *db8* wavelet from Fig. 2.8 is used. In all the subbands, the noise standard deviation σ_n is the same and is equal to the standard deviation of input noise σ . Let us start from a simplified version of the proposed method, in which there is *no* spatial adaptation in determining the probability of signal presence. In this case, μ_l is equal for all the coefficients in a given subband, and computed from their global statistics only: $\mu_l = \rho = P(H^1)/P(H^0)$. For various values of σ , the optimal T was as expected: $T = \sigma$; representative diagrams for different values of $\sigma = 15$, and $\sigma = 30$ are shown in the top of Fig. 5.5. Now we repeat the simulations with the full, spatially adaptive method, where μ_l is influenced also by the local spatial activity indicator z_l , using thus eqs. (5.3.6), (5.3.20) and (5.3.21). The representative diagrams, for the same two values of σ as above are shown in the bottom of Fig. 5.5. The window size was 5x5. One can see that the value $T = \sigma$ is optimum in this case as well. From these diagrams one can also quantify the benefit from computing μ_l from both the global and the local statistical properties with respect to the above simplification. The diagrams clearly show that the improvement is always present, and that it is image-dependent, ranging from 0.2dB only up to large differences of 1dB.

Now we investigate the influence of the threshold T when a *non-decimated* transform is used. We do not analyze the simplified version of the method anymore, but only the complete, spatially adaptive one. We use the "orthogonal" *db8* wavelet as above. The standard deviation of noise is therefore again the same in each subband and is equal to that of the input noise σ . Simulations have revealed an interesting result. The optimal value of T is again related to σ , but through a constant factor, larger than one. Two representative diagrams showing the resulting PSNR with respect to T are shown in Fig. 5.6. The optimal value is approximately 1.5σ . A possible explanation for this is that in a non-decimated wavelet transform, white input noise is transformed into a spatially correlated one, and the assumed statistical model should be seen as an approximation in this case. One should also note that the

difference in the resulting PSNR for $T = \sigma$ with respect to the optimum value of T is rather small (less than 0.16dB in all analyzed cases); the differences are hardly or even not visible. Finally, by comparing the diagrams from Fig. 5.6 with those from the bottom of Fig. 5.5, one can quantify the benefit of using a non-decimated representation with respect to the orthogonal one. The improvement is obviously large: in most cases, the PSNR has increased of ≈ 1.3 dB.

To summarize, the most important conclusions from this Section are

In an orthogonal wavelet transform the optimum (in the MSE sense) threshold that specifies the signal of interest is $T = \sigma$, where σ is the standard deviation of noise. This conclusion was drawn from simulations and has confirmed the expectations according to theory.

In a non-decimated wavelet transform the optimum threshold in the above sense is $T \approx 1.5\sigma$. Choosing a threshold equal to σ instead of an optimum one incurs a penalty, which is usually negligible.

Estimating μ_l from both local and global statistical properties of the wavelet coefficients, as compared to using the global statistics only, yields an image dependent improvement of up to 1dB.

When a non-decimated transform is used instead of the orthogonal one, the PSNR increases for more than 1 dB.

Using the proposed method with the orthogonal transform might be interesting in applications where both image denoising and compression are desired (see e.g., [Chang00a]). This subject is left for further research. Since our interest here is to develop high-performance image denoising techniques, we shall not further treat the orthogonal transform, which is inferior in this respect as compared to the non-decimated one. Another possibility is to use the orthogonal transform with the cycle spinning method (Section 2.2.6), but the latter was not implemented in our algorithms. All the subsequent results were obtained using the non-decimated wavelet transform.

5.3.6 Algorithm overview

The complete algorithm of the proposed method is

Compute the non-decimated wavelet transform

If σ is not available, estimate it using Eq (2.3.8)

Choose the window size (we use $W=7$)

For each orientation and scale:

- Estimate the parameters of $p_Y(y)$: Eqs (5.3.10) and (5.3.11) for the GL prior (or Eq (5.3.12) for the LP prior)
- Estimate $\rho = P(H^1)/P(H^0)$: Eq (5.3.15) for the GL prior or Eq (5.3.17) for the LP prior
- Split $p_Y(y)$ into $p_Y(y|H^0)$ and $p_Y(y|H^1)$, where $T = 1.5\sigma$:
 $p_Y(y|H^1) = 0$ for $y < T$ and $p_Y(y|H^1) \propto p_Y(y)$ for $y \geq T$
 $p_Y(y|H^0) \propto p_Y(y)$ for $y < T$ and $p_Y(y|H^0) = 0$ for $y \geq T$
- Find $p_W(w|H^0)$ and $p_W(w|H^1)$: Eqs (5.3.18) and (5.3.19)
- $p_\Omega(\omega|H^0) = 2p_W(w|H^0)$, and $p_\Omega(\omega|H^1) = 2p_W(w|H^1)$, $\omega > 0$
Find $p_Z(z|H^{1,0})$: Eq (5.3.21)
- For all the coefficients in the given subband, $l = 1, \dots, n$
 - Set $\xi_l = p_Y(y_l|H_l^1)/p_Y(y_l|H_l^0)$
 - Compute z_l from Eq (5.3.20)
 - Set $\eta_l = p_Z(z_l|H_l^1)/p_Z(z_l|H_l^0)$
 - Estimate the coefficient: $\hat{y}_l = \rho\xi_l\eta_l/(1 + \rho\xi_l\eta_l)w_l$

Apply the inverse wavelet transform

The characteristic parts from this algorithm are depicted in Fig. 5.7.

5.3.7 Results and Discussion

The performance of the proposed method will be demonstrated on different images with artificial noise, and compared both with the MRF-based approach from Chapter 4, and with other, state of the art methods from literature.

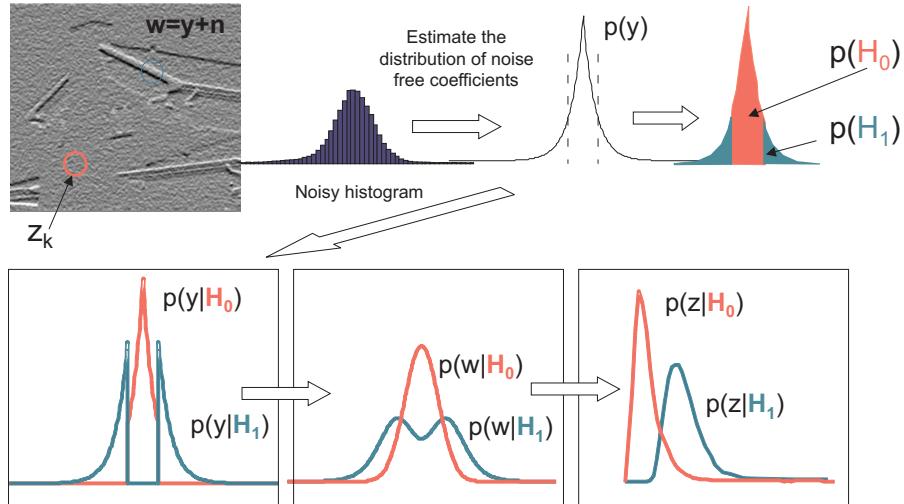


Figure 5.7: An illustration of the characteristic parts in the GenLik method.

Comparison with the MRF-based approach

In the previous Chapter, we have developed an advanced MRF based wavelet denoising method, which was denoted as MRF-JC algorithm. The new GENLIK method of the previous Section is obviously less complex, but aims also at using accurate statistical modeling. It is useful to summarize at this point the most important differences between the two algorithms:

	Dependencies accounted for	Prior knowledge based on
MRF-JC	intra <i>and</i> interscale	spatial <i>clustering</i> of important coefficients
GENLIK	intrасale	spatial <i>continuity</i> of coefficient magnitudes in a <i>local window</i> and the <i>global histogram</i>

Numerous experiments have been performed to compare the performance of the two methods. Interestingly enough, the results have shown that on certain types of images (as we next demonstrate) the MRF-JC performs best, while on others GENLIK is the best. On most images of natural scenes the differences are small and hardly visible.

Images on which the MRF-JC method yields a superior performance, with respect to GENLIK, (and all the other tested methods) are those

that are dominated by large areas of uniform (or slightly varying) intensity, intercepted by sharp edges and/or thin lines. Examples of such images are the house image from Fig. 5.8, and the artificial, hand-drawn image in the same figure, which consists of a flat background and of particularly thin lines. The advantage of the MRF-JC method is more pronounced at high noise levels: the background noise is strongly suppressed due to its powerful, joint conditional model; moreover, the anisotropic MRF model helps connecting thin lines that are largely disturbed by noise. On the extremely noisy house image, the result of the MRF-JC method is better than that of the GENLIK method for 0.8 dB, and on the hand-drawn image this difference is 0.5dB. The corresponding differences on both images were smaller when the input PSNR was larger. For comparison, we have included also the results of two classical methods: the spatially adaptive Wiener filter and the hard thresholding in the orthogonal DWT. The latter two methods are obviously inferior both in terms of PSNR and visually.

It is interesting that on images that are rich in natural textures, and especially at low to moderate noise levels (where those textures are not completely covered by noise), GENLIK usually outperforms the more complex MRF-JC. In this respect, a good example is the standard Lena image; for the input noisy image with $\sigma = 20$, shown in Fig. 5.9, the result of the GENLIK method is better than that of the MRF-JC by more than 1dB. In this case, one can also clearly see the difference in visual quality in Fig. 5.10; the most obvious differences are in the texture of the hat and in the face. Another example is the particularly textured 'Baboon' image from Fig. 5.11. For the input image from the bottom of Fig. 5.11, with $\sigma = 25$, the GENLIK method yields a result that is 0.7 dB better than that of the MRF-JC method. However, the differences in the visual appearance in Fig. 5.12, are now hardly noticeable. The latter conclusion is valid for many images of natural scenes: the two methods most often yield a similar visual appearance.

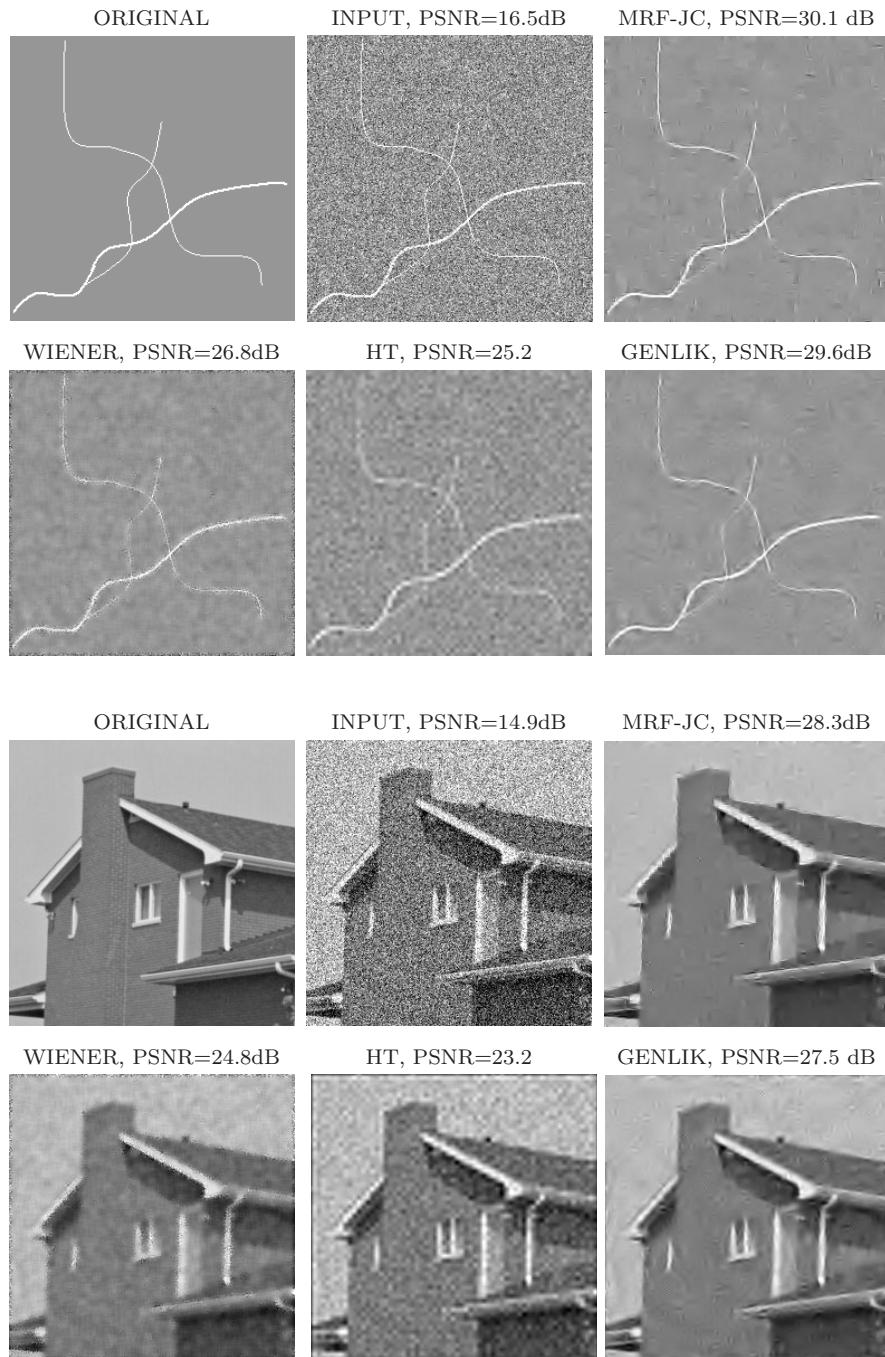


Figure 5.8: A comparison between the MRF-JC and GENLIK algorithms, and two classical methods: the spatially adaptive Wiener filter and the hard thresholding (HT) in the orthogonal DWT.



Figure 5.9: 512x512 Lena image (top) and the noisy image, with $\sigma = 20$ (bottom).



Figure 5.10: The results of the MRF-JC (top, PSNR=31.3dB) and the GENLIK (bottom, PSNR=32.5dB) algorithms applied to the noisy image from Fig. 5.9.

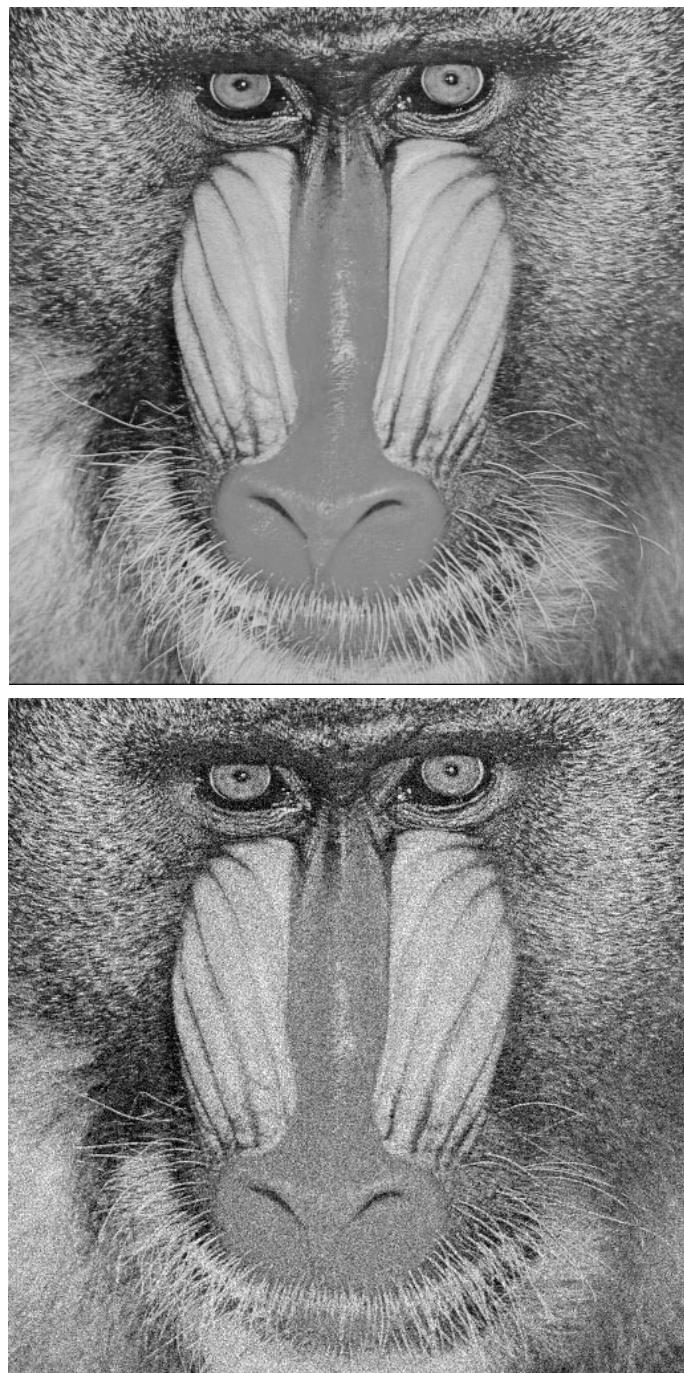


Figure 5.11: 512x512 Baboon image (top) and the noisy image, with $\sigma = 25$ (bottom).

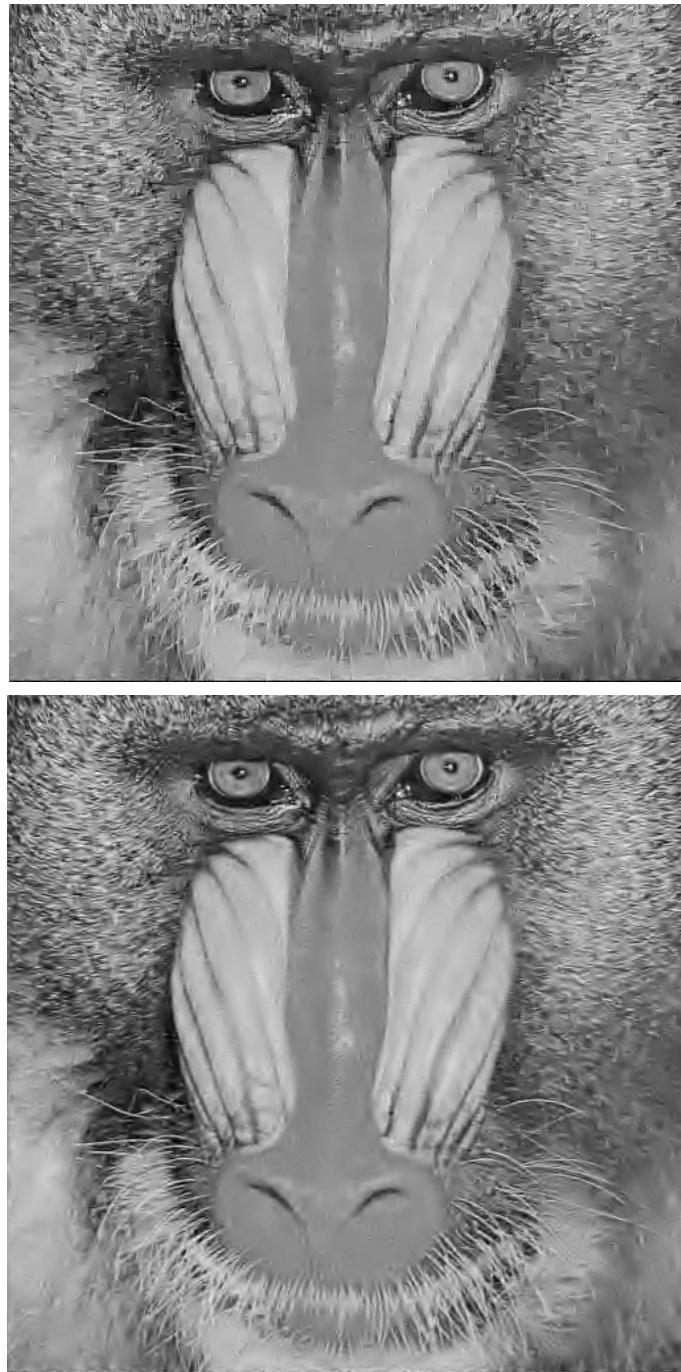


Figure 5.12: The results of the MRF-JC (top, PSNR=24.2dB), and the GENLIK (bottom, PSNR=24.9dB) algorithms applied to the noisy image from Fig. 5.11.

Comparison with other methods

In Table 5.1, the quantitative performance of the proposed GENLIK method is compared with the available state of the art results, on two standard 512x512 test images: ‘Lena’ and ‘Barbara’. The table lists also the results of a single resolution spatially adaptive method: Matlab’s spatially adaptive Wiener filter. The results of the proposed method are given for both the generalized Laplacian (GL) and the Laplacian prior (LP), the latter being slightly easier to implement. All the methods in the Table 5.1 are organized into three categories: “orthogonal WT”, (the orthogonal wavelet transform), “redundant WT” (including both the cycle spinning method and the non-decimated transform), and ”single resolution” methods.

The reference methods from Table 5.1 were recently published:

- Spatially Adaptive Wavelet Thresholding (SAWT) with the context modeling [Chang98, Chang00a]; its idea was explained in Section 2.4.2
- The Locally Adaptive Wiener filtering with the MAP estimation of the standard deviation of the wavelet coefficients (LAWMAP) [Mihcak99]; a brief explanation was given in Section 2.6.4
- The Spatially Adaptive denoising under Overcomplete Expansion (SAOE) [Li00], which was also mentioned in Section 2.6.4
- Redundant Hidden Markov Tree (RHMT) method of [Romberg99]; its concept was briefly explained in Section 2.6.5
- Local Contextual Hidden Markov Modeling (LCHMM) method of [Fan01]. Its main idea was also explained in 2.6.5

Three different wavelets have been used in the above methods, as indicated in front of each method in the Table 5.1. Most methods use orthogonal wavelets: the Daubechies *db8*, or the least asymmetrical *sym8* wavelet. In our experiments these two wavelets yield minor differences in PSNR (0.1-0.2dB at the most). The SAOE method is the only one among the ones listed, where biorthogonal wavelets, with a large number of vanishing moments were used to minimize the mean squared error.

The RHMT and LHCMM are more complex with respect to all the others in the Table 5.1, and are the only ones that include training procedures for estimating their parameters. Among those two, the LHCMM is more powerful (and also more complex), and makes use of both the local

spatial context and the inter-scale dependencies of the coefficients (see Fig. 2.21). These two schemes use the redundant wavelet representation that is obtained through a cycle spinning procedure (Section 2.2.6). On the two standard test images, the proposed GENLIK method clearly outperforms the RHMT approach and compares well with the LCHMM.

In [Mihcak99], the LAWMAP method was implemented only with the inferior orthogonal transform. To enable a fair comparison, we have implemented this method with the same non-decimated transform that appears in our algorithms as well. The corresponding method is in the Table 5.1, denoted as LAWMAP-ND. The SAWT method uses the same non-decimated transform as in our algorithms. The SAOE method uses a specific overcomplete expansion that is obtained in a similar fashion to cycle spinning: the difference is that instead of processing the shifted versions separately, all the coefficients are put together, and are treated as a single, overcomplete representation. The methods LAWMAP, SAOE, and SAWT are of similar complexity as the new GENLIK method. The new method achieves better results on the analyzed test images.

We shall further demonstrate the visual quality for the GENLIK method in comparison with the LAWMAP-ND. The latter method is the most relevant among the ones above for this comparison: the two methods use the same wavelet transform; they are of similar complexity and both employ the MMSE estimation using a local spatial window. Under this setup, the LAWMAP-ND applies a classical estimation, while the GENLIK method is based on the uncertainty of the presence of a signal of interest. By comparing the two methods, the advantage of the developed model may be clearly evaluated. The PSNR results in the Table 5.1 are in favor of the new method. The corresponding differences in visual quality are demonstrated in Fig. 5.13 and Fig. 5.15, for the above two test images. The new method yields an improved visual appearance. The sharpness and the details in the results are preserved equally well by both methods, but the new one introduces considerably less artifacts.

Table 5.1: A comparison of resulting PSNR [dB] for several state of the art wavelet methods, and a single resolution adaptive filtering. The wavelet type is listed in front of each method.

		Standard deviation of noise			
		10	15	20	25
		L E N A			
<i>db8</i>	orthogonal WT LAWMAP [Mihcak99]	34.3	32.3	31.0	30.0
	redundant WT				
<i>sym8</i>	LAWMAP-ND	35.1	33.3	31.9	30.9
<i>sym8</i>	SAWT [Chang98]	–	33.0	31.9	30.6
<i>bior10/18</i>	SAOE [Li00]	34.9	33.0	31.7	30.6
<i>db8</i>	RHMT [Romberg99]	34.6	32.6	31.2	30.1
<i>sym8</i>	LCHMM [Fan01]	35.0	33.0	31.7	30.6
<i>sym8</i>	GENLIK-GL	35.7	33.9	32.5	31.4
<i>sym8</i>	GENLIK-LP	35.7	33.8	32.2	31.2
	Single resolution				
	Adaptive Wiener	33.6	31.1	29.0	27.2
		B A R B A R A			
<i>db8</i>	orthogonal WT LAWMAP [Mihcak99]	32.6	30.2	28.6	27.4
	redundant WT				
<i>sym8</i>	LAWMAP-ND	33.1	30.8	29.1	27.8
<i>sym8</i>	SAWT [Chang98]	–	30.7	28.9	27.6
<i>bior10/18</i>	SAOE [Li00]	33.3	31.1	29.4	28.2
<i>db8</i>	RHMT [Romberg99]	32.8	30.3	28.6	27.7
<i>sym8</i>	LCHMM [Fan01]	33.6	31.4	29.7	28.5
<i>sym8</i>	GENLIK-GL	33.5	31.2	29.5	28.3
<i>sym8</i>	GENLIK-LP	33.5	31.3	29.6	28.3
	Single resolution				
	adaptive Wiener	29.9	28.3	26.8	25.5

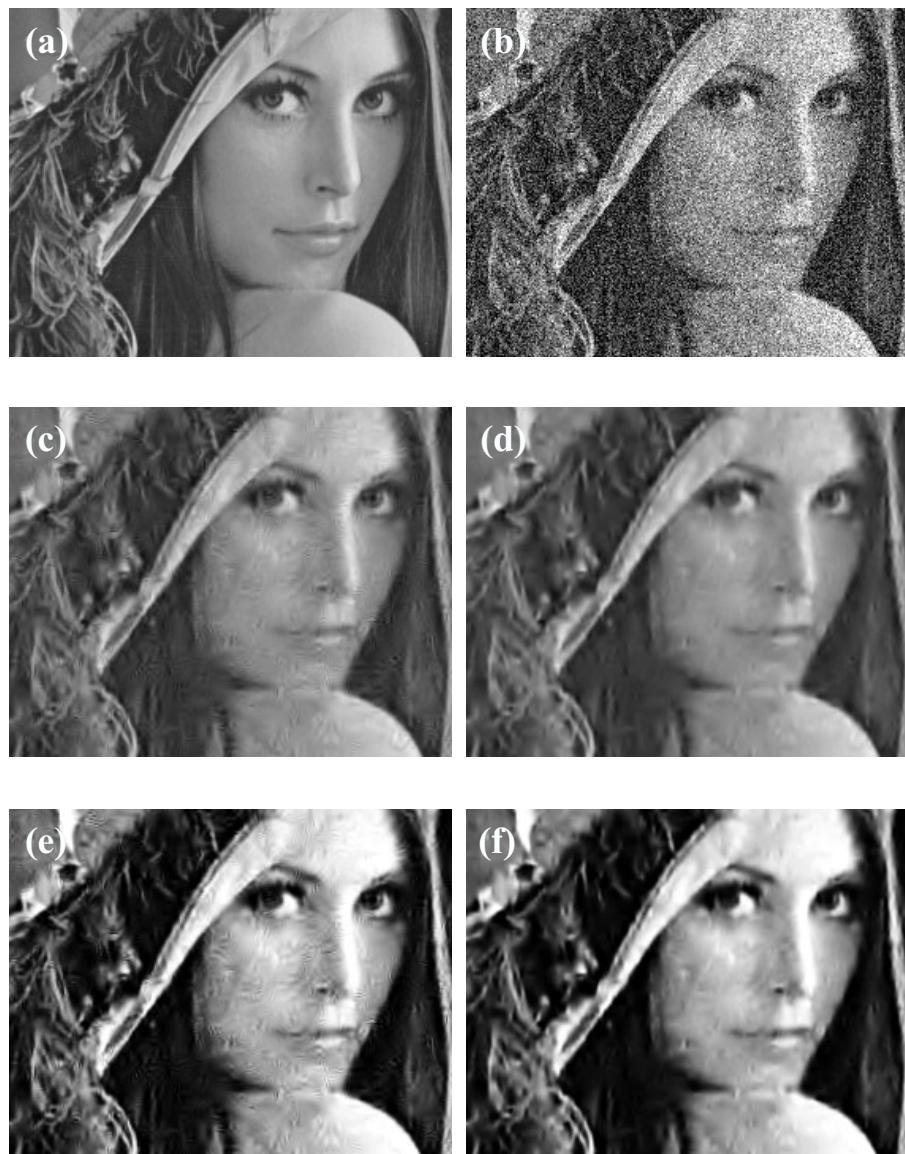


Figure 5.13: Enlarged parts of (a) the noise-free Lena image, (b) a noisy image standard deviation 40, (c) the result of the improved, redundant implementation of the LAWMAP [Mihcak99] method (LAWMAP-ND), PSNR=28.58 dB, and (d) the result of the GENLIK method PSNR=29.08 dB. (e) and (f) are histogram-equalized versions of the images (c) and (d), respectively.



Figure 5.14: 512x512 Barbara image (top) and the noisy image, with $\sigma = 25$ (bottom).



Figure 5.15: The results of the LAWMAP-ND (top, PSNR=27.8 dB) and the GENLIK (bottom, PSNR=28.3 dB) methods on the noisy image from Fig. 5.14.

Discussion

We developed a new joint detection and estimation method for wavelet domain image denoising. The method combines the MMSE criterion with an assumption that each wavelet coefficient represents a signal of interest with the probability $p < 1$. To estimate this probability, we have introduced an analytic model, which is adapted to both the global coefficients histogram and a local indicator of spatial activity. In particular, the local spatial activity indicator was chosen simply as the locally averaged magnitude, and a simple local model for its statistical distributions given the signal presence and given the signal absence was assumed. The resulting algorithm is of low complexity, both in its concept and the implementation. Furthermore, the method is fully automatic. The local window size is fixed and all the other parameters are computed directly from the input noisy image.

In comparison with the advanced MRF based approach from Chapter 4, the GENLIK algorithm is considerably less complex. For images that are dominated by large areas of uniform or slowly varying intensity, intercepted by sharp edges and thin lines, the MRF based method offers a superior performance. However, for most images of natural scenes, the two methods yield similar results, and on images that are rich with natural texture the GenLik method yields a superior performance in comparison with the MRF based one.

The new algorithm compares well with the best state of the art methods from literature of similar and of higher complexity. The advantage over related methods has been demonstrated both in terms of PSNR and visually.

Many extensions of the proposed method are possible. Using of a more complex indicator of the local spatial activity, which takes into account the inter-scale correlations as well, is likely to further improve the noise suppression performance. For such extensions, the appropriate pdf models $p(z_l|H_l^{0,1})$ should be developed. This and other aspects, like e.g., use of more exact expressions for $E(y_l|w_l, H_l^1)$ instead of the approximation $E(y_l|w_l, H_l^1) = w_l$ are subjects for further research.

5.4 An empirical GenLik Approach

Now we slowly turn towards various practical applications. The fully analytical method from the previous Section offers a good performance on various types of images, in the case of Gaussian noise. In numer-

ous practical applications one faces the suppression of Gaussian noise or noise which can be approximated by Gaussian. However, we shall also treat applications in which this is not the case. In order to apply the previous, analytical approach one needs first to estimate the parameters of the underlying distribution $p_Y(y)$ of the noise free coefficients from their noisy histogram. For other than Gaussian noise types this might be difficult. And we go one step further: suppose that a versatile algorithm is desired, which might be suboptimal, but applicable to various and possibly unknown noise distributions. This reasoning motivates the empirical approach that we introduce now.

We shall develop two practical algorithms, the use of which will be later illustrated in various applications: medical image denoising, image deblurring and speckle noise suppression in SAR images. This Section explains a global idea behind these algorithms and puts them in a common framework, from which other extensions may be foreseen.

5.4.1 The main idea and the global concept

A common concept of the algorithms that follow in the rest of this Chapter is an *empirical estimation* of the probabilities and the probability density functions that specify the generalized likelihood ratio estimator in Eq (5.3.3). These empirical estimates will be found using *masks* which roughly indicate the positions of important wavelet coefficients in a given detail image. The notion of masks may bring us back to the MRFs from Chapters 3 and 4, but now we shall focus our interest in a different direction. A simplified concept, which illustrates the use of masks for the empirical estimation of the required conditional densities is given in Fig. 5.16.

We shall find it convenient to represent the two hypotheses H_l^1 and H_l^0 , referring to the presence and to the absence of a signal of interest, by the events $X_l = 1$ and $X_l = 0$, respectively, where X_l is a binary random variable. With this notation we have $P(H_l^1) = P(X_l = 1)$, $p_{Wl}(w|H^1) = p_{Wl|X_l}(w_l|1)$, etc. The estimator from Eq (5.3.3) which we consider is thus

$$\hat{y}_l = \frac{\xi_l \mu_l}{1 + \xi_l \mu_l} w_l, \quad (5.4.1)$$

$$\xi_l = \frac{p_{Wl|X_l}(w_l|1)}{p_{Wl|X_l}(w_l|0)}, \quad \mu_l = \frac{P(X_l = 1|\mathcal{P})}{P(X_l = 0|\mathcal{P})}. \quad (5.4.2)$$

In estimating μ_l we may use a *continuous* indicator of local spatial activity z_l , as we have done in the analytical approach of Section 5.3,

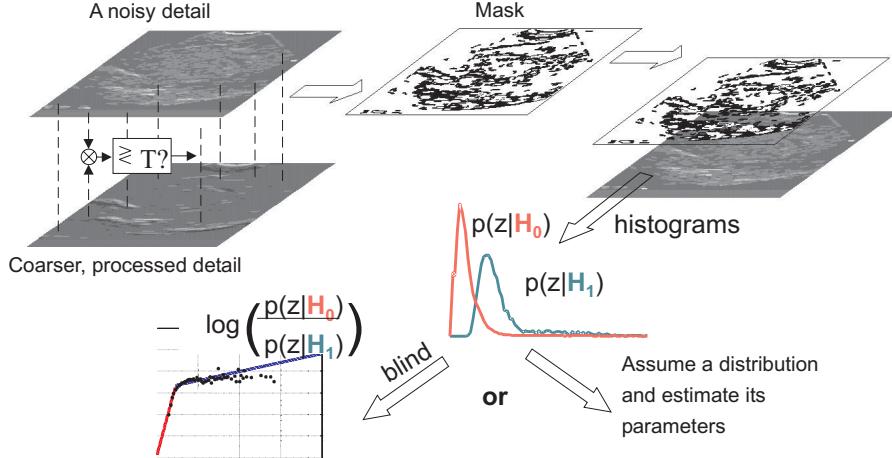


Figure 5.16: The idea of the empirical GenLik method.

where

$$\hat{\mu}_l = \rho \frac{p_{Z_l|X_l}(z_l|1)}{p_{Z_l|X_l}(z_l|0)}, \quad \rho = \frac{P(X_l = 1)}{P(X_l = 0)}. \quad (5.4.3)$$

An alternative, *discrete* indicator of the local spatial activity may be derived from the estimated mask $\hat{\mathbf{x}}$, as a function of neighboring labels $\hat{\mathbf{x}}_{\partial l}$

$$\hat{\mu}_l = \frac{P(X_l = 1|\hat{\mathbf{x}}_{\partial l})}{P(X_l = 0|\hat{\mathbf{x}}_{\partial l})}. \quad (5.4.4)$$

We have first used the latter case in Section 3.5 in the scope of an MRF based method. Now it has been placed into a broader framework, where MRF's are not necessarily used.

To estimate the above probabilities and pdf's empirically from a given detail image, an estimate of the mask $\hat{\mathbf{x}}$ is required. In this respect, we now introduce a simple and robust *coarse-to-fine* procedure:

$$\hat{x}_l = \begin{cases} 0, & \text{when } |w_{l,j}| |\hat{y}_{l,j+1}| < (K\hat{\sigma}_{n,j})^2, \\ 1, & \text{when } |w_{l,j}| |\hat{y}_{l,j+1}| \geq (K\hat{\sigma}_{n,j})^2, \end{cases} \quad (5.4.5)$$

where $\hat{\sigma}_{n,j}$ is an estimate of the standard deviation of noise at the resolution scale 2^j , and K is a parameter, which controls the notion of the “signal of interest”. In most applications K will be set to a fixed value. In some sensitive applications, like in medical ultrasound, a user interaction may be preferred. An alternative for this classification step is, e.g., the iterative correlation method of [Xu94], but that method imposes

some practical problems as we remarked in Section 2.5.4. We avoid here the iterative procedure, and yet (as it is demonstrated later) achieve a robust edge detection, due to a coarse-to-fine processing strategy: already processed, coarser detail coefficients $\hat{y}_{l,j+1}$ at the scale 2^{j+1} , are used to better detect the positions of important ones at the scale 2^j .

Having the estimate $\hat{\mathbf{x}} = \{\hat{x}_1 \dots \hat{x}_n\}$, let

$$S_0 = \{l : \hat{x}_l = 0\} \quad \text{and} \quad S_1 = \{l : \hat{x}_l = 1\}.$$

The empirical estimates $\hat{p}_{W_l|X_l}(w_l|0)$ and $\hat{p}_{W_l|X_l}(w_l|1)$ are found from the histograms of $\{w_l : l \in S_0\}$ and $\{w_l : l \in S_1\}$, respectively. Equivalently, one can find the empirical estimates of $p_{Z_l|X_l}(z_l|x_l)$. In practical algorithms, since the pdf's of the wavelet coefficients are highly symmetrical around zero, we shall compute the histograms of coefficient magnitudes $\omega_l = |w_l|$ and use

$$\hat{\xi}_l = \frac{p_{\Omega_l|X_l}(\omega_l|1)}{p_{\Omega_l|X_l}(\omega_l|0)}. \quad (5.4.6)$$

It is obvious that one cannot use directly the (normalized) histograms to compute the required conditional pdf's ratios: due to large errors in the tails, unreasonable results would be produced. Therefore, we shall use one of the following two strategies

- If no prior knowledge is available about the functional form of those densities: *piece-wise linear fitting of the log-ratio*. We develop a simple algorithm based on this idea in Section 5.5.
- If the functional forms of the densities are available: *MLE (maximum likelihood parameter estimation) of their parameters from the corresponding histograms*. In Section 5.6, we propose analytic models for the densities in case of SAR images and derive a simple and fast algorithm.

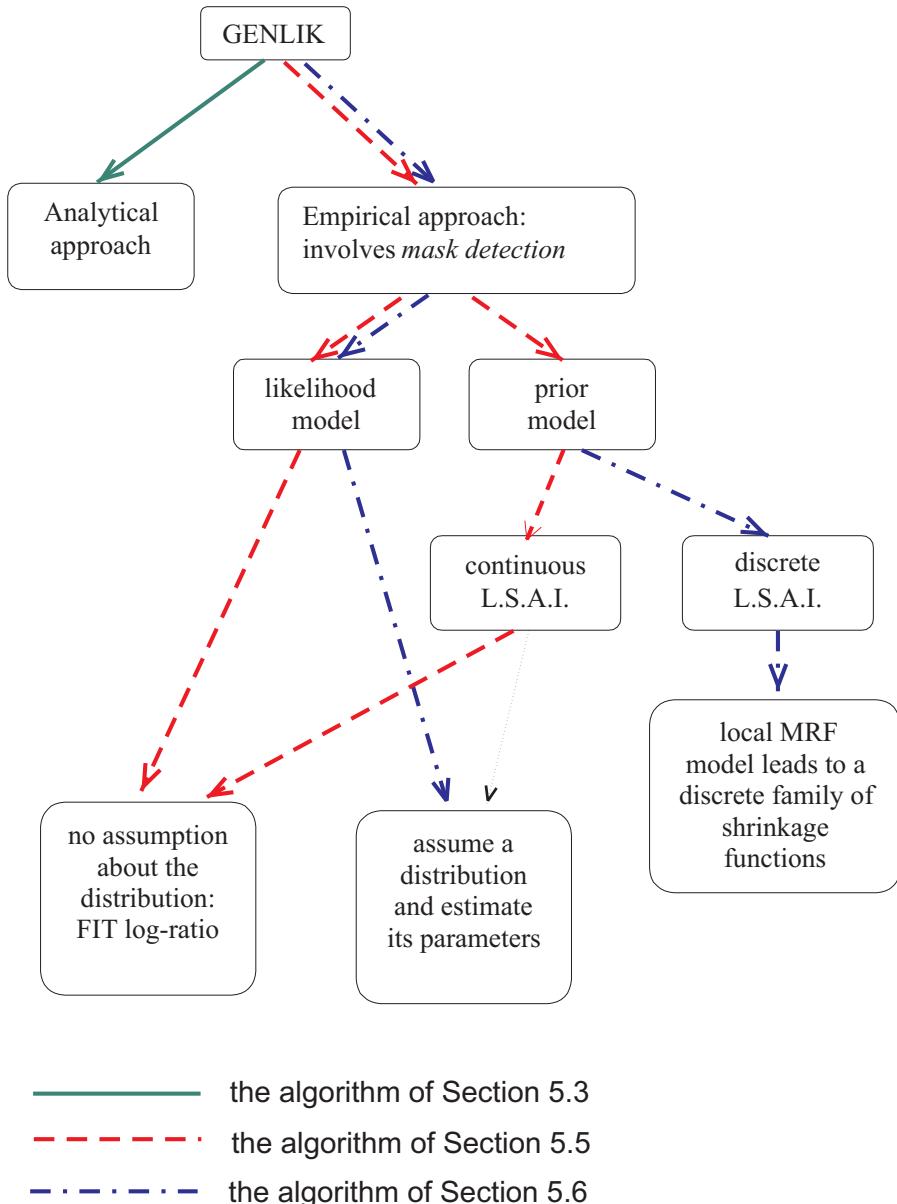


Figure 5.17: A summary of the developed GENLIK approaches.

5.5 A versatile algorithm for various noise types

In the framework of the previously described empirical generalized likelihood ratio approach, we now propose a versatile algorithm where no assumption is made about the type of underlying signal and noise distributions. Its application will be demonstrated on medical ultrasound images (correlated speckle noise), MRI images (Rician noise) and in image deblurring (suppression of artifacts produced by deconvolution).

The proposed algorithm starts from Eqs (5.4.1), (5.4.2) and (5.4.6). The required quantities therein will be obtained via the preliminary coefficient classification from Eq (5.4.5), according to which a mask $\hat{\mathbf{x}} = \{\hat{x}_1, \dots, \hat{x}_n\}$ is estimated for each detail image $\mathbf{w} = \{w_1, \dots, w_n\}$.

Reasoning that $P(X_l = 1)$ is the fractional number of labels for which $\hat{x}_l = 1$, we now estimate $\rho = P(X_l = 1)/P(X_l = 0)$ as

$$\hat{\rho} = \frac{\sum_{l=1}^n \hat{x}_l}{n - \sum_{l=1}^n \hat{x}_l}. \quad (5.5.1)$$

Further on, as was explained in Section 5.4.1, using $\hat{\mathbf{x}}$, and computing the corresponding histograms, we obtain empirical estimates $\hat{p}_{\Omega|X}(\omega|0)$, $\hat{p}_{\Omega|X}(\omega|1)$, $\hat{p}_{Z|X}(z|0)$, $\hat{p}_{Z|X}(z|1)$ of the required densities. Since those cannot be used directly in subsequent computations, and since we do not make any assumption about their analytic form (or approximation of it), we propose a simple solution. Note that our method does not use these densities directly, but rather their ratios: $\hat{\xi}(\omega) = \hat{p}_{\Omega|X}(\omega|1)/\hat{p}_{\Omega|X}(\omega|0)$ and $\hat{\eta}(z) = \hat{p}_{Z|X}(z|1)/\hat{p}_{Z|X}(z|0)$. Observing that the *logarithms* of both of these ratios can be approximated well by fitting a piece-wise linear curve as illustrated in Fig. 5.18, we approximate

$$\log(\hat{\xi}(m)) \simeq \begin{cases} a_1 + b_1 m, & \text{for } \hat{p}_{M|X}(m|0) > \hat{p}_{M|X}(m|1), \\ a_2 + b_2 m, & \text{for } \hat{p}_{M|X}(m|0) \leq \hat{p}_{M|X}(m|1), \end{cases} \quad (5.5.2)$$

$$\log(\hat{\eta}(e)) \simeq \begin{cases} c_1 + d_1 e, & \text{for } \hat{p}_{E|X}(e|0) > \hat{p}_{E|X}(e|1), \\ c_2 + d_2 e, & \text{for } \hat{p}_{E|X}(e|0) \leq \hat{p}_{E|X}(e|1), \end{cases} \quad (5.5.3)$$

In a summarized form, using the above three expressions, we estimate each wavelet coefficient as

$$\hat{y}_l = \frac{\rho \hat{\xi}(\omega_l) \hat{\eta}(z_l)}{1 + \rho \hat{\xi}(\omega_l) \hat{\eta}(z_l)} w_l. \quad (5.5.4)$$

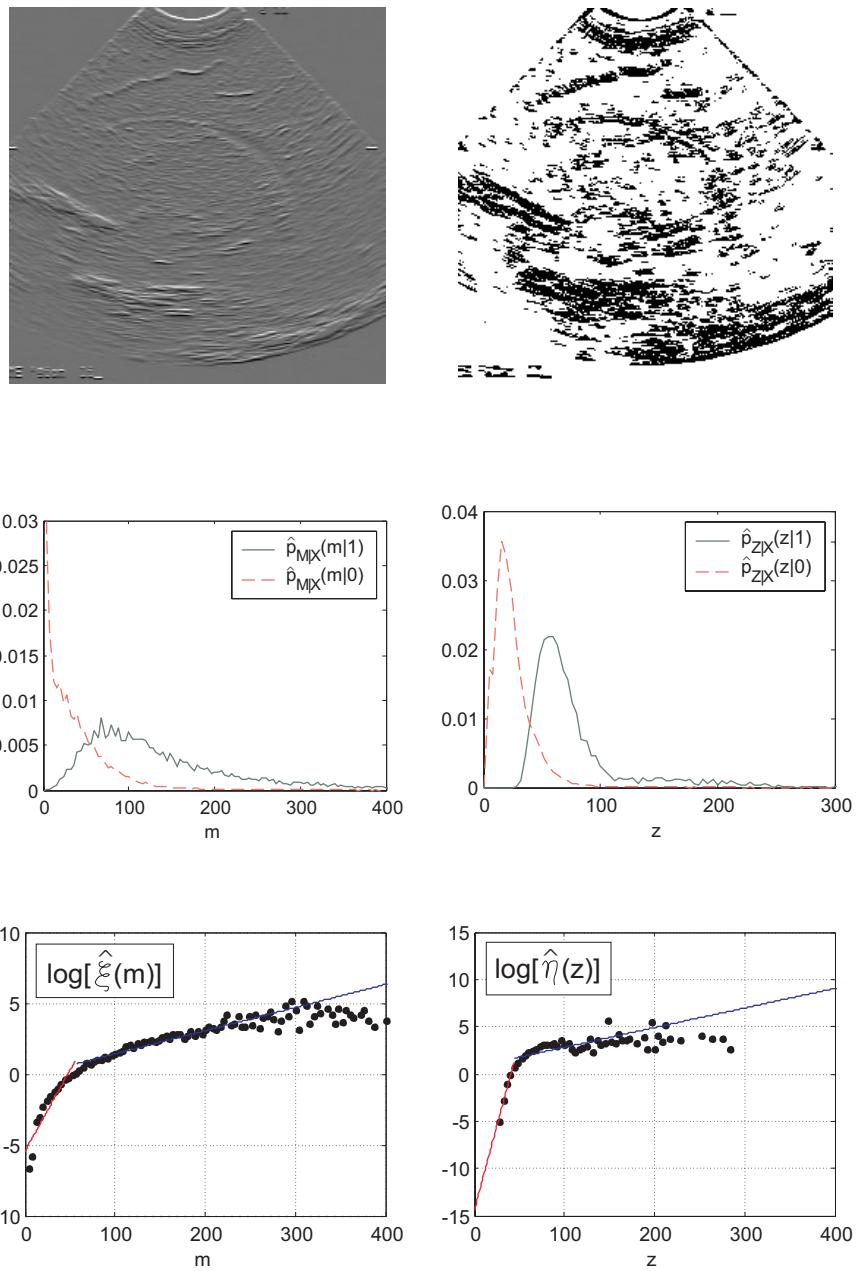


Figure 5.18: Examples of the empirical pdf's computed in the proposed method, for the top left ultrasound image from Fig. 5.19.

5.5.1 The practical algorithm

Here we first summarize the complete algorithm and then discuss some implementation details. The proposed algorithm is:

Compute the non decimated wavelet transform

Estimate the noise standard deviations $\sigma_{n,j}^d$ for each subband

Select a value of K to specify the level at which a signal becomes “of interest”

- For each scale and orientation

For all the spatial positions $l = 1, \dots, n$

If $|w_{l,j}\hat{y}_{l,j}| > (K\sigma_{n,j})^2$, set $\hat{x}_l = 1$; otherwise $\hat{x}_l = 0$
compute the L.S.A.I. z_l

Compute $\hat{\rho} = \sum_{l=1}^n \hat{x}_l / (n - \sum_{l=1}^n \hat{x}_l)$

With $S_0 = \{l : \hat{x}_l = 0\}$ and $S_1 = \{l : \hat{x}_l = 1\}$ compute:

$\hat{p}_{\Omega|X}(\omega|0) = \text{normalize}(\text{histogram}\{\omega_l : l \in S_0\})$

$\hat{p}_{\Omega|X}(\omega|1) = \text{normalize}(\text{histogram}\{\omega_l : l \in S_1\})$

$\hat{p}_{Z|X}(z|0) = \text{normalize}(\text{histogram}\{z_l : l \in S_0\})$

$\hat{p}_{Z|X}(z|1) = \text{normalize}(\text{histogram}\{z_l : l \in S_1\})$

Fit the log-ratio from (5.5.2):

Find: $R_\xi = \{l : \hat{p}_{\Omega|X}(\omega_l|0) \neq 0, \hat{p}_{\Omega|X}(\omega_l|1) \neq 0\}$ and
 $\xi_{\log} = \{\log(\hat{p}_{\Omega|X}(\omega_l|1)) - \log(\hat{p}_{\Omega|X}(\omega_l|0)) : l \in R_\xi\}$

Find the coefficients a_1, b_1 and a_2, b_2 such that

$a_1 + b_1\omega_l \simeq \xi_{\log,l}$, for $\xi_{\log,l} < 0$ and

$a_2 + b_2\omega_l \simeq \xi_{\log,l}$, for $\xi_{\log,l} > 0$ and use a function

$$\hat{\xi}(\omega) = \begin{cases} \exp(a_1 + b_1\omega), & \text{for } \hat{\xi}(\omega) < 1, \\ \exp(a_2 + b_2\omega), & \text{for } \hat{\xi}(\omega) \geq 1. \end{cases} \quad (5.5.5)$$

Fit the log-ratio from (5.5.3):

Apply the procedure from the previous item, using now $\hat{p}_{Z|X}(z|0)$ and $\hat{p}_{Z|X}(z|1)$, and leading to

$$\hat{\eta}(z) = \begin{cases} \exp(c_1 + d_1z), & \text{for } \hat{\eta}(z) < 1, \\ \exp(c_2 + d_2z), & \text{for } \hat{\eta}(z) \geq 1. \end{cases} \quad (5.5.6)$$

For $l = 1, \dots, n$: $\hat{y}_l = \hat{\rho}\hat{\xi}(\omega_l)\hat{\eta}(z_l)/(1 + \hat{\rho}\hat{\xi}(\omega_l)\hat{\eta}(z_l))w_l$

Apply the inverse wavelet transform

The new algorithm can be seen as a direct empirical extension of the analytical one from Section 5.3.6. Lacking prior knowledge about the underlying distribution of noise-free coefficients and noise, we now estimate all the “ingredients” in this method empirically.

Here we choose the local spatial activity indicator (L.S.A.I.) z_l as the averaged magnitude of eight neighboring coefficients of w_l , where the neighbors are the nearest coefficients in the same scale and the “parent” (i.e., the coefficient at the same spatial position at the first coarser scale). One can choose z_l as an *arbitrary function* of the neighboring wavelet coefficients in the same and/or adjacent resolution scales: all the other steps remain unchanged. It gives another degree of flexibility to this method. It might be interesting to experiment with other similar functions for z_l , where, e.g., the influence of the neighbors is weighted according to their distance from the current spatial position.

The proposed method can be implemented with arbitrary wavelets. Since the preliminary classification step involves products of the wavelet coefficients across scales, we use an “edge detection wavelet” (in particular, Mallat’s quadratic spline from Fig. 2.10). In various approaches where interscale coefficient products or ratios were used (e.g., [Xu94],[Hsung99],[Malfait97] and our own from Chapter 4) this wavelet was found the most appropriate.

5.5.2 On medical applications

A possible application where the proposed method may find its use is medical imaging. We shall demonstrate its application on medical ultrasound images and on medical MRI images. In the latter case a slight extension of the algorithm will be required to cope with the noise induced reduction of image contrast.

In medical images, noise suppression is a particularly delicate and difficult task. A trade off between noise reduction and the preservation of actual image features has to be made in a way that enhances the diagnostically relevant image content. Image processing specialists usually lack the biomedical expertise to judge the diagnostic relevance of the denoising results. For example, in ultrasound (US) images, even speckle noise may contain information useful to medical experts, since the speckles originating from different tissues have different properties. The use of speckled “texture” in US B-scans for diagnostics was discussed in, e.g., [Kossoff76, Sommer81]. Therefore, even though smoothing out speckle noise may increase the signal to noise ratio and may be useful for some

applications (e.g., segmentation) the denoised image may have less diagnostic relevance than the original, noisy one.

Another difficulty in dealing with biomedical images is, as pointed out in [Unser96], the extreme variability of signals and the necessity to operate on a case by case basis. This motivates the construction of robust and versatile denoising methods that are applicable in various circumstances, rather than being optimal under very specific conditions. The algorithm proposed above is robust and adapts to various types of image noise as well as to the *preference of the medical expert*: a single parameter (K) can be used to balance the preservation of (expert-dependent) relevant details against the degree of noise reduction.

In particular, the classification step of the proposed method involves a tunable parameter K , which can be related to the notion of the expert-defined “relevant image features”. Once this parameter is specified, all the other parameters are computed directly from the noisy image. In certain applications the optimal value of K can be selected as the one that maximizes the signal-to-noise ratio (SNR) and the algorithm can operate as fully automatic. However, in most medical applications the tuning of this parameter leading to a gradual noise suppression may be advantageous. Since the proposed method is simple and fast, such a tuning is feasible.

5.5.3 Application to medical ultrasound images

Ultrasound images are corrupted by speckle noise [Goodman76] that affects all coherent imaging systems. Within each resolution cell a number of elementary scatterers reflect the incident wave towards the sensor. The backscattered coherent waves with different phases undergo a constructive or a destructive interference in a random manner. The acquired image is thus corrupted by a random granular pattern, that hinders the interpretation of the image content and reduces detectability of the features of interest.

For an experienced radiologist, speckle noise, that is in medical literature also referred to as “texture” [Wagner83], may present useful diagnostic information [Kossoff76, Sommer81]. In this application, it is therefore advantageous to provide a user interactive denoising method, where the degree of speckle smoothing can be tuned. The examples in Fig. 5.19 illustrate the gradual suppression of speckle using the proposed method. The results in this figure correspond to the window size 5x5 and different values of the tuning parameter K in Eq (5.4.5). The

results demonstrate that the increase of K leads to a stronger suppression of the background texture and the enhancement of sharp intensity variations.

We investigate the quantitative performance of the method, using images with artificial speckle noise. A speckled image $\mathbf{v} = \{v_1, \dots, v_n\}$ is commonly modeled as [Achim01, Sattar97]

$$v_l = f_l \vartheta_l, \quad (5.5.7)$$

where $\mathbf{f} = \{f_1, \dots, f_n\}$ is a noise-free image, and $\vartheta = \{\vartheta_1, \dots, \vartheta_n\}$ is a unit mean random field. Realistic spatially correlated speckle noise ϑ_l in ultrasound images can be simulated by lowpass filtering a complex Gaussian random field and taking the magnitude of the filtered output [Sattar97]. We perform this lowpass filtering by averaging the complex values in the 3x3 sliding window. Such a short-term correlation was found sufficient to model the realistic images well [Achim01]. By changing the variance of the underlying complex Gaussian random field, we generate images with different levels of speckle noise. Two types of reference noise-free images are used: (1) real ultrasound images from Fig. 5.19, in which natural speckle noise was previously suppressed by the proposed method, and (2) a purely synthetic image in Fig. 5.20, which consists of regions with uniform intensity, sharp edges, and strong scatterers.

Regarding the quantitative noise suppression performance, the proposed method shows a stable behaviour with respect to the tuning parameter K . Fig. 5.21 demonstrates that for different noise levels and for different test images the same value of this parameter can be chosen to provide the maximum output SNR. It can be also seen that the window size 3x3 is optimal under the assumed speckle model.

We compare the performance of the proposed method with one conventional approach in speckle filtering: the homomorphic Wiener filter [Achim01]. In particular, we apply Matlab's spatially adaptive Wiener filter to the image logarithm and subsequently perform the exponential transformation of the filtered output. The window size of the Wiener filter was experimentally optimized to produce the maximum output SNR for each test image and for each noise level used in simulations. Our results clearly demonstrate that the proposed filter outperforms the homomorphic spatially adaptive Wiener filtering both in terms of SNR (Fig. 5.22) and in terms of the visual quality of results in Fig. 5.20.

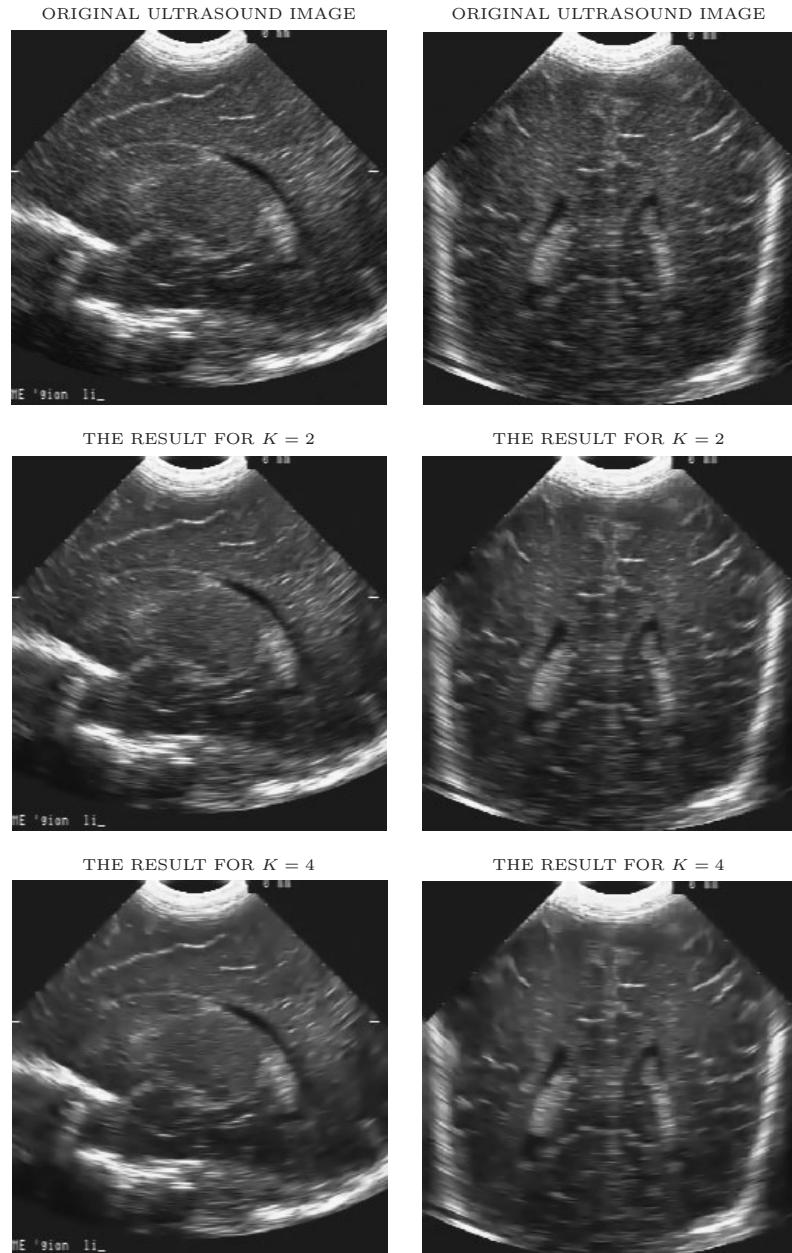


Figure 5.19: Original ultrasound images and the results of the proposed method for two different values of the parameter K , illustrating a gradual suppression of speckle.

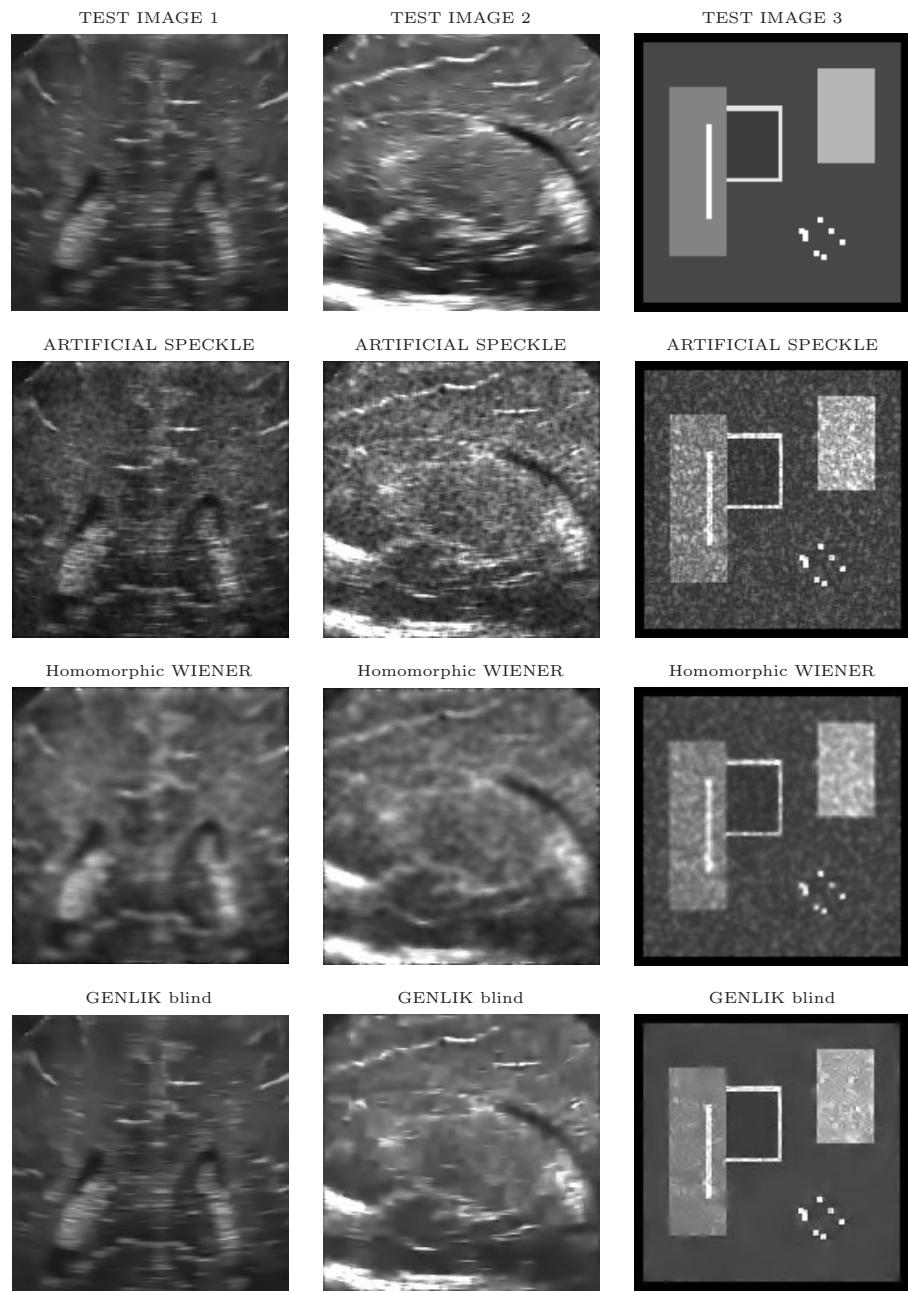


Figure 5.20: From top to bottom: test images, artificially speckled images, the results of the homomorphic spatially adaptive Wiener filter, and the results of the proposed method.

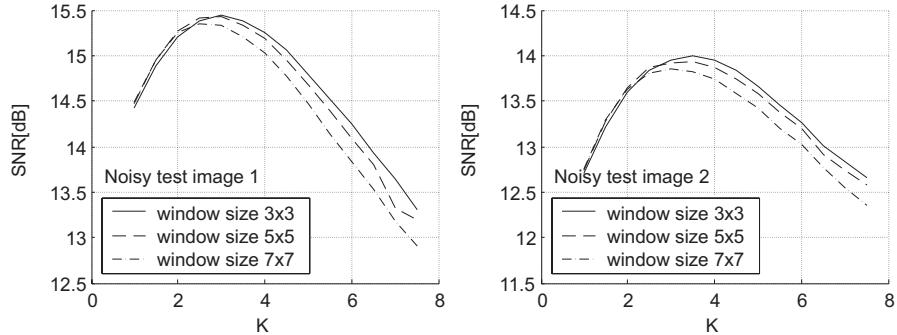


Figure 5.21: The performance of the proposed method as a function of the value of the parameter K . Left: noisy test image 1, input SNR=13.6 dB, Right: noisy test image 2, input SNR=11.6 dB.

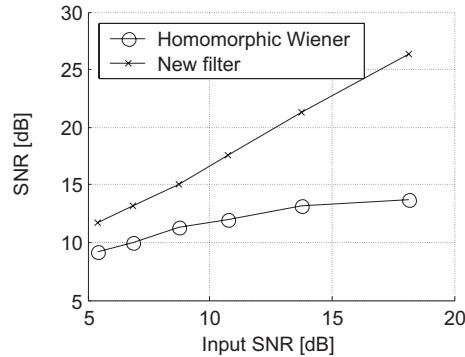


Figure 5.22: Comparison between the proposed filter and the homomorphic Wiener filter on the synthetic test image of Fig. 5.20.

5.5.4 The algorithm adapted for medical MRI images

In magnetic resonance imaging the practical limits of the acquisition time impose a trade-off between SNR and image resolution (see, e.g., [Gudbjartsson95, Macovski96, McVeigh85]). The acquisition time is limited in practice due to patient comfort and due to physical limitations (especially in dynamic applications, such as cardiac imaging and functional MRI [Nowak99]). Post-processing noise reduction is therefore often seen as the only means of achieving a desired MRI image quality. Previously proposed wavelet domain filtering methods for MRI images

were based on different thresholding schemes [Nowak99, Weaver92], including [Xu94], where the coefficient selection was based on inter-scale correlations. Only the first method among the above mentioned accounts explicitly for a specific noise type in MRI. In particular, Nowak in [Nowak99] explained that proper removal of noise from MRI images requires (1) a certain image transformation (squaring the magnitude image) prior to computing the wavelet transform, and (2) a modification of both the scaling and the wavelet coefficients. First we explain in some more detail the problem and then extend our algorithm from Section 5.5 along those lines.

Noise in MRI

The main source of noise in MRI images is the thermal noise in the patient [Edelstein86]. The MRI image is commonly reconstructed by computing the inverse discrete Fourier transform of the raw data [Macovski96]. The signal component of the measurements is present in both real and imaginary channels; each of the two orthogonal channels is affected by additive white Gaussian noise. The noise in the reconstructed complex-valued data is thus complex white Gaussian noise.

Most commonly, the *magnitude* of the reconstructed MRI image is used for visual inspection and for automatic computer analysis. Since the magnitude of the MRI signal is the square root of the sum of the squares of two independent Gaussian variables, it follows a Rician distribution. In low intensity (dark) regions of the magnitude image, the Rician distribution tends to a Rayleigh distribution [Papoulis84] and in high intensity (bright) regions it tends to a Gaussian distribution. A practical consequence is a reduced image contrast: noise increases the mean value of pixel intensities in dark image regions.

Due to the signal-dependent mean of the Rician noise, both the wavelet and the scaling coefficients of a noisy MRI image are *biased* estimates of their noise-free counterparts. In [Nowak99] it was shown that one can efficiently overcome this problem by filtering the *square* of the MRI magnitude image in the wavelet domain. In the squared magnitude image, data are non-central chi-square distributed, and the wavelet coefficients are no longer biased estimates of their noise-free counterparts. The bias still remains in the scaling coefficients, but is not signal-dependent and it can be easily removed: at the resolution scale 2^j , from each scaling coefficient $2^{j+1}\sigma_c$ should be subtracted, where σ_c^2 is the underlying complex Gaussian noise variance. This value is typically

estimated from the noisy image: MRI images include an empty region of air outside the patient; in the squared magnitude image, the average pixel value in those empty (border) regions is $2\sigma_c^2$.

The adapted algorithm and its results

According to the explanation given above, we now adapt the algorithm of Section 5.5.1 for the MRI images as:

- Compute the square of the MRI magnitude image;
- Compute the non-decimated wavelet transform with L decomposition levels (in practice, we used $L=4$);
- Estimate the wavelet coefficients as described in Section 5.5.1;
- Subtract $2^{L+1}\sigma_c$ from the scaling coefficients;
- Apply the inverse wavelet transform;
- Compute the square root of the image.

First, we illustrate the performance of the proposed method on an MRI image with artificial noise, and compare it with the spatially adaptive Wiener filter. For the reference noise-free image we have chosen a sufficiently clean, original MRI magnitude image of a human brain (the top left image in Fig. 5.23). In simulations, complex zero mean white Gaussian noise with standard deviation σ_c was added to this image. The top right image in Fig. 5.23 shows the noisy MRI magnitude for $\sigma_c = 30$. The results from Fig. 5.23 demonstrate that the proposed method suppresses noise better and preserves image sharpness better than the spatially adaptive Wiener filter.

The quantitative performance of the proposed method, with respect to the tuning parameter K is illustrated in Fig. 5.24(a). For different amounts of noise the optimal value of this parameter is in the range (1.8, 2), and the algorithm shows a stable behaviour with respect to K . In our experiments, on a number of different reference MRI images, the improvement over spatially adaptive Wiener filtering was at least 0.5 dB (for relatively clean images) and more than 3 dB for low SNR images. Fig. 5.24(b) illustrates such a comparison between the two filters for the reference image from Fig. 5.23.

The application of the proposed method to real noisy MRI images is demonstrated in Fig. 5.25. The MRI images were provided by the Ghent University hospital. The noise suppression in these images facilitates further automatic processing, like, e.g., the segmentation.

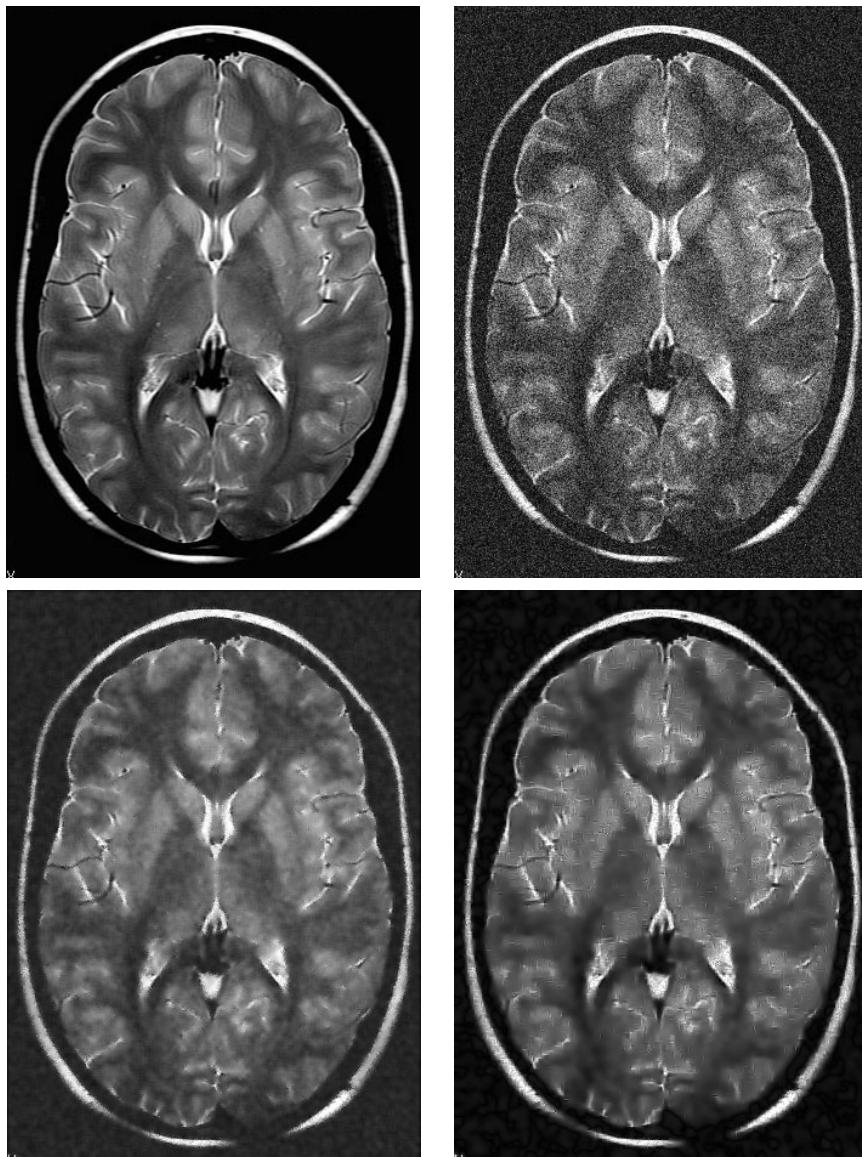


Figure 5.23: Top left: original MRI image. Top right: image with artificial Rician noise. Bottom left: the result of spatially adaptive Wiener filtering. Bottom right: the result of the proposed method.

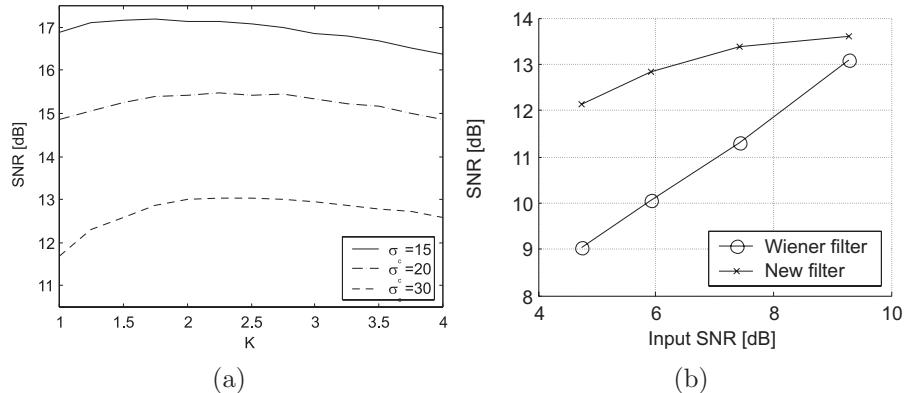


Figure 5.24: Quantitative performance of the proposed method. (a) Influence of the parameter K . (b) Noise suppression performance in comparison with the spatially adaptive Wiener filtering. Reference image is the top left MRI image from Fig. 5.23.

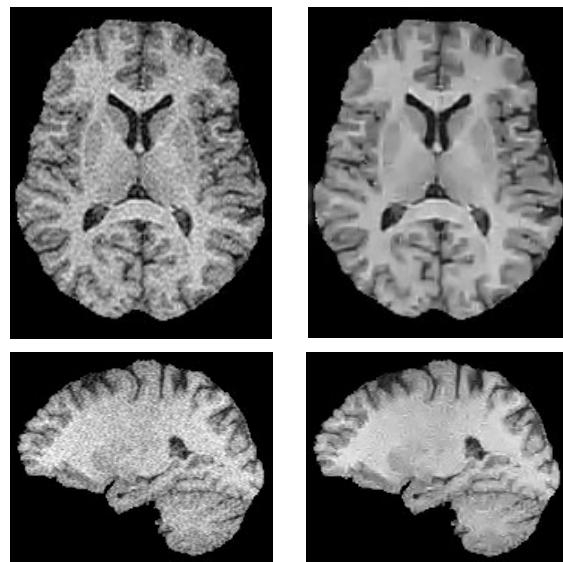


Figure 5.25: Application to realistic noisy MRI images. Left: original images, right: the results of the proposed method.

5.5.5 Application in image deblurring

Another application of the proposed method is in image deblurring. Sharpening of edges increases noise and causes various artifacts. Wavelet based denoising techniques are therefore often used as a regularization step in image deblurring [Mallat98, p.455]. We have tested our algorithm in combination with the wavelet based regularized deconvolution (WaRD) approach of [Neelamani99]. Instead of the classical hard thresholding that is implemented in the WaRD software of the Rice University [Neelamani99a], we have implemented the algorithm described in Section 5.5.1. Other wavelet domain deconvolution methods include [Donoho95c, Jalobeanu01, Kalifa99].

The WaRD approach of [Neelamani99] considers the following classical deconvolution scenario. Two degradations corrupt the observation v_l of the desired data f_l : convolution with a linear shift-invariant system having impulse response h and additive white Gaussian noise ϑ_l with variance σ^2 . The observed data are

$$v_l = (h \star f)_l + \vartheta_l, \quad l = 0, \dots, n - 1 \quad (5.5.8)$$

In the discrete Fourier transform (DFT) domain, defined as $H(\omega_k) = \sum_{l=0}^{n-1} h_l e^{-j\omega_k l}$, with $\omega_k = 2\pi k/N$, $k = 0, \dots, n - 1$, one equivalently has

$$V(\omega_k) = H(\omega_k)F(\omega_k) + \vartheta(\omega_k), \quad k = 0, \dots, n - 1 \quad (5.5.9)$$

If the system frequency response has no zeros, an unbiased estimate of f is obtained by a pure inversion: $\hat{F}(\omega_k) = H^{-1}(\omega_k)V(\omega_k)$. This inversion amplifies noise enormously at frequencies ω_k where $H(\omega_k)$ is small. Noise amplification can be alleviated by using an approximate, regularized inverse instead of a pure inverse. Here, the regularization assumes attenuation (shrinking) the frequency components $\hat{F}(\omega_k)$: the components at which the signal power is smaller are shrunk more; the aim is to provide a better solution by reducing noise in exchange for some distortion in the estimate. The WaRD method combines the Fourier domain regularization (FoRD) (actually Wiener filtering) followed by wavelet denoising. The FoRD estimate is given by $\hat{F}_{FoRD}(\omega_k) = H^{-1}(\omega_k)R_\alpha(\omega_k)V(\omega_k)$, where the frequency dependent weights R_α are

$$R_\alpha(\omega_k) = \frac{|H(\omega_k)|^2 |P_f(\omega_k)|^2}{|H(\omega_k)|^2 |P_f(\omega_k)|^2 + \alpha\sigma^2}. \quad (5.5.10)$$

$P_f(\omega_k)$ is the power spectral density [Papoulis84] of the input signal, and α is a regularization parameter, which controls the amount of noise suppression and the amount of signal distortion.



Figure 5.26: (a) Original image; (b) blurred image, with 9x9 boxcar blur; (c) the FoRD estimate of [Neelamani99a]; (d) the final result (the WaRD estimate) of [Neelamani99a] and (e) the proposed method applied to the FoRD estimate (c).

The WaRD software of [Neelamani99a] uses a small value of the regularization parameter ($\alpha = 0.1$) and suppresses the remaining noise by wavelet shrinkage. The input in the wavelet denoising procedure is \hat{f}_{FoRD} , which is the inverse Fourier transform of \hat{F}_{FoRD} .

While the original WaRD method uses a simple hard thresholding of the wavelet coefficients for denoising the FoRD estimate, other wavelet based denoising methods can be used in this approach as well. The results in Fig. 5.26 demonstrate that the proposed algorithm from Section 5.5.1 may be considered as a good candidate in this respect. Noise and artifacts are well suppressed and the sharpness of image details is well preserved. A more extended analysis is left for further research.

5.5.6 Discussion

In this Section, we proposed a new, robust and efficient wavelet domain denoising technique, which is applicable to various types of image noise. The method is simple to implement and is fast. It can be seen as a direct empirical extension of the analytical one from Section 5.3.6. Lacking prior knowledge about the underlying distribution of noise-free coefficients and noise, we estimate in this method all the involved probabilities and pdf's empirically. The method relies on the persistence of significant wavelet coefficients across scales and performs accordingly a preliminary coefficient classification, in a coarse-to-fine manner: already processed coefficients from the coarser scales are used to better detect the positions of edges in a given detail image. We demonstrated the application of the method to different types of noise in medical imaging and in image deblurring.

The proposed method is interesting for medical image denoising, since it accounts for the preference of the medical expert: a single parameter can be used to balance the preservation of (expert-dependent) relevant details against the degree of noise reduction. Such a user interaction is in the first place useful for speckle noise removal from the ultrasound images. The algorithm was adapted to cope with the noise in MRI images. In this case, a signal-dependent noise reduces the image contrast; both the wavelet and the scaling coefficients are biased counterparts of their noise-free versions. The proposed method is in this case applied to the squared magnitude image and the remaining, constant bias is removed from the scaling coefficients. The core of the algorithm remains the same.

5.6 SAR image despeckling

Active radar sensing is often a prime source of inventory information about remote and cloud-covered areas of the world. Due to its high penetration power, Synthetic Aperture Radar (SAR) acquires high resolution images in almost all atmospheric conditions. However, the automatic interpretation of SAR images is often extremely difficult due to speckle noise. Appearing as a random granular pattern, speckle seriously degrades the image quality and hampers the interpretation of image content.

In this Section, we propose and demonstrate a new, fast locally adaptive algorithm for speckle noise suppression in SAR images. First we explain the problem in some detail and we review briefly several well known filtering approaches.

5.6.1 Speckle noise in SAR images

Speckle noise in SAR images arises as a consequence of the coherent illumination used by radar. The principle of its formation is similar to that of the speckle formation in medical ultrasound (Section 5.5.3). Within each ground resolution cell a large number of elementary scatterers reflects the radar wave towards the sensor. The backscattered waves being coherent and having different phases undergo a constructive or a destructive interference in a random manner. For a surface that is rough on the scale of the radar wavelength, the number of elementary scatterers within a resolution cell is large enough to ensure the statistical independence in phase and amplitude of the elementary backscattered waves. The speckle is then fully developed and its formation is referred to as Rayleigh fading [Ulaby86a].

Usually, the speckle is partially suppressed already during the image formation process, by using the multilook technique: L “looks” at the same scene corresponding to, e.g., disjoint frequency bands are averaged. The L -look averaging reduces the standard deviation of speckle by a factor of \sqrt{L} , but it also deteriorates the resolution by a factor of L [Goodman76].

Fully developed speckle behaves as random multiplicative noise described by Eq (5.5.7). Within the homogeneous and the weakly textured areas of SAR images this multiplicative noise model is considered to be valid [Verhoest00b]: given the underlying scene reflectivity f_l , the observed pixel intensity is $v_l = f_l \vartheta_l$, where ϑ_l is a random variable. The spatial correlation of the speckle in SAR images is less prominent than

that in the medical ultrasound images, and is most often neglected in developing models for the filtering procedures [Foucher01, Walessa00].

It is widely accepted [Foucher01, Ulaby86b, Walessa00] that the observed intensity of an L -look SAR image conditioned on the underlying reflectivity is Gamma distributed: $p_{V|F}(v|f) \propto v^{L-1} \exp(-Lv/f)$. Under this model, the multiplicative noise is also Gamma distributed. Such models follow from the assumption that the real and the imaginary parts (i.e., the so called in-phase and quadrature components) of the speckle in a complex SAR image are zero mean Gaussian distributed. The mean-to-standard deviation ratio of the resulting noise density satisfies [Gagnon97]

$$\left(\frac{\bar{\vartheta}}{\sigma_\vartheta}\right)^2 = L = \text{constant}. \quad (5.6.1)$$

The usual way to characterize the speckle noise in a SAR image is to calculate Eq (5.6.1) over areas of uniform pixel intensity. The resulting estimate of L is often termed the *Equivalent Number of Looks* (ENL). Other noise distributions have been also used in literature, e.g., a χ^2 distribution with $2L$ degrees of freedom in [Lee89, Ulaby86a], and a Log-Normal distribution in [Gagnon97]. These models for the speckle noise density in SAR images are discussed in detail in [Verhoest00b, p.8-55].

Regardless of the particular functional form of the speckle noise pdf, a normalization to the unit mean $E(\vartheta) = \bar{\vartheta} = 1$ is largely employed in literature. In this case it can be shown that the following relations hold between the mean values \bar{f} and \bar{v} and the normalized standard deviations C_f and C_v [Foucher01, Kuan85]

$$\bar{f} = \bar{v}, \quad \text{and} \quad C_f^2 = \frac{C_v^2 - C_v^2}{1 + C_\vartheta^2}, \quad (5.6.2)$$

where $C_f = \sigma_f/\bar{f}$, $C_v = \sigma_v/\bar{v}$ and $C_\vartheta = \sigma_\vartheta/\bar{\vartheta} = 1/\sqrt{L}$. Numerous locally adaptive speckle filtering techniques, e.g., [Lee80, Kuan85, Lopes90, Lopes90a, Foucher01], use an estimate of C_v from a local window to distinguish homogeneous regions ($\hat{C}_v \leq C_\vartheta$) where $C_f = 0$ from heterogeneous areas ($\hat{C}_v > C_\vartheta$) where $C_f > 0$. Some of those and some other speckle filtering methods, are briefly reviewed in the next Section.

5.6.2 Filtering SAR images

A nice overview of various speckle filters and their comparative performance can be found in e.g., [Gagnon97, ShiFung94]. Here we briefly

review some of the well known filters. The indices that denote the spatial position are omitted for clarity: the corrupted pixel intensity is $v = f\vartheta$, where $\bar{\vartheta} = 1$. Many standard filters require knowledge of \bar{v} and σ_v . In practice, these quantities are estimated within a finite-size window. Noise standard deviation σ_ϑ (or the number of looks L) is either given as the input filter parameter or it is estimated from a uniform area in the image: $ENL = (\bar{v}/\sigma_v)^2 = 1/\sigma_\vartheta^2$.

Single-resolution speckle filters

Among the standard filters for speckle noise,[Frost82, Kuan85, Lee80] are best known. These filters apply a local MMSE estimation, employing the second-order sample statistics. Practically, the pixel intensity and the local mean [Kuan85, Lee80] or the neighboring pixel intensities [Frost82] are weighted, where the weighting factors depend on the local intensity variability.

In particular, the *Kuan filter* [Kuan85] considers the multiplicative noise model under the form $v = f + (\vartheta - 1)f$, from which the corresponding linear MMSE filter is deduced. Under the unit-mean noise assumption, the pixel value estimate \hat{f} is

$$\hat{f} = \bar{v} + \frac{\sigma_f^2(v - \bar{v})}{\sigma_f^2 + (\bar{y}^2 + \sigma_f^2)/L}, \text{ where } \sigma_f^2 = \frac{L\sigma_y^2 - \bar{y}^2}{L + 1}. \quad (5.6.3)$$

In the “pathological” cases, when the estimated $\sigma_f^2 < 0$, one puts $\hat{f} = \bar{v}$.

The *Lee filter* [Lee80] was earlier derived in a similar fashion, but by approximating the multiplicative noise model by a first-order Taylor expansion of v about f and ϑ . The corresponding filter is a special case of Eq (5.6.3), where the term σ_f^2/L is removed from the denominator.

The *Frost filter* [Frost82] is an adaptive Wiener filter which convolves the pixel values within a fixed size window with an exponential impulse response $\exp(-KC_v|t|)$, where $C_v = \sigma_v/\bar{v}$, t is a distance from the processed pixel, and K is a filter parameter. This response results from an autoregressive exponential model for the scene reflectivity f .

The enhanced versions of the above and similar filters (like the so-called *enhanced Lee* and the *enhanced Frost* filters [Lopes90a]) apply a combined filtering and a pixel classification: on the basis of the estimated local standard deviation, the image pixels are assigned into one of the three classes: *homogeneous*, *weakly textured* or *highly heterogeneous*. Supposedly homogeneous image segments are simply averaged, while the

highly heterogeneous ones are kept unmodified; only the remaining image segments (weakly textured and/or "edged") are adaptively filtered (with a given filter). Fig. 5.27 illustrates a SAR image and the three different classes of image segments mentioned above.

The first maximum a posteriori (MAP) approach to speckle filtering [Kuan87] was assuming a Gaussian probability density function (pdf) for the scene reflectivity, while the *Gamma-MAP* filter [Lopes90] assumes a more realistic Gamma distribution. The Gamma-MAP filter of [Lopes90] combines in practice the filtering with a pixel classification, as explained above, and estimates the pixel intensity as

$$\hat{f} = \begin{cases} \bar{v}, & \text{if } \sigma_v/\bar{v} \leq 1/\sqrt{L}, \\ \frac{(\alpha-L-1)\bar{v} + \sqrt{\bar{v}^2(\alpha-L-1)^2 + 4\alpha L g \bar{v}}}{2\alpha}, & \text{if } 1/\sqrt{L} < \sigma_v/\bar{v} < \sqrt{2/L}, \\ v, & \text{if } \sigma_v/\bar{v} \geq \sqrt{2/L}. \end{cases} \quad (5.6.4)$$

where

$$\alpha = \frac{L+1}{L(\sigma_v/\bar{v})^2 - 1}. \quad (5.6.5)$$

This filter is commonly used as a standard for comparison in recent speckle suppression literature. A more sophisticated and a high-complexity MAP filtering approach, which uses MRFs for texture modeling, was recently proposed in [Walessa00]. This approach retains image texture well but has problems with the preservation of the strong point-like targets¹. This is due to the fact that the underlying continuous MRF model penalizes strongly such highly localized intensity peaks; in their turn, these intensity peaks also hamper the estimation of the texture parameters in a given window. In order to solve this difficult problem, the approach of [Walessa00] removes the strong scatterers from the image and places them back after the image is processed.

Other speckle filters include: the *Oddy* filter [Oddy83] (it replaces the pixel intensity by the mean value of the pixel intensities within a varying-shape window); the *Adaptive Filter on Surfaces* (AFS) of [Alparone90] (employs a similar concept, where adaptive masks within a window are used for the selection of the pixel that are subsequently averaged); the *Crimmins* geometric filter [Crimmins85] (it represents an image as a 3-dimensional diagram; the morphological filtering is then performed on 2-dimensional slices, each of which is a binary image). The three filters mentioned above are usually outperformed by the ones that employ the

¹Strong scatterers have a great importance for the physical interpretation of the image content, because they correspond to high reflectivity objects.

local MMSE (Lee, Frost, Kuan) or local MAP estimation (Gamma MAP filter) (see, e.g., a comparative study in [Gagnon97]).

According to recent studies in [Foucher01, ShiFung94, Verhoest00b, Walessa00], the Gamma MAP filter from Eq (5.6.4) is currently among the best speckle filters, and we shall use it for comparisons with the new wavelet filter.

Wavelet domain speckle filters

For speckle noise, the standard wavelet thresholding is not effective. Therefore, usual wavelet domain despeckling techniques filter the image logarithm, since after the logarithmic transformation the speckle noise becomes approximately additive Gaussian. Such an approach was applied, e.g., in [Gagnon96] and [Odegard95]. Foucher *et al* [Foucher01] note that the logarithmic transformation itself leads to a biased estimation of the reflectivity, which calls for alternative approaches that filter the wavelet coefficients of the original, speckled image. Only a few such methods have been published. The approach of [Foucher96] basically applies the correlation method of [Xu94] (Section 2.5.4) to select the wavelet coefficients at different scales. A recently proposed multiscale MAP filtering technique of [Foucher01] performs a filtering combined with pixel classification (see Fig. 5.27): in supposedly homogeneous and highly heterogeneous regions no filtering is performed, but the wavelet coefficients are replaced by zero in the first case and left unmodified in the second case; in supposedly textured regions the MAP filtering is applied, with Pearson-type prior and noise distributions, leading to a third-degree equation that is numerically solved. In comparison with the Gamma MAP filter, the multiscale MAP filter of [Foucher01] preserves more detail in weakly textured areas, but it smooths more the strong reflectors. In general, the preservation of strong scatterers (point-like targets) appears a difficult problem in numerous SAR filtering methods, including the most complex ones like the earlier mentioned single-resolution MAP-MRF approach of [Walessa00]. A new wavelet filter that we introduce next has precisely the ability to preserve the strong scatterers well, while smoothing speckle in the homogeneous areas.



Figure 5.27: A SAR image and an illustration of a homogeneous, an edged and a highly heterogeneous region. Typical speckle filtering methods apply different types of filtering in these regions.

5.6.3 A new family of wavelet shrinkage functions

Most of the SAR speckle filters from the previous Section do not show a sufficient spatial adaptation to remove noise in the homogeneous areas while retaining details in the highly heterogeneous ones. To alleviate this problem, various “practical tricks” are employed, among which the most common one is a preliminary pixel classification into three disjoint classes and applying the actual filtering only in one of those, while simply averaging or leaving unchanged the pixels from the other classes.

In contrast, our approach achieves a “soft” spatial adaptation, from the homogeneous to the highly heterogeneous areas. We do not filter the wavelet coefficients in different regions according to different rules, but apply a unique shrinkage function, that automatically adapts to the local spatial activity in the image, as it is depicted in Fig. 5.28. We start from a global concept from Section 5.4, and in particular from the equations (5.4.1), (5.4.4) and (5.4.6). Fig. 5.17 explains in a summarized form both the common framework and the conceptual differences between the algorithm of this Section and the two others that we developed in this Chapter. The main novelty here is that we introduce functional forms for the densities $p_{\Omega_l|X_l}(\omega_l|0)$ and $p_{\Omega_l|X_l}(\omega_l|1)$ of the coefficient magnitudes. The parameters of the proposed densities are estimated quickly from

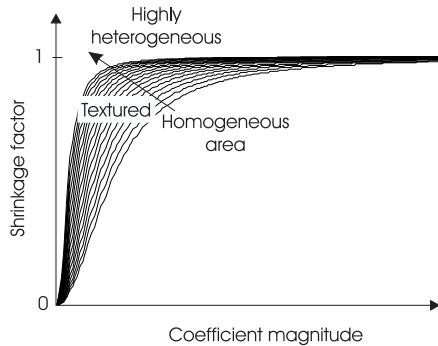


Figure 5.28: An illustration of the proposed wavelet shrinkage functions for SAR image despeckling. A unique shrinkage function is softly adapted to the local homogeneity in the image; “strong reflectors” are never smoothed.

the corresponding coefficient histograms. The discrete indicator of the local spatial activity (L.S.A.I.) that we use here, leads to a family of shrinkage functions (Fig. 5.28) similarly as we had earlier in Section 3.5.3 (Fig. 3.14). While the shrinkage functions from Section 3.5.3 were chosen heuristically, the new ones are adapted to the statistical properties of the coefficient magnitudes of SAR images.

An exponential-Gamma likelihood model for SAR images

The pdf models that we propose here result from our experiments on a number of different SAR images. By investigating the histograms of the wavelet coefficients in regions dominated by speckle noise on the one hand and in regions dominated by actual image transitions on the other hand, we observed that the two corresponding pdf’s follow well the scaled exponential and Gamma densities, respectively:

$$\hat{p}_{\Omega_l|X_l}(\omega|0) \simeq (1/a) \exp(-\omega/a), \quad (5.6.6)$$

$$\hat{p}_{\Omega_l|X_l}(\omega|1) \simeq (1/2b)(\omega/b)^2 \exp(-\omega/b). \quad (5.6.7)$$

The fit between these analytic models and the normalized coefficient histograms (computed from the masks in the new method) are illustrated in Fig. 5.30. The proposed pdf models involve only *one parameter* each; the maximum likelihood estimates (MLE) of these parameters are computed quickly from the observed data: if we denote, $S_0 = \{l : \hat{x}_l = 0\}$ and $S_1 = \{l : \hat{x}_l = 1\}$, one can show (see the Appendix B) that the MLE

estimates of the parameters of the proposed pdf models are

$$\hat{a} = (1/N_0) \sum_{i \in S_0} \omega_i, \quad N_0 = \#S_0, \quad (5.6.8)$$

$$\hat{b} = (1/3N_1) \sum_{i \in S_1} \omega_i, \quad N_1 = \#S_1, \quad (5.6.9)$$

where $\#S$ is the cardinality of the set S .

For the preliminary coefficient classification, from which the required parameters and the local spatial activity indicator (L.S.A.I.) are estimated, we use the coarse-to-fine procedure from Eq (5.4.5), with $K = 1$. For the simplicity of implementation, $\hat{\sigma}_{n,j}^d$ in Eq (5.4.5) is now estimated as (or strictly speaking, is replaced by) the median absolute deviation of the wavelet coefficients in the orientation subband d at the resolution scale 2^j . Examples of the estimated masks for two different SAR images are shown in Fig. 5.29. Even though the employed classification is rather simple, the edges and the heterogeneous regions are detected well.

From the detected masks in each subband, we estimate for each spatial position the prior probability ratio μ_l from Eq (5.4.4) using the same reasoning as in Section 3.5.2 leading to

$$\hat{\mu}_l = \exp\left(\sum_{k \in \partial l} \gamma(2\hat{x}_k - 1)\right), \quad \hat{x}_k \in \{0, 1\}. \quad (5.6.10)$$

Since the estimates \hat{x}_l rely on interscale coefficient correlations, μ_l exploits both the inter- and the intrascale dependencies between the wavelet coefficients.

To summarize, the proposed method estimates the wavelet coefficients as $\hat{g}_l = \hat{\xi}_l \hat{\mu}_l / (1 + \hat{\xi}_l \hat{\mu}_l) w_l$, where

$$\hat{\xi}_l = (a/2b)(\omega_l/b)^2 \exp(\omega_l/a - \omega_l/b), \quad (5.6.11)$$

and the parameters a and b are estimated using (5.6.8) and (5.6.9).

Fig. 5.31(a) illustrates the resulting shrinkage factor $\hat{\xi}_l \hat{\mu}_l / (1 + \hat{\xi}_l \hat{\mu}_l)$ without the influence of the local spatial neighborhood (i.e., for $\mu_l = 0$) in comparison with the standard hard and soft thresholding functions. For the two standard thresholding functions, the threshold T was determined as the value T for which $p_{\Omega|X}(T|0) = p_{\Omega|X}(T|1)$. In Fig. 5.31(a), one can see that the proposed shrinkage function suppresses high magnitude coefficients less than the classical soft-thresholding. This is very important for high-resolution SAR images, where strong scatterers need

to be preserved. Fig. 5.31(b), Fig. 5.31(c) illustrate the resulting families of shrinkage functions, when different relative importance γ is given to the local spatial context. One can deduce from these figures that a larger value of γ smooths the homogeneous areas better but also attenuates isolated strong targets in the homogeneous areas more. In all the subsequent results the value $\gamma = 0.2$ was used.

5.6.4 A practical SAR despeckling algorithm

Based on the description from the previous Section, now we can summarize the new algorithm as

Compute the non decimated wavelet transform

Specify the value of γ that controls the influence of the local spatial context (we use $\gamma = 0.2$)

- For each scale and orientation

 Compute $n_j = \text{Median}(|\mathbf{w}_j - \text{Median}(\mathbf{w}_j)|)/0.6745$

 For all the spatial positions $l = 1, \dots, n$

 If $|w_{l,j}\hat{y}_{l,j+1}| > (n_j)^2$, set $\hat{x}_l = 1$; otherwise set $\hat{x}_l = 0$

 Estimate the parameters of the conditional pdf models:

$$a = (1/N_0) \sum_{i \in S_0} |w_i|, \quad N_0 = \#S_0, \quad S_0 = \{l : \hat{x}_l = 0\}$$

$$b = (1/3N_1) \sum_{i \in S_1} |w_i|, \quad N_1 = \#S_1, \quad S_1 = \{l : \hat{x}_l = 1\}$$

 For $l = 1, \dots, n$:

 Estimate $\hat{\xi}_l = (a/b)(|w_l|/b)^2 \exp(|w_l|/a - |w_l|/b)$

 Estimate $\hat{\mu}_l = \exp\left(\gamma \sum_{k \in \partial l} \hat{x}_k\right)$

 Estimate the coefficient: $\hat{y}_l = \hat{\xi}_l \hat{\mu}_l / (1 + \hat{\xi}_l \hat{\mu}_l) w_l$

Apply the inverse wavelet transform

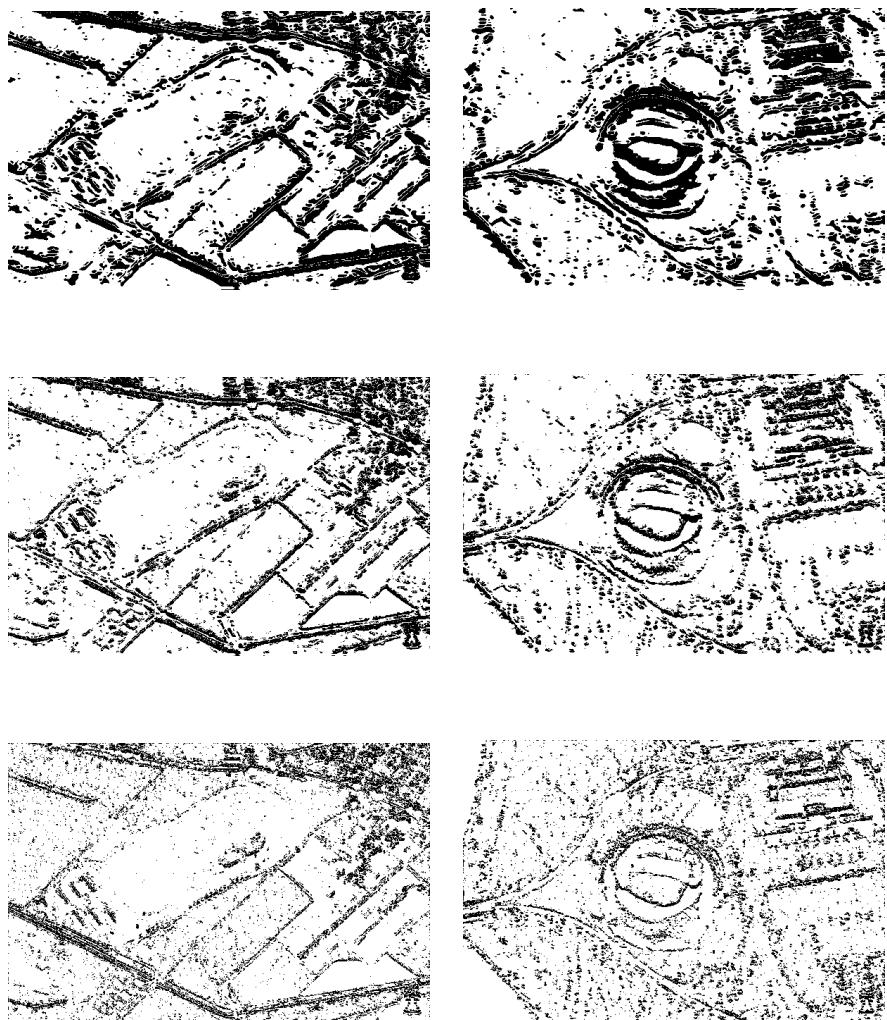


Figure 5.29: Examples of the detected masks for two original SAR images (from Fig. 5.27 and Fig. 5.34) in the proposed SAR denoising algorithm. From top to bottom, the resolution scales are 2^3 , 2^2 and 2^1 .

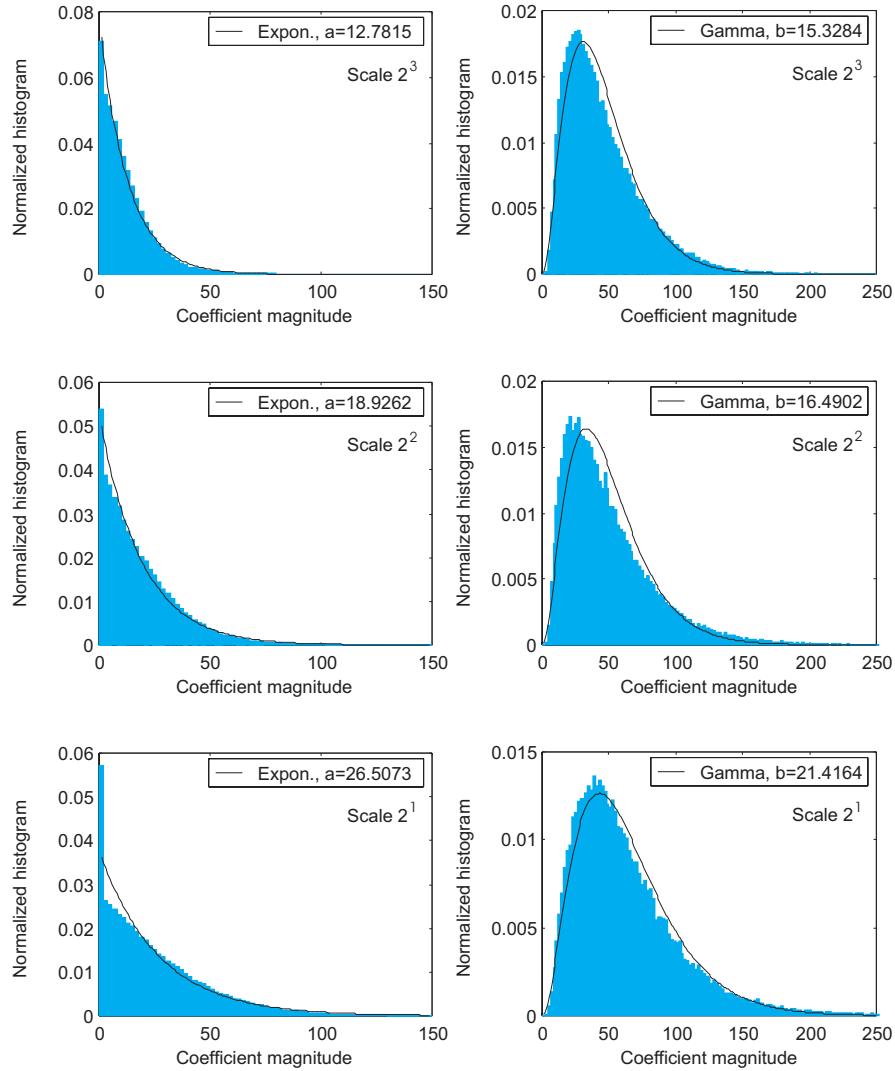


Figure 5.30: An illustration of the proposed functional forms $p_{\Omega_l|X_l}(\omega|0)$ (left) and $p_{\Omega_l|X_l}(\omega|1)$ (right) in comparison with the corresponding coefficient histograms. The diagrams correspond to the SAR image from Fig. 5.34 and to the three corresponding masks from Fig. 5.29.

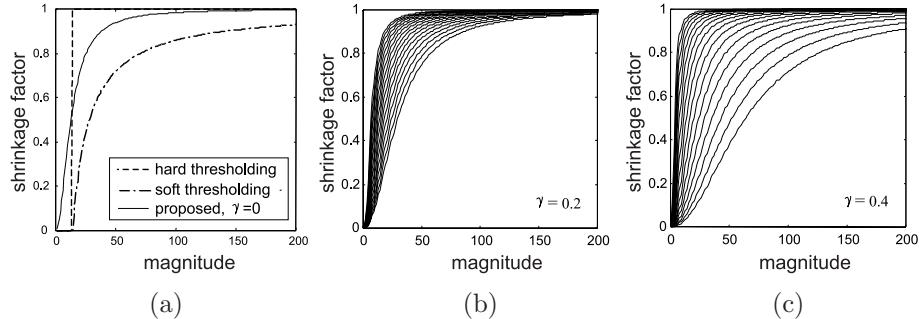


Figure 5.31: (a) The resulting shrinkage function for $\gamma = 0$, in comparison with standard hard- and soft-thresholding functions. (b) and (c) Families of resulting shrinkage functions, depending on the local spatial context, for $\gamma = 0.2$ and $\gamma = 0.4$, respectively.

5.6.5 Results and Discussion

First we demonstrate the performance of the proposed method on several representative SAR images and then compare the results with the Gamma MAP filter of [Lopes90] from Section 5.6.2.

The original image in Fig. 5.32 (800x600 pixels) was acquired by the German space agency (DLR)². The image is highly corrupted by noise. The proposed method strongly suppresses noise, while preserving details well; the image contrast has also improved after denoising. Strong reflectors are well preserved and they are even enhanced. For example, the four points arranged in the corners of a rectangle in the lower left part of the image are better recognized after denoising.

The SAR images of urban areas shown in Fig. 5.33 and Fig. 5.34 are provided by the Sandia National Laboratories. Both images are 400x600 pixels and are not as much corrupted by noise as compared to the previous case. The results in these figures confirm also the ability of the proposed method to preserve point-like targets while suppressing noise in the homogeneous areas. The SAR image from Fig. 5.35 (350x600 pixels, source Sandia) shows agricultural fields and is rich with natural texture. The result of denoising in this figure demonstrates that the proposed method preserves natural image texture rather well, even though it is not specifically designed to deal with textures.

²We processed this image at the Signal and Image Centre (SIC) of the Belgian Royal Military Academy. We thank the DLR and the SIC centre for kindly providing us this image for experiments.

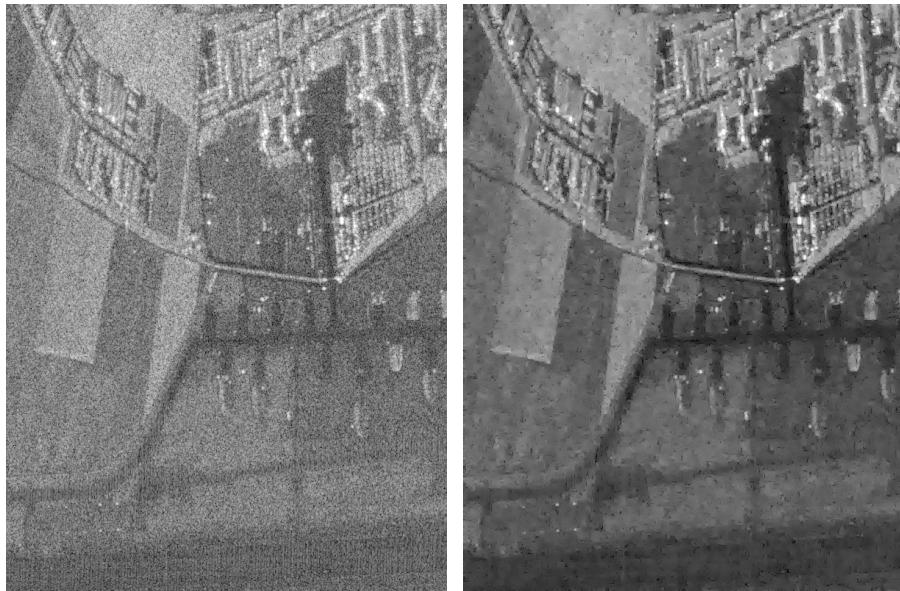


Figure 5.32: Original single-look SAR image (left) and the result of the proposed method (right).

In Fig. 5.36, another type of a SAR image is shown. The image (450x600 pixels) shows a river network with a lot of thin lines and tiny details. After denoising the tiny lines and other image details seem remarkably well preserved.

As a reference method for comparison, we use the Gamma MAP filter, which was explained briefly in Section 5.6.2. Fig. 5.37 shows parts of different SAR images and the results of the proposed filter in comparison with the corresponding results of the Gamma filter. The results demonstrate clearly that the proposed method preserves the image sharpness and point-like reflectors significantly better.

An objective evaluation of the results is not completed at this moment. Also, it would be interesting to investigate the use of the proposed method as a preprocesing tool for edge detection, segmentation or texture analysis in SAR images. Speckle noise complicates such procedures and their operation is expected to be largely facilitated after denoising. Another practical benefit from denoising SAR images is image compression. These subjects are left for further research.



Figure 5.33: Original SAR image (top) and the result of the proposed method (bottom).

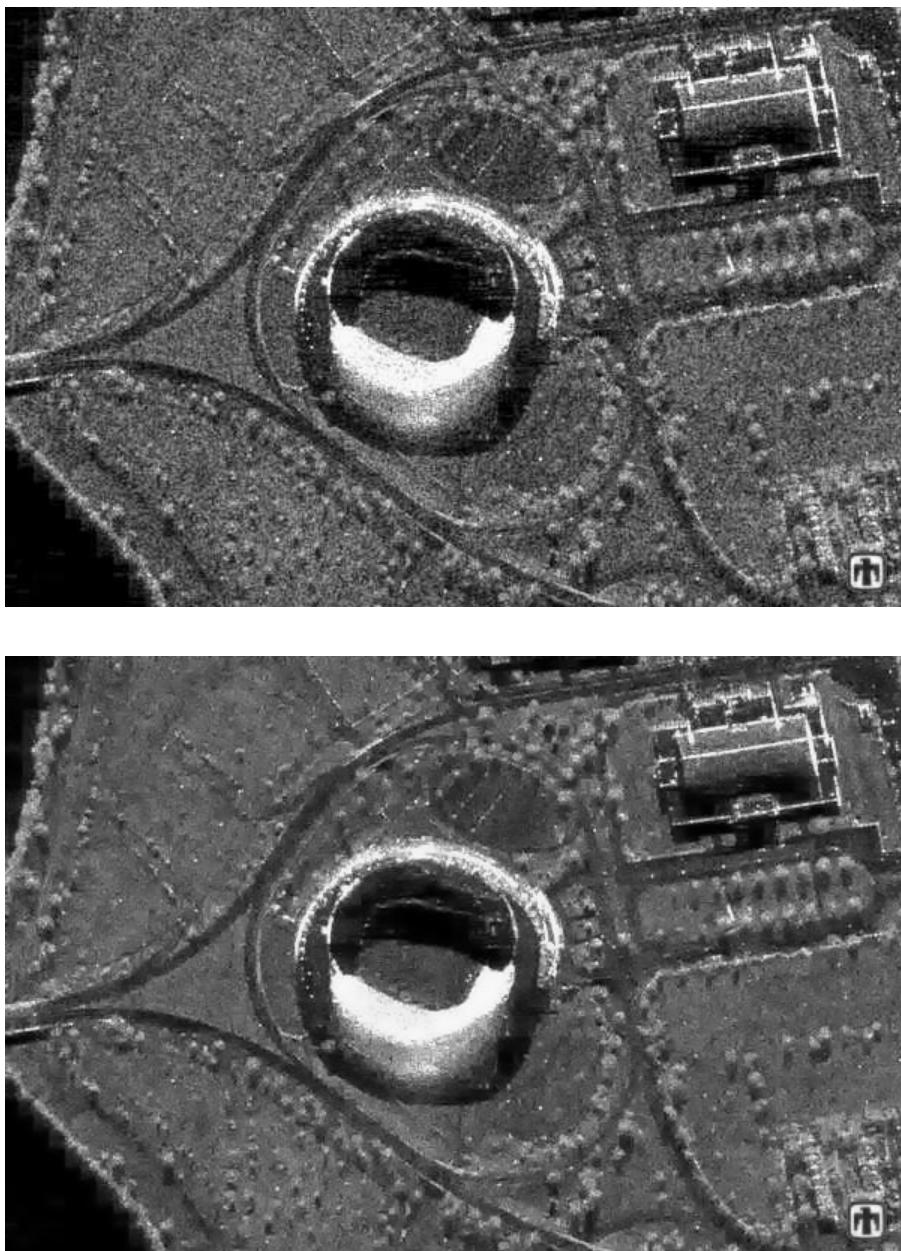


Figure 5.34: Original SAR image (top) and the result of the proposed method (bottom).

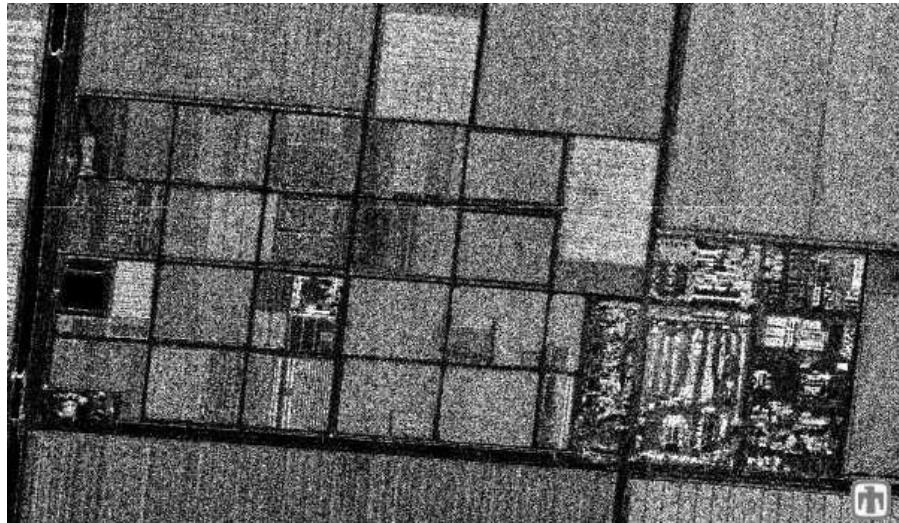


Figure 5.35: Original SAR image (top) and the result of the proposed method (bottom).

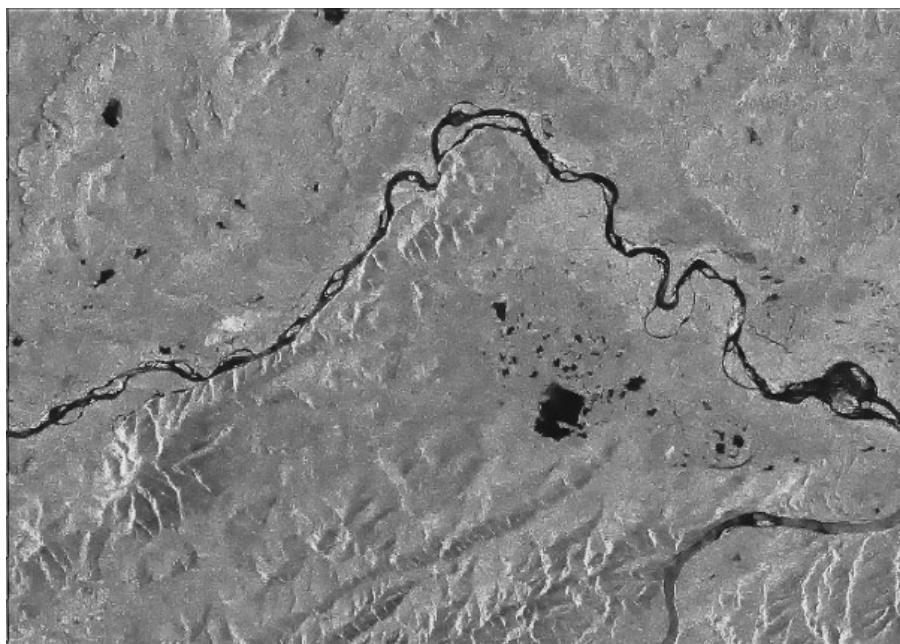
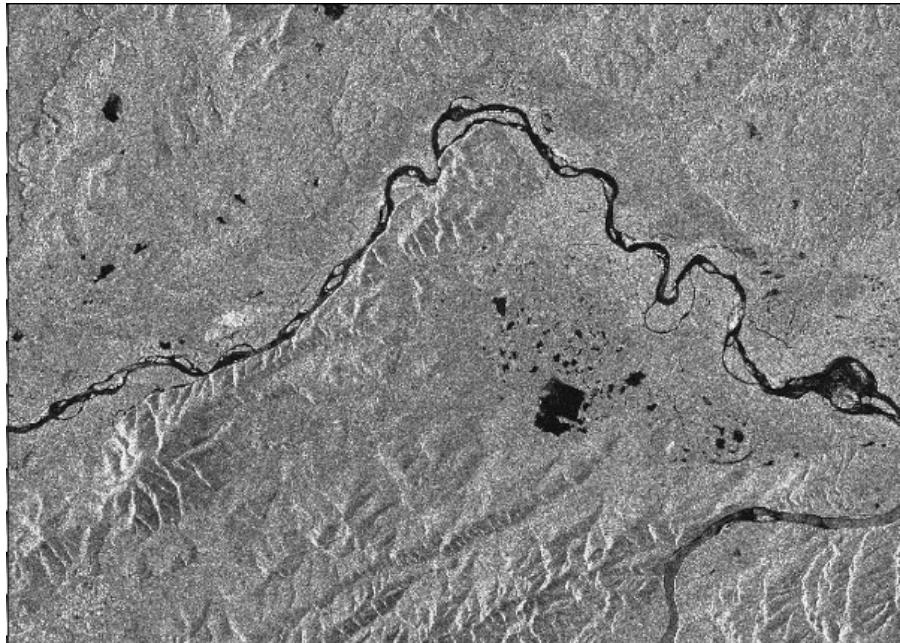


Figure 5.36: Original SAR image (top) and the result of the proposed method (bottom).

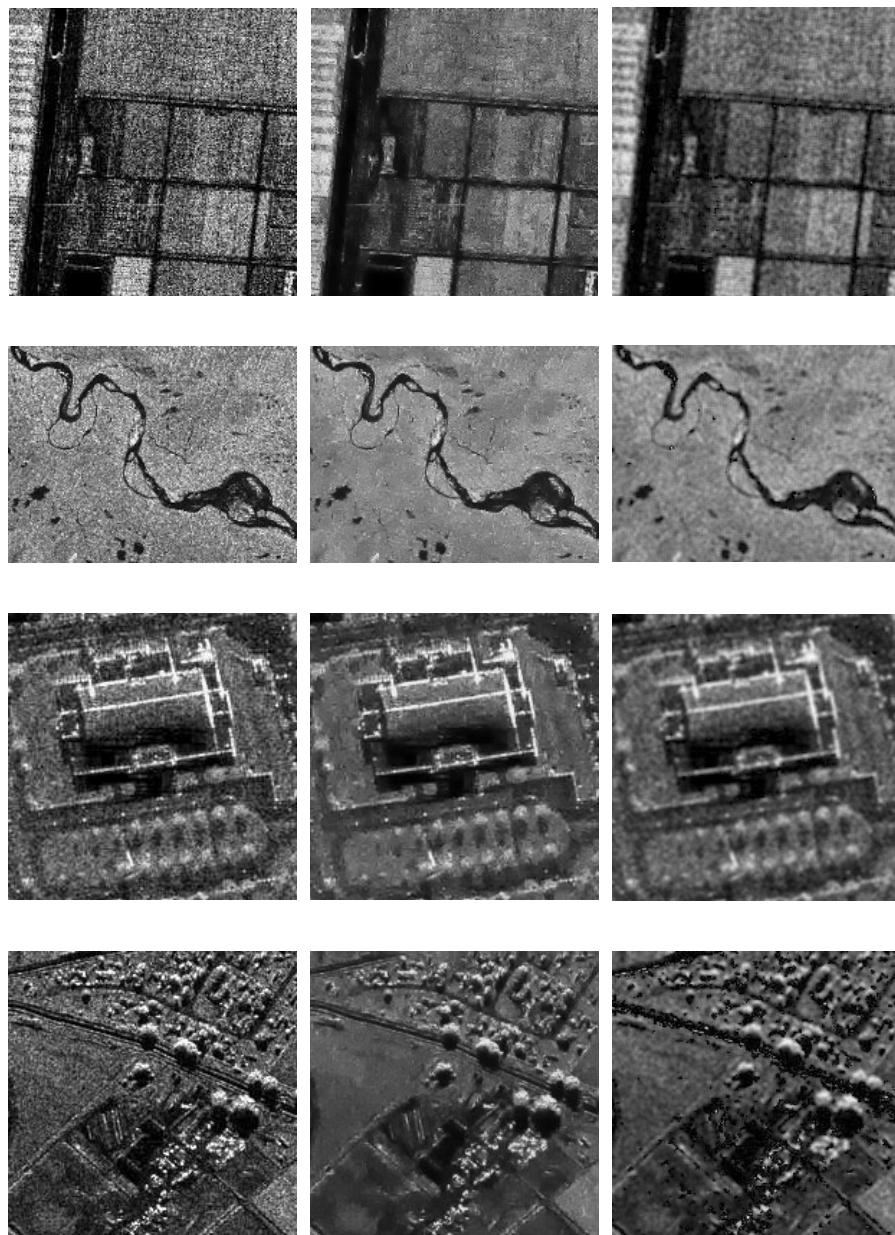


Figure 5.37: Parts of original SAR images (left), denoised images using the new method (middle) and the corresponding results using the Gamma MAP filter (right).

5.7 Summary and conclusions

In this Chapter, we proposed a new class of locally adaptive, denoising methods. In these developments, we started by upgrading our method from Section 3.5 and recognized a broader framework of the joint detection and estimation principles.

In Section 5.3, we proposed a new approach to estimating the probability of signal presence in a given wavelet coefficient, relying on the global and on the local statistical properties of image wavelet coefficients; its essence is expressed in Eqs (5.3.6) and (5.3.14). We integrate the underlying pdf of the noise free coefficients to estimate the global probabilities (of signal presence) and combine these in a Bayesian framework with the conditional pdf's of the local spatial activity indicator. We derived the analytic expressions for the proposed global prior probability ratio and for two common prior models (Section 5.3.3). In Section 5.3.4, we proposed a simple analytic model for the conditional densities of the locally averaged coefficient magnitudes, that we used as the local spatial activity indicators. In this way, we derived a practical, low complexity algorithm for Gaussian noise suppression; its characteristic parts are depicted in Fig. 5.7. The new algorithm compares well with the best state of the art methods of similar and of higher complexity. The advantage over related methods was demonstrated both in terms of PSNR and visually.

In Section 5.4, we proposed a framework for various empirical extensions of the new approach, which may lead to an increased flexibility in practice and adaptability to different noise types; our ideas in this respect were summarized in Fig. 5.16 and Fig. 5.17. Within this framework, we developed two other algorithms, described in Sections 5.5 and 5.6. The first one is versatile and applicable to various noise types; we demonstrated its applications to medical ultrasound and MRI images and its use in image deblurring, showing the advantages over some standard filtering techniques in these applications. The algorithm that we developed in Section 5.6 is aimed for speckle suppression in SAR images. We proposed functional forms for the conditional densities of the coefficient magnitudes in SAR images. The main advantage of the new method over the majority of existing techniques is that it preserves point reflectors well, while adapting automatically to the local homogeneity in the image.

Different parts of this Chapter are published in [Pizurica01a] and in [Pizurica01b] and submitted for publication [Pizurica02b, Pizurica02c].

Chapter 6

Application to humanitarian demining

This Chapter covers a particular application area: humanitarian demining. Mine detection is a challenging framework for developing and validating advanced sensor technologies and accompanying signal and image processing methods. Here we tackle a small segment of this complex problem, focusing on enhancement of images of landmines. In this respect, we consider techniques developed in the previous Chapters and also propose and discuss some other, simpler solutions. In less detail, we illustrate some other important aspects of humanitarian mine detection.

6.1 Introduction

When accessing a humanitarian demining problem, even from an image processing point of view, one cannot forget its *human* dimension. Currently, approximately 60 million antipersonnel mines are polluting and endangering the environment in some 60 countries [Verlinde01]. A number of specialized studies, e.g., [Brooks97, Bruschini96, Bruschini97, Daniels97, Nicoud97, Milisavljevic01, VanKempen98], is devoted to advanced technical solutions of the land mine problem.

Two basic types of landmines are: anti-tank mines and anti-personnel mines. Anti-tank mines are designed to destroy heavy vehicles and are usually laid to form regular patterns. The anti-personnel mines, designed to maim or kill people, are small (e.g., 5-6 centimeters in diameter) and light (often less than 50 g), and appear in many different shapes (see Fig. 6.1). A detailed description is in [Milisavljevic01]. The anti-



Figure 6.1: Examples of anti-personnel mines.

personnel mines are often scattered at random on purpose; apart from that, since they are small and light, erosion or floods can carry them over quite large distances, which makes their localization extremely difficult.

An important aspect of the humanitarian demining is the requirement that every single mine is located and destroyed. Most of such mine clearance procedures are in practice still carried out manually: the “detection” process relies on sticks for prodding the ground or at most on basic metal detectors. Since the metal detector cannot distinguish the metallic content of a landmine and other metallic pieces, like e.g., metallic debris, a number of false alarms is created; each of those needs to be carefully checked, which is time consuming. A special problem are the so-called plastic mines, which have minimum amounts of metal. Needless to say, a deminers’ work is extremely dangerous, current demining procedures are extremely slow, and new more efficient technological solutions are needed. In this respect, for locating minefields and individual landmines, advanced sensor technologies [Bruschini97, Nicoud97] are in (experimental) use or under development. A reliable interpretation of the acquired data and automatized mine detection techniques require various signal and image processing [Brooks97, VanKempen98] and data fusion [Milisavljevic99c, Milisavljevic01] techniques.

Our contribution in the scope of this topic is wavelet domain noise reduction. The goals are enhancing the image quality for visual inspection as well as facilitating subsequent steps in an automatic mine detection procedure. In particular, our research has been carried out within the Belgian joint research program on humanitarian demining

(HUDEM)[Acheroy98b], which has determined the choice of sensors for which we consider noise reduction. These sensors are: infrared cameras [Acheroy98a, Klein97] and ground (or surface) penetrating radar [Daniels96].

The Chapter is organized as follows. In Section 6.2, we first give a brief overview of the sensors that are considered among the most useful or promising ones for humanitarian mine detection. Then we describe briefly the Belgian joint humanitarian research program HUDEM [Acheroy98b, Verlinde01], of which our work is a part. In Section 6.3, we address the processing of infrared (IR) images of landmines. First, we describe in some detail the IR sensors and images available for our experiments. Then we propose a new, simple wavelet domain filtering technique for these images, and also apply one of the previously developed spatially adaptive methods from this thesis. Section 6.4 is structured similarly, but devoted to ground penetrating radar. First, we describe the sensor and the available images; then we apply a simple pyramidal filtering technique and spatially adaptive wavelet denoising to different types of GPR images. Section 6.5 concludes the Chapter.

6.2 Humanitarian demining technologies

Here we address some standard and emerging technologies in humanitarian mine detection. The aim is to give a reader a quick insight into this problem; for a deeper analysis we refer rather to a number of related, specialized publications.

6.2.1 Sensor technologies in humanitarian demining

A possible rough categorization of different sensor technologies in humanitarian demining is according to their range:

remote and close range

sensors, as depicted in Fig. 6.2. Remote sensing plays an important role in planning demining operations (acquiring information about the terrains, infrastructure and roads) and in the detection of *minefields*; in this category, Synthetic Aperture Radar (SAR), (polarimetric) infrared, multispectral and hyperspectral imaging are among the most often considered technologies [Bishop98]. In this respect, noise reduction techniques for SAR, including the one that we have developed in Section 5.6 can be seen as helpful in humanitarian demining as well.

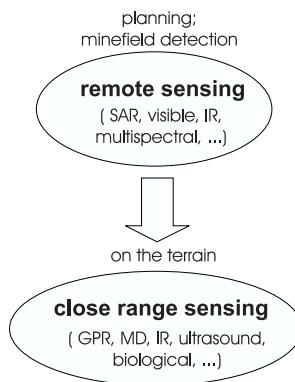


Figure 6.2: Various sensor technologies in humanitarian demining.

Close range sensors are employed for the detection of *individual landmines*. Since in humanitarian mine detection a great importance is given to the detection of each single landmine, the close range sensors are of particular importance in this application. A variety of such sensors are in use or under development. A concise review can be found, e.g., in [Bruschini97], and a detailed technical study is in [Milisavljevic01]. The description that follows here is far from being complete and remains at an introductory level. Two types of sensors, for which we develop noise reduction techniques, are described in more detail in Sections 6.3.1 and 6.4.1. We start with a brief overview of the sensors that are, according to recent studies the most promising ones.

Magnetometers and metal detectors

Magnetometers are frequently used for detection of unexploded ordnance (UXO) [Milisavljevic01]. These sensors are passive (i.e., do not radiate any energy) and measure the disturbance of the Earth's natural electromagnetic field. To increase the sensitivity of a magnetometer, a *gradiometer* is used. This apparatus measures the gradient variation of the disturbed magnetic field. Both magnetometer and gradiometer can detect only ferrous objects but no other metallic object. *Metal detectors* (MD) [Nicoud97] (or coil induction sensors) are among the oldest, simplest and best known mine detection sensors. The principle of operation is: the current through the primary (transmitter) coil generates a magnetic field, that induces a secondary electrical current in a metallic object; the induced current generates a secondary magnetic field that

is detected by the receiving coil of the detector. Usually, metal detectors produce numerous false alarms (due to various metallic pieces in the ground). Conventional metal detectors produce an acoustic signal. More advanced ones are aimed for use in combination with imaging devices [Bruschini99]; specific image processing techniques are there required, among which deconvolution [Druyts00, Verlinde01].

Trace explosive detectors

The sensors from this category are far out of the scope of this thesis, but are indeed considered among the most important ones in humanitarian mine detection. Actually, one of the most effective “sensors” of explosives are the *dogs*, thanks to their great olfactory power. Some of the obstacles faced in practice are an expensive and long training, and an inclination of dogs to tire quickly (after 15 to 30 minutes of work). Other biological sensors that are currently under research for explosive detection are *rodents*. The first results of such studies [Weetjens99] show that rodents are at least as sensitive as dogs, with the advantage of being lighter and much longer concentrated (8 hours). Alternatives are *artificial vapor sensors*, [Jankowski92, Rouhi97] which are already in use in chemical industry and in airports.

Bulk explosive detectors

Interest is growing in sensors that can detect explosives in bulk form, in any environment. Such sensor technologies include nuclear (thermal neutron activation, neutron backscatter, and X- ray backscatter) and NMR/NQR techniques [Bruschini97]. These techniques are used, e.g., in screening airport luggage and mail [Novakoff92]. Their application in mine detection is quite recent and appears promising [Engelbeen98], but expensive.

Thermal infrared sensors

Mines retain or release heat at a different rate than the surrounding environment. Consequently, a temperature difference exists between the soil above a mine and the soil close to it. Sensitive infrared (IR) cameras [Acheroy98a] detect thermal differences that are less than 0.1K. Maximum burial depth at which mines can still be detected by means of IR sensors is estimated between 10 and 15cm [Bruschini97], but is in practice often much smaller. Infrared cameras are mainly aimed at

detecting surface laid and shallow buried mines. We address this sensor in some more detail in Section 6.3.1.

Multi- and hyperspectral visible/infrared sensors

Hyperspectral stands for a sensor with more than twenty bands, while multispectral describes a sensor with twenty or less spectral bands. These sensors are aimed at detecting not just surface-laid, but buried mines as well. The underlying principle is detecting localized spectral differences in the scene, caused by the presence of mines. Immediately after the mine placement, there exist textural differences of the soil that can be detected with a broad band infrared (or visible during the day) instrument. In time, the surface textural differences dissipate. There are still likely to be compositional differences between the soil above the mine and the surrounding area (e.g., due to the fact that a mine prevents the vertical flow of soil moisture). These compositional differences of the soil result in a local difference in the spectral signature [DePersia95].

Ground penetrating radar

Ground penetrating radar (GPR) [Daniels96] is considered to be among the best sensors for detecting buried plastic and low-metal content mines [Milisavljevic01]. GPR emits electromagnetic waves into the soil and detects their reflections. The GPR technology is used in civil engineering, geology and archeology for detecting buried objects and studying soil [Peters94]. For mine detection, time domain ultra wide band (UWB) radars are used (some more technical details are given in Section 6.4.1). By moving the antenna it is possible to reconstruct an image representation of a vertical slice of the soil; further data processing makes it possible to also visualize the horizontal soil slices or to build 3-D representations [Scheers00]. The main problems are: the choice of the frequency band is a trade-off between resolution and penetration depth (higher frequencies do not propagate well through ground); shallow buried objects are not reliably detected due to strong surface-air reflections; discrimination between mines and other objects of minelike shape is difficult.

Acoustic sensors

Conventional ultrasound sensors [Donskoy98] emit a sound wave, with a frequency higher than 20 kHz, into a medium. This sound wave is reflected on boundaries between materials with different acoustical prop-

erties. Such systems penetrate well through very wet and heavy ground such as clay, and that makes them complementary to GPR (although they are also likely to experience problems at the air-ground interface [Bruschini97]). Accordingly, most of the work in this area is oriented towards underwater mine detection [Goo98].

Multisensor systems

Each of the sensors mentioned above works well under certain conditions but is less or not at all reliable under the others. An improved detection probability as well as a reduced false alarm rate should result from a multisensor data fusion [Milisavljevic99c]. In this respect, one of the most often analyzed combination of sensors is:

infrared camera(s) (IR),
ground penetrating radar (GPR) and
metal detector (MD)

Within this framework, noise reduction techniques are primarily aimed for the IR and GPR.

6.2.2 A practical research program

Among numerous others, the Belgian joint research program on humanitarian demining (Hudem) [Acheroy98b] has merged some of the sensor technologies mentioned above and accompanying signal and image processing and mechanization/robotization techniques. This research has been carried out at several Universities and laboratories: the Royal Military Academy (RMA), the “Vrije Universiteit Brussel” (VUB), the “Université libre de Bruxelles” (ULB), the “Universitaire Instelling Antwerpen” (UIA), the “Université catholique de Louvain” (UCL), the “Katholieke Universiteit Leuven” (KUL), the “Université de Liège” (ULg), the “Universiteit Gent” (RUG), and the “Facultés universitaires Notre-Dame de la Paix” (FUNDP). Different segments and some of the practical contributions of this research are summarized below:

Sensor development and modeling: modeling thermal radiations, detected by IR sensors [Schachne98]; modeling the electrical properties of soils [Storme98]; development of new small sized UWB GPR antennas for the detection of minelike objects [Scheers98, Scheers00]; applying new sensor technologies,



Figure 6.3: Acquisition of a data set of IR images of landmines. (a) Dummy minefields and (b) a test setup with two IR cameras.

like Nuclear Quadrupole Resonance (NQR) to mine detection [Engelbeen98].

Signal processing and data fusion: specific techniques for metal detector imaging [Druyts00]; noise reduction techniques for infrared images [Pizurica99c] and GPR images [Pizurica99a] of landmines; the analysis of IR image sequences by means of Karhoenen Loëve transform [Schachne98]; feature extraction and pattern recognition methods for GPR [Milisavljevic01, VanKempen98]; shape analysis applied to IR images of landmines [Milisavljevic99a]; fusion of the data acquired by IR, GPR and metal detector [Milisavljevic99c, Milisavljevic01].

Robotics: design of small platforms for sensors and development of navigation techniques, including the localization of sensors in the minefield [Verlinde01].

Within the framework defined above, our focus is on noise reduction and enhancement of IR and GPR images, on which we concentrate now.

6.3 Infrared image processing

Before describing the noise reduction methods for IR images, we explain briefly their formation and especially the type of images used in our experiments.

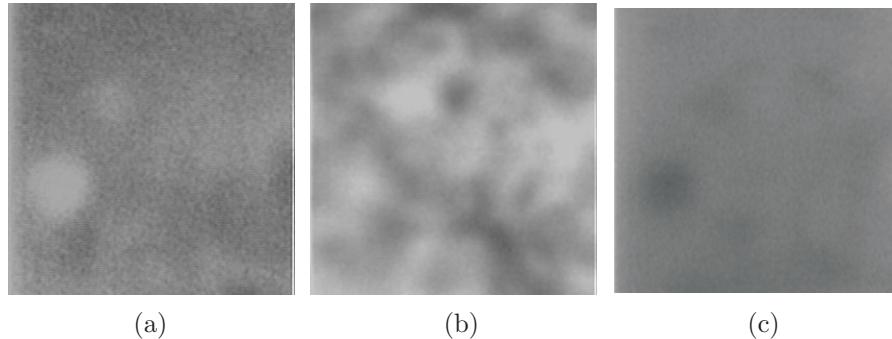


Figure 6.4: Examples of infrared images of the same landmine (in the lower left part) acquired at different hours of the day: (a) in the morning (around 11h), (b) in the afternoon (around 15h) and (c) in the evening (around 21h). Soil: sand, camera: $3\mu\text{m}$ - $5\mu\text{m}$.

6.3.1 Infrared images of landmines

Infrared sensors [Acheroy98a, Klein97] detect thermal radiation of objects. The thermal contrast of an object with respect to its environment is a function of two parameters: the *temperature* of the object and the *emittance* of its radiating surface. Within the infrared ($0.7\mu\text{m}$ - 1 mm) part of the spectrum, the IR sensors most often operate in one of the following *windows* where the IR transmittance of the atmosphere is high [Klein97]:

short-wavelength infrared: $0.87\mu\text{m}$ - $1.5\mu\text{m}$

mid-wavelength infrared: $3\mu\text{m}$ - $5\mu\text{m}$

long-wavelength infrared: $8\mu\text{m}$ - $12\mu\text{m}$

The choice of the band depends on the situation at hand, depending where a strong target signature exists, but also taking into account background characteristics, i.e., type of terrain (soil), spectral signatures, etc.

Normally, mines become hotter and colder faster than their surroundings, meaning that the daily evolution (so-called “diurnal cycle”) of IR mine signature exists: there are parts of the day when a mine appears on an IR image as lighter (warmer) than the surroundings, and other parts when it is darker (colder). In between the two extremes, there are times of the day when a mine cannot be noticed because it is in thermal equilibrium with its surrounding. Important factors that

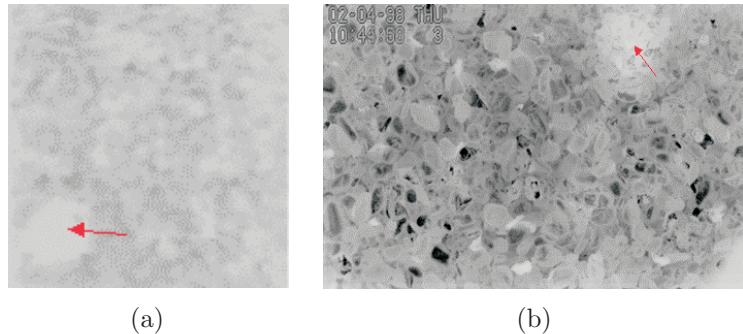


Figure 6.5: Two infrared images of the same landmine (buried in gravel) and acquired with (a) a $3\mu\text{m}$ - $5\mu\text{m}$ camera and (b) an $8\mu\text{m}$ - $12\mu\text{m}$ camera.

influence the thermal contrast, like the time of year, position of sun, humidity of air, etc. are discussed in detail in [Milisavljevic01].

A data set acquisition

The data set of IR images of landmines that was at our disposal was acquired by the researchers of the HUDEM project (including ourselves). The data acquisition was performed on realistic dummy minefields in Meerdaal, Belgium, which were installed for research purposes by the Belgian Bomb Removal Unit and by the Royal Military Academy. Those dummy minefields, a part of which is shown in the Fig. 6.3(a), cover four different types of soil: natural, sand, gravel and a mixture of sand and gravel. In each type of soil, we were taking infrared images of one particular mine with two infrared cameras (Fig. 6.3 (b)): a mid- and a long wavelength one, each 15 min during 24 hours. The data are available on <http://www.tdp.sai.jrc.it/APL-Database/Home/sigdata.htm>.

Images shown in Fig. 6.4 were taken by the $3\mu\text{m}$ - $5\mu\text{m}$ IR camera at different hours during the day: in the morning (left), in the afternoon (middle) and in the evening (right). The landmine was buried in sand at the depth of 5 cm. One can recognize in these images the “diurnal cycle” that was mentioned above: before noon (Fig. 6.4(a)) the highest pixel intensities correspond to the buried mine. In the early afternoon, the temperature of the surrounding soil equalizes with the landmine temperature; this is the so called “blind period” for IR cameras: one can no longer recognize the presence of the mine in Fig. 6.4(b). In the evening image from Fig. 6.4(c), the landmine that is cooling faster than

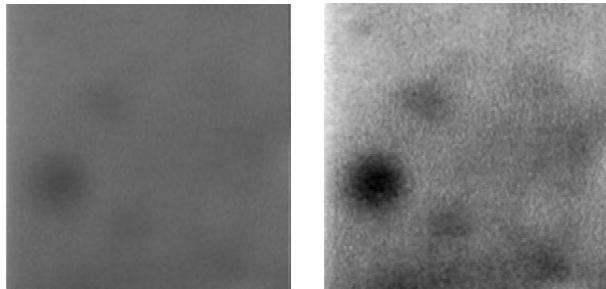


Figure 6.6: An original infrared image and the result of a simple contrast enhancement.

the surrounding is represented by the darkest pixel intensities. During the night, the second “blind period” occurs.

Fig. 6.5 shows examples of the IR images acquired in the gravel field with the two cameras. The position of the landmine (denoted by an arrow) appears different in the two images due to different positions of the cameras (see Fig. 6.3 (b)).

One can conclude from the above examples that the infrared cameras should be used during those periods of the day when the thermal contrast between mines and surrounding is expected to be the greatest. The acquired images during those periods need to be further enhanced in order to facilitate the interpretation of the data.

Contrast enhancement

A great step in improving the visual quality of an infrared image is often a simple contrast enhancement. Even though the results of our noise reduction/enhancement techniques would seem much more impressive as compared to the raw data visualized in Fig. 6.4, we first optimize the visualization of the data themselves. Namely, the acquisition system used in our experiments was introducing a systematic error: border pixels were falsely assigned extremely large values preventing a proper use of the available dynamic range for visualization. By removing this error and rescaling the pixel intensities linearly within the available grey scale range, one already obtains a great improvement in the visual appearance of the data, as demonstrated in Fig. 6.6. From here on, we proceed with demonstrating noise reduction techniques.

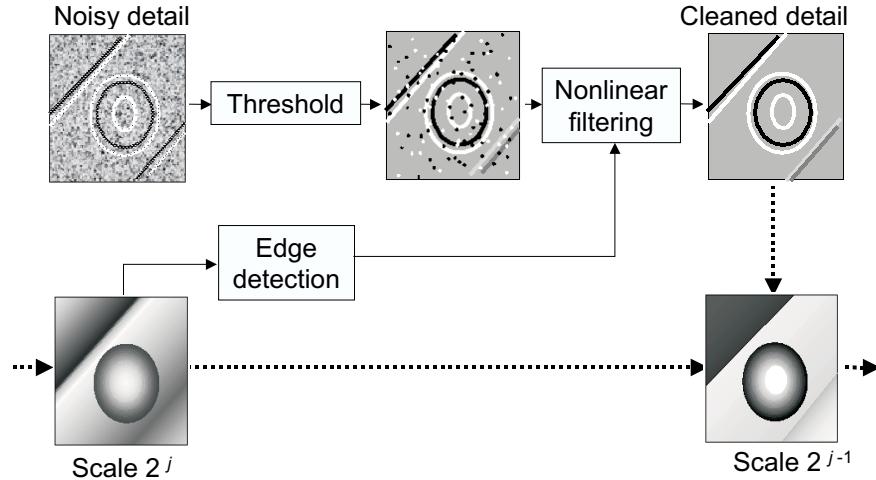


Figure 6.7: An illustration of a filtering step in the proposed multiscale edge reconstruction technique.

6.3.2 A new nonlinear wavelet denoising technique

Here we propose one simple wavelet domain noise reduction technique for IR images of landmines. The proposed approach exploits the *low-pass images* (scaling coefficients) in order to detect the positions of significant edges in noisy details. A related method of Sattar [Sattar97] (Section 6.4.2) uses a decimated representation with *quincunx* wavelets [Feilner01, Kovacevic92, Starck98], applies an edge detector to each low-pass image and selects simply those wavelet coefficients at the positions of the detected edges. Here we apply a slightly more sophisticated approach. We use a specific non decimated wavelet transform and introduce a new nonlinear filtering scheme. Firstly, a wavelet coefficient is selected based on two criteria: its magnitude and the *closeness* to the estimated edge positions. Secondly, in order to suppress the artifacts, the selected coefficients are filtered along the edges. Practically, we first remove the smallest wavelet coefficients and integrate the subsequent selection/filtering into one step. The basic concept is illustrated in Fig. 6.7. The details are outlined below.

A one-subband non decimated wavelet representation

To obtain a non decimated wavelet representation with one detail image at each scale, we simply rearrange the decomposition and the reconstruction stages of the two-subband decomposition [Mallat92b] from Fig. 2.11.

Let us denote the two dimensional convolution as

$$(a * b)_{m,n} \triangleq \sum_k \sum_l a_{k,l} \cdot b_{n-k,m-l}, \quad (6.3.1)$$

a separable filter as $ab_{m,n} \triangleq a_m b_n$ and a discrete Dirac function by δ_n . Like in Section 2.2.7, a^j denotes inserting $2^j - 1$ zeros between each two coefficients of the filter a . Now we can represent in a compact form the decomposition stage from Fig. 2.11 as:

$$\begin{aligned} w_{j+1,m,n}^1 &= (s_j * g^j \delta)_{m,n} \\ w_{j+1,m,n}^2 &= (s_j * \delta g^j)_{m,n} \\ s_{j+1,m,n} &= (s_j * h^j h^j)_{m,n} \end{aligned} \quad (6.3.2)$$

and the reconstruction stage as

$$s_{j,m,n} = (w_{j+1}^1 * q^j r^j)_{m,n} + (w_{j+1}^2 * r^j q^j)_{m,n} + (s_{j+1} * \tilde{h}^j \tilde{h}^j)_{m,n}. \quad (6.3.3)$$

The wavelet coefficients $w_{j+1,m,n}^1$ and $w_{j+1,m,n}^2$ represent the horizontal and the vertical edge components at the position (m, n) . To obtain the representation with one detail image, that is appropriate for our noise reduction algorithm, we actually treat the total bandpass content from (6.3.3) as detail coefficients, denoting

$$d_{j+1,m,n} = (w_{j+1}^1 * q^j r^j)_{m,n} + (w_{j+1}^2 * r^j q^j)_{m,n}. \quad (6.3.4)$$

Accordingly, the new decomposition and reconstruction stages are

$$s_{j+1,m,n} = (s_j * h^j h^j)_{m,n}, \quad (6.3.5)$$

$$d_{j+1,m,n} = (s_j * k^j r^j)_{m,n} + (s_j * r^j k^j)_{m,n}, \quad (6.3.6)$$

$$s_{j,m,n} = w_{j+1,m,n} + (s_{j+1} * \tilde{h}^j \tilde{h}^j)_{m,n}, \quad (6.3.7)$$

where the new filter k_n is

$$k_n = (g * q)_n. \quad (6.3.8)$$

A decomposition and a reconstruction step in this scheme are depicted in Fig. 6.8.

Algorithm

Like in other algorithms from this thesis, we now denote the spatial position with a single index $l \triangleq (m, n)$. The new algorithm removes noise according to the scheme in Fig. 6.8 in a coarse-to-fine manner. A specific nonlinear filter is introduced to suppress the noise-induced coefficient variation along the edges. The complete algorithm is:

Decompose an original image into lowpass and detail images at N successive resolution scales applying recursively Eq. (6.3.5) and Eq. (6.3.6)

For all resolution levels $j = N$ to $j = 1$:

- Apply the Canny edge detector [Canny86] to the lowpass image $\hat{\mathbf{s}}_j$; it yields a binary edge map $\mathbf{e}_j = \{e_{j,1}, \dots, e_{j,n}\}$, $e_{j,l} \in \{0, 1\}$

For all spatial positions $l = 1, \dots, n$ (where $l \triangleq (m, n)$)

Set to zero $d_{j,l}$ if $d_{j,l} < \hat{\sigma}_j$, where $\hat{\sigma}_j$ is an estimate of the noise standard deviation

Apply a nonlinear filter:

$$\hat{d}_{j,l} = x_{j,l} \cdot \frac{\sum_{k \in N_l} x_{j,k} \cdot d_{j,k}}{\sum_{k \in N_l} x_{j,k}} \quad (6.3.9)$$

where N_l is a square window centered at l and $x_{j,l} = 1$ if the pixel l is located in the specified neighborhood of the estimated edges and $x_{j,l} = 0$, otherwise. Practically:

$$x_{j,l} = 1 \text{ if } \sum_{k \in N_l} e_{j,k} > 1, \quad \text{and} \quad x_{j,l} = 0, \quad \text{otherwise.}$$

The filter (6.3.9) removes those detail coefficients that are not located in the vicinity of the detected edges, while the remaining coefficients are filtered along the edge segments.

- Reconstruct $\hat{\mathbf{s}}_{j-1}$ at the next finer resolution scale from $\hat{\mathbf{s}}_j$ and $\hat{\mathbf{d}}_j$ using Eq. (6.3.7).

The image $\hat{\mathbf{s}}_0$ reconstructed from the processed detail $\hat{\mathbf{w}}_1$ and the lowpass image $\hat{\mathbf{s}}_1$ is the final result of this denoising algorithm.

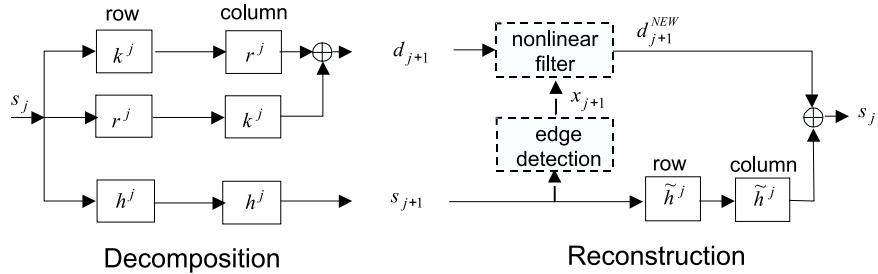


Figure 6.8: A schematic representation of the proposed multiscale edge reconstruction technique.

In [Pizurica99d], we demonstrated the advantage of this algorithm over standard soft thresholding (implemented with the same wavelet representation) on images with artificial Gaussian noise. On infrared images of landmines from our data set, this simple technique offers a significant improvement, as it is illustrated in Fig. 6.9. The background noise is strongly suppressed and the presence of the object of interest is enhanced.

One should note that noise suppression is achieved here by a “severe” suppression of all the coefficients that are not located in the vicinity of the edges detected from the lowpass images. This is useful for images where a uniform-intensity object needs to be distinguished from a background, but this method is not as favorable in cases where fine image details need to be preserved.

6.3.3 Wavelet denoising with spatial priors

In some cases (e.g., a surface laid or shallow buried mine and/or use of some more sophisticated sensors) it is possible to reveal more relevant details from the infrared image; those details can be important, e.g., to determine the type of the landmine. Also, infrared cameras can be used to scan larger areas, often covered by different types of vegetation. In these cases, a more sophisticated denoising approach is needed to reconstruct the available information from noisy data, without loosing relevant details.

In the previous Chapters, we developed several new spatially adaptive denoising techniques; here we demonstrate the results of one of those methods on the available IR images from our data set. In particular, we apply a wavelet denoising method that combines the prior models

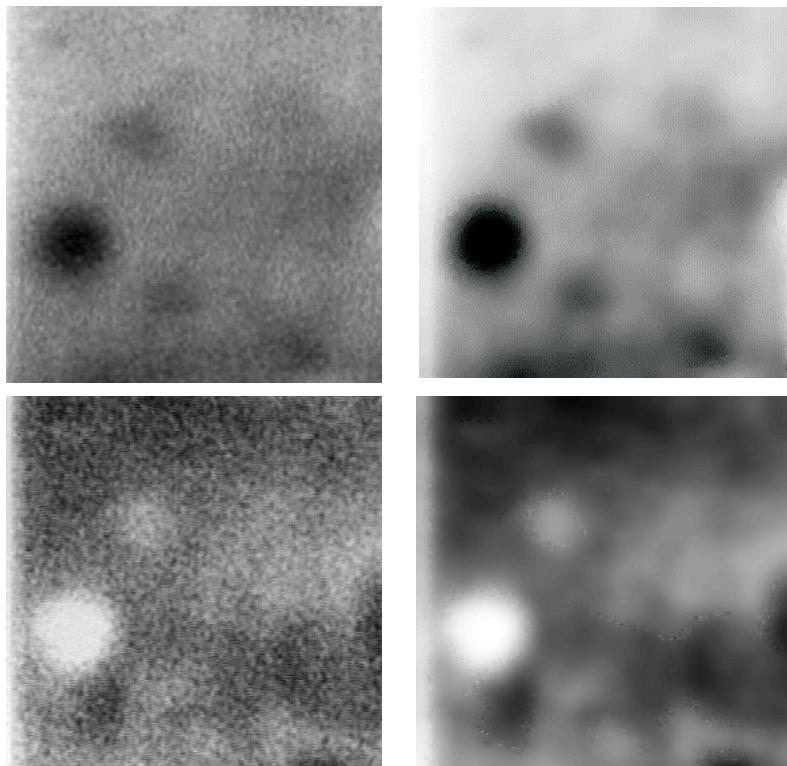


Figure 6.9: Left: preprocessed IR images. Right: results of the proposed nonlinear wavelet filtering technique.

for spatial clustering from Section 3.5.2, with the heuristic shrinkage families from Section 3.5.3. The significance measure is the coefficient magnitude and the parameters are $\alpha = 1$, $\gamma = 0.2$ and $\delta = 0.5$. The results are demonstrated in Fig. 6.10. The original IR images in this figure were acquired in gravel and preprocessed in order to enhance the contrast. Noise suppression enhances the presence of the buried mine; structural information in the background is well preserved.

6.4 Processing of GPR images

Here we first briefly explain the principles of Ground Penetrating Radar imaging and the GPR images used in our experiments. Then we demonstrate the application of noise reduction techniques to GPR scans taken with a commercial and an experimental radar in different soils.

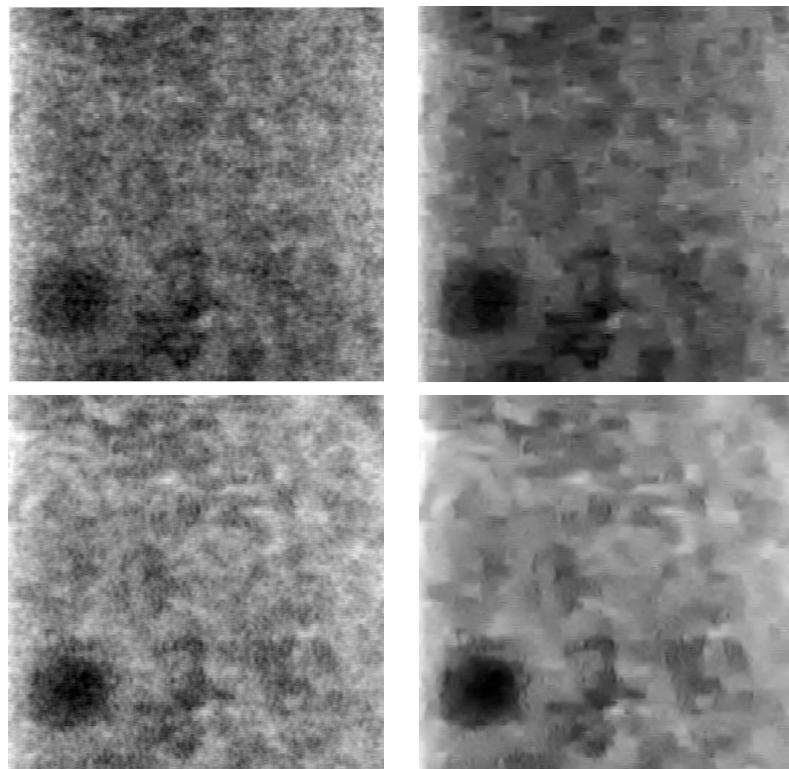


Figure 6.10: Left: preprocessed IR images. Right: results of the wavelet denoising using spatial priors.

6.4.1 GPR images of landmines

Ground Penetrating Radar (GPR) [Daniels96, Peters94] transmits short radar pulses, which propagate through the ground and are partially reflected at permittivity discontinuities. The amplitude of a reflected wave is recorded as a function of time at the receiving antenna. The one dimensional signal representing the reflected wave amplitude versus time (or, with a good approximation, versus depth inside the ground) is usually called an A-scan. The collection of A-scans taken at equidistant points along a given direction on the surface forms a two dimensional signal called a B scan, while horizontal slices are denoted as C-scans. Most often B-scans are visualized and used for object recognition. Buried mines or mine-like targets may produce multiple reflections resulting in characteristic, nested *hyperbole* signatures [Capineri98] in the B-scans

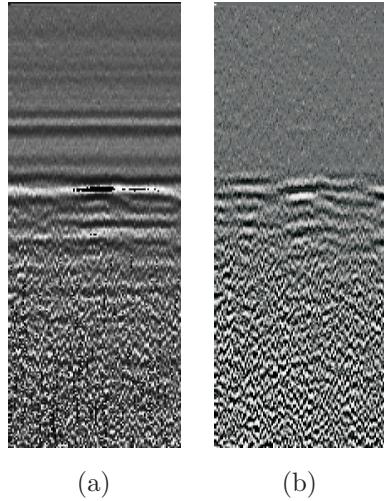


Figure 6.11: (a) A GPR image of a mine buried in natural soil and (b) the result after the “background removal”.

(see Fig. 6.11 and Fig. 6.12).

In estimating the depth of the object one usually neglects the propagation of the wave through air (since the GPR antenna is placed near to the ground). The depth to the top of the object is estimated by dividing the two-way travel time to the object by twice the velocity of the electromagnetic wave through the ground v_g [Peters94]. In practice, one estimates the travel time from the number of samples acquired between the signal emission and the first reflection peak from the object [VanKempen99]. The propagation (phase) velocity of a wave in ground is estimated as $v = c/\sqrt{\epsilon_r}$, where c is the speed of light in vacuum and ϵ_r is the relative permittivity of the material.

Data set in our experiments

For our experiments, we have used two types of GPR B-scans, the examples of which are shown in Fig. 6.11(a) and Fig. 6.12(a), respectively. A collection of scans as in Fig. 6.11(a), was provided by the Belgian Royal Military Academy and acquired by an experimental UWB GPR system developed by Scheers *et al* [Scheers98], which transmits ultra-short electromagnetic pulses (< 200 ps) at a 2.5 GHz central frequency. Images were acquired in natural soil in the test minefields from Fig. 6.3, and are 256x109 pixels. The second type of scans (Fig. 6.12(a)) was provided by

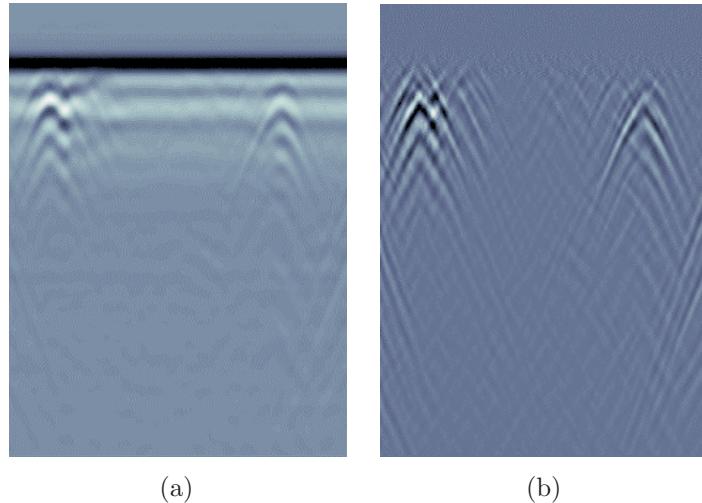


Figure 6.12: (a) A GPR image of mines and mine-like objects buried in pure sand (b) and the result after the “background removal”.

the Vrije Universiteit Brussels and was acquired by a commercial GPR system; in this case, several objects were buried in a sand box, and each scan is 512x232 pixels.

The horizontal lines in the original GPR scans are a consequence of strong surface reflections. To enhance the data of interest one needs to remove this non useful signal. The corresponding procedure is commonly called *background removal*, and can be based on different filtering techniques. We used a common and simple horizontal filtering [Fritzche95], which consists of subtracting the average trace from each row. The results of such background removal procedure for the two scans are shown in Fig. 6.11(b) and Fig. 6.12(b). In these pre-filtered images one can recognize the hyperbole signatures of the buried objects.

The GPR scans from the natural soil (Fig. 6.11(b)) are quite noisy; in Section 6.4.2, we propose one simple noise reduction technique. The GPR scans taken in the sand box (Fig. 6.11(b)) do not contain a lot of noise. In Section 6.4.3, we illustrate a suppression of the background clutter in these scans by means of spatially adaptive wavelet denoising.

Signal and image processing methods

Different signal and image processing techniques for GPR scans of landmines are reviewed, e.g., in [Brooks97, Verlinde01], and here we mention

a few. In [VanKempen98] a method was developed for signal detection in GPR A-scans using feature extraction techniques and binary hypothesis testing. Other types of GPR A-scan processing include the envelope detection using Hilbert transform [Milisavljevic01] (aimed for visual enhancement of horizontal slices), and use of the fast Fourier transform [Witten98] or wavelet transform [Scheers00] for object detection. The extraction of the hyperbole signatures of mines and minelike objects from GPR B-scans using randomized Hough transform (RHT) are treated, e.g., in [Capineri98, Milisavljevic01]. A 3D visualisation method was presented in [Milisavljevic99b] and 3D segmentation was discussed in [VanKempen99]. We focus here on applying wavelet based noise reduction techniques to GPR B-scans.

6.4.2 A nonlinear pyramidal filtering technique

For noise reduction in highly corrupted GPR scans from Fig. 6.11 we adapted a nonlinear filtering method of Sattar [Sattar97]. A pyramidal multiresolution representation used in this approach is a specific type of the non separable 2D representation, often denoted as the quincunx scheme [Kovacevic92, Starck98].

A decomposition step is depicted in Fig. 6.13(a). The image u_{m-1} at the resolution scale $m - 1$, is decomposed into a coarser image u_m and a detail image d_m . The coarser image is obtained by applying a lowpass filter, and subsampling the output by the factor 2 in each dimension. The subsampling is performed by removing every second pixel in the image. In order to make the pixels fit a cartesian grid, the image is not only subsampled but also rotated by 45°. The detail image d_m has the same dimension as u_{m-1} , and is obtained by subtracting an interpolated version of u_m from u_{m-1} . The produced sequence of lowpass and detail images is illustrated in Fig. 6.14.

Algorithm

Noise removal in this algorithm is a simplified form of the approach from Section 6.3.2, with the additional difference that it is performed in a decimated representation. One stage of the reconstruction process is represented in Fig. 6.13(b). The lowpass image at the resolution scale m is interpolated to match the dimension of the detail image d_m . To determine the positions of meaningful edges in the noisy detail, an edge detector is applied to the interpolated lowpass image, producing a binary mask x_m . The binary mask at a given position has value 1 or

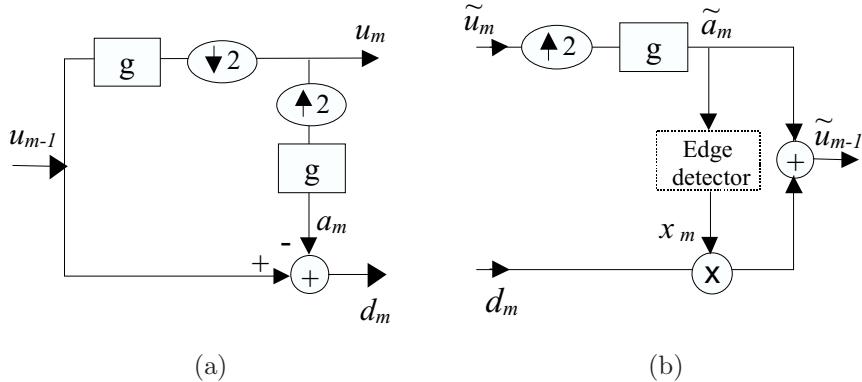


Figure 6.13: Pyramidal multiresolution technique [Sattar97]. (a) One decomposition stage and (b) one reconstruction stage.

0 depending on whether the corresponding pixel belongs to an edge or not. The detail image is multiplied by this mask, in order to remove the noise. The reconstruction consists of combining this modified detail image with the interpolated lowpass image.

The lowpass filter used in [Sattar97] was nonseparable, and in [Pizurica99a] we use a separable filter, which is much faster in a real implementation. The spectral characteristic of the filter is

$$G_s(\omega_1, \omega_2) = G(\omega_1)G(\omega_2), \quad G(\omega) = \begin{cases} 1, & |\omega| < B_s, \\ 1/2, & |\omega| = B_s \\ 0, & \text{otherwise,} \end{cases} \quad (6.4.1)$$

with $B_s = 2\pi/(3\sqrt{2})$. The actual filter is approximated using a two dimensional Hamming window of size 11x11.

For the edge detection we use the LoG approach as in [Sattar97]. The LoG filter corresponds to a bandpass filter, whose passband is determined by the spatial dispersion σ_G of the underlying Gaussian pulse. After filtering with the LoG filter, the zero crossings of the output are declared to be edges only when the brightness gradient exceeds a given threshold T . In [Pizurica99a], we investigated the influence of the LoG filter bandwidth as well as the threshold: for the considered GPR images, highly corrupted by speckle noise, the best results were obtained for $\sigma_G = 3$. For $\sigma_G \geq 3$, the change of this parameter in a wide range of values made no significant difference on the output image. Also, in this range the threshold selection did not appear to be critical. For $\sigma_G < 3$,

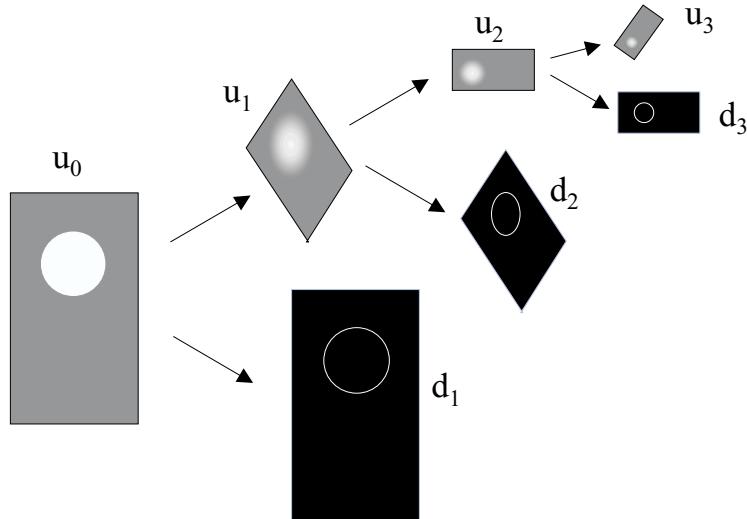


Figure 6.14: An illustration of the multiresolution pyramidal decomposition.

the edge detector overreacted to noisy regions without any real edges.

Results

As the input image for this multiscale denoising technique, we use a preprocessed (horizontally filtered) GPR scan. An input image (the same one as in Fig. 6.11(b)) is shown in Fig. 6.15(a). The result of the above described denoising technique is given in Fig. 6.15(b). One can see that noise is completely suppressed, and that the characteristic “hyperbole” signature of a landmine can be recognized.

The results of several other standard techniques are shown in Fig. 6.15 as well. Simple median filtering (Fig. 6.15(c)) does not provide satisfactory results on this type of images. The Lee filter (Fig. 6.15(d)) performs slightly worse than the proposed multiresolution technique. We also show the result of the classical wavelet thresholding [Donoho95a] applied to the image logarithm as in [Gagnon96, Odegard95]. Fig. 6.15(e) illustrates that in case of GPR images, this technique is not applicable.

6.4.3 Application of wavelet denoising with spatial priors

In images of commercial ground penetrating radars, especially if they are taken in sand (Fig. 6.12) there is not much noise present; the main

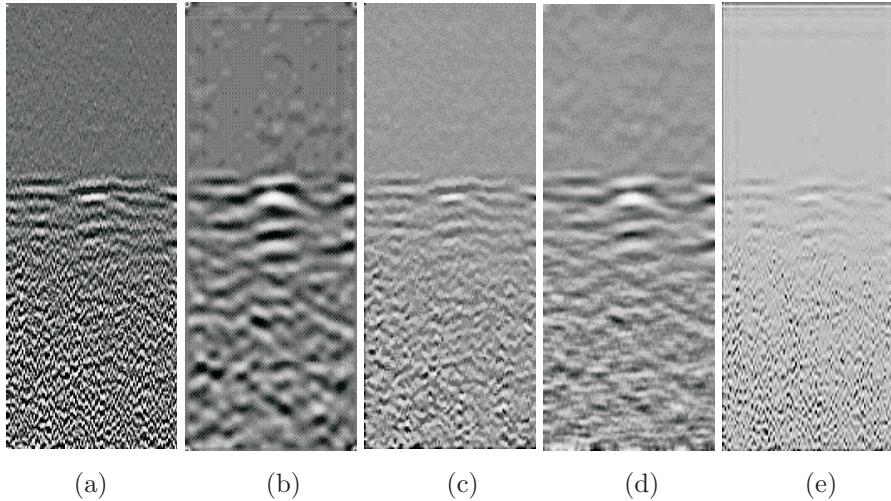


Figure 6.15: (a) An original GPR image after horizontal filtering. (b) The result of the nonlinear pyramidal filtering technique. (c) Median filter 3x3. (d) The Lee filter 7x7. (e) Soft thresholding of the wavelet coefficients applied to the logarithmically transformed input image.

goal of preprocessing techniques here is the removal of the “background clutter” (reflections that are not produced by the object of interest) in order to enhance the hyperbole signatures. Wavelet domain filtering techniques are sometimes seen as a solution to this problem [Brooks97]. We note here that removal of the background clutter from GPR images is not a typical noise reduction problem: in our opinion, most of the sophisticated noise reduction techniques should actually preserve this pronounced textural structure.

In our experiments, the simple nonlinear filtering techniques from Sections 6.4.2 and 6.3.2 suppress the background clutter at the expense of oversmoothed edges; these methods are less appropriate for enhancing relatively clean GPR scans. Among the methods that we developed in the previous Chapters, for this application seems suitable the method which uses spatial priors and heuristic shrinkage families (Section 3.5.3) and the method from Section 5.5.1, where the notion of a “signal of interest” is user defined. We did not test the latter one yet.

The algorithm of Section 3.5.3 was applied with the same parameters as in the case of infrared images in Section 6.3.3 ($\alpha = 1$, $\gamma = 0.2$ and $\delta = 0.5$). The results are shown in Fig. 6.16. From these and similar results on other GPR scans, we conclude that the proposed wavelet

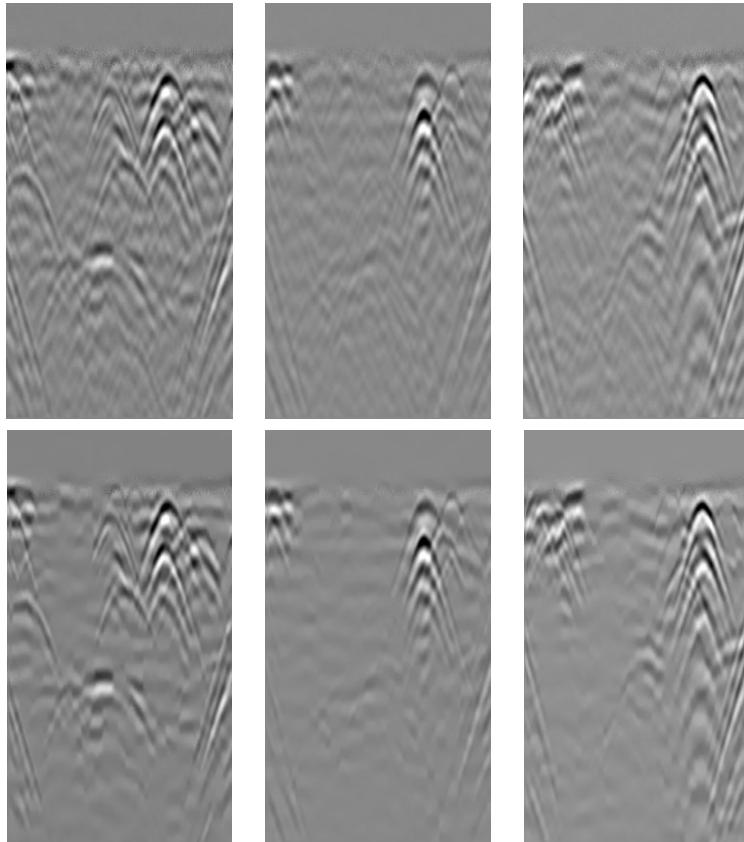


Figure 6.16: Top: scans of a commercial GPR from a sand box, after horizontal filtering. Bottom: the results of wavelet denoising with spatial priors.

denoising method enhances the characteristic target signatures. The validity of these results in practice needs to be further investigated.

6.5 Summary and conclusions

Different image processing techniques are needed in different steps of humanitarian mine detection. The algorithm that we developed in Section 5.6 fits in this framework, since SAR is one of the most frequently considered sensors for minefield detection. Our interest in this Chapter was denoising and enhancement of images of individual landmines, acquired by close range sensors that are often used in advanced demini-

ning concepts: IR cameras and GPR. Noise reduction techniques should provide images that are more suitable for visual inspection and for further processing steps, in case of an automatic detection procedure. At the moment we do not have actual validation of results in the sense of preprocessing for automatic detection procedures. The achieved visual enhancement of images is in most cases evident.

The application of nonlinear filtering techniques to infrared (Section 6.3.2) and to GPR images (Section 6.4.2) and the corresponding results presented in this Chapter resulted from our early research in this field [Pizurica98a, Pizurica98b, Pizurica99a, Pizurica99e]. Even though less advanced with respect to recent methods that we developed in the previous Chapters, these techniques provide a significant enhancement of certain types of images that are of interest in this application. We demonstrated also the application of the proposed methods that use spatial context information; the results of this research on various landmine images were published in [Pizurica99b, Pizurica99c, Pizurica99f]. These techniques suppress well noise in infrared images retaining well details; also this approach seems promising for the background clutter suppression in GPR scans. Further efforts are needed towards validation of these results in practical demining procedures.

Chapter 7

Conclusions

“In research, the horizon recedes as we advance, and is no nearer at 60 than it was at 20. As the power of endurance weakens with age, the urgency of pursuit grows more intense... and research is always incomplete.” Mark Pattison (1875)

In this thesis, we addressed wavelet domain image denoising adapting to the local and to the global image context. Throughout the text, we tried to present numerous original interpretations, pictorial explanations and discussions broadening our viewpoints on this topic.

In our original contributions, we paid a great deal of attention to accurate statistical modeling of the wavelet coefficients and their significance measures; also, we addressed deeply the statistical modeling of spatial interactions and of local spatial activity indicators. In making our way through this research field, we were inspired and almost excited by the concepts of uncertainty of signal presence and the corresponding joint detection and estimation (JDE) principles. We admired the pioneering work of Middleton and Esposito in this field. Despite their great potentials, the JDE concepts are not explored enough in wavelet based denoising. In this sense we contributed here a new class of low complexity locally adaptive methods which employ such concepts.

We tried to motivate the proposed algorithms in terms of specific optimization criteria, e.g., minimizing the mean squared error or maximizing the posterior probability of a solution given unknown uncorrupted data. From time to time, while bridging theory and applications, we employed heuristics too. In doing so, we were not abandoning our theoretical principles, but rather bringing them to practice. For example, the simple coarse-to-fine mask detection method from Section 5.4.1 helped

us to empirically estimate marginal densities of the coefficients and of local spatial activity indicators in cases where prior knowledge about their true functional forms was not or only partially accessible.

The main novelties and contributions presented in this thesis are the following. In Chapter 2, we systematized different wavelet denoising methods; with new pictorial interpretations, we tried to summarize the main ideas, similarities and differences between some representative methods. In Chapter 3, we described a more general framework of the wavelet based denoising using MRF priors, where the existing methods appear as one possibility. We then proposed a new approach in Section 3.5; the main idea of this approach is to compute from the MAP mask estimate a local spatial activity indicator; its value dictates the choice of a particular curve in a family of shrinkage characteristics; in this way, local spatial context information refines a wavelet shrinkage estimator. A practical algorithm developed from this concept was found useful in hydrologic applications. In Chapter 4, we presented our main contributions to Bayesian denoising using MRF priors: statistical characterization of different significance measures of wavelet coefficients; their objective performance evaluation; new, joint inter and intrascale significance measures and new anisotropic MRF prior model. Those developments led to generalizing and to improving previous methods from this branch. In Chapter 5, we developed a new locally adaptive denoising approach, upgrading further our concept of using local spatial activity indicators, from Section 3.5. First, we proposed an original idea to estimate the probability of the presence of a signal of interest from the global coefficient histogram and from a local spatial activity indicator; the resulting method is among the best state of the art ones. Further on, we introduced empirical extensions of the proposed approach, opening possibilities for various flexible algorithms that adapt to different noise types. Within this empirical framework, we developed other two practical algorithms and demonstrated their usefulness in practical applications.

We covered a range of application domains, image and noise types. Some of these were treated in more detail while others were not, because we merely wanted to illustrate the versatility of the developed algorithms. An extensive performance evaluation was performed for the suppression of artificially added *Gaussian noise* on various representative test images. In this respect, we proposed two new advanced algorithms: one employing Markov Random Fields for spatial clustering (Section 4.4) and the other employing a low complexity generalized

likelihood approach (Section 5.3). The first one usually yields a better performance on images that are highly corrupted by noise and/or dominated by large areas of uniform intensity intercepted by sharp edges and thin lines. The second approach usually offers advantages in case of low to moderate noise levels and images rich with natural textures. The versatile noise suppression algorithm that we developed in Section 5.5, was shown to be effective in denoising medical *ultrasound* and *magnetic resonance* images; also, preliminary results have shown that this method can be useful as a regularization step in *image deblurring*. The algorithm for speckle noise suppression in *Synthetic Aperture Radar* images that we developed in Section 5.6 was shown to preserve tiny details and point-like reflectors in these high-resolution images remarkably well, especially given the simplicity of the method both in its concept and realization. The enhancement of SAR images is important in *humanitarian mine detection*, which is the application that motivated our research in this field. We demonstrated the application of the developed noise suppression methods on some close-range sensors in humanitarian demining as well. In particular, we addressed in Chapter 6 denoising *infrared* and *Ground Penetrating Radar* images of landmines.

Apart from that, there is much room for further developments and new applications of the algorithms presented in this thesis. The anisotropic MRF model that we developed in Section 4.3 can be used in applications other than denoising as well; in particular, it seems suitable for detecting edges in (noisy) images and we intend to investigate its use for an automatic road detection in SAR images. It would be also interesting to apply this anisotropic MRF model with other and possibly larger subneighborhoods than those illustrated in Fig. 4.12. Also, the proposed generalized likelihood ratio denoising approach from Chapter 5, calls for further research. For example, in the analytical method from Section 5.3 use of more complex indicators of the local spatial activity, which take into account the inter-scale correlations as well, is likely to further improve the noise suppression performance. For such extensions, appropriate conditional pdf models should be developed. Next, from the proposed empirical approach that was summarized in Fig. 5.17, we have drawn two practical algorithms which are only special cases in this framework. We shall also analyze in more detail some application areas, like, e.g., image deblurring, that was here only briefly illustrated. It would be also interesting to investigate use of the proposed noise suppression algorithm for SAR images, as a tool for a lossy compression of those large images.

Appendix A

Deriving the global prior probability ratio

This appendix explains in detail how we derive the global ratio of prior probabilities $\rho = P(H^1)/P(H^0)$ in Section 5.3.3, for the Laplacian and for the generalized Laplacian prior $p_Y(y)$ distributions of noise free wavelet coefficients Y .

Our starting point was depicted in Fig. 5.3: the global probability of the presence of the signal of interest $P(H^1)$ is

$$P(H^1) = \int_{-\infty}^{-T} p_Y(y)dy + \int_T^{\infty} p_Y(y)dy, \quad (\text{A.0.1})$$

where T specifies the notion of a signal of interest. The ratio $\rho = P(H^1)/P(H^0)$ then becomes

$$\rho = \frac{1 - \int_{-T}^T p_Y(y)dy}{\int_{-T}^T p_Y(y)dy}. \quad (\text{A.0.2})$$

Now we proceed deriving this expression for the two prior models $p_Y(y)$ mentioned above.

Deriving ρ for the Laplacian prior

For the Laplacian prior

$$p_Y(y) = \frac{1}{2s} \exp(-|y/s|), \quad (\text{A.0.3})$$

we have

$$\int_{-T}^T p_Y(y) dy = \frac{1}{s} \int_0^T \exp(-(y/s)) dy = 1 - \exp\left(\frac{T}{s}\right). \quad (\text{A.0.4})$$

By substituting the previous expression into Eq (A.0.2), we obtain

$$\rho = \frac{\exp\left(\frac{T}{s}\right)}{1 - \exp\left(\frac{T}{s}\right)}, \quad (\text{A.0.5})$$

as was given in Eq (5.3.17).

Deriving ρ for the Generalized Laplacian prior

For the generalized Laplacian prior

$$p_Y(y) = \frac{\nu}{2s\Gamma(\frac{1}{\nu})} \exp(-|y/s|^\nu), \quad (\text{A.0.6})$$

the integral in Eq (A.0.2) becomes

$$\int_{-T}^T p(y) dy = \frac{\nu}{2s\Gamma(\frac{1}{\nu})} \int_{-T}^T \exp(-|y/s|^\nu) dy = \frac{\nu}{s\Gamma(\frac{1}{\nu})} \int_0^T \exp(-(y/s)^\nu) dy.$$

By introducing the change of variables:

$$t = \left(\frac{y}{s}\right)^\nu,$$

it follows

$$dt = \frac{\nu}{s} \left(\frac{y}{s}\right)^{\nu-1} dy = \frac{\nu}{s} t^{\frac{\nu-1}{\nu}} dy, \quad \text{i.e.} \quad dy = \frac{s}{\nu} t^{\frac{1}{\nu}-1} dt.$$

From this, it follows that

$$\int_{-T}^T p_Y(y) dy = \frac{1}{\Gamma(\frac{1}{\nu})} \int_0^{\left(\frac{T}{s}\right)^\nu} t^{\frac{1}{\nu}-1} e^{-t} dt = \Gamma_{inc}\left(\left(\frac{T}{s}\right)^\nu, \frac{1}{\nu}\right),$$

where

$$\Gamma_{inc}(x, a) = \frac{1}{a} \int_0^x t^{a-1} e^{-t} dt \quad (\text{A.0.7})$$

is the *incomplete gamma function*. The global probability ratio ρ from Eq (A.0.2) now becomes

$$\rho = \frac{1 - \Gamma_{inc}\left(\left(\frac{T}{s}\right)^\nu, \frac{1}{\nu}\right)}{\Gamma_{inc}\left(\left(\frac{T}{s}\right)^\nu, \frac{1}{\nu}\right)}, \quad (\text{A.0.8})$$

as was given in Eq (5.3.15).

Appendix B

MLE parameter estimation

This Appendix provides details of the mathematical derivation in the proposed SAR despeckling algorithm from Section 5.6.3. In that method, we assume functional forms of the conditional densities of the coefficient magnitudes, given noise and given noisy edges, respectively. To estimate the parameters of those densities directly from the observed noisy image, we first detect binary masks and accordingly compute the histograms of the magnitudes of the coefficients representing noise and noisy edges, respectively. From the corresponding histograms, we then compute the maximum likelihood estimates (MLE) of the required parameters. In particular, the functional form of the densities are given by Eqs (5.6.6) and (5.6.7) and the MLE estimates of their parameters by Eqs (5.6.8) and (5.6.9), respectively. Here we show how we derived those MLE parameter estimates.

General considerations

In explaining the concept of the MLE parameter estimation, we closely follow [Duda73]. Let ω denote a realization of a continuous random variable with the probability density function

$$p(\omega; \theta_1, \dots, \theta_k), \quad (\text{B.0.1})$$

where $\theta_1, \dots, \theta_k$ are the parameters. Suppose that $\omega_1, \omega_2, \dots, \omega_n$ are n independent observations drawn from the corresponding distribution. The *likelihood function* of the parameters with respect to the set of observations (samples) is

$$p(\omega_1, \dots, \omega_n; \theta_1, \dots, \theta_k) = \prod_{i=1}^n p_\Omega(\omega_i; \theta_1, \dots, \theta_k). \quad (\text{B.0.2})$$

The *maximum likelihood estimate* of $\boldsymbol{\theta} = [\theta_1, \dots, \theta_k]$ is by definition that value $\hat{\boldsymbol{\theta}}$ that maximizes Eq (B.0.2). In practical calculations, it is usually easier to work with the logarithm of the likelihood function (log-likelihood):

$$L(\omega_1, \dots, \omega_n; \theta_1, \dots, \theta_k) = \ln p(\omega_1, \dots, \omega_n; \theta_1, \dots, \theta_k). \quad (\text{B.0.3})$$

The parameter vector $\hat{\boldsymbol{\theta}}$ that maximizes the log-likelihood also maximizes the likelihood. The MLE estimates of the parameters $\theta_1, \dots, \theta_k$ are thus obtained by solving the following k equations

$$\frac{\partial L(\omega_1, \dots, \omega_n; \theta_1, \dots, \theta_k)}{\partial \theta_j} = 0, \quad j = 1, \dots, k. \quad (\text{B.0.4})$$

Now we turn to our practical problem, i.e., to deriving the parameters of the two pdf models from Section 5.6.3.

B.1 Exponential distribution

In Section 5.6.3, we model the conditional pdf of the coefficient magnitudes dominated by speckle noise as $p_{\Omega_l|X_l}(\omega|0) \simeq (1/a) \exp(-\omega/a)$. Here we start from an equivalent model, with the parameter $\theta = 1/a$, which simplifies the following derivation.

For the exponential density

$$p(\omega; \theta) = \theta \exp(-\theta\omega), \quad (\text{B.1.1})$$

the likelihood of the parameter θ with respect to the set of samples $\omega_1, \dots, \omega_n$ is

$$p(\omega_1, \dots, \omega_n; \theta) = \prod_{i=1}^n \theta \exp(-\theta\omega_i) = \theta^n \exp\left(-\theta \sum_{i=1}^n \omega_i\right). \quad (\text{B.1.2})$$

Taking the logarithm of the above expression, we obtain the log-likelihood

$$L(\omega_1, \dots, \omega_n; \theta) = n \ln(\theta) - \theta \sum_{i=1}^n \omega_i. \quad (\text{B.1.3})$$

The MLE estimate of the parameter θ follows from

$$\left(\frac{\partial L(\omega_1, \dots, \omega_n; \theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n \omega_i \right)_{\theta=\hat{\theta}} = 0, \quad (\text{B.1.4})$$

which leads to

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \omega_i}. \quad (\text{B.1.5})$$

Now, if instead of $\{\omega_1, \dots, \omega_n\}$, our set of observations is $\{\omega_l : l \in S_0\}$, and if $N_0 = \#S_0$ is the cardinality of the set S_0 , then the MLE estimate of the parameter $a = 1/\theta$ follows from (B.1.5) as

$$\hat{a} = (1/N_0) \sum_{i \in S_0} \omega_i, \quad N_0 = \#S_0, \quad (\text{B.1.6})$$

which is the expression given in Eq (5.6.8).

B.2 Generalized Gamma distribution

In Section 5.6.3, we model the conditional pdf of the coefficient magnitudes dominated by useful signal as $\hat{p}_{\Omega_l|X_l}(\omega|1) = (1/2b)(\omega/b)^2 \exp(-\omega/b)$. We start again from an equivalent model, with the parameter $\theta = 1/b$, which simplifies the following derivation.

For a generalized Gamma density of the form

$$p(\omega; \theta) = \frac{\theta}{2}(\theta\omega)^2 \exp(-\theta\omega), \quad (\text{B.2.1})$$

the likelihood of the parameter θ with respect to the set of observations $\{\omega_1, \dots, \omega_n\}$ is

$$p(\omega_1, \dots, \omega_n; \theta) = \prod_{i=1}^n \frac{\theta}{2}(\theta\omega_i)^2 \exp(-\theta\omega_i). \quad (\text{B.2.2})$$

The logarithm of the above expression leads to the log-likelihood

$$L(\omega_1, \dots, \omega_n; \theta) = n \ln\left(\frac{\theta}{2}\right) + 2n \ln(\theta) + 2n \ln(\omega_i) - \theta \sum_{i=1}^n \omega_i. \quad (\text{B.2.3})$$

The MLE estimate of the parameter θ follows from

$$\left(\frac{\partial L(\omega_1, \dots, \omega_n; \theta)}{\partial \theta} = \frac{3n}{\theta} - \sum_{i=1}^n \omega_i \right)_{\theta=\hat{\theta}} = 0, \quad (\text{B.2.4})$$

leading to

$$\hat{\theta} = \frac{3n}{\sum_{i=1}^n \omega_i}. \quad (\text{B.2.5})$$

In Section 5.6.3, instead of $\{\omega_1, \dots, \omega_n\}$, our set of observations drawn from the corresponding density is $\{\omega_l : l \in S_1\}$, and we are seeking the estimate of $b = 1/\theta$. If $N_1 = \#S_1$ is the cardinality of the set S_1 , then the MLE estimate of the parameter $b = 1/\theta$ follows from (B.2.5) as

$$\hat{b} = (1/3N_1) \sum_{i \in S_1} \omega_i, \quad N_1 = \#S_1, \quad (\text{B.2.6})$$

which is the expression given in Eq (5.6.9).

Bibliography

- [Aach96a] T. Aach and D. Kunz, “Spectral estimation filters for noise reduction in x-ray fluoroscopy imaging,” in *Proc. European Conf. on Signal Proc. EUSIPCO96*, Trieste, 10-13 Sep, 1996.
- [Aach96b] T. Aach and D. Kunz, “Anisotropic spectral magnitude estimation filters for noise reduction and image enhancement,” in *Proc. IEEE Internat. Conf. on Image Proc. ICIP96*, pp. 335–338, Lausanne, Switzerland, 16-19 Sep, 1996.
- [Abdou79] I.E. Abdou and W.K. Prat, “Quantitative design and evaluation of enhancement/thresholding edge detectors,” *Proc. IEEE*, 67:753–763, May 1979.
- [Abend65] K. Abend, T.J. Harley, and L.N. Kanal, “Classification of binary random patterns,” *IEEE Trans. Inform. Theory*, 11:538–544, 1965.
- [Abramovich98] F. Abramovich, T. Sapatinas, and B. W. Silverman, “Wavelet thresholding via a Bayesian approach,” *J. of the Royal Statist. Society B*, 60:725–749, 1998.
- [Acheroy98a] M. Acheroy, ”L’infrarouge thermique, principes et applications au déminage humanitaire,” *Revue HF*, 3:13–24, 1998.
- [Acheroy98b] M. Acheroy, Y. Baudoin, and M. Piette, “Belgian project on humanitarian demining,” in *Proc. Internat Symp. on Climbing and Walking robots CLAWAR98*, pp. 215–218, Brussels, Belgium, 26-28 Nov, 1998.
- [Achim01] A. Achim, A. Bezerianos, and P. Tsakalides, “Novel Bayesian multiscale method for speckle removal in medical ultrasound images,” *IEEE Trans. Medical Imaging*, 20(8):772-783, Aug 2001.

- [Ahuja81] N. Ahuja, A. Rosenfeld, “Mosaic models for textures,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 3:1–11, Jan 1981.
- [Alparone90] L. Alparone, F. Boragine, and S. Fini, “Parallel architectures for the postprocessing of SAR images,” in *Proc. SPIE Conf.*, vol. 1360, pp. 790–802, 1990.
- [Andrews74] D. Andrews and C. Mallows, “Scale mixtures of normal distributions,” *J. Royal. Statist. Soc.*, 36, 1974.
- [Antonini92] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, “Image coding using wavelet transform,” *IEEE Trans. Image Proc.*, 1:205–220, Feb 1992.
- [Banham96] M.R. Banham, A.K. Katsaggelos, “Spatially adaptive wavelet-based multiscale image restoration,” *IEEE Trans. Image Proc.*, 5:619–634, 1996.
- [Barten99] P.G.J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, SPIE Press, Bellingham, WA, 1999.
- [Basseville92] M. Basseville et al, “Modeling and estimation of multiresolution stochastic processes,” *IEEE Trans. Inform. Theory*, 38(2):766–784, Mar 1992.
- [Battle87] G. Battle, “A block spin construction of ondellettes, Part 1: Lemarie functions,” *Commun. Math. Phys.*, 110:601–615, 1987.
- [Besag74] J. Besag, “Spatial Interaction and Statistical Analysis of Lattice Systems,” *J. Royal Statistical Society B*, 6:192–236, 1974.
- [Besag86] J. Besag, “On the statistical analysis of dirty pictures,” *J. Royal Statistical Society B*, 48:259–302, 1986.
- [Bijaoui96] A. Bijaoui, E. Slezak, F. Rue, and E. Lega, “Wavelets and study of the distant universe,” *Proc. IEEE*, 84(4):670–679, Apr 1996.
- [Bishop98] P. K. Bishop, K. M. Perry, and M. A. Poulter, “Airborne minefield detection,” in *Proc. IEE Internat. Conf. on the Detection of Abandoned Landmines*, pp. 213–221, Edinburgh, UK, 12–14 Oct, 1998.

- [Blake89] A. Blake, “Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 11:2–12, 1989.
- [Bouman91] C. Bouman and B. Liu, “Multiple resolution segmentation of textured images,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 13:99–113, 1991.
- [Brooks97] J.W. Brooks, “A survey of modern signal processing methods and their application to sustainable humanitarian demining,” in *Sustainable Humanitarian Demining: Trends, Techniques, and Technologies*, E. Pass, editor, Mid Valley Press, Verona, Virginia, 1997, pp. 196–205.
- [Bruschini96] C. Bruschini, B. Gros, F. Guerne, P.-Y. Pièce, and O. Carmona, “Ground penetrating radar and induction coil sensor imaging for antipersonnel mine detection,” in *Proc. Internat. Conf. on Ground Penetrating Radar GPR96*, Sendei, Japan, 30 Sep - 3 Oct, 1996.
- [Bruschini97] C. Bruschini and B. Gros, “A survey of current sensor technology research for the detection of landmines,” in *Sustainable Humanitarian Demining: Trends, Techniques, and Technologies*, E. Pass, editor, Mid Valley Press, Verona, Virginia, 1997, pp. 172–187.
- [Bruschini99] C. Bruschini, K. De Bruyn, H. Sahli, and J. Cornelis, “Study on the State of the Art in the EU related to humanitarian demining technology, products and practice,” the final report of the EU on humanitarian demining (EUDEM), Brussels, July 1999.
- [Bryant79] D.J. Bryant and D.W. Bouldin, “Evaluation of edge operators using relative and absolute grading,” in *Proc. IEEE Comp. Soc. Conf. on Pattern Recognition and Image Proc.*, Chicago, pp. 138–145, 1979.
- [Burt83] P. J. Burt and E. H. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Trans. Commun.*, 31:532–540, Apr 1983.
- [Camus95] R. Birk, W. Camus, E. Valenti, and W. McCandless, “Synthetic aperture radar imaging systems,” *IEEE AES Systems Magazine*, 10:15–23, 1995.

- [Candes99] E.J. Candès, “Ridgelets and their derivatives: representation of images with edges,” in *Curves and Surfaces*, L.L. Schumaker *et al*, editors, Vanderbilt University Press, Nashville, TN, 1999.
- [Candes00] E.J. Candès and D.L. Donoho, “Curvelets, multiresolution representation, and scaling laws”, in *Proc. SPIE 4119, Wavelet Applications in Signal and Image Processing VIII*, A. Aldroubi, A. F. Laine, M. A. Unser eds., 2000.
- [Canny86] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 8:679–698, Nov 1986.
- [Capineri98] L. Capineri, P. Grande, and J.A.G. Temple, “Advanced image-processing technique for real-time interpretation of ground penetrating radar images,” *Int. J. Imaging Syst. Technol.*, 9:51–59, 1998.
- [Cetin94] A. E. Cetin and R. Ansari, “Signal recovery from wavelet transform maxima,” *IEEE Trans. Signal Proc.*, 42:194–196, 1994.
- [Chambolle98] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, “Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage,” *IEEE Trans. Image Proc.*, 7(3):319–335, Mar 1998.
- [Chang98] S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising”, in *Proc. IEEE Internat. Conf. on Image Proc. ICIP*, Chicago, IL, Oct 1998.
- [Chang00a] S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising ,” *IEEE Trans. Image Proc.*, 9(9):1522–1531, Sep 2000.
- [Chang00b] S. G. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression ,” *IEEE Trans. Image Proc.*, 9(9):1532–1546, Sep 2000.
- [Chellappa83] R. Chellappa, Y.-H. Hu, and S.-Y. Kung, “On two-dimensional Markov Spectral Estimation,” *IEEE Trans. Acoustics, Speech, Signal Proc.*, 31:836–841, Aug 1983.

- [Chellappa85] R. Chellappa, “Two-dimensional discrete gaussian markov random field models for image processing,” *Progress in Pattern Recognition*, L. N. Kanal and A. Rosenfeld, editors, North-Holland 1985.
- [Chipman97] H.A. Chipman, E.D. Kolaczyk, and R.E. McCulloch, “Adaptive bayesian wavelet shrinkage,” *J. of the Amer. Statist. Assoc.*, 92:1413–1421, 1997.
- [Chui92] C. Chui, *Wavelets*, Academic Press, 1992.
- [Clyde96] M. Clyde, H. DeSimone, and G. Parmigiani, “Prediction via orthogonalized model mixing,” *J. of the Amer. Statist. Society*, 91(435):1197–1208, 1996.
- [Clyde98] M. Clyde, G. Parmigiani, and B. Vidakovic, “Multiple shrinkage and subset selection in wavelets,” *Biometrika*, 85(2):391–401, 1998.
- [Cohen92] A. Cohen, I. Daubechies, and J. Feauveau, “Biorthogonal bases of compactly supported wavelets,” *Commun. on Pure and Appl. Math.*, 45:485–560, 1992.
- [Cohen96] A. Cohen and J. Kovacevic, “Wavelets: the mathematical background,” *Proc. IEEE*, 84:514–522, 1996.
- [Cohen00] I. Cohen, “On speech enhancement under signal presence uncertainty,” in *Proc. Internat Conf. on Acoust. Speech and Signal Proc. ICASSP*, Salt Lake City, Utah, 2000.
- [Coifman92] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, “Wavelet analysis and signal processing,” in *Wavelets and Their Applications*, B. Ruskai *et al*, editors, pp. 153–178, Jones and Bartlett, Boston, 1992.
- [Coifman95] R.R. Coifman and D.L. Donoho, “Translation-invariant de-noising,” in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, editors, pp. 125–150, Springer Verlag, New York, 1995.
- [Comer99] M. L. Comer and E. J. Delp, “Segmentation of textured images using a multiresolution Gaussian autoregressive model,” *IEEE Trans. Image Proc.*, 8:408–420, 1999.
- [Crimmins85] T.R. Crimmins, “Geometric filter for speckle reduction,” *Appl. Optics*, 24:1438–1443, 1985.

- [Cross83] G. R. Cross and A. K. Jain, “Markov random field texture models,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 5(1):25–39, Jan 1983.
- [Crouse98] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, “Wavelet-based statistical signal processing using hidden markov models,” *IEEE Trans. Signal Proc.*, 46:886–902, 1998.
- [Cvetkovic95] Z.Cvetkovic and M.Vetterli, “Discrete-time wavelet extrema representation: design and consistent reconstruction,” *IEEE Trans. Signal Proc.*, 43(3):681-694 Mar 1995.
- [Daniels96] D.J. Daniels, *Surface-Penetrating Radar*, The Institution of Electrical Engineers, London, United Kingdom, 1996.
- [Daniels97] D.J. Daniels, “Radar techniques for mine detection,” in *Sustainable Humanitarian Demining: Trends, Techniques, and Technologies*, E. Pass, editor, Mid Valley Press, Verona, Virginia, 1997, 188–205.
- [Daoudi98] K. Daoudi, J. Lévy-Véhel and Y. Meyer, “Construction of continuous functions with prescribed local regularity,” *Constructive approximation*, 14(3):349–385, 1998.
- [Daubechies88] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Comm. Pure & Appl. Math.*, 41:909–996, 1988.
- [Daubechies92] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.
- [Daubechies96] I. Daubechies, “Where do wavelets come from? - a personal point of view,” *Proc. IEEE*, 84(4):510–513, Apr 1996.
- [DePersia95] A.T. DePersia, A. Bowman, P. Lucey, and E.M. Winter, “Phenomenology considerations for hyperspectral mine detection,” in *Proc. SPIE*, vol. 2496, pp. 159–167.
- [Derin84] H. Derin, H. Elliott, R. Cristi, and D. Geman, “Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 6:707–720, 1984.
- [Derin87] H. Derin and H. Elliott, “Modeling and segmentation of noisy and texture images using gibbs random fields,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 9:39–55, 1987.

- [Donoho92a] D.L. Donoho and I.M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika*, 8:425–455, 1994.
- [Donoho92b] D. L. Donoho and I. M. Johnstone, “Minimax estimation via wavelet shrinkage,” Technical Report, Statistics, Stanford, 1992.
- [Donoho95a] D. L. Donoho, “De-Noising by Soft-Thresholding,” *IEEE Trans. Inform. Theory*, 41:613–627, May 1995.
- [Donoho95b] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *J. Amer. Stat. Assoc.*, 90(432):1200–1224, Dec 1995.
- [Donoho95c] D. L. Donoho, “Nonlinear solution of linear inverse problems by Wavelet-Vaguelette Decomposition,” *Appl. Comp. Harm. Anal.*, 2:101126, 1995.
- [Donskoy98] D.M. Donskoy, “Nonlinear seismo-acoustic technique for land mine detection and discrimination,” in *Proc. IEE Internat. Conf. on the Detection of Abandoned Land Mines*, pp. 244–248, Edinburgh, UK, 1998.
- [Druyts00] P. Druyts, L. Merlat, and M. Achery, “Modeling considerations for imaging with a standard metal detector,” in *Proc. SPIE, Detection and Remediation of Mines and Minelike Targets V*, vol. 4038, pp. 1431–1451, Orlando, FL, USA, Apr 2000.
- [Dubes85] R. C. Dubes and A. K. Jain, ‘Random Field Models in Image Analysis,’ *Progress in Pattern Recognition*, L. N. Kanal and A. Rosenfeld, editors, North-Holland 1985.
- [Duda73] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.
- [Duskunovic00a] I. Duskunovic, A. Pizurica, W. Philips, and I. Lemahieu, “Implementation of a new filter based on simplified spatial rules for removing Gaussian noise,” in *Proc. 2nd IEEE Benelux Signal Processing Symposium SPS*, Hilvarenbeek, The Netherlands, 1999.
- [Duskunovic00b] I. Duskunovic, G. Stippel, A. Pizurica, W. Philips, and I. Lemahieu, “A new restoration method and its application to speckle images,” *Proc. IEEE International Conf. on Image Proc. ICIP*, pp. 273–276, Vancouver, BC, Canada, Sep, 2000.

- [Duskunovic00c] I. Duskunovic, A. Pizurica, G. Stippel, W. Philips, and I. Lemahieu, "Wavelet based denoising techniques for ultrasound images," in *Proc. World Congress on Medical Physics and Biomedical Engineering*, Chicago, USA, July 2000.
- [Edelstein86] W. A. Edelstein, G. Glover, C. Hardy, and R. Redington, "The intrinistic signal-to-noise ratio in NMR imaging," *Magn. Reson. Med.*, 3:604–618, 1986.
- [Engelbeen98] A. Engelbeen, "Nuclear quadrupole resonance mine detection," in *Proc. Internat. Symposium on Climbing and Walking Robots*, pp. 249–253, Brussels, Belgium, Nov 1998.
- [Ephraim84] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimation," *IEEE Trans. Acoust. Speech and Signal Proc.*, 32(6):1109–1121, Dec 1984.
- [Esteban77] D. Esteban and C. Galland, "Applications of quadrature mirror filters to split band voice coding schemes," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, May 1977.
- [Fan00] G. Fan and X.-G. Xia, "Wavelet-based image denoising using hidden markov models," in *Proc. IEEE International Conf. on Image Proc. ICIP*, Vancouver, BC, Canada, 2000.
- [Fan01] G. Fan and X.-G. Xia, "Image denoising using local contextual hidden markov model in the wavelet domain," *IEEE Signal Processing Lett.* 8(5):125–128, May 2001.
- [Farge96] M. Farge, N. Kevlahan, V. Perrier, and E. Goirand, "Wavelets and Turbulence," *Proceedings IEEE*, 84(4):639–669, April 1996.
- [Feilner01] M. Feilner, M. Jacob, M. Unser, "Orthogonal quincunx wavelets with fractional orders," in *Proc. IEEE Internat. Conf. on Image Proc.*, pp. 606-609, Thessaloniki, Greece, October 7-10, 2001,
- [Figueiredo97] M. A. T. Figueiredo, J. M. N. Leitão, "Unsupervised Image Restoration and Edge Location Using Compound Gauss-Markov Random Fields and the MDL Principle," *IEEE Trans. Image Proc.*, 6:1089–1102, 1997.
- [Foucher96] S. Foucher, G. B. Bénié, and J. M. Boucher, "Unsupervised multiscale speckle filtering," in *Proc. Internat. Conf. Image Proc. ICIP*, Lausanne, Switzerland, 16-19 Sep. 1996, pp. 391–394.

- [Foucher01] S. Foucher, G. B. Bénié, and J. M. Boucher, “Multiscale MAP filtering of sar images,” *IEEE Trans. Image Proc.*, 10(1):49–60, Jan 2001.
- [Fritzche95] M. Fritzche, “Detection of buried ladmines using ground penetrating radar,” in *Detection Technologies for Mines and Mine-like Targets, Proc. SPIE*, vol. 2496, pp. 100–109, 1995.
- [Frost82] V.S. Frost, J.A. Stiles, K.S. Shanmugan, and J.C. Holtzman, “A model for radar images and its application to adaptive digital filtering of multiplicative noise,” *IEEE Trans. Pattern Anal. and Machine Intell.*, 4(2):157–166, Mar 1982.
- [Gagnon96] L. Gagnon, F. Drissi Smaili, “Speckle noise reduction of airborne SAR images with symmetric Daubechies wavelets,” *SPIE Proc. no. 2759, conference “Signal and Data Processing of Small Targets”*, Orlando 1996.
- [Gagnon97] L. Gagnon and A. Jouan, “Speckle filtering of SAR images - a comparative study between complex-wavelet-based and standard filters” *SPIE Proc. no. 3169, conference “Wavelet Applications in Signal and Image Processing V”*, San Diego 1997.
- [Gao99] D.J. Gao, “Duality-mathematics,” in *Wiley Encyclopedia of Electrical and Electronics Engineering*, J.G. Webster, editor, John Wiley & Sons, Inc., vol. 6, 1999, pp. 68–77.
- [Geman84] S. Geman and D. Geman, “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 6:721–741, 1984.
- [Geronimo94] J. Geronimo, D. Hardin, and P. R. Massupust, “Fractal functions and wavelet expansions based on several functions,” *J. of Approx. Theory*, 78:373–401, 1994.
- [Gidas89] B. Gidas, “A renormalization group approach to image processing problems,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 11:164–180, 1989.
- [Goo98] G.-I. Goo, Y. Yang, W.W.L. Au, and L.S. Couchman, “Detection and identification of underwater objects in acoustic daylight,” in *Proc. SPIE Conf. on Detection Technologies for Mines and Mine-like Targets*, vol. 3392, pp. 218–233, Orlando, USA, 1998.

- [Goodman76] J. W. Goodman, “Some fundamental properties of speckle,” *J. Opt. Soc. Am.*, 66:1145–1150, 1976.
- [Graps95] A. Graps, “An introduction to wavelets,” *IEEE Computational Science and Engineering*, 2(2), 1995.
- [Grosman84] A. Grosman and J. Morlet, “Decomposition of Hardy functions into square integrable wavelets of constant shape,” *SIAM J. Math.*, 15:723–736, 1984.
- [Gudbjartsson95] H. Gudbjartsson and S. Patz, “The Rician distribution of noisy MRI data,” *Magn. Reson. Med.*, 34:910–914, 1995.
- [HessNielsen96] N. Hess-Nielsen and M. V. Wickerhauser, “Wavelets and time-frequency analysis,” *Proc. IEEE*, 84(4):523–540, Apr 1996.
- [Hilton97] M. L. Hilton and R. T. Ogden, “Data analytic wavelet threshold selection in 2-d denoising,” *IEEE Trans. Signal Proc.*, 45:496–500, 1997.
- [Holschneider89] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, “Wavelets, time-frequency methods and phase space”, *Real Time Algorithm for Signal Analysis with the Help of the Wavelet Transform*, pp. 289–297, Springer-Verlag, Berlin, 1989.
- [Hsung99] T.-C. Hsung, D. P.-K. Lun, and W. -C. Siu, “Denoising by singularity detection,” *IEEE Trans. Signal Proc.*, 47(11):3139–3144, Nov 1999.
- [Jaffard91] S. Jaffard, “Pointwise smoothness, two-microlocalization and wavelet coefficients,” *Publicacions Matemàtiques*, 35:155-168, 1991.
- [Jain81] A. K. Jain, “Advances in mathematical models for image processing,”, *Proc. IEEE*, 69:502–528, May 1981.
- [Jain91] A. K. Jain and F. Farrokhnia, “Unsupervised texture segmentation using gabor filters”, *Pattern Recognition*, 24(12):1167–1186, 1991.
- [Jalobeanu01] A. Jalobeanu, “Modèles, estimation bayésienne et algorithmes pour la déconvolution d’images satellitaires et aériennes,” Ph.D. thesis, INRIA Sophia-Antipolis, France, 2001.

- [Jankowski92] P.Z. Jankowski, A.G. Mercado, and S.F. Hallowell, “FAA explosive vapor/particle detection technology,” in *Proc. SPIE*, vol. 1824, pp. 13–24, 1992.
- [Jansen97] M. Jansen, M. Malfait, A. Bultheel, “Generalized Cross Validation for Wavelet Thresholding,” *Signal Processing*, 56:33–44, Jan 1997.
- [Jansen99] M. Jansen and A. Bultheel, “Geometrical Priors for Noise-free Wavelet Coefficient Configurations in Image De-noising,” in *Bayesian inference in wavelet based models*, P. Müller and B. Vidakovic, editors, Springer Verlag 1999, pp. 223–242.
- [Jansen01a] M. Jansen and A. Bultheel, “Empirical Bayes approach to improve wavelet thresholding for image noise reduction,” *J. of the Amer. Statist. Assoc.*, 96(454):629–639, June 2001
- [Jansen01b] M. Jansen, *Noise Reduction by Wavelet Thresholding*, Springer-Verlag, New York, 2001.
- [Johnstone96] I. Johnstone, B. W. Silverman, “Wavelet threshold estimators for data with correlated noise,” *J. of the Royal. Statist. Society B*, 59:319–351.
- [Kalifa99] J. Kalifa and S. Mallat, “Thresholding Estimators for Inverse Problems and Deconvolutions,” submitted to *Annals of Statistics*, 1999.
- [Kashiap83] R. L. Kashiap and R. Chellappa, “Estimation and choice of neighbors in spatial-interaction models of images,” *IEEE Trans. Inform. Theory*, 29(1):60–72, Jan 1983.
- [Kato93] Z. Kato, M. Berthod, and J. Zerubia, “A hierarchical Markov random field model and multi-temperature annealing for parallel image classification,” Rapport de recherche No 1938, Institut National de Recherche en Informatique, Robotique, image et vision, 1994.
- [Khotanzad99] A. Khotanzad and J. Bennett, “A spatial correlation based method for neighbor set selection in random field image models,” *IEEE Trans. Image Proc.*, 8:734–740, 1999.
- [Kitchen81] L. Kitchen and A. Rosenfeld, “Edge evaluation using local edge coherence,” *IEEE Trans. SMC*, 94:597–605, Sep 1981.

- [Klein97] L.A. Klein, *Millimeter-Wave and Infrared Multisensor Design and Signal Processing*, Artech House, Inc.Norwood, USA, 1997.
- [Kossoff76] G. Kossof, W.J. Garret, D. A. Carpenter, J. Jellins, and M. J. Dadd, “Principles and classification of soft tissues by gray scale echography,” *Ultrasound Med. Biol.*, 2:89–105, 1976.
- [Kovacevic92] J. Kovacevic and M. Vetterli, “Nonseparable multidimensional perfect reconstruction filter banks and wavelet for R_n ,” *IEEE Trans. Inform. Theory*, 38(2):533–556, Mar 1992.
- [Krishnamachari97] S. Krishnamachari and R. Chelappa, “Multiresolution Gauss-Markov random field models for texture segmentation,” *IEEE Trans. Image Proc.*, 6:251–267, 1997.
- [Kuan85] D. T. Kuan, A. A. Sawchuk, T. C. Strand, P. Chavel, “Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 7(2):165–177, Mar 1985.
- [Kuan87] D. T. Kuan, A. A. Sawchuk, T. C. Strand, P. Chavel, “Adaptive restoration of images with speckle,” *IEEE Trans. Acoust., Speech, Signal Proc.*, 35:373–383, Mar 1987.
- [Lee80] J. S. Lee, “Digital image enhancement and noise filtering by use of local statistics,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 2(2):165–168, Mar 1980.
- [Lee86] J. -S. Lee, “Speckle suppression and analysis for synthetic aperture radar images,” *Opt. Engineering*, 25(5):636–643, May 1986.
- [Lee89] J. -S. Lee and I. Jurkevich, “Segmentation of SAR Images,” *IEEE Trans. Geosc. Remote Sens.*, 27(6):674–680.
- [Lemarie86] P. G. Lemarié, “Ondelettes à localisation exponentielle,” *J. Math Pures et Appl.*, 67:227–236, 1988.
- [Levers92] V.F. Leavers, *Shape Detection in Computer Vision Using the Hough Transform*, Springer, London, 1992.
- [LevyVehel01] J. Lévy-Véhel and E. Lutton, “Evolutionary signal enhancement based on Hölder regularity analysis,” in *Proc. European Workshop on Evolutionary Computation in Image Anal. and Signal Proc. EVOIASP2001*, Lake Como, Italy, Springer Verlag, LNCS 2038, 2001.

- [Li95] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, 1995.
- [Li00] X. Li and M. Orchard, Spatially adaptive denoising under overcomplete expansion, in Proc. IEEE Internat. Conf. on Image Proc. ICIP, Vancouver, Canada, Sept. 2000.
- [Lillesand94] Lillesand, T.M. and Kiefer, R.W., *Remote Sensing and Image Interpretation*, John Wiley, New York, 1994.
- [Liu00] J. Liu and P. Moulin, “Analysis of interscale and intrascale dependencies between image wavelet coefficients”, in Proc. Int. Conf. on Image Proc., ICIP, Vancouver, Canada, Sep. 2000.
- [Lopes90] A. Lopes, R. Touzi and E. Nezry, “Structure Detection and Statistical Adaptive Speckle Filtering in SAR Images”, *Int. J. Remote Sensing*, 14(9):1735–1758, 1993.
- [Lopes90a] A. Lopes, R. Touzi and E. Nezry, “Adaptive speckle filters and Scene heterogeneity”, *IEEE Transaction on Geosc. and Remote Sens.*, 28(6):992–1000, Nov 1990.
- [Luettgen93] M. R. Luettgen, W. C. Karl, A. S. Willsky, and R. R. Tenney, “Multiscale Representations of Markov Random Fields,” *IEEE Trans. Signal Proc.*, 41(12):3377–3396, 1993.
- [MacKay99] D.J.C. MacKay, “Introduction to Monte Carlo methods,” in *Learning in Graphical Models*, M. I. Jordan, editor, MIT Press, 1999.
- [Macovski96] A. Macovski, “Noise in MRI,” *Magn. Reson. Med.*, 36:494–497, 1996.
- [Malfait97] M. Malfait, D. Roose, “Wavelet-based image denoising using a markov random field a priori model,” *IEEE Trans. Image Proc.*, 6(4):549–565, Apr 1997.
- [Malfait95] M. Malfait, *Stochastic Sampling and Wavelets for Bayesian Image Analysis*, PhD thesis, Department of Computer Science, K.U.Leuven, Belgium, 1995.
- [Mallat89a] S. Mallat, “Multifrequency channel decomposition of images and wavelet models,” *IEEE Trans. Acoust., Speech, Signal Proc.*, 37(12):2091–2110, Dec 1989.

- [Mallat89b] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 11(7):674–693, July 1989.
- [Mallat92a] S. Mallat, W. L. Hwang, “Singularity Detection and Processing with Wavelets,” *IEEE Trans. Inform. Theory*, 38(2):617–643, Mar 1992.
- [Mallat92b] Mallat, S. Zhong, “Characterization of Signals from Multiscale Edges,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 14(7):710–732, July 1992.
- [Mallat96] S. Mallat, “Wavelets for Vision,” *Proc. IEEE*, 84(4):604–614, Apr 1996.
- [Mallat98] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.
- [Malyshev91] V. A. Malyshev and R. A. Minlos, *Gibbs Random Fields. Method of Cluster Expansions*, Kluwer, Academic Publishers, 1991.
- [Marr82] D. Marr, *Vision*, W. H. Freeman and Co., New York, 1982.
- [McAulay80] R. J. McAulay and M. L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” *IEEE Trans. Acoust. Speech and Signal Proc.*, 28(2):137–144, Apr 1980.
- [McVeigh85] E. R. McVeigh, R. M. Henkelman, and M. J. Bronskill, “Noise and filtration in magnetic resonance imaging,” *Med. Phys.*, 3:604–618, 1985.
- [Meyer86] Y. Meyer, “Principe d’incertitude, bases hilbertiennes et algèbres d’opérateurs,” *Bourbaki seminar*, no. 662, 1985-1986.
- [Meyer93] Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, 1993.
- [Middleton68] D. Middleton and R. Esposito, “Simultaneous optimum detection and estimation of signals in noise,” *IEEE Trans. Inform. Theory*, 14(3):434–443, May 1968.
- [Mihaeck99] M.K. Mihaeck, I. Kozintsev, K. Ramchandran, and P. Moulin, “Low-complexity image denoising based on statistical modeling of wavelet coefficients,” *IEEE Signal Proc. Lett.*, 6(12):300–303, Dec 1999.

- [Milisavljevic99a] N. Milisavljevic, "Comparison of three methods for shape recognition in the case of mine detection," *Pattern Recognition Lett.*, 20:1079–1083, 1999.
- [Milisavljevic99b] N. Milisavljevic, B. Scheers, Y. Yvinec, and M. Achery, "3D Visualization of Data Acquired by Laboratory UWB GPR in the Scope of Mine Detection," in *Proc. of Mine Identification Novelties Euroconference MINE'99*, pp. 149–154, Florence, Italy, 1999.
- [Milisavljevic99c] N. Milisavljevic, I. Bloch, and M. Achery, "Modeling, combining and discounting mine detection sensors within Dempster-Shafer framework," in *Proc. SPIE Conf. on Detection Technologies for Mines and Minelike Targets*, vol. 4038, pp. 1461–1472, Orlando, USA, 2000.
- [Milisavljevic01] N. Milisavljevic, *Analysis and Fusion Using Belief Function Theory of Multisensor Data for Close-Range Humanitarian Mine Detection*, Ph.D. thesis, ENST, Paris, 2001.
- [Moulin99] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors," *IEEE Trans. Inform. Theory*, 45:909–919, Apr 1999.
- [Nason94] G. P. Nason, "Wavelet regression by cross-validation," Technical Report 447, Department of Statistics, Stanford, 1994.
- [Nason96] G. Nason, "Wavelet shrinkage using cross-validation," *J. Roy. Stat. Soc. B.*, 58:463–479, 1996.
- [Neelamani99] R. Neelamani, H. Choi, and R. G. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems," submitted to *IEEE Trans. Image Proc.*, available at <http://www-dsp.rice.edu/software/WaRD/>
- [Neelamani99a] R. Neelamani, "Software for Image Deconvolution using Wavelet-based Regularized Deconvolution (WaRD)", available at <http://www-dsp.rice.edu/software/WaRD/>
- [Nicoud97] J.D. Nicoud, "Validation of anti-personnel mine sensors," in *Sustainable Humanitarian Demining: Trends, Techniques, and Technologies*, E. Pass, editor, Mid Valley Press, Verona, Virginia, 1997, pp. 206–212.

- [Nielsen96] N. H. Nielsen and M. V. Wickerhauser, “Wavelets and time-frequency analysis,” *Proc. IEEE*, 84(4):523–541, Apr 1996.
- [Novakoff92] A. K. Novakoff, “FAA bulk technology overwiev for explosive detection,” in Proc. SPIE, vol. 1824, pp. 2–12, 1992.
- [Nowak97] R. D. Nowak, “Optimal signal estimation using cross-validation,” *IEEE Signal Proc. Lett.*, 4:23–25, Jan 1997.
- [Nowak99] R. D. Nowak, “Wavelet-based Rician noise removal for magnetic imaging,” *IEEE Trans. Image Proc.*, 8(10):1408–1419, Oct 1999.
- [Novak99a] R. D. Nowak, R. G. Baranuik, “Wavelet-domain filtering for photon imaging systems,” *IEEE Trans. Image Proc.*, 8(5):666:678, May 1999.
- [Oddy83] C.J. Oddy and A. J. Rye, “Segmentation of SAR images using a local similarity rule,” *Pattern Recognition Lett.*, 1:1457–1462, 1987.
- [Odegard95] J. E. Odegard, H. Guo, M. Lang, C. S. Burrus, R. O. Wells, “Wavelet based SAR speckle reduction and image compression”, in *Proc. SPIE Symposium on OE/Aerospace Sensing and Dual Use Photonics*, pp. 17–21 April 1995, Orlando, Florida.
- [Papoulis84] A. Papoulis, *Probability, random variables, and stochastic processes*, McGraw-Hill, 1984.
- [Pennec00] E. Le Pennec and S. Mallat, “Image compression with geometrical wavelets,” in *Proc. IEEE International Conf. on Image Proc. ICIP*, Vancouver, BC, Canada, 2000.
- [Pentland84] A. P. Pentland, “Fractal-based description of natural scenes,” *IEEE Trans. Pattern Anal. and Machine Intel.*, 6:661–674, Nov 1984.
- [Peters94] L. Peters Jr, J.L. Daniels, and J.D Young, “Ground penetrating radar as a subsurface environmental sensing tool,” *Proc. IEEE*, 8(12):1802–1822, Dec 1994.
- [Pizurica98a] A. Pizurica, W. Philips, I. Lemahieu and M. Achery, “The application of a nonlinear multiscale method to GPR image processing,” in *Proc. International Conf. on Signal and Image Processing SIP98*, pp. 332–335, Las Vegas, Nevada, USA, Oct. 1998.

- [Pizurica98b] A. Pizurica, "Multiresolution techniques for image restoration in mine detection problems," in *Proc. Internat. Symp. on Mobile, Climbing and Walking Robots CLAWAR'98*, pp. 225–229, Brussels, Belgium, Nov. 1998.
- [Pizurica99a] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, "Speckle noise reduction in GPR images," in *Proc. Internat. Symp. on Pattern Recognition, "In Memoriam Pierre Devijver"*, pp. 162–165, Brussels, Belgium, Feb. 1999.
- [Pizurica99b] A. Pizurica and W. Philips, "Enhancement of GPR in infrared images of landmines," in *Proc. Internat. Symp. on Humanitarian Demining, HUDEM*, Brussels, Belgium, Apr. 1999.
- [Pizurica99c] A. Pizurica, W. Philips, I. Lemahieu, M. Acheroy, "Image de-noising in the wavelet domain using prior spatial constraints," *Proc. IEE Conf. on Image Proc. and its Applications IPA*, pp. 216–219, Manchester, UK, 1999.
- [Pizurica99d] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, "Image denoising using a multiscale nonlinear filtering technique," in *Proc. Internat. Symp. on Intelligent Vision Systems ACIVS*, pp. 9–13, Baden-Baden, Germany, 5-6 Aug 1999.
- [Pizurica99e] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, "Enhancement of noisy images using a multiscale edge reconstruction technique," in *Proc. Internat. Conf. on Signal and Image Processing SIP*, pp. 144–148, Nassau, Bahamas, Oct. 1999.
- [Pizurica99f] A. Pizurica, "An application of a wavelet-based denoising method to infrared images of landmines," in *Proc. European Workshop PHOTOMEC'99 - ETE'99*, pp. 75-79, Liege, Belgium, Nov. 1999.
- [Pizurica00a] A. Pizurica, N. E. C. Verhoest, W. Philips, and F. P. De Troch, "Detecting variable source areas from temporal radar imagery using advanced image enhancement techniques," in *Proc. IEEE International Geosc. and Remote Sens. Symp. IGARSS*, Honolulu, Hawaii, July 2000.
- [Pizurica00b] A. Pizurica, N. E. C. Verhoest, W. Philips, F. P. De Troch, I. Duskunovic and M. Acheroy, "An Application of a wavelet-based denoising method to temporal radar imagery," in *Proc. Internat.*

- Symp. on Intelligent Vision Systems ACIVS*, pp. 36-40, Baden-Baden, Germany, Aug 2000.
- [Pizurica00c] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, “A wavelet-based image denoising technique using spatial priors,” in *Proc. IEEE Internat. Conf. on Image Proc. ICIP*, pp. 296–299, Vancouver, BC, Canada, 2000.
- [Pizurica00d] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, “An image denoising technique using wavelets and spatial priors,” in *Proc. 2nd IEEE Benelux Signal Processing Symposium SPS*, Hilvarenbeek, The Netherlands, 2000.
- [Pizurica00e] A. Pizurica, “Wavelets in image denoising,” Invited presentation, *Wavelet Seminar*, sponsored by the Dutch science foundation NWO, University of Groningen, Groningen, The Netherlands, 17 Nov 2000.
- [Pizurica01a] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, “De-speckling SAR images using wavelets and a new class of adaptive shrinkage estimators,” in *Proc. IEEE Internat. Conf. on Image Processing ICIP*, Thessaloniki, Greece, Oct. 2001.
- [Pizurica01b] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, “Multiresolution image restoration: a filtering technique adapting to unknown type of noise,” in *Proc. PRORISC IEEE Benelux Workshop on Circuits, Systems and Signal Processing*, STW Technology Foundation, Veldhoven, the Netherlands, 28-30 Nov 2001.
- [Pizurica01c] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, “The application of Markov random field models to wavelet-based image denoising,” in *Imaging and Vision Systems: Theory, Assessment and Applications*, J. Blanc-Talon and D. Popescu, editors, NOVA Science Books, Huntington, USA, 2001.
- [Pizurica02a] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, “A joint inter- and intrascale statistical model for wavelet based Bayesian image denoising ”, *IEEE Trans. Image Proc.*, May 2002, in print.
- [Pizurica02b] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, “A wavelet domain joint detection and estimation method for image denoising,” submitted to *Signal processing*, Oct 2001, available at <http://telin.rug.ac.be/~sanja>

- [Pizurica02c] A. Pizurica, W. Philips, I. Lemahieu, and M. Achery, “A versatile wavelet domain noise filtration technique for medical imaging,” submitted to *IEEE Trans. Medical Imaging*, Nov 2001, available at <http://telin.rug.ac.be/~sanja>
- [Portilla01] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain,” in *Proc. IEEE Internat. Conf. on Image Proc. ICIP*, Thessaloniki, Greece, October 7-10, 2001.
- [Ramchandran96] K. Ramchandran, M. Vetterli, and C. Herley, “Wavelets, subband coding, and best bases,” *Proc. IEEE*, 84(4):541–560, Apr 1996.
- [Romberg99] J. K. Romberg, H. Choi, and R. G. Baraniuk, “Bayesian tree structured image modeling using wavelet-domain hidden Markov model,” in *Proc. SPIE*, vol. 3816, pp. 31-44, Denver, CO, July 1999.
- [Romberg00] J. Romberg, H. Choi, R. Baraniuk, and N. Kingsbury, “Multiscale classification using complex wavelets and hidden Markov tree models,” in *Proc. IEEE Internat. Conf. on Image Proc. ICIP*, Vancouver, BC, Canada, 2000.
- [Romberg01] J. K. Romberg, H. Choi, and R. G. Baraniuk, “Bayesian tree structured image modeling using wavelet-domain hidden Markov models,” *IEEE Trans. Image Proc.*, 10(7): 1056–1068, July 2001.
- [Rooms01] F. Rooms, A. Pizurica, W. Philips, “Estimating image blur in the wavelet domain,” in *Proc. PRORISC IEEE Benelux Workshop on Circuits, Systems and Signal Processing*, STW Technology Foundation, Veldhoven, the Netherlands, 28-30 Nov, 2000.
- [Rosenfeld70] A. Rosenfeld, “A nonlinear edge detection technique,” *Proc. IEEE*, 58:814-816, May 1970.
- [Rosenfeld71] A. Rosenfeld and M. Thurston, “Edge and curve detection for visual scene analysis,” *IEEE Trans. Comput.*, 20(5):562–569, May 1971.
- [Rouhi97] A.M. Rouhi, “Land mines: horrors begging for solutions,” *Chem. & Eng. News*, 75(10):14–22, Mar 1997. Available at <http://pubs.acs.org/hotartcl/cenear/970310/land.html>

- [Ruggeri99] F. Ruggeri and B. Vidakovic, “A Bayesian decision theoretic approach to wavelet thresholding,” *Statist. Sinica*, 9(1):183–197, 1999.
- [Sattar97] F. Sattar, L. Floreby, G. Salomonsson, B. Lvstrm, “Image enhancement based on a nonlinear multiscale method,” *IEEE Trans. Image Proc.*, 6(6):888–895, June 1997.
- [Schachne98] M. Schachne, L. Van Kempen, D. Milojevic, H. Sahli, Ph. Van Ham, M. Acheroy, and J. Cornelis, “Mine detection by means of dynamic thermography,” in *Proc. IEE Internat. Conf. on the Detection of Abandoned Land Mines*, pp. 124–127, Edinburgh, UK, 1998.
- [Scheers98] B. Scheers, M. Piette, and A. Vander Vorst, “The detection of AP mines using UWB GPR,” in *Proc. IEE Internat. Conf. on the Detection of Abandoned Land Mines*, pp. 50–54, Edinburgh, UK, 1998.
- [Scheers00] B. Scheers, M. Piette, M. Acheroy, and A. Vander Vorst, “A laboratory UWB GPR system for landmine detection,” in *Proc. Internat. Conf. on Ground-penetrating radar GPR2000*, Sydney, Australia, June 2001.
- [Schroder95] P. Schröder, “Spherical wavelets: texture processing,” in *Rendering Techniques '95*, P. Hanrahan and W. Purgathofer, editors, Springer-Verlag, Vienna/New York, 1995.
- [Schroder96] P. Schröder, “Wavelets in computer graphics,” *Proc. IEEE*, 84(4):615–625, Apr 1996.
- [Shapiro93] J. M. Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans. Signal Proc.*, 41:3445–3462, Dec 1993.
- [ShiFung94] Z. Shi and K. B. Fung, “A comparison of digital speckle filters,” in *Proc. IEEE Internat. Geosc. and Remote Sens. Symp. IGARSS*, pp. 2129–2133, Pasadena, USA, 1994.
- [Simoncelli96] E. P. Simoncelli and E. H. Adelson, “Noise removal via Bayesian wavelet coring,” in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, pp. 379–382, Lausanne, Switzerland, 1996.

- [Simoncelli99] E. P. Simoncelli, “Modeling the joint statistics of image in the wavelet domain,” in *Proc. SPIE Conf. on Wavelet Applications in Signal and Image Processing VII*, vol. 3813, Denver, CO, 1999.
- [Sommer81] F.G. Sommer, L.F. Joynt, B.A. Carroll, and A. Macowski, “Ultrasonic characterization of abdominal tissues via digital analysis of backscattered waveforms,” *Radiology*, 141:811-817, Dec 1981.
- [Starck98] J.L. Starck, “Image Processing and Data Analysis: the Multiresolution Approach”, Cambridge University Press, 1998.
- [Stein81] C. Stein, “Estimation of the mean of a multivariate normal distribution,” *Annals of Statistics*, 9(6):1135–1151, 1981.
- [Stippel00] G. Stippel, I. Duskunovic, A. Pizurica, F. Rooms, W. Philips, and I. Lemahieu, “A speckle suppression method for medical ultrasound images based on local statistics,” in *Proc. STW/ProRISC 2000*, pp. 525–530, Dec 2000.
- [Storme98] M. Storme, I. Huynen, and A. Vander Vorst, “Characterization of wet soils from 2 to 18 GHz, experimental results,” in *Proc. Internat. Symp. on Climbing and Walking Robots CLAWAR98*, pp. 237–239, Brussels, Belgium, Nov 1998.
- [Strang89] G. Strang “Wavelets and dilation equations,” *SIAM Review.*, 31:613–627, 1989.
- [Strang94] G. Strang, “Wavelets,” *American Scientist.*, 82:250–255, Apr 1994.
- [Strang95] G. Strang, and V. Strela, “Short wavelets and matrix dilation equations,” *IEEE Trans. Signal Proc.*, 43:108–115, 1995.
- [Strang96] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, USA, 1996.
- [Strela00] V. Strela, J. Portilla, and E. Simoncelli, “Image denoising using a local Gaussian scale mixture model in the wavelet domain,” in *Proc. SPIE, 45th Annual Meeting*, San Diego, July 2000.
- [Strickland93] R.N. Strickland and D.K. Cheng, “Adaptable edge quality metric,” *Optical Engineering*, 32(5):944-951, May 1993.

- [Sweldons94] W. Sweldons, “The lifting scheme: a custom-design construction of biorthogonal wavelets ,” Tech. Report 1994:7, Indust. Math. Initiative, Dept. Mathematics, University S. Carolina, 1994.
- [Sweldons95] W. Sweldons, “The lifting scheme: a construction of second generation wavelets,” Tech. Report 1995:6, Indust. Math. Initiative, Dept. Mathematics, University S. Carolina, 1995.
- [Sweldons96] W. Sweldons, “Wavelets: what next,” *Proc. IEEE*, 84(4):680–686, Apr 1996.
- [Taswell98] C. Taswell, “The what, how and why of wavelet shrinkage denoising,” Computational Toolsmiths, Stanford, CA 94309-9925, Technical Report CT-1998-09, 1998.
- [Ulaby86a] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing, Active and Passive, Volume III: from Theory to Applications*, Artech House, 1986.
- [Ulaby86b] F. T. Ulaby, F. Kouyate, B. Brisco, and T.H.L. Williams, “Textural information in SAR images,” *IEEE Trans. Geosci. Remote Sens.*, 24:235–245, Mar 1986.
- [Unser96] M. Unser, and A. Aldroubi, “A review of wavelets in biomedical applications ,” *Proc. IEEE*, 84(4):626–634, Apr 1996.
- [VanKempen98] L. Van Kempen, H. Sahli, E. Nyssen, and J. Cornelis, “Signal processing and pattern recognition methods for AP mine detection and identification,” in *Proc. IEE Internat. Conf. on the Detection of Abandoned Landmines*, pp. 81–85, Edinburgh, UK, 12-14 Oct 1998.
- [VanKempen99] L. van Kempen and H. Sahli, “Ground penetrating radar data processing: a selective survey of the state of the art literature,” Technical Report IRIS-TR-0060, Vrije Universiteit Brussel, Brussels, Belgium, 1999.
- [VanMeirvenne00] M. Van Meirvenne and P. Goovaerts, “Accounting for spatial dependence in the processing of multitemporal SAR images using factorial kriging analysis,” *Int. J. of Remote Sensing*, (submitted)
- [VanTrees68] H.L. Van Trees, *Detection, Estimation and Modulation Theory*, John Wiley, New York, 1968.

- [Verhoest98] Verhoest, N.E.C. and Troch, P.A. and Paniconi, C. and De Troch, F.P., “Mapping basin scale variable source areas from multi-temporal remotely sensed observations of soil moisture behaviour”, *IEEE Trans. Geosc. Remote Sens.*, vol. 34:3235–3244, May 1998.
- [Verhoest00a] N. E. C. Verhoest, A. Pizurica, W. Philips, and F. P. De Troch, “The application of wavelet-based filtering techniques for retrieving bio-physical parameters from multi-temporal ERS images,” in *Proc. ERS-ENVISAT Symp. “Looking down to Earth in the New Millennium”*, Gothenburg, 16-20 Oct. 2000.
- [Verhoest00b] N.E.C. Verhoest, *Retrieval of Soil Moisture Information from Synthetic Aperture Radar Data at the Point and at the Catchment Scale*, Ph.D. Thesis, Ghent University, Ghent, Belgium, 2000.
- [Verlinde01] P. Verlinde, M. Achteroy, and Y. Baudoin, “The Belgian humanitarian demining project and the European research context,” *Journal of Mechanical and Environmental Engineering*, 46(2):69–98, 2001.
- [Vetterli95] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice-Hall, 1995.
- [Vidakovic94] B. Vidakovic, “Nonlinear wavelet shrinkage with bayes rules and bayes factors,” *J. of the American Statistical Association*, 93:173–179, 1998.
- [Vidakovic98] B. Vidakovic, “Wavelet-based nonparametric Bayes methods”, in *Practical Nonparametric and Semiparametric Bayesian Statistics*, Lecture Notes in Statistics 133, Dey, Mller and Sinha, editors, Springer-Verlag, pp. 133–155, 1998.
- [Xu94] Y. Xu, J. B. Weaver, D. M. Healy, J. Lu, “Wavelet transform domain filters: a spatially selective noise filtration technique,” *IEEE Trans. Image Proc.* 3:747–758, Nov 1994.
- [Xue01] J.-H. Xue, W. Philips, A. Pizurica, and I. Lemahieu, “A novel method for adaptive enhancement and unsupervised segmentation of MRI brain images”, in *Proc. IEEE Internat. Conf. on Acoust., Speech and Signal Proc. ICASSP*, Salt Lake City, Utah, USA, May 2001.
- [Wagner83] R. F. Wagner, “Statistics of speckle in ultrasound B-scans,” *IEEE Trans. Sonics and Ultrasonics*, 30(3):156–163, May 1983.

- [Walessa00] M. Walessa and M. Datcu, "Model-based despeckling and information extraction from SAR Images," *IEEE Trans. Geosc. and Remote Sensing*, 38(5):2258-2269, Sep. 2000.
- [Weaver92] J. Weaver, Y. Xu, D. Healy, and J. Driscoll, "Filtering MR images in the wavelet transform domain," *Magn. Reson. Med.*, 21:288-295, 1991.
- [Weetjens99] B. Weetjens, C. Cox, and H. Van Krunkelsven, "Explosive trace vapor detection by rodents," in *Proc. Internat. Symp. on Humanitarian Mine Detection HUDEM*, Brussels, Belgium, 1999.
- [Weyrich98] N. Weyrich and G. T. Warhola, "Wavelet shrinkage and generalized cross validation for image denoising," *IEEE Trans. Image Proc.*, 7:82–90, Jan 1998.
- [Wickerhauser94] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, IEEE Press, Piscataway, NJ, 1994.
- [Winkler95] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer-Verlag (1995).
- [Witkin83] A. Witkin, "Scale space filtering," in *Proc. Int. Joint. Conf. Artificial Intell.*, Espoo, Finland, June 1983.
- [Witten98] T.R. Witten, "Present state-of-the-art in ground penetrating radars for mine detection," in *Proc. SPIE Conf. on Detection Technologies for Mines and Minelike Targets*, vol. 3392, pp. 576–585, Orlando, USA, 1998.
- [Wornell96] G. W. Wornel, "Emerging applications of multirate signal processing and wavelets in digital communications," *Proc. IEEE*, 84(4):586–603, Apr 1996.
- [Yoo99] Y. Yoo, A. Ortega, and B. Yu, "Image subband coding using context-based classification and adaptive quantization," *IEEE Trans. Image Proc.*, 8:1702–1715, Dec 1999.
- [Youla82] D. C. Youla and H. Webb, "Image restoration by the method of convex projections," *IEEE Trans. Med. Imag.*, vol. 1:81–101, Oct 1982.