



UNIVERSITY OF CAGLIARI

Department of Electrical and Electronic Engineering

**WAVELET BASED MULTIMEDIA
COMMUNICATIONS**

Ph.D. Thesis

Maurizio Murrone

Advisor: **Prof. Daniele D. Giusto**

December 2000

PREFACE

The spreading of the telecommunications market during the last decades has caused an increasing interest in the signal-processing field. In particular, the problem of signal storing and transmission pushes the birth of more and more sophisticated compression techniques and the continuous refinement of the already existing ones. In the mean time, the diffusion of the Internet causes an increasing development of refined image and video zooming methods, in order to guarantee a good quality in information transmission and to allow fast browsing and progressive image and video delivery.

This thesis focuses on the use of the wavelet representation of signals in several applications, from still image compression to data transmission over real channels to video zooming. The purpose of the thesis is to show the potentiality of this representation, which, born and formalized in the late '80s, has rightly gained the interest of the signal processing community, till to be inserted in the new standards for visual communications, i.e. JPEG2000 and MPEG-4.

One of the main reason of the increasing popularity of this transform is its amazing similarity to the human visual system (HVS): the representation in different frequency bands at different scales (with different levels of details) is, as a matter of fact, peculiar to the behavior of the human eye. In fact, the wavelet transform is an intermediate representation between the spatial and the frequency representations and contains both their advantages.

The rapid growth and development of wavelet representations have been caused and supported by several independent studies and discoveries. The idea of multiresolution was born out of the work done by Burt and Adelson [1] images obtained from any initial image by scaling it with a certain operator are related together by very simple causality relations. A disadvantage of such a representation is that it is not possible to know if a similitude among the details of an image at different resolution levels is a characteristic of the image itself or if it is due to the redundancy of the representation. Moreover, this representation does not introduce any selectivity in the spatial orientation during the decomposition process. The wavelet decomposition allows instead the separation of the informative content of the image itself along the different spatial orientations and at different resolutions. In [2] the possibility that the effects of the quantization noise could be limited by signal coding in the wavelet domain was discussed. Later on, in [3] the use of wavelets for image encoding was tested and it was shown how this approach could allow to take into account in a clever manner the characteristics of the HVS in order to render the quantization noise less annoying. Another important discovery for the development of these techniques is the one of the quadrature mirror filters (QMFs) [4], which allow signal decomposition and reconstruction without aliasing even if the filters are not ideal. The basic idea in the construction of the QMFs is the use of partially overlapping analysis and synthesis filters designed so that the aliasing is cancelled during reconstruction.

The wavelet representation is strictly related to the idea of multiresolution: a signal at a given resolution (or scale) can be represented by its approximation at lower resolution and by the difference between this latter approximation and the initial one. Such difference is contained in the detail signal. Repeating the process till a fixed resolution, the wavelet representation of a signal is obtained as composed by its approximation at the lowest resolution and by the set of all the extracted detail signals. The same concept is used for example in pattern recognition: the vision system tries to classify an object by an approximation; if the classification fails, more details are added so that a more accurate vision of the relevant object

can be obtained. This procedure can be repeated till object recognition [5]. In mathematical terms, a multiresolution is composed by the projections of a function into different subspaces and each projection gives more and more precise details about the function.

Multiresolution techniques are particularly suitable for images, since concepts as resolution and scale are rather intuitive. An important characteristic of these techniques is their property of successive approximation: as frequency increases, higher resolution images are obtained, as happens in the HVS. The decomposition into subbands itself, each of which relevant to a certain frequency band, corresponds to the classification of information more or less important for the HVS. Thus some subbands should be transmitted more accurately than others should and this permits the use of very efficient bit allocation techniques without compromising the quality of the reconstructed signal. Moreover, it is well known that the sensitivity of the HVS to the luminance decreases in the regions of the image with high-contrast details. In these regions, large quantization errors cannot be perceived by the HVS. The human eye, as a matter of fact, behaves as a bidimensional lowpass filter and the highest spatial frequencies are non-perceptible and thus do not need to be encoded.

The statistical properties of the wavelet coefficients, besides, are more useful than the ones of the original signal, allowing even a hard quantization without decreasing significantly the quality of the signal itself. The quantization in the frequency domain, as a matter of fact, permits the masking of the quantization errors during the reconstruction.

The wavelet representation of a signal is also the core for some advanced digital transmission schemes. In particular, both linear and quadrature Fractal Modulation algorithm have been investigated in the past.

The wavelet transform and some of its application in the signal-processing field are the subject of this thesis. It is divided into the following sections:

The wavelet transform is introduced in the first chapter as a mathematical tool to represent a signal into a different domain than the original one. The wavelet representation of monodimensional signals is examined in paragraph 1.2, then the extension to the bidimensional case is taken into consideration in paragraph 1.3 and finally the case of multidimensional signals is described in paragraph 1.4.

The use of the wavelet representation for data compression purposes is examined in the second chapter as regards pictorial data (paragraph 2.2) and remote sensing data (paragraph 2.3). A clever exploitation of the statistical properties of the wavelet coefficients allows a hard adaptive quantization that permits to obtain high compression levels without perceptual or significant annoying artifacts in the reconstructed images.

In chapter three, we present a performance analysis of fractal modulation transmission over a AWGN fast-fading channel. A quadrature transmission scheme is simulated and compared with frequently used transmission systems achieving better results in terms of error robustness and low complexity.

The fourth chapter deals with wavelets and data zooming: data zooming is performed exploiting iterated function systems (I.F.S.), both in the spatial and in the temporal domain, in order to reconstruct the entire sequence by starting from some selected Key-Frames. The use of wavelet-based analysis allows a large decrease of time processing, as well as the opportune post-processing based on overlapped range blocks coding reduces blockness effects due to the fractal coding scheme.

Finally some conclusions and further developments on the whole work are drawn.

Acknowledgements

I would like to acknowledge everyone in some way contributed to the work under this thesis. It would be tiresome to list everyone, but I wish to thank in particular some people, who gave really a strong contribution for this work, in one way or another.

First of all, my advisor, Prof. Daniele Giusto, who supported and assisted me with precious advice during all these years, Prof. Gianni Vernazza, who drove me at the beginning of this work and Prof. Francesco De Natale, who helped with some good and precious tips. Particular thanks to Prof. Maria Petrou, who supervised and assisted me during my early work at CVSSP, University of Surrey, and to Prof. Yao Wang, Polytechnic of Brooklyn, NY, who helped in the final developing of the work.

Many thanks to my colleagues, Gigi, Cristina, Marcella, Cristian, with whom I could enjoy most of my work and spend beautiful working days, and to all the staff of the DIEE the CVSSP, and the Polytecnic.

Last but not least, special thanks to my family and Pavla, who gave me support, encouragement and much more.

I apologize with everyone who has not been listed here but gave me even a little encouragement or advice; to everyone my sincere thanks.

Maurizio

CONTENTS

Preface	I
Contents	V
List of acronyms	1
References	2
Chapter 1	3
Mathematical overview	3
1.1 Introduction	3
1.2 Orthogonal wavelet representations of monodimensional signals	4
1.2.1 Multiresolution approximations of $L^2(\mathbf{R})$	4
1.2.2 Realization of a multiresolution representation	6
1.2.3 The detail signal	8
1.2.4 Realization of an orthogonal wavelet representation	10
1.2.5 Signal reconstruction from an orthogonal wavelet representation	12
1.3 Orthogonal wavelet representations of bidimensional signals	13
1.3.1 Multiresolution approximations of $L^2(\mathbf{R}^2)$	13
1.3.2 The detail signals	14
1.3.3 Realization of an orthogonal wavelet representation	16
1.3.4 Signal reconstruction from an orthogonal wavelet representation	16
1.4 Extension of the orthogonal wavelet representation to multidimensional signals ...	19
1.4.1 Multiresolution approximations of $L^2(\mathbf{R}^n)$	19
1.4.2 The detail signals	19
1.4.3 Realization of an orthogonal wavelet representation	21
1.4.4 Signal reconstruction from an orthogonal wavelet representation	22

Chapter 2	24
Wavelets and data compression	24
2.1 Introduction	24
2.2 Pictorial data	25
2.2.1 Introduction	25
2.2.2 The encoding of the low-pass image	25
2.2.2.1 DPCM encoding	25
2.2.2.2 Bidimensional DPCM encoding.....	26
2.2.2.3 DPCM low-pass image encoding	27
2.2.3 Classification and encoding of the <i>active zones</i>	27
2.2.3.1 Classification of the <i>active zones</i>	28
2.2.3.2 Encoding of the <i>active zones</i>	30
2.2.4 Encoding of the wavelet coefficients.....	34
2.2.4.1 VQ for wavelet coefficients encoding	34
2.2.4.2 Codebook generation	35
2.2.4.3 Adaptive encoding of the wavelet coefficients	37
2.2.5 Results.....	38
2.2.5.1 Identification and use of a unique mask.....	38
2.2.5.2 Identification and use of a mask for each orientation.....	44
2.2.5.3 Comparisons between AZWVQ with one mask and with three masks.....	50
2.2.6 Conclusions	51
2.3 Remote sensing data.....	53
2.3.1 Introduction	53
2.3.2 The Free-Angle concept.....	55
2.3.3 Filter Bank 1D wavelet implementation.....	56
2.3.4 Outline of the encoding algorithm.....	57
2.3.5 Active zone classification	59
2.3.6 Mask extraction	60
2.3.7 Wavelet coefficient encoding.....	60
2.3.8 Experimental results	60
2.3.9 Conclusions	61

References	62
Chapter 3.....	64
Wavelets and data zooming	64
5.1 Introduction	64
3.2 The fractal coding	65
3.3 Classical fractal decoder.....	68
3.3.1 Variants at the classical fractal decoder.....	69
3.3.2 Extension to 3D fractal encoders.....	73
3.4 The wavelet transform.....	76
3.4.1 Statistical properties of the wavelet coefficients.....	77
3.4.1 The classification of the active zones	78
3.4.3. The extraction of the mask.....	79
3.4.4 The encoding of the wavelet coefficients	80
3.4.5 The wavelet representation for video sequences.....	80
3.5 Temporal subsampling with <i>key frames</i> extraction and use.....	82
3.6 Selection of the active scenario	83
3.7 Tests and results.....	87
3.8 Conclusions and further developments	89
Chapter 4.....	92
4.1 Introduction	92
4.2 Homogeneous signals and wavelet representation	93
4.2.1 Iterative synthesis of homogeneous signals	95
4.3 Fractal modulation transmitter and receiver.....	96
4.4 Experiments set-up.....	98
4.4.1 Channel Model	99
4.4.2 Quadrature scheme transmitter/receiver	99
4.5 Performance analysis	101
4.6 Conclusion.....	106
Appendix A	109
Some wavelet filters	109
A.1 Haar filters	109
A.2 Daubechies filters.....	109
A.3 Lemariè-Battle filters	110

Appendix B.....	113
Statistical properties of the wavelet transform.....	113

LIST OF ACRONYMS

AC	alternating current
AWRI	adaptive wavelet coding of residual information
AZWVQ	active zones wavelet coefficients vector quantization
CR	compression ratio
DC	direct current
DCT	discrete cosine transform
DPCM	differential pulse code modulation
DST	discrete sine transform
HVS	human visual system
IFS	iterated function system
JPEG	joint photograph expert group
LBG	Linde Buzo and Gray
MAP	maximum a posteriori
MPEG	moving picture experts group
MSE	mean square error
PDF	probability density function
PSNR	peak signal-to-noise ratio
QMF	quadrature mirror filter
ROI	region of interest
RS	remote sensing
VLC	variable length code
VQ	vector quantization
WVQ	wavelet coefficients vector quantization

References

- [1] P.J. Burt, E.H. Adelson, "The Laplacian pyramid as compact image code," IEEE Trans. Comm., Vol. 31, pp. 532-540, 1983.
- [2] C. Galand, "Codage en sous-bandes: théorie et applications à la compression numérique du signal de parole," Thesis, University of Nice, France, 1983.
- [3] E.H. Adelson, R. Hingorani, E. Simoncelli, "Orthogonal pyramid transforms for image coding," Visual Communications and Image Processing II, Cambridge, MA, Proc. SPIE, Vol. 845, pp. 50-58, 1987.
- [4] G. Strang, T. Nguyen, "Wavelets and filter banks," Wellesley-Cambridge Press, MA, USA, 1996.
- [5] H.J. Barnard, "Image and video coding using a wavelet decomposition," Ph.D. Thesis, Delft University of Technology, The Netherlands, 1994.

Chapter 1

MATHEMATICAL OVERVIEW

1.1 Introduction

The wavelet transform of a signal is a multiresolution orthogonal representation [1]: it is composed by the set of the signal approximations at several resolutions or, in an equivalent manner, by the approximation of the signal at a fixed resolution and the set of the detail signals. Each detail signal contains the difference of information between the approximations at two adjacent resolutions and can be extracted by decomposing the signal in an orthogonal wavelet basis. Such decomposition is obtained through a pyramidal algorithm and defines an orthogonal multiresolution representation named *wavelet representation*.

The wavelet representation of a signal is thus composed by an approximation of the signal at J resolution and by the set of the detail signals that, for each resolution $j < J$, contain the differences between the approximations of the signal at resolution j and $j+1$. Such representation has the same number of data as the original signal and it is a reversible representation: the reconstruction of the signal from its wavelet representation is done through a pyramidal algorithm.

The reversibility of the process is related to the use of a bank of quadrature mirror filters (QMFs) to perform the decomposition, since they allow the perfect reconstruction¹ of the signal without distortions nor aliasing.

The monodimensional model is easily extendible to multidimensional signals: the wavelet representation of a multidimensional signal is again composed by its approximation at resolution J and by the set of the detail signals that, for each resolution $j < J$, contain the differences between the approximations of the signal at resolution j and $j+1$. However, in the multidimensional case these signals differentiate themselves one from the other for the spatial orientation they privilege.

The separable multiresolution approximations allow the decomposition of a multidimensional signal by applying the monodimensional method along each direction with the same QMFs. From this point of view the wavelet representation of a signal can be seen as its decomposition in a set of independent frequency channels, each of which is characterized by a particular spatial orientation. The independence of the frequency channels is guaranteed by the orthogonality of the representation. Again, the wavelet representation has the same number of samples as the original signal and it is a reversible transformation: the reconstruction of the signal is obtained starting from its wavelet representation through a pyramidal algorithm.

¹ Actually, in order to obtain the perfect reconstruction of a signal, it should be necessary to represent the coefficients of its wavelet representation with an infinite precision

1.2 Orthogonal wavelet representations of monodimensional signals

1.2.1 Multiresolution approximations of $L^2(\mathbf{R})$

Let $f(x) \in L^2(\mathbf{R})$ be a monodimensional measurable and with finite energy signal and let A_{2^j} be an operator approximating the signal at resolution 2^j , characterized by the following properties:

1. A_{2^j} is a linear operator: if $A_{2^j}f(x)$ is the approximation of $f(x)$ at resolution 2^j , a successive approximation at resolution 2^j does not modify $A_{2^j}f(x)$. $A_{2^j}f(x)$ is thus a **projection** operator on the vectorial space V_{2^j} in $L^2(\mathbf{R})$, this space being the set of all the approximations of functions of $L^2(\mathbf{R})$ at resolution 2^j .
2. The function $A_{2^j}f(x)$ is, among all the functions that approximate $f(x)$ at resolution 2^j , the more similar to it, that is:

$$\|g(x) - f(x)\| \geq \|A_{2^j}f(x) - f(x)\| \quad \forall g(x) \in V_{2^j} \quad (1.1)$$

so the operator A_{2^j} is an **orthogonal projection** over the vectorial space V_{2^j} .

3. **Causality property:** the approximation of a signal at resolution 2^{j+1} contains all the information embodied in the approximation of the same signal at lower resolution. Since A_{2^j} is the orthogonal projection operator over the vectorial space V_{2^j} , this property can be expressed in this way:

$$V_{2^j} \subset V_{2^{j+1}} \quad \forall j \in \mathbf{Z} \quad (1.2)$$

4. The operation of approximation is similar at all resolutions. The spaces of the approximated functions are obtained one from the other by scaling each approximated function according to the ratio between the resolution levels, that is:

$$g(x) \in V_{2^j} \Leftrightarrow g(2x) \in V_{2^{j+1}} \quad \forall j \in \mathbf{Z} \quad (1.3)$$

5. The approximation $A_{2^j}f(x)$ is characterized by 2^j samples per length unit. If $f(x)$ is shifted by a length proportional to 2^{-j} , also its approximation at resolution 2^j is shifted of the same quantity and is characterized by the same shifted samples.

For the causality property, it is sufficient to express the 5. at resolution $j=0$:

- Approximation shifting:

$$A_1 f_k(x) = A_1 f(x - k) \quad \text{with } f_k(x) = f(x - k) \quad \forall k \in \mathbf{Z} \quad (1.4)$$

- Discrete characterization: there exists an isomorphism I from V_1 to $I^2(\mathbf{Z})$ with

$$I^2(\mathbf{Z}) = \left\{ (\alpha_i)_{i \in \mathbf{Z}} : \sum_{i=-\infty}^{+\infty} |\alpha_i|^2 < \infty \right\} \quad (1.5)$$

- Samples shifting:

$$I(A_1 f(x)) = (\alpha_i)_{i \in \mathbf{Z}} \Leftrightarrow I(A_1 f_k(x)) = (\alpha_{i-k})_{i, k \in \mathbf{Z}} \quad (1.6)$$

By approximating a signal some information about the signal itself is lost, but as resolution tends to $+\infty$ the approximated signal converges to the original one. Vice-versa, as resolution tends to 0, the approximated signal contains less and less information and converges to 0. In mathematical terms, since the signal approximated at resolution 2^j is the orthogonal projection of the original signal over the vectorial space V_{2^j} , this is equivalent to:

$$\lim_{j \rightarrow +\infty} V_{2^j} = \bigcup_{j=-\infty}^{+\infty} V_{2^j} \text{ dense in } L^2(\mathbf{R}) \quad \text{and} \quad \lim_{j \rightarrow -\infty} V_{2^j} = \bigcap_{j=-\infty}^{+\infty} V_{2^j} = 0 \quad (1.7)$$

The set of all vectorial spaces $(V_{2^j})_{j \in \mathbf{Z}}$ that satisfy the properties (1.2)-(1.7) defines a multiresolution approximation of $L^2(\mathbf{R})$.

The set of all operators A_{2^j} that satisfy properties 1.-5. gives an approximation of any function in $L^2(\mathbf{R})$ at resolution 2^j .

An orthonormal basis for the set of vectorial spaces $(V_{2^j})_{j \in \mathbf{Z}}$, that allows the representation of any function in those spaces and therefore its multiresolution approximation, is given by the following theorem.

Theorem 1.1 If $(V_{2^j})_{j \in \mathbf{Z}}$ defines an approximation of $L^2(\mathbf{R})$, then there exists a unique function $\phi(x)$ in $L^2(\mathbf{R})$, named *scaling function*, such that, being $\phi_{2^j}(x) = 2^j \phi(2^j x)$ its dilation by a factor of $(2^j)_{j \in \mathbf{Z}}$, the set $(\sqrt{2^{-j}} \phi_{2^j}(x - 2^{-j} n))_{n \in \mathbf{Z}}$ defines an orthonormal basis of V_{2^j} .

It is then possible to obtain an orthonormal basis for each vectorial space V_{2^j} by scaling the corresponding scaling function by a coefficient 2^j and shifting the resulting function by a factor proportional to 2^{-j} . The multiplicative factor is inserted for normalization reasons. The scaling function has good localization properties in the time domain as well as in the frequency domain. This permits the construction of good orthonormal bases.

Let now $f(x)$ be the signal to be approximated at resolution 2^j .

As previously explained, this is the orthogonal projection of the signal over the space V_{2^j} , and so it can be obtained by decomposing the signal through the basis provided by the previous theorem:

$$\forall f(x) \in L^2(\mathbf{R}) \quad A_{2^j} f(x) = 2^{-j} \sum_{n=-\infty}^{+\infty} \langle f(u), \phi_{2^j}(u - 2^{-j} n) \rangle \phi_{2^j}(x - 2^{-j} n) \quad (1.8)$$

The approximation of the signal at resolution 2^j is then defined by the set of scalar products:

$$A_{2^j}^d f = \langle f(u), \phi_{2^j}(u - 2^{-j} n) \rangle_{n \in \mathbf{Z}} \quad (1.9)$$

$A_{2^j}^d f$ is the discrete approximation of $f(x)$ at resolution 2^j .

Each scalar product can then be saw as a convolution product calculated in $2^{-j} n$:

$$\langle f(u), \phi_{2^j}(u - 2^{-j} n) \rangle = \int_{-\infty}^{+\infty} f(u) \phi_{2^j}(u - 2^{-j} n) \, du = (f(u) * \phi_{2^j}(-u))(2^{-j} n) \quad (1.10)$$

so $A_{2^j}^d f$ can be written as:

$$A_{2^j}^d f = \left((f(u) * \phi_{2^j}(-u)) \right)_{n \in \mathbb{Z}} (2^{-j} n) \quad (1.11)$$

Since $\phi(x)$ is a low-pass filter, $A_{2^j}^d f$ is a low-pass filtering of the signal followed by a uniform sampling with step 2^j . So, by approximating the signal at resolution 2^j , the details of the signal lower than 2^{-j} are removed and the higher frequencies are deleted.

1.2.2 Realization of a multiresolution representation

For the causality property, from the discrete approximation of a signal at a given resolution all the discrete approximations of the signal itself at lower resolutions can be obtained. Let's suppose, for normality reasons, that the maximum resolution at which a device can measure a signal is unitary. This means that all the discrete approximations $A_{2^j}^d f$ of a signal f for each $j < 0$ can be obtained from the measure of the signal itself. The following iterative algorithm can be used to calculate such approximations.

Let $(V_{2^j})_{j \in \mathbb{Z}}$ be a multiresolution representation and let $\phi(x)$ be the corresponding scaling function. Since V_{2^j} is contained in $V_{2^{j+1}}$ and the functions $(\phi_{2^j}(x - 2^{-j} n))_{n \in \mathbb{Z}}$ belong to V_{2^j} , these can be expanded through an orthonormal basis $V_{2^{j+1}}$. Since for the theorem 1.1 $(\sqrt{2^{-j-1}} \phi_{2^{j+1}}(x - 2^{-j-1} k))_{k \in \mathbb{Z}}$ is an orthonormal basis of $V_{2^{j+1}}$, the above sentence can be so expressed:

$$\phi_{2^j}(x - 2^{-j} n) = 2^{-j-1} \sum_{k=-\infty}^{+\infty} \langle \phi_{2^j}(u - 2^{-j} n), \phi_{2^{j+1}}(u - 2^{-j-1} k) \rangle \phi_{2^{j+1}}(x - 2^{-j-1} k) \quad (1.12)$$

Through an easy change of variables, it can then be obtained:

$$2^{-j-1} \langle \phi_{2^j}(u - 2^{-j} n), \phi_{2^{j+1}}(u - 2^{-j-1} k) \rangle = \langle \phi_{2^{-1}}(u), \phi(u - (k - 2n)) \rangle \quad (1.13)$$

By scalar multiplying $f(x)$ for the two members of the (1.12), it is obtained:

$$\langle f(u), \phi_{2^j}(u - 2^{-j} n) \rangle = \sum_{k=-\infty}^{+\infty} \langle \phi_{2^{-1}}(u), \phi(u - (k - 2n)) \rangle \langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1} k) \rangle \quad (1.14)$$

Let now \mathbf{H} be the discrete filter which impulsive response is given by:

$$h(n) = \langle \phi_{2^{-1}}(u), \phi(u - n) \rangle \quad \forall n \in \mathbb{Z} \quad (1.15)$$

and $\tilde{\mathbf{H}}$ the corresponding mirror filter with impulsive response $\tilde{h}(n) = h(-n)$.

From (1.14) it follows:

$$A_{2^j}^d f = \langle f(u), \phi_{2^j}(u - 2^{-j} n) \rangle = \sum_{k=-\infty}^{+\infty} \tilde{h}(2n - k) \langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1} k) \rangle \quad (1.16)$$

that is, $A_{2^j}^d f$ can be obtained by $A_{2^{j+1}}^d f$ by convolution with $\tilde{\mathbf{H}}$ and downsampling by a factor of 2. All the discrete approximations of the signal $f(x)$ at resolutions 2^{-j} , with $j < 0$, can then be obtained from its approximation at resolution 1, by iteratively repeating this algorithm. This process is called *pyramidal transformation*.

As already said, an approximation at resolution 2^j is characterized by $2^j N$ samples, being N the number of the samples at unitary resolution given by the acquisition device. In order to avoid border effects, the original signal is supposed to be symmetrical with respect to $n = 0$ and $n = N$, that is to say that the samples are:

$$\alpha_n = \begin{cases} \alpha_{1-n} & \text{for } -N < n < 1 \\ \alpha_n & \text{for } 1 \leq n \leq N \\ \alpha_{2N-n} & \text{for } N < n < 2N \end{cases}$$

As set out by theorem 1.1, if $(V_{2^j})_{j \in \mathbb{Z}}$ defines an approximation of $L^2(\mathbf{R})$, then there exists a unique function $\phi(x)$ in $L^2(\mathbf{R})$, named *scaling function*, such that, being $\phi_{2^j}(x) = 2^j \phi(2^j x)$ its dilation by a factor of $(2^j)_{j \in \mathbb{Z}}$, $(\sqrt{2^{-j}} \phi_{2^j}(x - 2^j n))_{n \in \mathbb{Z}}$ defines an orthonormal basis of V_{2^j} . The scaling function characterizes then the corresponding multiresolution approximation. In figure 1.1 a scaling function and its Fourier transform are reported.

Some conditions about regularity of such a function should be imposed:

- continuity;
- differentiability;
- asinhtotic behavior:

$$|\phi(x)| = O(x^{-2}) \text{ and } |\phi'(x)| = O(x^{-2}) \quad (1.17)$$

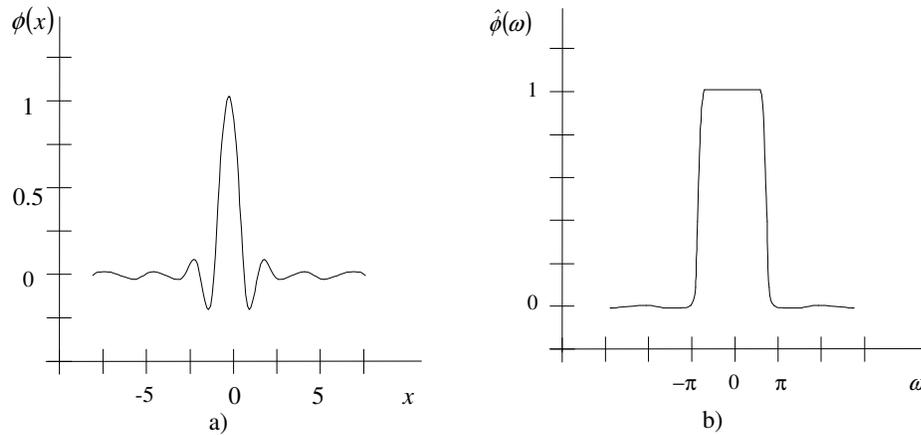


Figure 1.1

The scaling function that characterizes Lemarié-Battle multiresolution approximation described in appendix A (a) and its Fourier transform (b)

It holds also the following:

Theorem 1.2 Let $\phi(x)$ be a scaling function and \mathbf{H} the filter with impulsive response $h(n) = \langle \phi_{2^{-1}}(u), \phi(u - n) \rangle$. If $\mathbf{H}(\omega)$ is the Fourier series obtained by $h(n)$, that is:

$$\mathbf{H}(\omega) = \sum_{n=-\infty}^{+\infty} h(n) e^{-in\omega} \quad (1.18)$$

it satisfies the following properties:

1. $|\mathbf{H}(0)| = 1$ and $|h(n)| = O(n^{-2})$ at ∞ ;

2. $|\mathbf{H}(\omega)|^2 + |\mathbf{H}(\omega + \pi)|^2 = 1$

Vice-versa, if $\mathbf{H}(\omega)$ is the Fourier series that satisfies the 1. and 2. and it holds:

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

then the function defined by:

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \mathbf{H}(2^{-p} \omega) \quad (1.19)$$

is the Fourier transform of the scaling function.

The filters that satisfy the 2. are named *conjugate filters*. Given a conjugate filter that satisfies the three above properties, it is possible to calculate the Fourier transform of the corresponding scaling function by using the (1.19). The choice of the filter can be done so that scaling functions with good localization in the time and frequency domains can be found.

Further information about conjugate filters and their synthesis can be found in [2], [3] and [4].

1.2.3 The detail signal

The difference of information between the approximation of the function $f(x)$ at resolution 2^{j+1} and that at resolution 2^j is contained in the detail signal at resolution 2^j . As already said, the approximations of a function at resolutions 2^{j+1} and 2^j are the orthogonal projections of the function itself on the vectorial spaces $V_{2^{j+1}}$ and V_{2^j} respectively. As defined, the detail signal at resolution 2^j will be the orthogonal projection of the original signal on the orthogonal complement of V_{2^j} on $V_{2^{j+1}}$. Let O_{2^j} be such a space, so that:

$$O_{2^j} \text{ is orthogonal to } V_{2^j} \quad (1.20)$$

$$O_{2^j} \oplus V_{2^j} = V_{2^{j+1}} \quad (1.21)$$

In order to find the detail signal at resolution 2^j it is then necessary to find an orthonormal basis in O_{2^j} . It can be shown that such a basis can be built starting from a function $\psi(x)$:

Theorem 1.3 Let $(V_{2^j})_{j \in \mathbb{Z}}$ be a multiresolution approximation of $L^2(\mathbf{R})$, $\phi(x)$ the corresponding scaling function and \mathbf{H} the corresponding conjugate filter. Let then $\psi(x)$ be a function which Fourier transform is:

$$\hat{\psi}(\omega) = \mathbf{G}(\omega/2) \hat{\phi}(\omega/2) \quad \text{with } \mathbf{G}(\omega) = e^{-j\omega} \overline{\mathbf{H}(\omega + \pi)} \quad (1.22)$$

If $\psi_{2^j} = 2^j \psi(2^j x)$ is the dilation of $\psi(x)$ by a factor of 2^j , then:

$$\left(\sqrt{2^{-j}} \psi_{2^j}(x - 2^{-j} n) \right)_{n \in \mathbb{Z}} \text{ defines an orthonormal basis of } V_{2^j} \quad (1.23)$$

$$\left(\sqrt{2^{-j}} \psi_{2^j}(x - 2^{-j} n) \right)_{(n,j) \in \mathbb{Z}} \text{ defines an orthonormal basis of } L^2(\mathbf{R}) \quad (1.24)$$

$\psi(x)$ is called *wavelet*. In figure 1.2 the wavelet corresponding to the scaling function in figure 1.1 and its Fourier transform are reported.

It is then possible to find an orthonormal basis of \mathcal{O}_{2^j} by scaling the corresponding wavelet by a factor of 2^j and shifting it by a factor proportional to 2^{-j} .

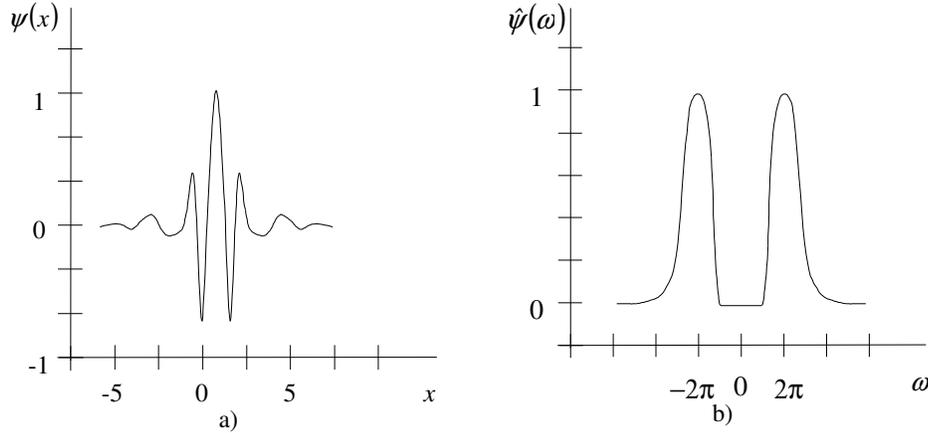


Figure 1.2

The wavelet corresponding to the scaling function in figure 1.1 (a) and its Fourier transform (b)

The wavelet decomposition of a signal is an intermediate representation between the Fourier and the spatial representations. Thanks to this double localization, in the Fourier and the spatial domains, it is possible to locally characterize the regularity of a function through the coefficients of its decomposition in an orthonormal wavelet basis. For instance, it is possible to determine the multiple differentiability of a function in a precise point starting from the velocity of asymptotical slope of its wavelet coefficients. For further information about the properties of the wavelet transform see [5].

Given a filter $H(\omega)$ that satisfies the theorem 1.2, it is therefore possible to calculate the Fourier transform of the corresponding orthonormal wavelet with (1.19) and (1.22). This process, however, does not guarantee that the scaling function and the corresponding wavelet have good properties regarding frequency separation and good time localization. The problem of finding good wavelet representations is still an open task.

A consequence of the theorem 1.3 is that, if $P_{\mathcal{O}_{2^j}}$ is the operator of orthogonal projection onto \mathcal{O}_{2^j} , the orthogonal projection of a function $f(x)$ on \mathcal{O}_{2^j} can be written in this way:

$$P_{\mathcal{O}_{2^j}} f(x) = 2^{-j} \sum_{n=-\infty}^{+\infty} \langle f(u), \psi_{2^j}(u - 2^{-j}n) \rangle \psi_{2^j}(x - 2^{-j}n) \quad (1.25)$$

$P_{\mathcal{O}_{2^j}} f(x)$ is then characterized by a set of scalar products:

$$D_{2^j} f = \left(\langle f(u), \psi_{2^j}(u - 2^{-j}n) \rangle \right)_{n \in \mathbb{Z}} \quad (1.26)$$

$D_{2^j} f$ is the discrete detail signal of $f(x)$ at resolution 2^j . It contains the difference of information between the discrete approximation of $f(x)$ at resolution 2^{j+1} , $A_{2^{j+1}}^d f$, and that at lower resolution, $A_{2^j}^d f$.

Again, it can be easily shown that each of these scalar products corresponds to the convolution of $f(x)$ with $\psi_{2^j} = 2^j \psi(2^j x)$ calculated in $2^{-j} n$, and then the discrete detail signal at resolution 2^j corresponds to the uniform sampling of such convolution by a step 2^j :

$$\mathbf{D}_{2^j}^d f = \left((f(u) * \phi_{2^j}(-u))(2^{-j} n) \right)_{n \in \mathbf{Z}} \quad (1.27)$$

Since the wavelet is a band-pass filter with band $[-2\pi, -\pi] \cup [\pi, 2\pi]$, the detail signal can be interpreted as the result of the filtering of the original signal in the band $[-2^{-j+1}\pi, -2^{-j}\pi] \cup [2^{-j}\pi, 2^{-j+1}\pi]$.

It can be demonstrated by induction that the discrete original signal at unitary resolution is represented by the set:

$$\left\{ \mathbf{A}_{2^{-j}}^d f, (\mathbf{D}_{2^j} f)_{-j \leq j \leq -1} \right\} \quad \text{for each } J > 0 \quad (1.28)$$

Such set of signals constitutes the orthogonal wavelet representation of the original signal $f(x)$, and defines the decomposition of the signal in a set of independent frequency channels. The independence of the bands is assured by the orthogonality of the wavelet functions.

1.2.4 Realization of an orthogonal wavelet representation

In order to find the wavelet representation of a signal a pyramidal algorithm is used. As in paragraph 1.2.2, it can be demonstrated that the detail signal at resolution 2^j is obtained by convolution of the discrete approximation of the original signal at resolution 2^{j+1} with an appropriate filter.

Since the function $\psi_{2^j}(x - 2^{-j} n)$ belongs to the space $\mathbf{O}_{2^j} \subset \mathbf{V}_{2^{j+1}}$, it can be decomposed in an orthonormal basis of $\mathbf{V}_{2^{j+1}}$, which, for the theorem 1.1, is constituted by

$$\left(\sqrt{2^{-j+1}} \phi_{2^{j+1}}(x - 2^{-j+1} k) \right)_{k \in \mathbf{Z}} :$$

$$\psi_{2^j}(x - 2^{-j} n) = 2^{-j-1} \sum_{k=-\infty}^{+\infty} \langle \psi_{2^j}(u - 2^{-j} n), \phi_{2^{j+1}}(u - 2^{-j+1} k) \rangle \phi_{2^{j+1}}(x - 2^{-j+1} k) \quad (1.29)$$

As for (1.13), with a change of variables, it can be obtained:

$$2^{-j-1} \langle \psi_{2^j}(u - 2^{-j} n), \phi_{2^{j+1}}(u - 2^{-j+1} k) \rangle = \langle \psi_{2^{-1}}(u), \phi(u - (k - 2n)) \rangle \quad (1.30)$$

Then, by scalar multiplying $f(x)$ for the (1.29):

$$\langle f(u), \psi_{2^j}(u - 2^{-j} n) \rangle = \sum_{k=-\infty}^{+\infty} \langle \psi_{2^{-1}}(u), \phi(u - (k - 2n)) \rangle \langle f(u), \phi_{2^{j+1}}(u - 2^{-j+1} k) \rangle \quad (1.31)$$

Let now \mathbf{G} be the discrete filter which impulsive response is given by:

$$g(n) = \langle \psi_{2^{-1}}(u), \phi(u - n) \rangle \quad \forall n \in \mathbf{Z} \quad (1.32)$$

and $\tilde{\mathbf{G}}$ the corresponding mirror filter with impulsive response $\tilde{g}(n) = g(-n)$.

From (1.31) it can be found:

$$\mathbf{D}_{2^j} f = \langle f(u), \psi_{2^j}(u - 2^{-j} n) \rangle = \sum_{k=-\infty}^{+\infty} \tilde{g}(2n - k) \langle f(u), \phi_{2^{j+1}}(u - 2^{-j+1} k) \rangle \quad (1.33)$$

that is to say that $D_{2^j} f$ can be obtained from $A_{2^{j+1}}^d f$ by convolution with \tilde{G} and downsampling by a factor of 2.

The orthogonal wavelet representation of a discrete signal can be then obtained by iteratively decomposing its discrete approximations $A_{2^{j+1}}^d f$ in $A_{2^j}^d f$ and $D_{2^j} f$ for $J \leq j \leq -1$ with the described pyramidal algorithm. Figure 1.3 shows the basic step of the algorithm.

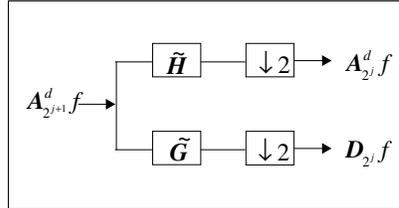


Figure 1.3
Decomposition algorithm

The signal, as measured by the acquisition device, has a finite number of samples and, in order to avoid border problems, it is supposed to be symmetrical with respect to the first and the last samples. If N is the number of samples of such signal, $A_{2^j}^d f$ and $D_{2^j} f$ have $2^j N$ samples each and the wavelet representation $\{A_{2^{-j}}^d f, (D_{2^j} f)_{-j \leq j \leq -1}\}$ has the same number of samples as the original signal $A_1^d f$, thanks to the orthogonality property.

The energy of the samples of $D_{2^j} f$ gives indications about the irregularity of the signal at resolution 2^{j+1} : as a matter of fact, the detail signal $D_{2^j} f$, as defined, has large amplitude if the approximations of the signal at resolutions 2^{j+1} and 2^j are very different one from the other.

For the (1.22) the impulsive responses of the filters H and G are related from the following:

$$g(n) = (-1)^{1-n} h(1-n) \quad (1.34)$$

The filters H and G are called Quadrature Mirror Filters (QMFs) since they are symmetrical with respect to $\pi/2$ (that is, 1/4 of the normalized sampling frequency).

Since H is a low-pass and G a high-pass filter, the approximations of a signal at several resolutions reproduce the informative content of the signal at low frequencies, while that corresponding to the high frequencies is contained in the detail signals.

Figure 1.4 shows the frequency behavior of the QMFs.

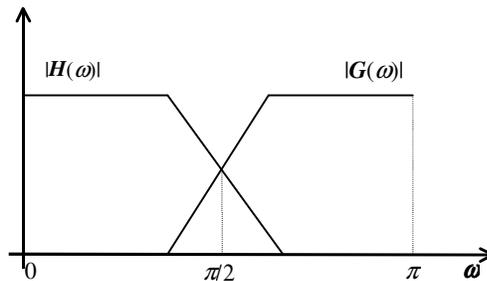


Figure 1.4
QMFs frequency responses

1.2.5 Signal reconstruction from an orthogonal wavelet representation

The original representation of a signal $f(x)$ can be reconstructed from a wavelet representation of the signal through a pyramidal algorithm.

For the (1.21), an orthonormal basis of $V_{2^{j+1}}$ can be derived from the bases of the spaces V_{2^j} and O_{2^j} , which, for the theorems 1.1 and 1.3, are given respectively by $(\sqrt{2^{-j}}\phi_{2^j}(x-2^{-j}n))_{n \in \mathbb{Z}}$ and $(\sqrt{2^{-j}}\psi_{2^j}(x-2^{-j}n))_{n \in \mathbb{Z}}$. So, the set $(\sqrt{2^{-j}}\phi_{2^j}(x-2^{-j}n), \sqrt{2^{-j}}\psi_{2^j}(x-2^{-j}n))_{n \in \mathbb{Z}}$ constitutes an orthonormal basis of $V_{2^{j+1}}$ and the function $\phi_{2^{j+1}}(x-2^{-j-1}n)$ can be then decomposed in such a basis:

$$\begin{aligned} \phi_{2^{j+1}}(x-2^{-j-1}n) &= 2^{-j} \sum_{k=-\infty}^{+\infty} \langle \phi_{2^j}(u-2^{-j}k), \phi_{2^{j+1}}(u-2^{-j-1}n) \rangle \phi_{2^j}(x-2^{-j}k) + \\ &+ 2^{-j} \sum_{k=-\infty}^{+\infty} \langle \psi_{2^j}(u-2^{-j}k), \phi_{2^{j+1}}(u-2^{-j-1}n) \rangle \psi_{2^j}(x-2^{-j}k) \end{aligned} \quad (1.35)$$

By scalar multiplying $f(x)$ for the previous expression it is obtained:

$$\begin{aligned} \langle f(u), \phi_{2^{j+1}}(u-2^{-j-1}n) \rangle &= \\ &= 2^{-j} \sum_{k=-\infty}^{+\infty} \langle \phi_{2^j}(u-2^{-j}k), \phi_{2^{j+1}}(u-2^{-j-1}n) \rangle \langle f(u), \phi_{2^j}(u-2^{-j}k) \rangle + \\ &+ 2^{-j} \sum_{k=-\infty}^{+\infty} \langle \psi_{2^j}(u-2^{-j}k), \phi_{2^{j+1}}(u-2^{-j-1}n) \rangle \langle f(u), \psi_{2^j}(u-2^{-j}k) \rangle \end{aligned} \quad (1.36)$$

This latter expression can be written for the (1.15) and (1.32) in the following way:

$$\begin{aligned} \langle f(u), \phi_{2^{j+1}}(u-2^{-j-1}n) \rangle &= \\ &= 2 \sum_{k=-\infty}^{+\infty} h(n-2k) \langle f(u), \phi_{2^j}(u-2^{-j}k) \rangle + 2 \sum_{k=-\infty}^{+\infty} g(n-2k) \langle f(u), \psi_{2^j}(u-2^{-j}k) \rangle \end{aligned} \quad (1.37)$$

$A_{2^{j+1}}^d f$ is then reconstructed from $A_{2^j}^d f$ and $D_{2^j} f$ by inserting a zero between a sample and the successive one and by convoluting the resulting signals respectively with the filters H and G .

Ideally the analysis and synthesis filter bank should consist in a set of non-overlapping filters with unitary gain impulsive responses, but adjacent, in order to avoid frequency gaps in the subband signals. In real filters the overlapping causes the aliasing effect in the filtered signal. The QMFs permit a reconstruction without aliasing in absence of quantization errors, due to the fact that the reconstruction filters are designed so as to cancel the aliasing and to minimize the total gain and the phase distortion [6]. The non-ideality of the filters is then compensated during the reconstruction.

The signal $A_1^d f$ is reconstructed by iteratively repeating such process for $J \leq j \leq 0$ and the continuous approximation $A_1 f(x)$ can be obtained by the (1.9). As in decomposition, also in the reconstruction phase a pyramidal algorithm is applied.

Figure 1.5 shows the basic step of the algorithm.

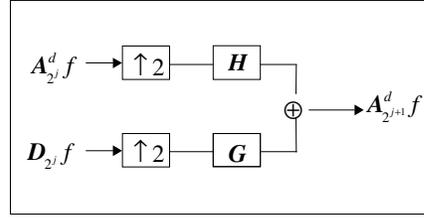


Figure 1.5
Reconstruction algorithm

1.3 Orthogonal wavelet representations of bidimensional signals

1.3.1 Multiresolution approximations of $L^2(\mathbf{R}^2)$

The model described for monodimensional signals can be easily extended to the bidimensional case for applications in the image-processing field. In this case the signal is a function $f(x, y)$ belonging to the space $L^2(\mathbf{R}^2)$ of the functions with finite energy. A multiresolution approximation of $L^2(\mathbf{R}^2)$ should satisfy the bidimensional extension of the properties (1.2)-(1.7). If $(V_{2^j})_{j \in \mathbf{Z}}$ is such multiresolution approximation, the approximation of $f(x, y)$ at resolution 2^j is given by its orthogonal projection on the space V_{2^j} .

Theorem 1.1 can be easily extended to the bidimensional case in the following way:

Theorem 1.4 If $(V_{2^j})_{j \in \mathbf{Z}}$ defines a multiresolution approximation of $L^2(\mathbf{R}^2)$, there exists then a unique function $\Phi(x, y)$ in $L^2(\mathbf{R}^2)$, named *scaling function*, such that, being $\Phi_{2^j}(x, y) = 2^{2j} \Phi(2^j x, 2^j y)$ its dilation by a factor of $(2^j)_{j \in \mathbf{Z}}$, the set $(2^{-j} \Phi_{2^j}(x - 2^{-j} n, y - 2^{-j} m))_{(n, m) \in \mathbf{Z}^2}$ defines an orthonormal basis of V_{2^j} .

An orthonormal basis for each vectorial space V_{2^j} can be then obtained by scaling the corresponding scaling function by a coefficient 2^j and shifting the resulting function by a factor proportional to 2^{-2j} . The multiplicative factor is inserted for normalization reasons. A particular case of these kind of approximations is constituted by the *separable multiresolution approximations*, that is to say that each vectorial space V_{2^j} can be decomposed in the tensorial product between two identical vectorial subspaces that constitute in turn a multiresolution approximation of $L^2(\mathbf{R})$. In this case it can be demonstrated that the scaling function $\Phi(x, y)$ is the product between the scaling functions corresponding to the vectorial subspaces constituting a multiresolution approximation of $L^2(\mathbf{R})$, so an orthogonal basis of V_{2^j} for such approximations can be obtained starting from monodimensional scaling functions by opportunely scaling and shifting them. The use of separable multiresolution approximations is very efficient in computational terms as it allows a reduction in the number of operations per sample from N^2 to N , if $N \times N$ is the dimension of the used filters. More importance is given to horizontal and vertical orientations of the image by using a separable multiresolution approximation.

If $A_1^d f$ is the approximation of an image at unitary resolution and N the number of its pixels, the approximation at resolution 2^j consists of $2^j N$ pixels. Border effects are avoided by supposing the image symmetric with respect to horizontal and vertical borders.

1.3.2 The detail signals

As in the monodimensional case, the detail signal at resolution 2^j is defined as the difference of information between the approximations of the original signal at resolution 2^{j+1} and 2^j , so it is given by the orthogonal projection of the signal on the orthogonal complement of V_{2^j} in $V_{2^{j+1}}$, indicated by O_{2^j} .

By extending theorem 1.3, the following holds:

Theorem 1.5 Let $(V_{2^j})_{j \in \mathbb{Z}}$ be a separable multiresolution approximation of $L^2(\mathbf{R}^2)$ and $\Phi(x, y) = \phi(x) \cdot \phi(y)$ the corresponding scaling function. Let then $\psi(x)$ be the monodimensional wavelet function associated to the scaling function $\phi(x)$.

Then, defined the following wavelet functions:

$$\Psi^1(x, y) = \phi(x) \cdot \psi(y), \quad \Psi^2(x, y) = \psi(x) \cdot \phi(y), \quad \Psi^3(x, y) = \psi(x) \cdot \psi(y)$$

an orthonormal basis of O_{2^j} is given by:

$$\left\{ \begin{array}{l} (2^{-j} \Psi_{2^j}^1(x - 2^{-j} n, y - 2^{-j} m)) \\ (2^{-j} \Psi_{2^j}^2(x - 2^{-j} n, y - 2^{-j} m)) \\ (2^{-j} \Psi_{2^j}^3(x - 2^{-j} n, y - 2^{-j} m)) \end{array} \right\}_{(n, m) \in \mathbb{Z}^2} \quad (1.38)$$

and an orthonormal basis of $L^2(\mathbf{R}^2)$ is given by:

$$\left\{ \begin{array}{l} (2^{-j} \Psi_{2^j}^1(x - 2^{-j} n, y - 2^{-j} m)) \\ (2^{-j} \Psi_{2^j}^2(x - 2^{-j} n, y - 2^{-j} m)) \\ (2^{-j} \Psi_{2^j}^3(x - 2^{-j} n, y - 2^{-j} m)) \end{array} \right\}_{(n, m, j) \in \mathbb{Z}^3} \quad (1.39)$$

The difference of information between the approximations of the original signal $f(x, y)$ at resolutions 2^{j+1} and 2^j is given by the orthogonal projection of the signal on O_{2^j} and it is characterized by the scalar products between $f(x, y)$ and the vectors that constitute an orthonormal basis of O_{2^j} . For the previous theorem, this difference of information is given by the three detail images:

$$\begin{aligned} D_{2^j}^1 f &= \left\langle f(x, y), \Psi_{2^j}^1(x - 2^{-j} n, y - 2^{-j} m) \right\rangle_{(n, m) \in \mathbb{Z}^2} \\ D_{2^j}^2 f &= \left\langle f(x, y), \Psi_{2^j}^2(x - 2^{-j} n, y - 2^{-j} m) \right\rangle_{(n, m) \in \mathbb{Z}^2} \\ D_{2^j}^3 f &= \left\langle f(x, y), \Psi_{2^j}^3(x - 2^{-j} n, y - 2^{-j} m) \right\rangle_{(n, m) \in \mathbb{Z}^2} \end{aligned} \quad (1.40)$$

As in the monodimensional case, the scalar products that define the approximation and the detail signals of an image at resolution 2^j are given by the uniform sampling of convolution products that, for the expressions of the wavelet functions, are:

$$\begin{aligned}
 \mathbf{A}_{2^j}^d f &= \left((f(x, y) * \phi_{2^j}(-x)\phi_{2^j}(-y))(2^{-j}n, 2^{-j}m) \right)_{(n,m) \in \mathbb{Z}^2} \\
 \mathbf{D}_{2^j}^1 f &= \left((f(x, y) * \phi_{2^j}(-x)\psi_{2^j}(-y))(2^{-j}n, 2^{-j}m) \right)_{(n,m) \in \mathbb{Z}^2} \\
 \mathbf{D}_{2^j}^2 f &= \left((f(x, y) * \psi_{2^j}(-x)\phi_{2^j}(-y))(2^{-j}n, 2^{-j}m) \right)_{(n,m) \in \mathbb{Z}^2} \\
 \mathbf{D}_{2^j}^3 f &= \left((f(x, y) * \psi_{2^j}(-x)\psi_{2^j}(-y))(2^{-j}n, 2^{-j}m) \right)_{(n,m) \in \mathbb{Z}^2}
 \end{aligned} \tag{1.41}$$

The approximated and the detail images at different resolutions are then obtained by filtering the images separately along the abscissas and the ordinates axes. From this point of view, the wavelet representation of an image can be interpreted as its decomposition in a set of independent frequency channels, characterized by a particular spatial orientation. Again, the independence between the bands is guaranteed by the orthogonality of the function.

Since the scaling function $\phi(x)$ is a low-pass filter and the wavelet $\psi(x)$ a band-pass one, the approximation of $\mathbf{A}_{2^{j+1}}^d f$ at lower resolution contains the information corresponding to the low frequencies, while the detail images $\mathbf{D}_{2^j}^1 f$, $\mathbf{D}_{2^j}^2 f$ and $\mathbf{D}_{2^j}^3 f$ put in evidence, respectively, the vertical high frequencies (horizontal edges), the horizontal high frequencies (vertical edges) and the high frequencies in both the directions (diagonal edges).

The set:

$$\left(\mathbf{A}_{2^{-j}}^d f, \left(\mathbf{D}_{2^j}^1 f \right)_{-j \leq j \leq -1}, \left(\mathbf{D}_{2^j}^2 f \right)_{-j \leq j \leq -1}, \left(\mathbf{D}_{2^j}^3 f \right)_{-j \leq j \leq -1} \right)_{j > 0} \tag{1.42}$$

defines an orthogonal wavelet representation in two dimensions.

This set of $3^j + 1$ images (see figure 1.6) represents completely $\mathbf{A}_1^d f$ and contains the same number of pixels of the original image, thanks to the orthogonality of the representation. In particular, the image $\mathbf{A}_{2^{-j}}^d f$ is the approximation of the original image at lowest resolution and it is composed by $2^{-2j}N$ pixels, and $\mathbf{D}_{2^j}^1 f$, $\mathbf{D}_{2^j}^2 f$ and $\mathbf{D}_{2^j}^3 f$ are the detail signals at resolution 2^j at the three different orientations. Each of these detail images is constituted by $2^{-2j}N$ pixels.

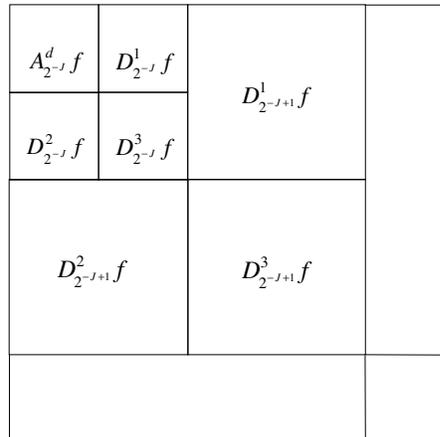


Figure 1.6
Wavelet representation of an image

1.3.3 Realization of an orthogonal wavelet representation

Also in the bidimensional case a pyramidal algorithm is used to build the orthogonal wavelet representation of the signal. By repeating the process applied in paragraph 1.3.2, the bidimensional wavelet transform is calculated through a separable extension of the monodimensional decomposition algorithm.

At each step $A_{2^{j+1}}^d f$ is decomposed in $A_{2^j}^d f$, $D_{2^j}^1 f$, $D_{2^j}^2 f$ and $D_{2^j}^3 f$:

- $A_{2^j}^d f$ is obtained by convoluting the rows of $A_{2^{j+1}}^d f$ and the monodimensional filter \tilde{H} , sub-sampling the result by a factor of 2, and convoluting the columns of the so-obtained signal and the monodimensional filter \tilde{H} , sub-sampling again by a factor of 2;
- $D_{2^j}^1 f$ is obtained by convoluting the rows of $A_{2^{j+1}}^d f$ and the monodimensional filter \tilde{H} , sub-sampling the result by a factor of 2, and convoluting the columns of the so-obtained signal and the monodimensional filter \tilde{G} , sub-sampling again by a factor of 2;
- $D_{2^j}^2 f$ is obtained by convoluting the rows of $A_{2^{j+1}}^d f$ and the monodimensional filter \tilde{G} , sub-sampling the result by a factor of 2, and convoluting the columns of the so-obtained signal and the monodimensional filter \tilde{H} , sub-sampling again by a factor of 2;
- $D_{2^j}^3 f$ is obtained by convoluting the rows of $A_{2^{j+1}}^d f$ and the monodimensional filter \tilde{G} , sub-sampling the result by a factor of 2, and convoluting the columns of the so-obtained signal and the monodimensional filter \tilde{G} , sub-sampling again by a factor of 2.

Figure 1.7 shows the basic step of this algorithm.

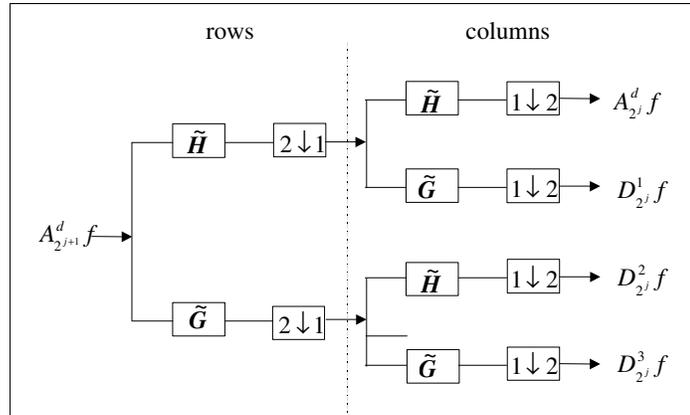


Figure 1.7

Decomposition algorithm in two dimensions

The filters used in this decomposition are the QMFs defined by the (1.15) and (1.32). The wavelet representation of an image is obtained by repeating such process for $-J \leq j \leq -1$ and corresponds to a decomposition through separable quadrature mirror filters.

1.3.4 Signal reconstruction from an orthogonal wavelet representation

Also the reconstruction algorithm can be obtained as a bidimensional extension of the process described in paragraph 1.2.5.

At each step $A_{2^{j+1}}^d f$ is recomposed from $A_{2^j}^d f$, $D_{2^j}^1 f$, $D_{2^j}^2 f$ and $D_{2^j}^3 f$ (see figure 1.8):

- an all zeroes row is inserted between the rows of $A_{2^j}^d f$, the result is then convoluted with the monodimensional filter H , an all zeroes column is then inserted between the columns of the so-obtained image and finally it is convoluted with the monodimensional filter H ;
 - an all zeroes row is inserted between the rows of $D_{2^j}^1 f$, the result is then convoluted with the monodimensional filter G , an all zeroes column is then inserted between the columns of the so-obtained image and finally it is convoluted with the monodimensional filter H ;
 - an all zeroes row is inserted between the rows of $D_{2^j}^2 f$, the result is then convoluted with the monodimensional filter H , an all zeroes column is then inserted between the columns of the so-obtained image and finally it is convoluted with the monodimensional filter G ;
 - an all zeroes row is inserted between the rows of $D_{2^j}^3 f$, the result is then convoluted with the monodimensional filter G , an all zeroes column is then inserted between the columns of the so-obtained image and finally it is convoluted with the monodimensional filter G .
- All the so-obtained images are added together and, if necessary, the result is amplified (depending on the used representation).

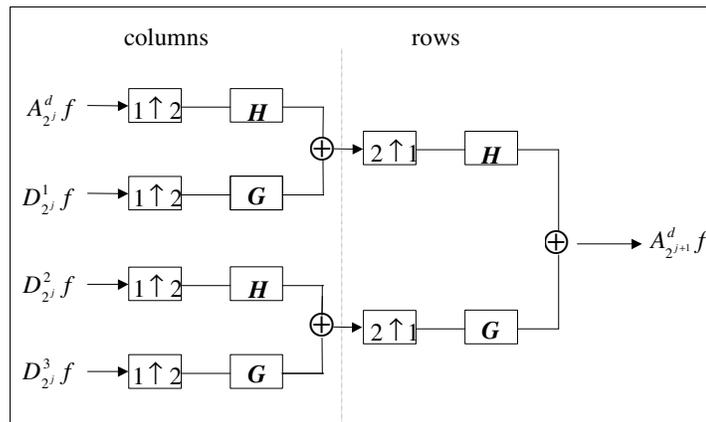


Figure 1.8
Reconstruction algorithm in two dimensions

The image $A_1^d f$ is reconstructed from its wavelet representation by repeating the illustrated process for $-J \leq j \leq -1$.

1.4 Extension of the orthogonal wavelet representation to multidimensional signals

1.4.1 Multiresolution approximations of $L^2(\mathbf{R}^n)$

The model described in 1.2 and 1.3 can be further extended for multidimensional signals for specific applications in the image-processing field. As in the bidimensional case, the signal is a function $f(x_1, x_2, \dots, x_n)$ belonging to the space of the finite energy functions $L^2(\mathbf{R}^n)$. A multiresolution approximation of $L^2(\mathbf{R}^n)$ should satisfy the multidimensional extension of the properties (1.2)-(1.7). If $(V_{2^j})_{j \in \mathbb{Z}}$ is such multiresolution approximation, the approximation of $f(x_1, x_2, \dots, x_n)$ at resolution 2^j is given by its orthogonal projection on the space V_{2^j} .

Again, Theorem 1.1 can be easily extended to the multidimensional case in the following way:

Theorem 1.6 If $(V_{2^j})_{j \in \mathbb{Z}}$ defines a multiresolution approximation of $L^2(\mathbf{R}^n)$, there exists then a unique function $\Phi(x_1, x_2, \dots, x_n)$ in $L^2(\mathbf{R}^n)$, named scaling function, such that, being $\Phi_{2^j}(x_1, x_2, \dots, x_n) = 2^{2^j} \Phi(2^j x_1, 2^j x_2, \dots, 2^j x_n)$ its dilation by a factor of $(2^j)_{j \in \mathbb{Z}}$, $(2^{-j} \Phi_{2^j}(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n))_{(m_i) \in \mathbb{Z}^n}$ defines an orthonormal basis of V_{2^j} .

An orthonormal basis for each vectorial space V_{2^j} can be then obtained by scaling the corresponding scaling function by a coefficient 2^j and shifting the resulting function of a factor proportional to 2^{-nj} . The multiplicative factor is inserted for normalization reasons.

Also for multidimensional signals, a particular case of these kind of approximations is constituted by the separable multiresolution approximations, that is to say that each vectorial space V_{2^j} can be decomposed in the tensorial product between n identical vectorial subspaces that constitute in turn a multiresolution approximation of $L^2(\mathbf{R})$. In this case the scaling function $\Phi(x_1, x_2, \dots, x_n)$ is the product between the scaling functions corresponding to the vectorial subspaces constituting a multiresolution approximation of $L^2(\mathbf{R})$, so an orthogonal basis of V_{2^j} for such approximations can be obtained starting from monodimensional scaling functions by opportunely scaling and shifting them.

As already said, the use of separable multiresolution approximations is very efficient in computational terms as it allows a reduction in the number of operations per sample from N^n to N , if $N \times N \times \dots \times N$ is the dimension of the used filters.

1.4.2 The detail signals

As in the monodimensional case, the detail signal at resolution 2^j is defined as the difference of information between the approximations of the original signal at resolution 2^{j+1} and 2^j , so it is given by the orthogonal projection of the signal on the orthogonal complement of V_{2^j} in $V_{2^{j+1}}$, indicated by O_{2^j} .

By extending theorem 1.3, the following holds:

Theorem 1.7 Let $(\mathbf{V}_{2^j})_{j \in \mathbb{Z}}$ be a separable multiresolution approximation of $L^2(\mathbf{R}^n)$ and $\Phi(x_1, x_2, \dots, x_n) = \phi(x_1) \cdot \phi(x_2) \cdot \dots \cdot \phi(x_n)$ the corresponding scaling function. Let then $\psi(x)$ be the monodimensional wavelet function associated to the scaling function $\phi(x)$. Then, defined the following wavelet functions:

$$\begin{aligned} \Psi^1(x_1, x_2, \dots, x_n) &= \phi(x_1) \cdot \phi(x_2) \cdot \dots \cdot \phi(x_{n-1}) \cdot \psi(x_n) \\ \Psi^2(x_1, x_2, \dots, x_n) &= \phi(x_1) \cdot \phi(x_2) \cdot \dots \cdot \psi(x_{n-1}) \cdot \phi(x_n) \\ &\dots \\ &\dots \\ \Psi^k(x_1, x_2, \dots, x_n) &= \psi(x_1) \cdot \psi(x_2) \cdot \dots \cdot \psi(x_{n-1}) \cdot \psi(x_n) \end{aligned} \quad (1.43)$$

with $k = 2^n - 1$, an orthonormal basis of \mathbf{O}_{2^j} is given by:

$$\left\{ \begin{array}{l} \left(2^{-j} \Psi_{2^j}^1(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right) \\ \left(2^{-j} \Psi_{2^j}^2(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right) \\ \dots \\ \dots \\ \left(2^{-j} \Psi_{2^j}^k(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right) \end{array} \right\}_{\{m_i\} \in \mathbb{Z}^n} \quad (1.44)$$

and an orthonormal basis of $L^2(\mathbf{R}^n)$ is given by:

$$\left\{ \begin{array}{l} \left(2^{-j} \Psi_{2^j}^1(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right) \\ \left(2^{-j} \Psi_{2^j}^2(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right) \\ \dots \\ \dots \\ \left(2^{-j} \Psi_{2^j}^k(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right) \end{array} \right\}_{(\{m_i\}, j) \in \mathbb{Z}^n} \quad (1.45)$$

The difference of information between the approximations of the original signal $f(x_1, x_2, \dots, x_n)$ at resolutions 2^{j+1} and 2^j is given by the orthogonal projection of the signal on \mathbf{O}_{2^j} and it is characterized by the scalar products between $f(x_1, x_2, \dots, x_n)$ and the vectors that constitute an orthonormal basis of \mathbf{O}_{2^j} . For the previous theorem, this difference of information is given by the $k = 2^n - 1$ detail images:

$$\begin{aligned} \mathbf{D}_{2^j}^1 f &= \left\langle f(x_1, x_2, \dots, x_n), \Psi_{2^j}^1(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right\rangle_{\{m_i, j\} \in \mathbb{Z}^n} \\ \mathbf{D}_{2^j}^2 f &= \left\langle f(x_1, x_2, \dots, x_n), \Psi_{2^j}^2(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right\rangle_{\{m_i, j\} \in \mathbb{Z}^n} \\ &\dots \\ &\dots \\ \mathbf{D}_{2^j}^k f &= \left\langle f(x_1, x_2, \dots, x_n), \Psi_{2^j}^k(x_1 - 2^{-j} m_1, x_2 - 2^{-j} m_2, \dots, x_n - 2^{-j} m_n) \right\rangle_{\{m_i, j\} \in \mathbb{Z}^n} \end{aligned} \quad (1.46)$$

As in the monodimensional case, the scalar products which define the approximation and the detail images of an image at resolution 2^j are given by the uniform sampling of convolution products that, for the expressions of the wavelet functions, are:

$$\begin{aligned}
\mathbf{A}_{2^j}^d f &= \left((f(x_1, x_2, \dots, x_n) * \phi_{2^j}(-x_1) \phi_{2^j}(-x_2) \dots \phi_{2^j}(-x_{n-1}) \phi_{2^j}(-x_n)) (2^{-j} m_1, 2^{-j} m_2, \dots, 2^{-j} m_{n-1}, 2^{-j} m_n) \right)_{\{m_i\} \in \mathbb{Z}^n} \\
\mathbf{D}_{2^j}^1 f &= \left((f(x_1, x_2, \dots, x_n) * \phi_{2^j}(-x_1) \phi_{2^j}(-x_2) \dots \phi_{2^j}(-x_{n-1}) \psi_{2^j}(-x_n)) (2^{-j} m_1, 2^{-j} m_2, \dots, 2^{-j} m_{n-1}, 2^{-j} m_n) \right)_{\{m_i\} \in \mathbb{Z}^n} \\
&\dots \\
&\dots \\
\mathbf{D}_{2^j}^k f &= \left((f(x_1, x_2, \dots, x_n) * \psi_{2^j}(-x_1) \psi_{2^j}(-x_2) \dots \psi_{2^j}(-x_{n-1}) \psi_{2^j}(-x_n)) (2^{-j} m_1, 2^{-j} m_2, \dots, 2^{-j} m_{n-1}, 2^{-j} m_n) \right)_{\{m_i\} \in \mathbb{Z}^n}
\end{aligned} \tag{1.47}$$

The approximated and the detail images at different resolutions are then obtained by filtering the images separately along all the axes. From this point of view, the wavelet representation of a multidimensional signal can be interpreted as its decomposition in a set of independent frequency channels, characterized by a particular spatial orientation. Again, the independence between the bands is guaranteed by the orthogonality of the functions.

Since the scaling function $\phi(x)$ is a low-pass filter and the wavelet $\psi(x)$ a band-pass one, the approximation of $\mathbf{A}_{2^{j+1}}^d f$ at lower resolution contains the information corresponding to the low frequencies, while the detail images $\mathbf{D}_{2^j}^i f$, with $i=1,2,\dots,k$ and $k=2^n-1$, put in evidence the high frequencies (edges) along the corresponding directions. The set:

$$\left(\mathbf{A}_{2^j}^d f, \mathbf{D}_{2^j}^1 f \right)_{-J \leq j \leq -1}, \dots, \left(\mathbf{D}_{2^j}^k f \right)_{-J \leq j \leq -1} \Big|_{J > 0, k=2^n-1} \tag{1.48}$$

defines an orthogonal wavelet representation in n dimensions.

This set of $k \cdot J + 1$ images represents completely $\mathbf{A}_1^d f$ and contains the same number of pixels as the original image, thanks to the orthogonality of the representation. In particular, the image $\mathbf{A}_{2^{-J}}^d f$ is the approximation of the original image at lowest resolution and it is composed by $2^{-nJ} N$ pixels, and $\mathbf{D}_{2^j}^i f \Big|_{i=1,2,\dots,k}$ are the detail signals at resolution 2^j at the different orientations. Each of these detail images is constituted by $2^{-nj} N$ pixels.

1.4.3 Realization of an orthogonal wavelet representation

Again, a pyramidal algorithm is used to build the orthogonal wavelet representation of the signal. By repeating the process applied in paragraph 1.3.2, the multidimensional wavelet transform is calculated through a separable extension of the monodimensional decomposition algorithm.

At each step $\mathbf{A}_{2^{j+1}}^d f$ is decomposed into $\mathbf{A}_{2^j}^d f, \mathbf{D}_{2^j}^1 f, \dots, \mathbf{D}_{2^j}^k f$ (see figure 1.9):

- $\mathbf{A}_{2^j}^d f$ is obtained by successively convoluting the rows of $\mathbf{A}_{2^{j+1}}^d f$ along each of the different directions x_1, x_2, \dots, x_n with the monodimensional filter $\tilde{\mathbf{H}}$, sub-sampling each time the result by a factor of 2;
- $\mathbf{D}_{2^j}^1 f$ is obtained by successively convoluting the rows of $\mathbf{A}_{2^{j+1}}^d f$ along x_1, x_2, \dots, x_{n-1} with the monodimensional filter $\tilde{\mathbf{H}}$ and x_n with the monodimensional filter $\tilde{\mathbf{G}}$, sub-sampling each time the result by a factor of 2;

-
- $D_{2^j}^k f$ is obtained by successively convoluting the rows of $A_{2^{j+1}}^d f$ along each of the different directions x_1, x_2, \dots, x_n with the monodimensional filter \tilde{G} , sub-sampling each time the result by a factor of 2.

The wavelet representation of an image is obtained by repeating such process for $-J \leq j \leq -1$ and corresponds to a decomposition through separable quadrature mirror filters.

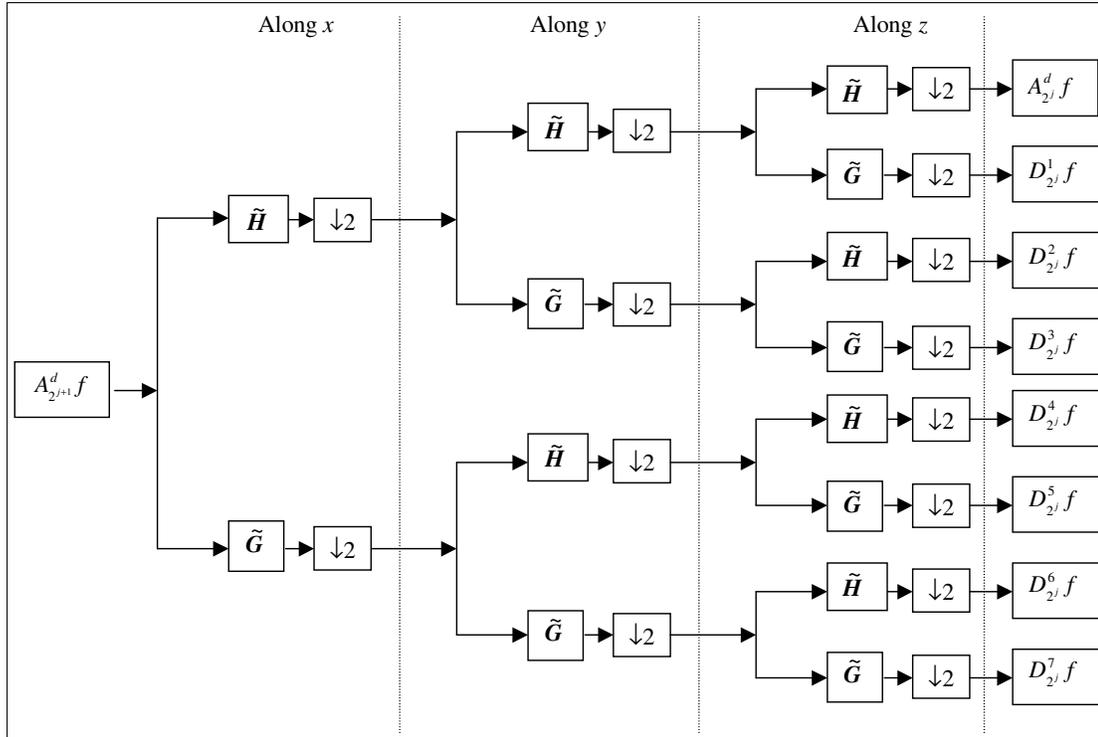


Figure 1.9
Decomposition algorithm in three dimensions

1.4.4 Signal reconstruction from an orthogonal wavelet representation

Also the reconstruction algorithm can be obtained as a multidimensional extension of the process described in paragraph 1.2.5.

At each step $A_{2^{j+1}}^d f$ is recomposed from $A_{2^j}^d f, D_{2^j}^1 f, \dots, D_{2^j}^k f$: an all zeroes row is inserted between the rows of the signal, each time along the corresponding direction. The result is then convoluted with the monodimensional filter H or G . All the so-obtained signals are added together and, if necessary, the result is amplified (depending on the used representation).

The image $A_1^d f$ is reconstructed from its wavelet representation by repeating the illustrated process for $-J \leq j \leq -1$.

References

- [1] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. on Pattern Analysis and Machine Intelligence, Tech. Rep. MS-CIS-87-22, Un. Of Penn., USA, 1989.
- [2] D. Esteban, C. Galand, "Applications of quadrature mirror filters to split band voice coding schemes," Proc. ICASSP, pp. 191-195, 1977.
- [3] G. Pirani, V. Zingarelli, "Analytical formula for the design of quadrature mirror filters," IEEE Trans. on ASSP, Vol. 32, pp. 645-648, 1984.
- [4] C. Millar, C. Paul, "Recursive quadrature mirror filters; criteria, specification and design method," IEEE Trans. on ASSP, Vol. 33, pp. 413-420, 1985.
- [5] M. Vetterli, J. Kovacevic, "Wavelets and subband coding," Prentice Hall, U.S.A, 1995.
- [6] P.P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques," IEEE ASSP Magazine, Vol. 4, No. 3, pp. 4-20, 1987.

Chapter 2

WAVELETS AND DATA COMPRESSION

2.1 Introduction

The most of the popularity of the wavelet transform to the image processing community is surely due to its impressive suitability to data compression: the statistical properties of the wavelet coefficients make it possible to adapt the level of compression according to the required quality, similarly to what happens with JPEG. Besides, the similarity of the wavelet representation with the HVS permits an adaptive quantization of the coefficients according to their informative content.

Moreover, the characteristics of the statistical distribution of wavelet coefficients permit the localization, inside each subband, of regions with higher informative content (*active zones*), where quantization should be more accurate. So, there should be an analysis phase before encoding in order to decide which subbands to encode and which not. The encoding of regions with low information content is not important for the quality of the reconstructed image. Classification of active zones is thus the tool for obtaining high compression levels without precluding the possibility of getting reconstructed images with good quality.

In this chapter the power of the wavelet transform for data compression purposes is examined as regards two real applications: one is the compression of pictorial data, which is becoming more and more interesting for commercial purposes, as regards both storing and transmission (with the spreading use and diffusion of CDs and the Internet); the other is the compression of remote sensing data, which represents a useful tool for space applications (both data storing and transmission for satellite systems).

The proposed coding technique is based on the following idea. Since quantization errors in the low-pass image are propagated during reconstruction, it is necessary to apply a lossless technique for LL subband encoding, as DPCM. The active zones of the other subbands can instead be encoded with a lossy technique (as VQ) with parameters chosen according to the characteristics of the subbands themselves (orientation, resolution, statistical properties).

The effectiveness of the proposed technique is analyzed from different point of views, according to the kind of applications for which the data are used. So, results on pictorial data are analyzed as regards especially the perceptual quality of the images, while for remote sensing data the quality of the reconstructed images is analyzed as regards the accuracy of their classification.

In both the presented applications, an adaptive technique for the selection of the “useful” wavelet coefficients has been used.

2.2 Pictorial data

2.2.1 Introduction

Wavelet decomposition is a very efficient tool for image coding, as it allows data compression also at very low bitrate without introducing annoying blocking effects. This is especially true for pictorial data, since their informative content is strongly related to the visual perception.

The reason of the popularity of this kind of representation is related to the statistical properties of the wavelet coefficients: their distribution at a fixed orientation and resolution is very similar to a Gaussian distribution with zero mean and a very small variance. The rapid slope of this distribution permits the distinction, into each subband, of a region characterized by a great energetic and informative content (*active zone*). The coefficients corresponding to almost uniform regions (whose values are rather near to the mean) do not carry high informative content and thus they do not need an accurate coding (they can even be discarded during the encoding process without significantly decreasing the performance of the method). This fact allows the achieving of a higher compression level without appreciably worsening the quality of the reconstructed image.

Once the significant regions have been detected for each subband, they are encoded with vector quantization (VQ). Several methods are presented in literature for wavelet-based image compression and, among them, a number of approaches using VQ of wavelet coefficients in different subbands has been proposed [1,2]. This technique, when applied in the spatial domain, can present annoying artifacts in the reconstructed image due to an inevitable blocking effect. The application of this method in the frequency domain avoids this drawback, since the reconstruction process masks the quantization errors.

One or more masks containing the information regarding the shape and the position of the active zones should be sent to the decoder in order to reconstruct the image.

The encoding of the low-pass image should be very accurate, since the quantization errors in this image can propagate during the reconstruction, so it is done with a lossless technique.

The following steps then compose the encoding process [3]:the image is decomposed into N levels, according to the target compression factor;

- the resulting low-pass image (in the LL subband) is error-free encoded (with DPCM);
- classified wavelet coefficients into each subband only are encoded by adaptive VQ.

2.2.2 The encoding of the low-pass image

The image in the LL subband is a low-pass version and maintains the same characteristics as the original image. A classical image coding technique can be used for this image. As already said, the use of a lossy technique (as, for example, VQ) can cause annoying artifacts due to the propagation of the quantization errors, even if the low-pass is reconstructed with an acceptable quality. The use of a lossless technique, as the DPCM, assures a better quality, since the only introduced errors are due to the conversion from floating point numbers to unsigned char ones. The error deriving from this transformation is not significant.

The description that follows holds for gray level images, but it could be easily extended to other kind of signals.

2.2.2.1 DPCM encoding

The basic idea of the DPCM encoding technique is to encode the difference of values between two neighboring pixels instead of the whole value of each pixels, exploiting the correlation

usually present in most of the natural images. Of course, the more an image presents gradual variations along the scanning direction, the more this technique will be efficient.

The signal is processed in a raster scan mode. The first pixel is entirely encoded with 8 bits. The successive one can be then encoded with n bits if the associated gray level value is no greater or lower than $2^{n-1} - 1$ with respect to the previous one. If this is not the case, an error signal should be sent and the entire value of the current pixel should be encoded.

Table 2.1 reports the possible codes for a DPCM encoder with 3 bits: the possible configurations are 8, one of these is necessary in order to identify the possible error, one other to code the possibility of no variation and the other 6 for the other cases.

Code	Interpretation
000	no variation
001	variation +1
010	variation +2
011	variation +3
100	variation -1
101	variation -2
110	variation -3
111	larger variation

Table 2.1 An example of a 3 bit (monodimensional) DPCM encoder

2.2.2.2 Bidimensional DPCM encoding

The monodimensional DPCM technique is not very efficient for image coding, since it does not take advantage of the correlation between near pixels in all the directions (horizontal, vertical and diagonal). As a matter of fact, only the possible correlation between pixels that are neighboring along the scanning direction is exploited even if they could have no relation (as happens for pixels belonging to the borders of the image).

So the bidimensional DPCM technique defined in [4] has been used: the value associated to each pixel is estimated according to the values of the near pixels and the prediction error is encoded.

First, a 1-D DPCM is applied to the first row and to the first column of the image. Then each pixel value is predicted by a linear combination of the values of its three near pixels and the prediction error is coded. Among all the possible combinations, it has been found that the best estimations are obtained with the following ones:

$$\hat{x}_{i,j} = 0.75 \cdot x_{i-1,j} - 0.50 \cdot x_{i-1,j-1} + 0.75 \cdot x_{i,j-1} \quad (2.1)$$

$$\hat{x}_{i,j} = 0.75 \cdot x_{i-1,j} - 0.53125 \cdot x_{i-1,j-1} + 0.75 \cdot x_{i,j-1} \quad (2.2)$$

The techniques of this kind are called of third order, since the estimation of the pixel is made by a linear combination of three values.

The DPCM encoding is now applied on the error:

$$\mathcal{E}_{x_{i,j}} = \hat{x}_{i,j} - x_{i,j} \quad (2.3)$$

Now, n bits are sufficient to encode an error lower than $2^{n-1} - 1$.

Table 2.2 reports the possible codes for a DPCM encoder with 3 bits.

Code	Interpretation
000	Null error
001	Error = +1
010	Error = +2
011	Error = +3
100	Error = -1
101	Error = -2
110	Error = -3
111	Larger error

Table 2.2 An example of a 3-bit bidimensional DPCM encoder

2.2.2.3 DPCM low-pass image encoding

The length of the DPCM code is adaptively chosen in order to minimize the total number of bits.

Table 2.3 contains the results on some well-known images, reported in figure 2.1, decomposed into 3 and 4 levels.



Figure 2.1
The test images

Image	Dimension	Original	Encoded	Codelength
<i>Baboon</i>	64x64 pixels	32768 bits	26778 bits	6 bits
	32x32 pixels	8192 bits	6530 bits	6 bits
<i>Carmen</i>	64x64 pixels	32768 bits	29889 bits	6 bits
	32x32 pixels	8192 bits	7457 bits	6 bits
<i>Lena</i>	64x64 pixels	32768 bits	27642 bits	6 bits
	32x32 pixels	8192 bits	7378 bits	6 bits
<i>Masquerade</i>	64x64 pixels	32768 bits	29553 bits	7 bits
	32x32 pixels	8192 bits	7641 bits	7 bits
<i>Model</i>	64x64 pixels	32768 bits	27122 bits	6 bits
	32x32 pixels	8192 bits	7417 bits	7 bits

Table 2.3 Results of 2D DPCM encoding of LL subband of the images in figure 2.1

2.2.3 Classification and encoding of the *active zones*

The statistical properties of the wavelet coefficients (see appendix B) allow the selection of the coefficients according to their informative content, which is correlated to their energetic content. The selected regions are called *active zones*: they are encoded separately from the background and their shape and position is encoded by a binary mask, constituted by the logic values that determine the belonging of the corresponding block of pixels.

Figure 2.2 shows a subband of the wavelet decomposition of *Lena* image.

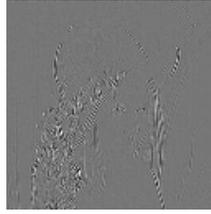


Figure 2.2
LH2 subband of the wavelet decomposition of the *Lena* image

Two different strategies have been used: one mask containing shape and position of all the active zones or three different masks, one for each orientation, are extracted.

The active coefficients are encoded by quadtree or run length. Alternatively, binary streams can be transmitted without encoding and the whole mask has a fixed size of 1024 bits.

2.2.3.1 Classification of the *active zones*

After an extensive experimental analysis on different images, a heuristic algorithm for identifying the active zones has been designed: an accurate analysis of the quantized wavelet coefficients histogram is performed into each subband in order to apply a thresholding process to the histogram. The histogram is first regularized, to avoid falling into a local minimum during the thresholds searching, then its maximum value is found. Since the histograms are not exactly symmetrical, the two thresholds cannot be taken symmetrically with respect to the mean. The thresholding is thus done along the ordinates axis instead of the abscissas one. In fact, since the histogram can be modeled as unimodal, with the maximum usually located around the zero value, two thresholds can be found proceeding into the two directions starting from this maximum, so that they include a histogram area around the 95% of the total (see figure 2.3). The search of these thresholds is made adaptively, in the sense that at each step the threshold, whose movement minimizes the variation of the included area, is shifted.

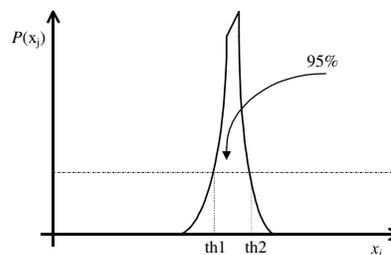


Figure 2.3
Selection of the two thresholds

Going into further details, the procedure is applied into each subband and is composed by the following steps:

Histogram regularization

This is a useful shrewdness in order to avoid the “falling” into a local minimum during the research of the best thresholds. As a matter of fact, this could cause the blocking of the shifting of a threshold and two thresholds strongly asymmetrical with respect to the mean could be thus selected.

The mean value of 4 among the 5 values that are around the current one, after discarding the minimum, is substituted to it.

So, for each value i ($2 \leq i \leq 253$):

- take into account the values between $i - 2$ and $i + 2$;
- discard the minimum among them;
- calculate the average value among the 4 remaining ones;
- substitute the calculated mean value for i .

In this way, the histogram is regularized from possible local oscillations.

Maximum detection

The maximum is the central value among the 5 ones containing the maximum area.

In order to detect it, the histogram is scanned and, for each value i , the area contained between $i - 2$ and $i + 2$ is computed with opportune weights: the larger weight is assigned to the central value and the other weights are distributed in a symmetrical and inversely proportional manner with respect to the distance from the central value.

Thresholds searching

Once the maximum has been fixed, two thresholds are chosen with the following algorithm (see figure 2.4):

a value that is lower than the maximum is fixed in the ordinates axis and the two corresponding values in the abscissas axis are selected;

the area contained between them is computed:

if the area is lower than the 95% of the total, the threshold corresponding to the larger value is shifted away from the other;

if the area is larger than the 95%, the thresholds are drawn near with the same criterion;

the process is repeated until two thresholds, such that the area included between them is the minimum value larger than the 95% of the total one, are found;

the initial value (to which the initial values of the thresholds correspond) can be chosen for example as half the value of the maximum.

Active zones selection

At the end of this process the distinction between active zone and not is made in this way: the pixels whose values are included between the two thresholds are classified as belonging to the non-active zone (their values are near the mean and they constitute an enough uniform zone of the image), while pixels whose values are not included between the two thresholds are classified as belonging to the active zone, since their values constitute the distribution tails and are quite different from the mean value.

Extraction of the mask

After the classification process, for each subband a binary-value “mask”, that contains the information corresponding to the position and the shape of the active zone, is extracted. The mask is then scaled down and logically summed with the one obtained with the same procedure at the lower resolution and at the same orientation.

In order to avoid the loss of any of the element of the active zone, if there is at least one active coefficient into each of the 2x2 blocks, the whole block is classified as active. There is a lower limit to the possible resolution of the mask, since at low-resolution level the subbands have a high energetic content and the active zone would fill completely the mask, limiting in

this way the global compression level. A good compromise between the quality and the compression level has led to the choice of a mask at resolution $j = -3$. For example, the final mask for 256x256 images is constituted by 32x32 binary values.

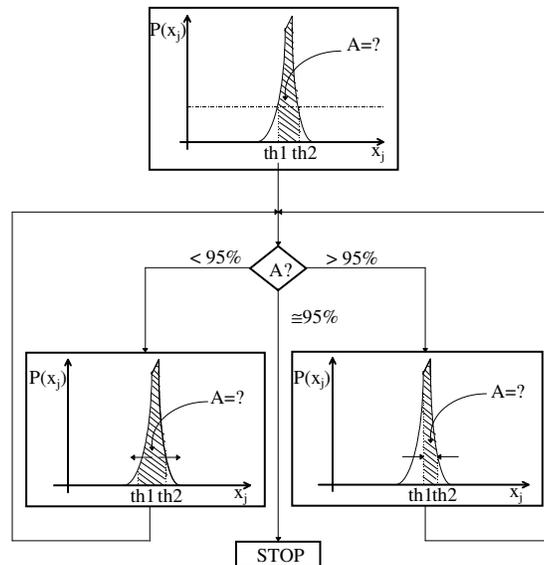


Figure 2.4
The thresholds searching algorithm

At the end, three masks, each of which contains the information regarding the active zones of the subbands at the same orientation, are extracted. In alternative, a unique mask can be obtained by the logic sum of these three ones. Figure 2.5 shows the extracted masks containing the active zones of *Lena* image.



Figure 2.5
The masks of the active zones of *Lena* image for:
(a) LH orientation, (b) HL orientation, (c) HH orientation and (d) total mask

2.2.3.2 Encoding of the active zones

The final mask (or the final masks) is encoded by quadtree or optimized run length, according to the convenience.

Quadtree

The quadtree is an iterative algorithm that is very efficient for binary image encoding [5]. The basic idea is the following: if the image is uniform, one bit is sufficient for its encoding (e.g. 0=black, 1=white), otherwise the image is decomposed into 4 blocks and each of them is analyzed in the same way.

A tree is thus generated. Each node corresponds to a block with decreasing dimensions and a bit is associated to each node to code if the corresponding block is uniform or not. If it is not, the block is divided into 4 sub-blocks until all uniform blocks are found.

Once all uniform blocks (leaves) are found, a bit encodes the associated binary value (e.g. 0=black, 1=white).

The worst case is when the algorithm terminates with all blocks constituted by only 1 pixel. In this case, if the original image is constituted by $N \times N$ pixels (with $N = 2^n$), $\sum_{i=0}^n 4^i + 4^n$ bits are necessary to encode the image. This method is convenient for images with uniform regions. Figure 2.6 shows an example of quadtree and figure 2.7 the basic algorithm.

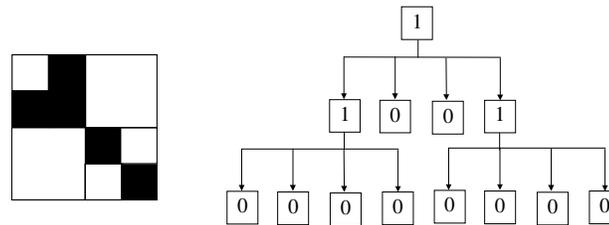


Figure 2.6

An example of quadtree (0=uniform block, 1=non-uniform block)

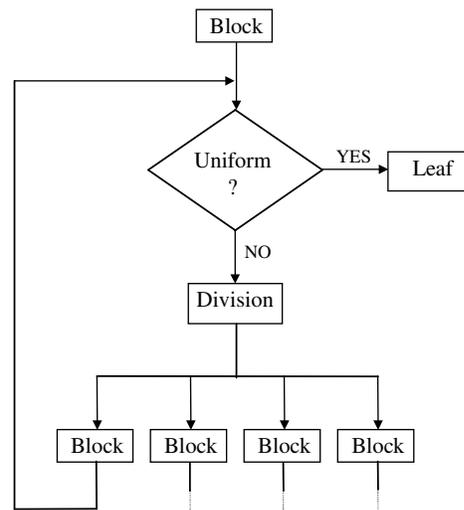


Figure 2.7

Quadtree algorithm

Table 2.4 shows the results obtained with the quadtree encoding of the three masks and of the total one on some test images.

image	# bits LH mask	# bits HL mask	# bits HH mask	# bits total mask
<i>Agave</i>	1234	1101	1297	905
<i>Airplane</i>	961	737	583	772
<i>Baboon</i>	1346	1339	1374	1059
<i>Carmen</i>	1059	1080	1066	835
<i>Cat</i>	1451	1360	1213	1269
<i>Lena</i>	1122	1031	982	863
<i>Masquerade</i>	1535	1346	1374	1437
<i>Model</i>	877	856	737	716
<i>Peppers</i>	1220	1458	1259	1255
<i>Tiffany</i>	1157	1234	1206	1024

Table 2.4 Results of quadtree encoding of the test masks

Run length

The run length is a technique often used for black and white text and fax transmission and is based on the coding of binary streams along a prefixed scanning direction. The encoding of consecutive 0s or 1s sequences (*runs*) is made with codewords whose first bit indicates the type of sequence and the following ones binary encode the length of the run, that is, the number of equal consecutive bits. Figure 2.8 shows the basic algorithm of run length encoding.

The performance of this technique depends on the average length of 0 or 1 runs in the image to be encoded, that is on the source statistics, and on the scanning direction.

In order to optimize the compression, both scanning directions (horizontal and vertical) are investigated.

The used algorithm is the following:

for each scanning direction:

calculate the average lengths of 0 and 1 runs;

initialize the lengths of the runs with the found average values;

for each run:

calculate the number of bits necessary to encode the present sequences;

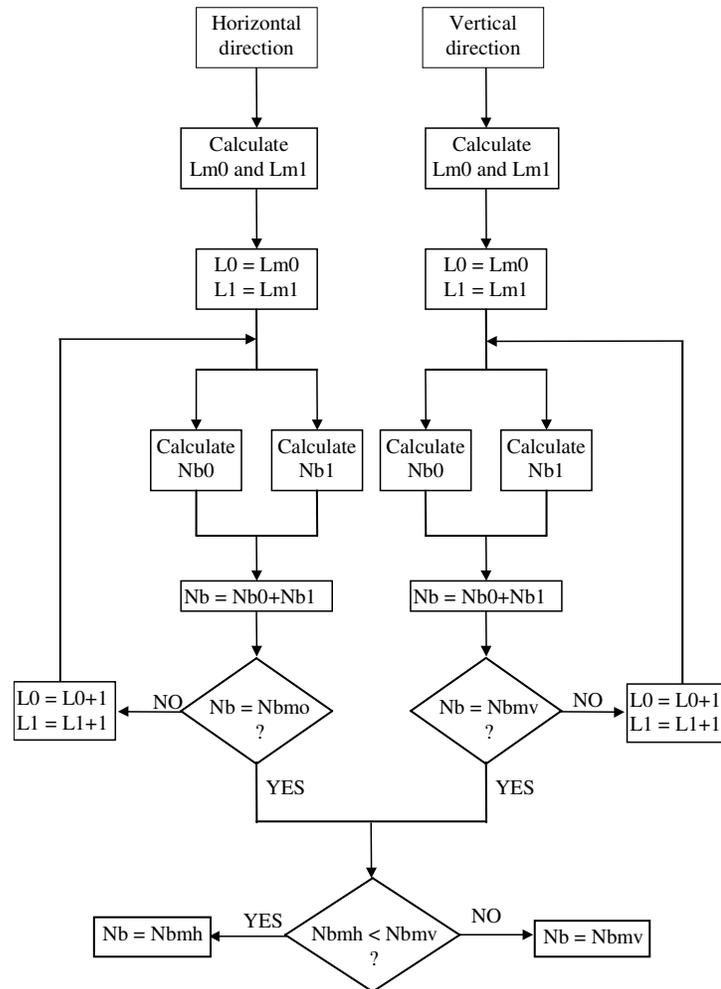
increase the length of the run and continue until the minimum number of bits is found;

the minimum found value in both the directions is chosen.

Table 2.5 shows the results obtained with the runlength encoding of the three masks and of the total one on some test images.

The choice of the encoding technique for the mask is made according to the maximum value of compression that can be obtained. In some cases it could be even more convenient to send the mask without any of the two described techniques.

Table 2.6 shows the results obtained for the test masks.

**Figure 2.8**

Run length algorithm

(L=length, m=minimum, N=number, b=bit, h=horizontal, v=vertical)

image	#bits LH mask		#bits HL mask		#bits HH mask		#bits total mask	
	H	V	H	V	H	V	H	V
<i>Agave</i>	1131	1492	1332	1307	1256	1343	964	1214
<i>Airplane</i>	1065	1263	672	826	602	732	658	1031
<i>Baboon</i>	1492	1508	1488	1468	1520	1512	1126	1074
<i>Carmen</i>	1251	1230	1372	1155	1115	1115	868	1050
<i>Cat</i>	1512	1536	1532	1416	1460	1440	1373	1341
<i>Lena</i>	1261	1294	1392	1055	1287	1060	974	968
<i>Masquerade</i>	1508	1656	1652	1468	1564	1512	1537	1493
<i>Model</i>	1002	956	1210	899	858	750	812	680
<i>Peppers</i>	1285	1300	1564	1428	1396	1440	1233	1221
<i>Tiffany</i>	1285	1264	1436	1267	1298	1327	1149	993

Table 2.5 Results of runlength encoding of the test masks

image	LH mask		HL mask		HH mask		#bits total mask	
	technique	#bits	technique	#bits	technique	#bits	technique	#bits
<i>Airplane</i>	quadtree	961	run H	672	quadtree	583	run H	658
<i>Lena</i>	-----	1024	-----	1024	quadtree	982	quadtree	863
<i>Model</i>	quadtree	877	quadtree	859	quadtree	737	run V	680

Table 2.6 Best encoding results for the test masks

2.2.4 Encoding of the wavelet coefficients

Some common properties arise from the analysis of the statistical distribution of the wavelet coefficients of the images constituting the training set (see appendix B). This permits the use of an adaptive VQ, with parameters chosen according to the resolution and orientation of the corresponding subband.

In each subband the coefficients belonging to the non-active zone are not coded because their informative content is very low (they constitute the background of the subband). The coefficients of the active zone, on the other hand, have the main part of the energetic and informative content of the subband, so their encoding must be accurate.

These coefficients are coded by VQ with parameters chosen starting from energetic considerations. As a matter of fact, the HH subbands are in general poorer of informative content than LH and HL ones. Moreover, a subband at low resolution is more energetic than subbands at the same orientation but at higher resolutions. In fact, as the resolution reduces, the energetic content of a subband increases and thus quantization must be more accurate.

The parameters of the VQ (codebooks and codevectors dimensions) are chosen accordingly to the variation of MSE value during the LBG algorithm. Codebooks are specific for each subband (at each resolution and orientation) and are generated from the active zones of the subbands of the images belonging to the training set.

2.2.4.1 VQ for wavelet coefficients encoding

The VQ is a multidimensional extension of the scalar quantization [6]. In the latter one, the set of all the possible input values is divided into a finite number of intervals, each of which is represented by an opportune value. This value is the output associated to every value falling into the corresponding interval.

In VQ, the same association is made in a N -dimensional space: the space is divided into $N_c = 2^{NR}$ subsets, a representative value (*codevector*) is associated to each of them, and a block constituted by N pixels can be thus encoded by $N \cdot R$ bits, constituting the address of the corresponding codevector. R is the average number of bits per pixel (*bitrate*) and the set of all the codevectors is called *codebook*.

The process consists of 3 main steps:

- codebook generation;
- encoding, with an opportune metric;
- decoding, with a look-up table.

The first phase strongly influences the performance of the method. The most used algorithm has been ideated by Linde, Buzo and Gray, and it is well known as LBG algorithm [7].

Once the algorithm has been generated, the encoding is simply done by substituting to each block the nearest codevector, according to a prefixed metric. For each block the information to be transmitted is the address of the codevector with minimum distance.

A set of representative images, decomposed into blocks of N pixels, is used in order to build the codebook (*training set*). As in the scalar case, the quantizer introduces a distortion

corresponding to the distance from the input vector X and the code one \hat{X} : in order to minimize the global distortion, the main problem is then the opportune choice of the partition of the original space and of the corresponding codevectors.

The most used metric $d(X, \hat{X})$ is the mean square error (MSE), defined as:

$$\text{MSE} = E\{[o(m,n) - c(m,n)]^2\} \quad (2.4)$$

being $o(m,n)$ the original image and $c(m,n)$ the encoded one. The corresponding distortion is defined by:

$$d(X, \hat{X}) = \frac{(X, \hat{X})(X, \hat{X})^T}{N} = \frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N} \quad (2.5)$$

and the global distortion is the mean of the distortion on the single blocks.

2.2.4.2 Codebook generation

LBG algorithm

Codebooks are generated starting from a training set composed by representative images. Obviously, the optimum codebook for an image is the one created starting from the image itself (*local codebook*), but this should be generated each time for each image and moreover it should be sent to the decoder, limiting in this way the compression level. Again a compromise between quality and compression should be found.

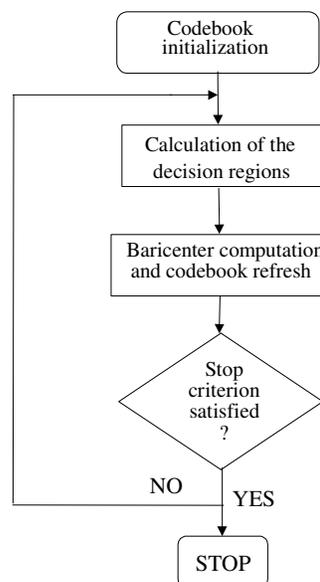


Figure 2.9
LBG algorithm for the codebook construction

The LBG algorithm (see figure 2.9) is composed of the following steps:

1. The codebook is initialized with a set of vectors belonging to the training set opportunely chosen (the choice of the initial vectors strongly influences the performance of the

- method). Let $\hat{X}_i^{(0)}$, for $i = 1, \dots, N_c$, be the initial codebook, d a distortion measure and ε a prefixed threshold. The number of iterations l is initialized to 1 and $D^{(0)}$, the mean distortion of all the training vectors, is initialized to ∞ .
2. A codevector $\hat{X}_i^{(l)}$ is associated to each training vector belonging to the corresponding decision region, according to a criterion of minimum average distortion. The codebook is refreshed by substituting the new vector $\hat{X}_i^{(l)}$ that minimizes the quantization error for the corresponding decision region, to each codevector $\hat{X}_i^{(l-1)}$. If the MSE criterion is applied, the codevector that minimizes the distortion is the baricenter of the training vectors belonging to the corresponding decision region, that is, each of its components is the mean of the corresponding components of the training vectors.
 3. The mean distortion $D^{(l)}$ obtained by the encoding process of the training set is calculated. The algorithm terminates if $\frac{D^{(l-1)} - D^{(l)}}{D^{(l-1)}} \leq \varepsilon$, otherwise the number of iterations increases and the step 2 is repeated. The maximum number of iterations should be prefixed in order to avoid a too long processing.

Figure 2.10 shows the 8 images constituting the training set. All images have 512x512 pixels and are quantized with 256 gray levels.

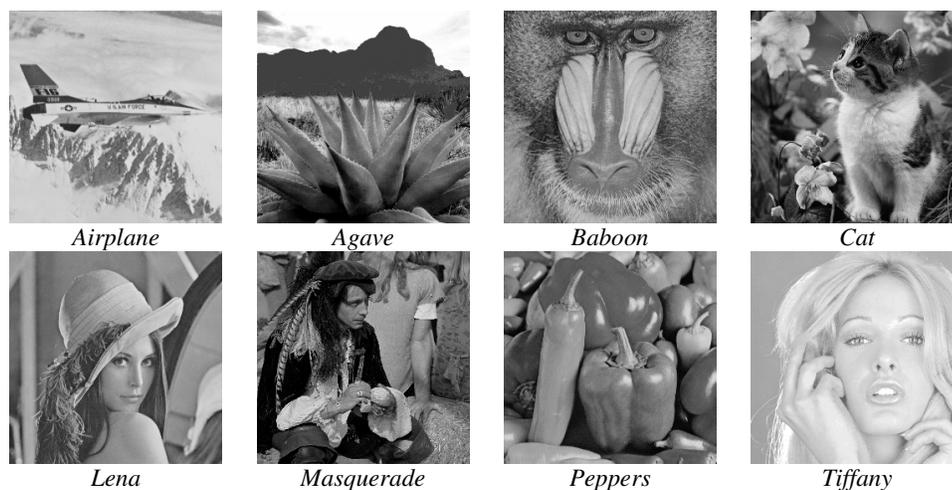


Figure 2.10
The training set

Construction of a multiresolution codebook

The wavelet representation allows the generation of specific codebooks for each orientation and resolution. Each of these is generated by the LBG algorithm and uses a training set of images with opportune orientation and resolution. For each input vector, only the specific sub-codebook is analyzed, lightening the computational burden of the encoding process.

The parameters of each sub-codebook can be chosen according to the requirements. For example, a region that is not very important in the reconstruction of the image can be quantized with a codebook with a low number of vectors, increasing in this way the compression level without damaging too much the quality.

2.2.4.3 Adaptive encoding of the wavelet coefficients

The encoding of the active zones is done, for each subband, with codebooks generated starting from a training set of the active zones of images at the same orientation and resolution. The parameters of each sub-codebook are chosen such that the compression does not preclude a good reconstruction quality: subbands constituted by coefficients whose distribution has a low variance can be heavily encoded without damaging too much their informative content, while a more accurate quantization is done on the regions characterized by wider distributions.

The following characteristics arise from the analysis of the distributions of the wavelet coefficients into each subband (see appendix B):

- for a fixed orientation, as the resolution decreases the variance of the distribution increases;
- for a prefixed resolution, the subband with diagonal orientation has the distribution with the minimum variance with respect to the other orientation.

Thus, the subband that can then be encoded in the heaviest manner is:

- the one with diagonal orientation for a fixed resolution;
- the one with higher resolution for a fixed orientation.

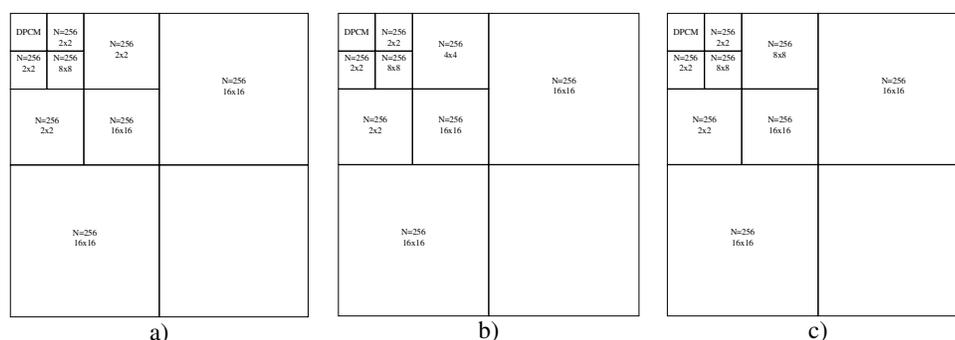
It is then evident that the subbands at low resolution and with horizontal and vertical orientations are richer of active zones (and thus have higher informative content) than the others.

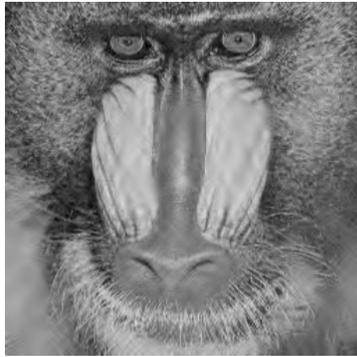
At the same resolution, the subbands with vertical orientation are the more energetic ones.

This accurate analysis has led to the choice of the following parameters:

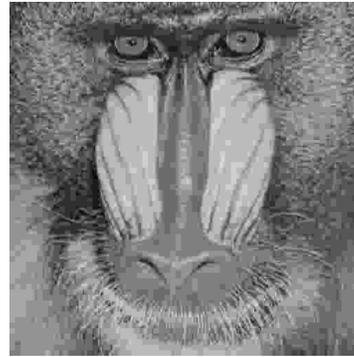
- all codebooks are composed by 256 codevectors;
- the encoding of the subbands with higher resolution at diagonal orientation does not significantly weigh on the quality of the reconstructed image and thus this is not done;
- codevectors with dimensions 16x16 are used to encode the subbands with higher resolution.

Figure 2.11 shows the used bit allocation schemes.





AZWVQ
($r_b=0.132$ bpp, PSNR=26.89 dB)



JPEG
($r_b=0.132$ bpp, PSNR=24.39 dB)

a)



AZWVQ
($r_b=0.329$ bpp, PSNR=32.90 dB)



JPEG
($r_b=0.329$ bpp, PSNR=33.30 dB)



AZWVQ
($r_b=0.128$ bpp, PSNR=30.50 dB)



JPEG
($r_b=0.128$ bpp, PSNR=28.43 dB)

b)



AZWVQ
($r_b=0.329$ bpp, PSNR=26.77 dB)



JPEG
($r_b=0.329$ bpp, PSNR=26.35 dB)



AZWVQ
($r_b=0.13$ bpp, PSNR=25.43 dB)



JPEG
($r_b=0.13$ bpp, PSNR=23.36 dB)

c)

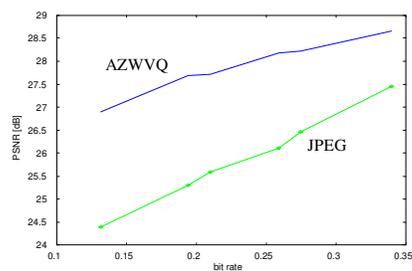
Figure 2.12

Results of AZWVQ vs. JPEG at some compression ratios for
(a) *Baboon*, (b) *Lena* and (c) *Masquerade* images

The quality of the images encoded by AZWVQ is higher than the one with JPEG both in quantitative (PSNR) and in qualitative terms: as a matter of fact, a notable artifact caused by JPEG is the blockiness distortion, which is absent in AZWVQ. The images encoded by AZWVQ present an effect known as *ringing* around the edges. This is caused by the loss of information at high frequencies, due to the hard quantization of these zones. The main advantage of the use of the VQ in the frequency domain is that a high compression level can be obtained by exploiting the statistical properties of the wavelet coefficients without having the well-known artifacts of this technique.

Baboon

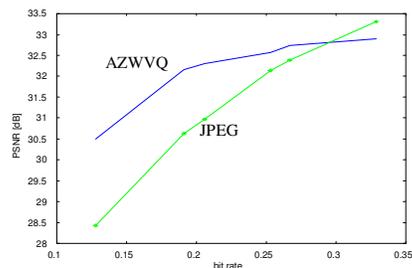
schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.340	23.46	28.65 dB	27.45 dB
b	0.275	28.99	28.22 dB	26.46 dB
c	0.259	30.77	28.18 dB	26.10 dB
d	0.210	37.92	27.72 dB	25.58 dB
e	0.194	41.04	27.69 dB	25.30 dB
f	0.132	60.18	26.89 dB	24.39 dB



a)

Lena

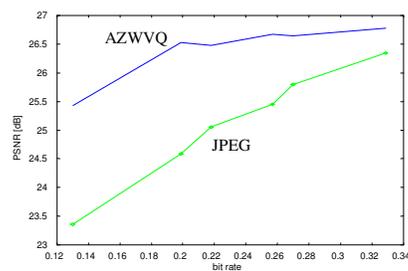
schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.329	24.27	32.90 dB	33.30 dB
b	0.267	29.93	32.73 dB	32.38 dB
c	0.253	31.56	32.56 dB	32.14 dB
d	0.206	38.79	32.31 dB	30.97 dB
e	0.191	41.81	32.15 dB	30.62 dB
f	0.128	62.34	30.50 dB	28.43 dB



b)

Masquerade

schema	bitrate [bpp]	CR	PSNR [dB]	
			AZWVQ	JPEG
a	0.329	24.24	26.77	26.35
b	0.270	29.52	26.64	25.80
c	0.257	31.01	26.67	25.45
d	0.218	37.57	26.48	25.05
e	0.199	40.01	26.52	24.59
f	0.130	61.10	25.43	23.36



c)

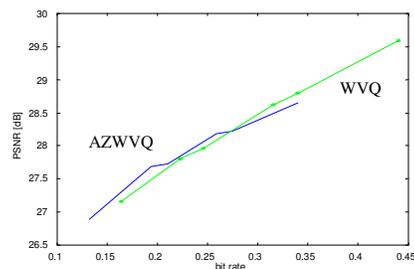
Table 2.7 Quantitative results of AZWVQ vs. JPEG at some compression ratios for (a) *Baboon*, (b) *Lena* and (c) *Masquerade* images

AZWVQ vs. WVQ

The main innovative aspect of the presented technique is in the selection of the active zones into the subbands. In order to highlight the advantage of the AZWVQ with respect to the use of the same technique without the classification of the active zones (Wavelet coefficient Vector Quantization, WVQ), comparisons between the results obtained on some test images with both the techniques are reported in table 2.8.

Baboon

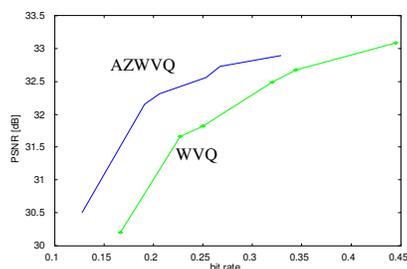
schema	AZWVQ			WVQ		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.340	23.46	28.65	0.441	18.13	29.586
b	0.275	28.99	28.22	0.340	23.5	28.797
c	0.259	30.77	28.18	0.316	25.29	28.615
d	0.210	37.92	27.72	0.246	32.47	27.959
e	0.194	41.04	27.69	0.223	35.82	27.81
f	0.132	60.18	26.89	0.164	48.85	27.16



a)

Lena

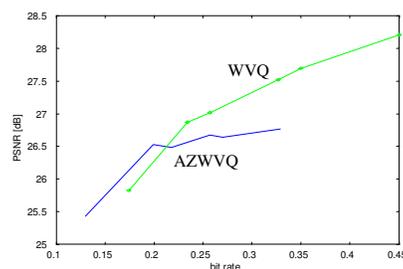
schema	AZWVQ			WVQ		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.329	24.27	32.90	0.445	18	33.077
b	0.267	29.93	32.73	0.344	23.26	32.680
c	0.253	31.56	32.56	0.320	25	32.485
d	0.206	38.79	32.31	0.250	32	31.824
e	0.191	41.81	32.15	0.227	35.24	31.664
f	0.128	62.34	30.50	0.167	47.96	30.204



b)

Masquerade

schema	AZWVQ			WVQ		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.329	24.24	26.77	0.452	17.7	28.21
b	0.270	29.52	26.64	0.351	22.79	27.70
c	0.257	31.01	26.67	0.327	24.47	27.52
d	0.218	37.57	26.48	0.257	31.13	27.02
e	0.199	40.01	26.52	0.234	34.19	26.87
f	0.130	61.10	25.43	0.174	46	25.83



c)

Table 2.8 Quantitative results of AZWVQ vs. WVQ at some compression ratios for (a) *Baboon*, (b) *Lena* and (c) *Masquerade* images

AZWVQ presents advantages with respect to WVQ both in terms of compression and of quality of the reconstructed image. As regards the compression, this result is quite obvious: only a part of the subbands of the image (corresponding to the 60-70 % of the total) is quantized.

As concerns the quality, in AZWVQ the codebook is built on a training set composed by the active zones of the 8 input images, which are the richest of informative content. Regions constituted by almost uniform zones are not considered in the generation of the codebook, as they can be reconstructed by substitution with the mean value, and a more specific codebook is thus built.

Images that do not belong to the training set

The performance of the VQ is strongly influenced by the images in the training set. The best results for an image are obtained if the training set is composed only by that image. In this case, however, the encoding of different images would require every time the generation of a different codebook.

The generation of the most general codebook requires several images in the training set so that an enough complete statistics can be considered in order to guarantee a good quality in the encoding of images that do not belong to the training set itself.

The results obtained on two images that do not belong to the training set are reported in figure 2.13 and table 2.9 for a more complete evaluation of the performance of the proposed method. Again, comparisons with the standard JPEG are reported.



AZWVQ
($r_b=0.291$ bpp, PSNR=24.87 dB)



JPEG
($r_b=0.291$ bpp, PSNR=28 dB)

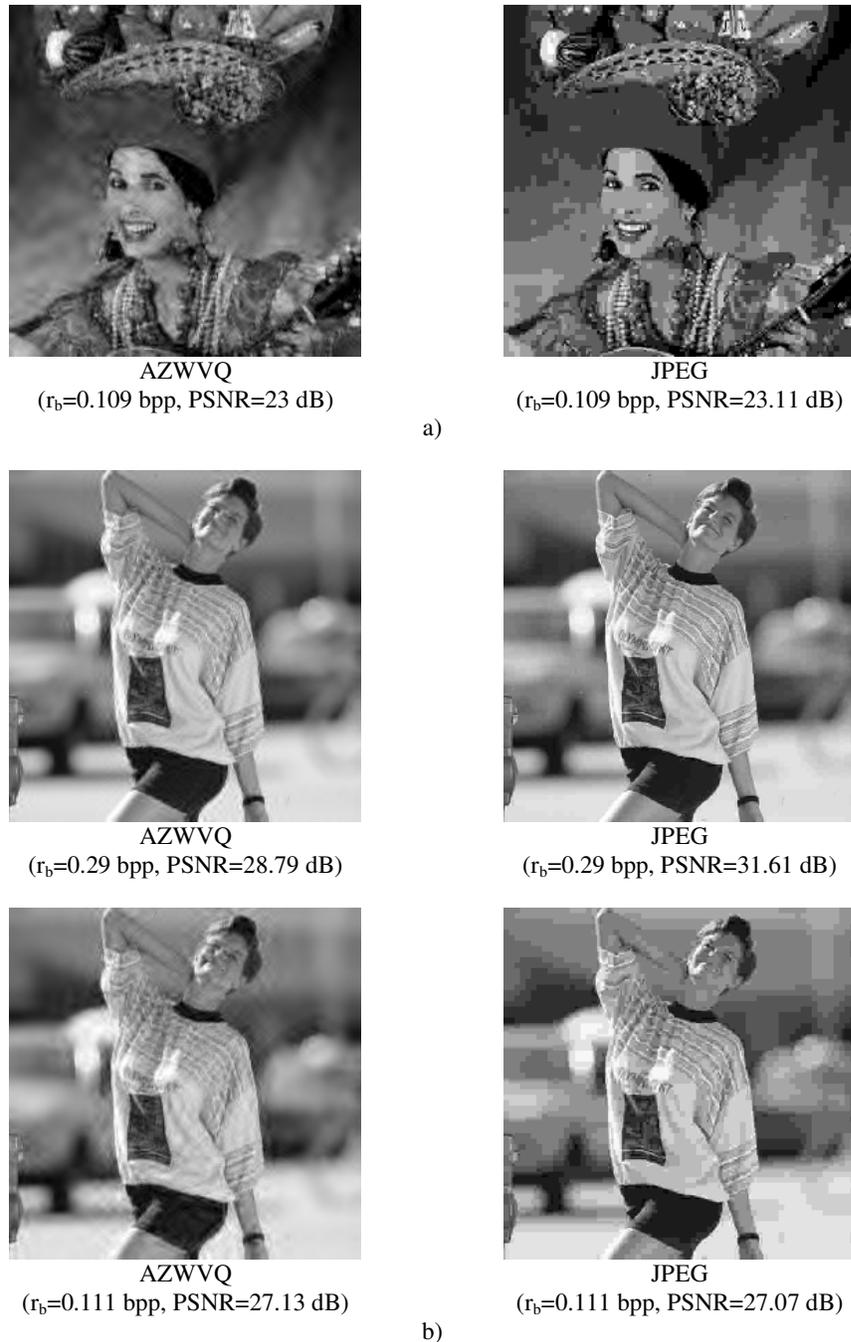


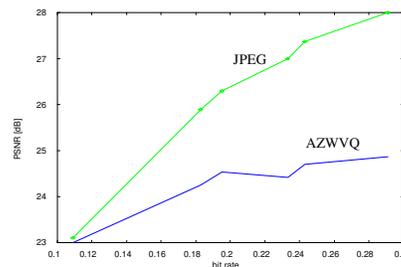
Figure 2.13
Results of AZWVQ vs. JPEG at some compression ratios for
(a) *Carmen* and (b) *Model* images

The performance of the method for images that do not belong to the training set can be bad for two main reasons:

- it is difficult to compress the image, apart from the training set, due to the presence of high frequencies in the image itself;
- the training set is not enough general and does not contain any image with characteristics similar to the one to be encoded.

Carmen

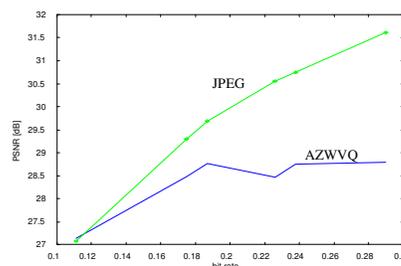
schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.291	27.47	24.87 dB	28.00 dB
b	0.243	32.89	24.71 dB	27.37 dB
c	0.233	34.33	24.42 dB	27.00 dB
d	0.195	40.98	24.53 dB	26.30 dB
e	0.183	43.66	24.25 dB	25.94 dB
f	0.109	73.25	23.00 dB	23.11 dB



a)

Model

schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.290	27.52	28.79 dB	31.61 dB
b	0.238	33.52	28.74 dB	30.75 dB
c	0.226	35.29	28.46 dB	30.55 dB
d	0.187	42.63	28.76 dB	29.68 dB
e	0.175	45.54	28.48 dB	29.30 dB
f	0.111	71.66	27.13 dB	27.07 dB



b)

Table 2.9 Quantitative results of AZWVQ vs. JPEG at some compression ratios for (a) *Carmen* and (b) *Model*

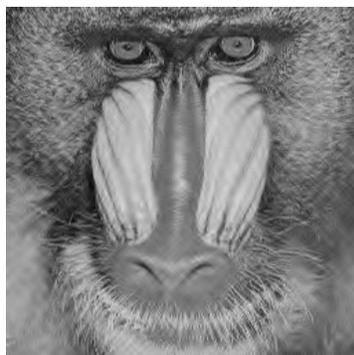
From the comparison with other techniques (e.g. JPEG) it is possible to understand if the encoding difficulty is due to the kind of image or to the training set. In the latter case, it is necessary to generate a more general codebook.

Anyway, from the performed tests it has been found that the AZWVQ is advantageous with respect to JPEG for high levels of compression.

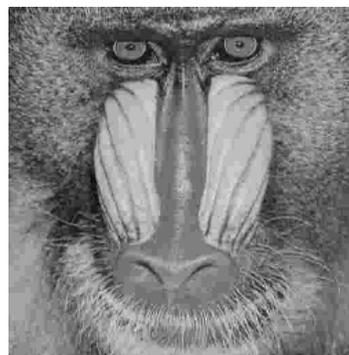
2.2.5.2 Identification and use of a mask for each orientation

Images belonging to the training set

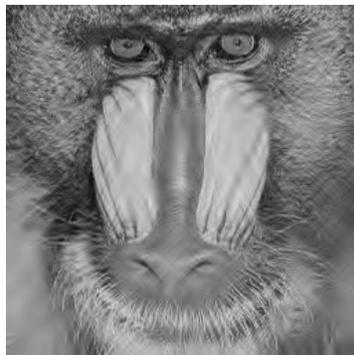
Results obtained with the use of three different masks on the images *Lena*, *Baboon* and *Masquerade* are reported in figure 2.14 and table 2.10 and compared to the JPEG at the same compression ratios.



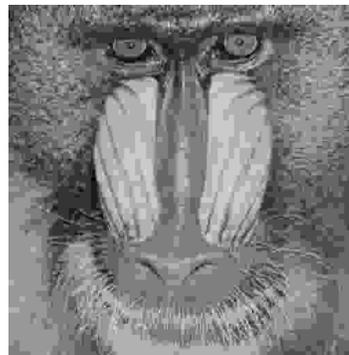
AZWVQ
($r_b=0.249$ bpp, PSNR=27.49 dB)



JPEG
($r_b=0.249$ bpp, PSNR=26.04 dB)



AZWVQ
($r_b=0.102$ bpp, PSNR=26.37 dB)



JPEG
($r_b=0.102$ bpp, PSNR=23.56 dB)

a)



AZWVQ
($r_b=0.251$ bpp, PSNR=31.94 dB)



JPEG
($r_b=0.251$ bpp, PSNR=32.14 dB)

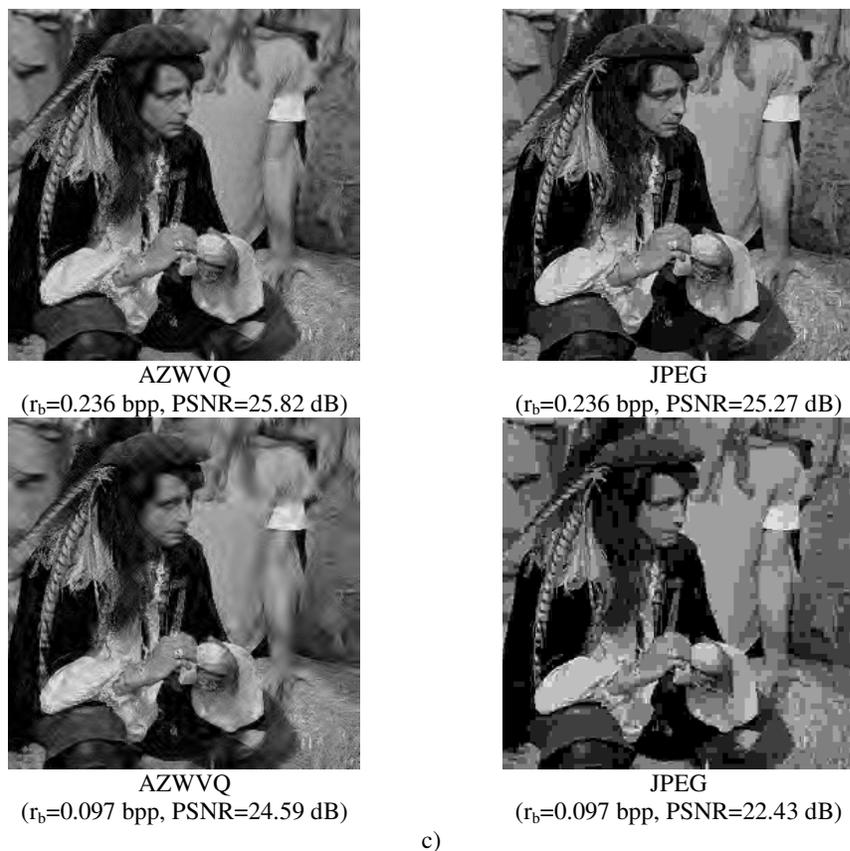


AZWVQ
($r_b=0.1$ bpp, PSNR=29.56 dB)



JPEG
($r_b=0.1$ bpp, PSNR=28.43 dB)

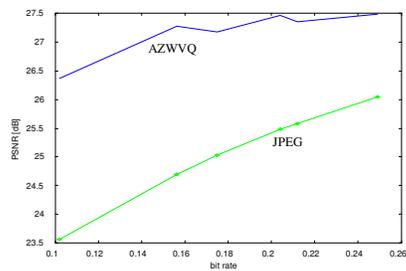
b)

**Figure 2.14**

Results of AZWVQ vs. JPEG at some compression ratios for
(a) *Baboon*, (b) *Lena* and (c) *Masquerade* images

Baboon

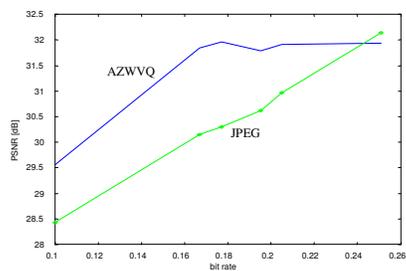
schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.249	32.03	27.49 dB	26.04 dB
b	0.212	37.60	27.36 dB	25.58 dB
c	0.204	39.06	27.46 dB	25.48 dB
d	0.175	45.51	27.18 dB	25.03 dB
e	0.156	51.02	27.28 dB	24.70 dB
f	0.102	77.83	26.37 dB	23.56 dB



a)

Lena

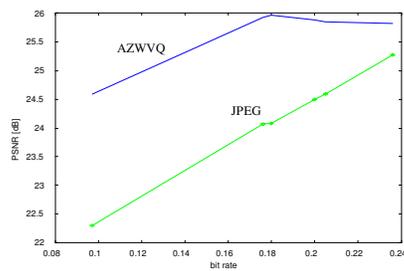
schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.251	31.77	31.94 dB	32.14 dB
b	0.205	38.87	31.91 dB	30.97 dB
c	0.195	40.86	31.79 dB	30.62 dB
d	0.177	45.00	31.96 dB	30.30 dB
e	0.167	47.68	31.84 dB	30.15 dB
f	0.100	80.00	29.56 dB	28.43 dB



b)

Masquerade

schema	bitrate [bpp]	CR	PSNR	
			AZWVQ	JPEG
a	0.236	33.79	25.82 dB	25.27 dB
b	0.205	38.87	25.84 dB	24.59 dB
c	0.200	40.00	25.88 dB	24.50 dB
d	0.176	45.25	25.93 dB	24.06 dB
e	0.180	44.44	25.97 dB	24.08 dB
f	0.097	81.81	24.59 dB	22.43 dB



c)

Table 2.10 Quantitative results of AZWVQ vs. JPEG at some compression ratios for (a) *Baboon*, (b) *Lena* and (c) *Masquerade* images

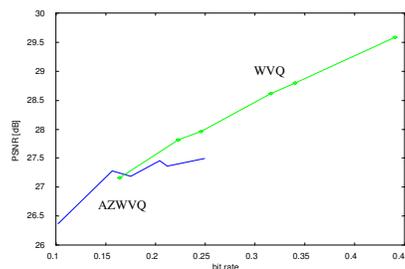
The performance of the AZWVQ with different masks for each orientation is still better than JPEG, especially at high levels of compression. The visual quality of the so-encoded images is even better than the found PSNR.

AZWVQ vs. WVQ

Comparisons between AZWVQ and WVQ with the use of three masks are reported in table 2.11. Again, AZWVQ gives better results than WVQ at the same compression ratios.

Baboon

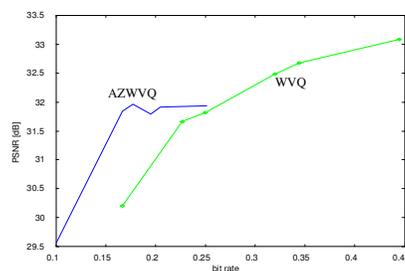
schema	AZWVQ			WVQ		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.249	32.03	27.49	0.441	18.125	29.586
b	0.212	37.60	27.36	0.340	23.5	28.797
c	0.204	39.06	27.46	0.316	25.29	28.615
d	0.175	45.51	27.18	0.246	32.47	27.959
e	0.156	51.02	27.28	0.223	35.82	27.81
f	0.102	77.83	26.37	0.164	48.85	27.16



a)

Lena

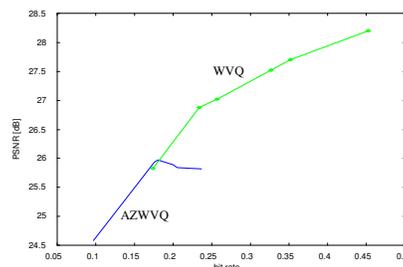
schema	AZWVQ			WVQ		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.251	31.77	31.94	0.445	18	33.077
b	0.205	38.87	31.91	0.344	23.26	32.68
c	0.195	40.86	31.79	0.32	25	32.485
d	0.177	45.00	31.96	0.25	32	31.824
e	0.167	47.68	31.84	0.227	35.24	31.664
f	0.100	80.00	29.56	0.167	47.96	30.204



b)

Masquerade

schema	AZWVQ			WVQ		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.236	33.79	25.82	0.452	17.7	28.209
b	0.205	38.87	25.84	0.351	22.79	27.70
c	0.200	40.00	25.88	0.327	24.47	27.52
d	0.176	45.25	25.93	0.257	31.13	27.02
e	0.180	44.44	25.97	0.234	34.19	26.87
f	0.097	81.81	24.59	0.174	46	25.83



c)

Table 2.11 Quantitative results of AZWVQ vs. WVQ at some compression ratios for (a) *Baboon*, (b) *Lena* and (c) *Masquerade* images

Images that do not belong to the training set

As already said, the performance of VQ is strongly influenced by the images used in for the generation of the codebook. In order to evaluate in a more complete manner the performance of the proposed method, results obtained on images that do not belong to the training set are reported in figure 2.15 and table 2.12 in comparison with JPEG.



AZWVQ
($r_b=0.248$ bpp, PSNR=24.61 dB)



JPEG
($r_b=0.248$ bpp, PSNR=27.37 dB)

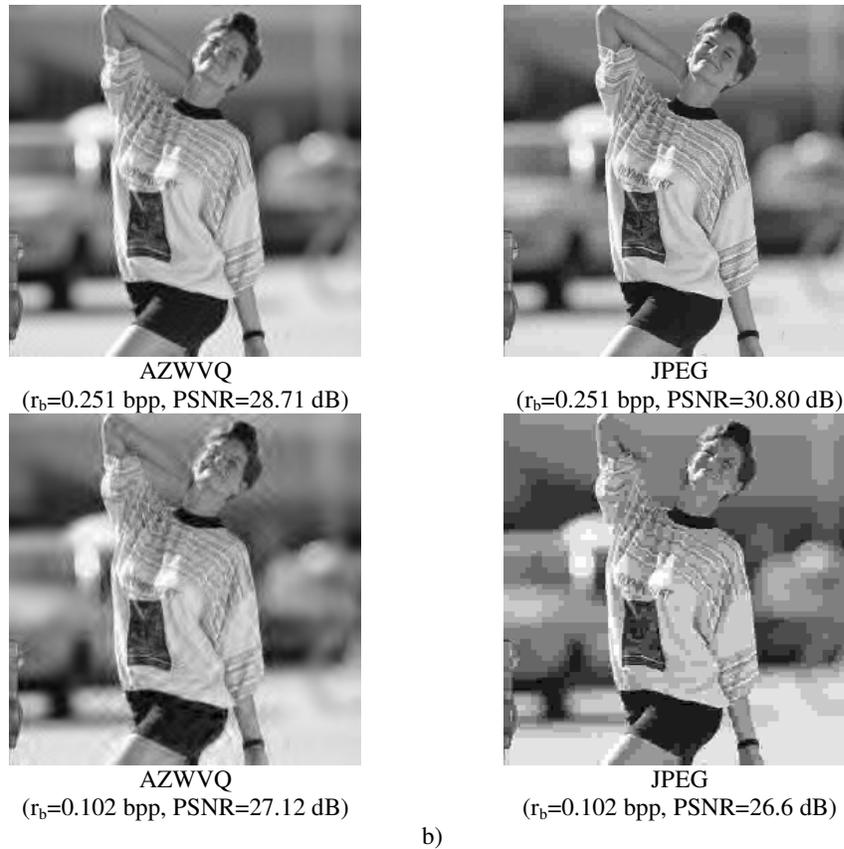


AZWVQ
($r_b=0.104$ bpp, PSNR=22.83 dB)



JPEG
($r_b=0.104$ bpp, PSNR=23 dB)

a)



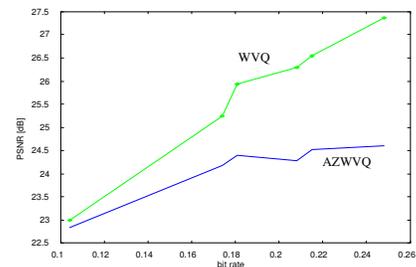
b)

Figure 2.15
Results of AZWVQ vs. JPEG at some compression ratios for (a) *Carmen* and (b) *Model* images

Carmen

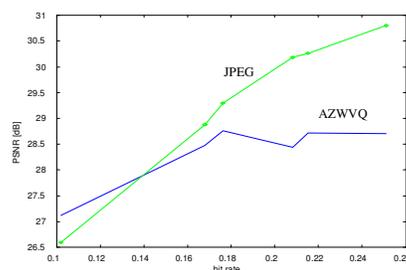
schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.248	32.16	24.61 dB	27.37 dB
b	0.215	37.20	24.52 dB	26.55 dB
c	0.208	38.46	24.28 dB	26.30 dB
d	0.181	44.20	24.39 dB	25.94 dB
e	0.174	45.98	24.17 dB	25.25 dB
f	0.104	76.93	22.83 dB	23.00 dB

a)



Model

schema	bitrate [bpp]	CR	PSNR AZWVQ	PSNR JPEG
a	0.251	31.87	28.71 dB	30.80 dB
b	0.215	37.21	28.72 dB	30.26 dB
c	0.208	38.46	28.44 dB	30.18 dB
d	0.176	45.45	28.76 dB	29.30 dB
e	0.168	47.62	28.48 dB	28.88 dB
f	0.102	78.43	27.12 dB	26.60 dB



b)

Table 2.12 Quantitative results of AZWVQ vs. JPEG at some compression ratios for (a) *Carmen* and (b) *Model* images

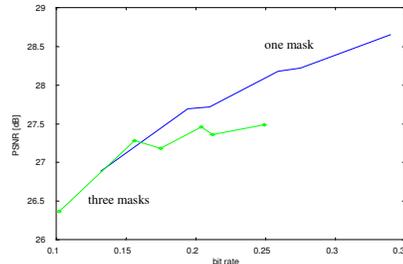
From the results, it arises that the AZWVQ results in a better quality than JPEG only at high compression levels. One application of the proposed method is then image coding at low bitrate. In order to increase the performance of the technique, it is necessary to enlarge the training set to have a wider statistics.

2.2.5.3 Comparisons between AZWVQ with one mask and with three masks

Finally, a comparison between the performance of the proposed method with the identification and use of a unique mask for all the orientations and with three different masks, specific for each orientation, is shown in table 2.13.

Baboon

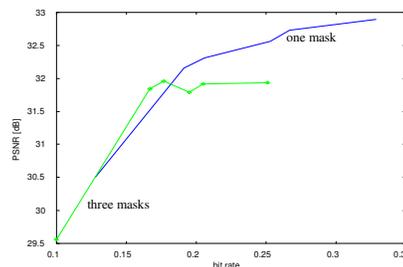
schema	One mask			Three masks		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.340	23.46	28.65	0.249	32.03	27.49
b	0.275	28.99	28.22	0.212	37.60	27.36
c	0.259	30.77	28.18	0.204	39.06	27.46
d	0.210	37.92	27.72	0.175	45.51	27.18
e	0.194	41.04	27.69	0.156	51.02	27.28
f	0.132	60.18	26.89	0.102	77.83	26.37



a)

Lena

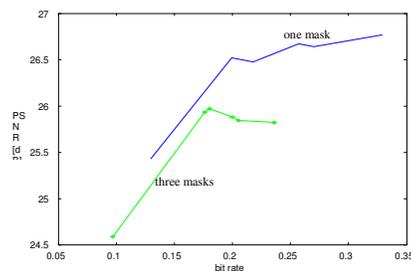
schema	One mask			Three masks		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.329	24.27	32.90	0.251	31.77	31.94
b	0.267	29.93	32.73	0.205	38.87	31.91
c	0.253	31.56	32.56	0.195	40.86	31.79
d	0.206	38.79	32.31	0.177	45.00	31.96
e	0.191	41.81	32.15	0.167	47.68	31.84
f	0.128	62.34	30.50	0.100	80.00	29.56



b)

Masquerade

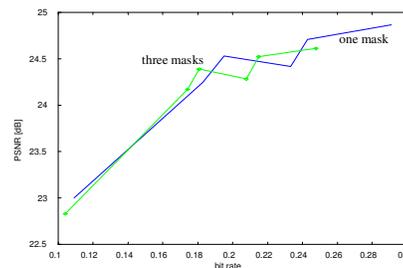
schema	One mask			Three masks		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.329	24.24	26.77	0.236	33.79	25.82
b	0.270	29.52	26.64	0.205	38.87	25.84
c	0.257	31.01	26.67	0.200	40.00	25.88
d	0.218	37.57	26.48	0.176	45.25	25.93
e	0.199	40.01	26.52	0.180	44.44	25.97
f	0.130	61.10	25.43	0.097	81.81	24.59



c)

Carmen

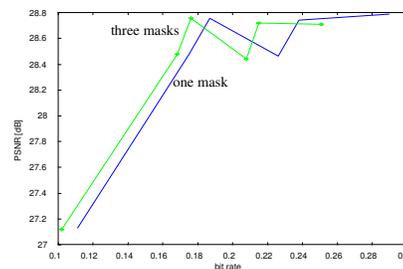
schema	One mask			Three masks		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.291	27.47	24.87	0.248	32.16	24.61
b	0.243	32.89	24.71	0.215	37.20	24.52
c	0.233	34.33	24.42	0.208	38.46	24.28
d	0.195	40.98	24.53	0.181	44.20	24.39
e	0.183	43.66	24.25	0.174	45.98	24.17
f	0.109	73.25	23.00	0.104	76.93	22.83



d)

Model

schema	One mask			Three masks		
	bitrate [bpp]	CR	PSNR [dB]	bitrate [bpp]	CR	PSNR [dB]
a	0.290	27.52	28.79	0.251	31.87	28.71
b	0.238	33.52	28.74	0.215	37.21	28.72
c	0.226	35.29	28.46	0.208	38.46	28.44
d	0.187	42.63	28.76	0.176	45.45	28.76
e	0.175	45.54	28.48	0.168	47.62	28.48
f	0.111	71.66	27.13	0.102	78.43	27.12



e)

Table 2.13 Quantitative results of AZWVQ with one mask vs. AZWVQ with three masks at some compression ratios for (a) *Baboon*, (b) *Lena*, (c) *Masquerade*, (d) *Carmen* and (e) *Model* images

AZWVQ with one unique mask gives better PSNRs at the same compression ratios than AZWVQ with three different masks. On the other hand, the use of three different masks, each specific for an orientation, allows the reaching of very high compression levels without an excessive degrading of the encoded images. As a matter of fact, for each orientation only regions characterized by a high energetic content in the specific orientation are encoded.

2.2.6 Conclusions

The performance of the described method (AZWVQ) has been compared with that of the standard JPEG and with that of the WVQ (i.e. vector quantization of all wavelet coefficients in each subband). The method was tested on a large variety of images.

In the first case the proposed method gives better results than JPEG, both in quantitative terms (PSNR) and in qualitative terms (there is no blockness distortion). In the second case the classification of active zones permits the reaching of higher compression levels and better quality.

The results obtained on two well-known images are reported in comparison with those of JPEG at the same bitrate.

Images encoded with the described method have not the blocking effect that is typical of the standard JPEG. There is a ringing effect in the borders caused by the loss of information in these zones, but, in spite of this, the images are more pleasant than JPEG ones.

In summary, the performance of the proposed approach is better than JPEG one, especially at low bitrates. High compression levels can be reached with this method without damaging the informative content of the images, because the classification of the active zones permits an accurate encoding of only a part of each subband, increasing the compression factor but maintaining a good quality in the reconstructed image.

2.3 Remote sensing data

2.3.1 Introduction

Remote sensing is comprised of all information acquisition techniques by means of a device that is not directly in contact with the object under analysis. However, usually this term refers to the investigation of the earth surface for several applications (meteorological, agricultural, pollution and, more in general, territorial control).

One of the major bottlenecks in remote sensing of the Earth's surface by satellites is the transmission of the obtained images to the ground station: it is often the case, particularly for low orbit satellites, that they remain in contact with a single tracking station for only a few minutes per orbit during which they can download only a very limited number of images, while they can capture and store hundreds in a single orbit. To solve this problem in an inexpensive way, one has to process and encode the images on-board so that they can be transmitted with fewer bits. Depending on the application for which these images are to be used, one may be interested only in a subset of the imaged regions. For example, in an agricultural application, one is not interested in urban and forested regions.

On the other hand, urban and forested regions appear in such images as highly textured regions and they require a large number of bits to be encoded.

The image coding task can be split into two successive steps, the first being the search for a hierarchy of messages, inside a set containing original information: the most important ones are to be coded with higher fidelity. The second step does not involve any analysis of original information, and is only devoted to efficiently coding extracted messages.

The classical image coding approach focuses on this second aspect of the problem, and many optimal codes to minimize the amount of information required for message transmission have been presented in the literature; moreover, this approach is the basis for the standards used in currently used apparatuses for image/video transmission. These first-generation techniques show good performance, in terms of low-distortion, at medium-high bit-rates, but present unacceptable degradations when stressed to low bit-rates (e.g., annoying blocking effects in DCT-based coding).

For some years, researchers addressed the problem of organizing hierarchically messages making up an image, i.e., they started to analyze the semantics of pictorial information [10,11]. Results obtained by this strategy show that pictures can be compressed even hundreds of times, while maintaining the overall semantics of a scene. Later on, other approaches aimed at mixing these two philosophies were presented in the literature (for instance, [9]), aimed at combining the best of both approaches and achieving good semantics in some image regions as well as good high fidelity in the other ones. Different approaches for adaptively coding an image may be used, as for instance two-sources coding using adaptive sampling and interpolation [8], but they are not tailored to a specific application where regions of interest may be defined by an end-user.

A region-based approach seems to be a good solution to the problem above described, and the last task is to choose the two coding methods that should be used for coding high-interest data or the background ones. The approach known as Clipping Service (which is now called Intelligent Bandwidth Compression) is using automatic target detection and recognition techniques to transmit or store only those parts of an image that are of interest [16]; Clipping Service is so named because the concept is similar to that of a newspaper clipping service that cuts out only those articles of interest and sends them to the customer.

Wavelet decomposition offers good capabilities for implementing both methods, as it allows the location of the relevant transformed coefficients for each defined region in an image. The separable wavelet transform has been used successfully by numerous authors for lossy image compression [17], [18], [19], and [20].

In this paper, we propose the segmentation and identification of such regions with the help of the texture boundary detection algorithm proposed in [12], and the subsequent encoding of the segmented regions independently from the rest.

As the identified regions are of irregular shape, all encoding methods that rely on block coding and in general 2D coding are excluded. We propose, therefore, the conversion of the regions to be encoded into 1D strings of data, which are encoded with the help of the 1D Wavelet transform.

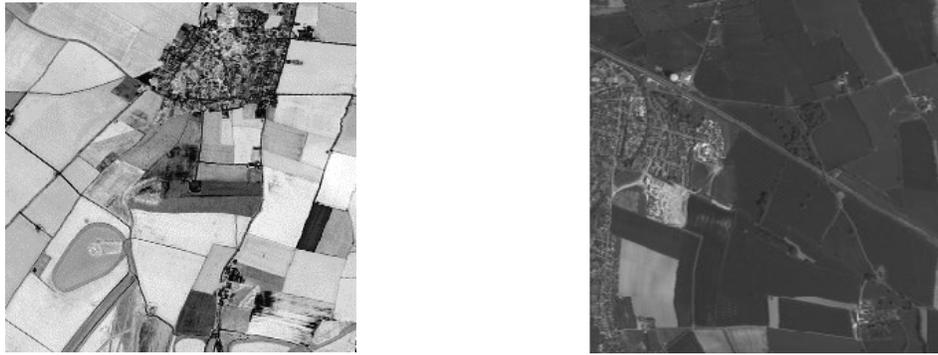


Fig.2.16. Original test images.

We perform experiments using 10m and 20m resolution images showed in Fig.2.16, which include some urban regions and we demonstrate the effectiveness of the method by comparing it to a no region-based algorithm based on a 2D wavelet code, and evaluating the improvement in the reconstruction achieved for fixed compression ratio. In this case the aim is to achieve as much compression as possible without significantly altering the image; of course “significant” is the keyword here. The determination of what constitutes a significant difference between an original image and a compressed and reconstructed version of the original is application dependent. However, if archiving is the goal, the future application may be unknown. Therefore, for the purposes of this paper, we will assume that archiving is the objective since it is the most general purpose application for compression. As a measure of quality of reconstruction we have used the well known computational metric of the minimum square error (MSE) or the equivalent peak signal to noise ratio (PSNR), with the definitions:

$$MSE = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{\gamma^2}{MSE} \right)$$

In section 2.3.2 we shall introduce the concept of free angle In section 2.3.3 we shall first present a review of the Wavelet transform and its statistical proprieties. We shall present the

details of our algorithm in section 2.3.4, our experimental results in section 2.3.5, and our conclusions in section 2.3.6.

2.3.2 The Free-Angle concept

Upon edge detection, a region that is highly textured is characterized by a high density of edge points. Surrounding non-textured regions, by contrast, show regular strings of edge points around well-defined objects. For example, in rural scenes, they may form the boundaries of fields or lakes. The information about edge location inside the textured region is not of interest to us, but we can use the difference in the edge characteristics between textured and non-textured regions to locate boundaries between them.

Consider one edge pixel in the edge map produced by edge detector. We can consider its “view” in the two-dimensional (2D) world of the image plane, where other edge pixels are obstructions. The edge pixels in a highly textured region are therefore like trees in a forest, whose field of view is severely limited. Edge pixels in more open regions of the image, such as in a arable land, will be able to “see” much more openly. To express this mathematically, we introduce the concept of the free angle. This is an angle with its vertex at an edge pixel, which is equal to the field of view of the edge pixel in any particular direction. It is the angle formed by lines joining the edge pixel with two successive neighboring edge pixel once they have been ranked in order of rotation angle about the edge pixel. An edge pixel, therefore, may have associated with it several free angles, possibly a different one for “looking” through any pair of successive neighbors. We choose the largest of these angles and assign it to the edge pixel. This angle is expected to be large for edge pixels that happen to be in open regions of the image and small for those in the middle of textured regions.

In principle, the edge pixels that restrict a field of view can be arbitrarily far away in the image plane. In the definition of the free angle, it is included the concept of a neighborhood. Thus, to compute the free angle, we have first to define the neighborhood of the pixel of interest. The maximum free angle at an edge pixel gives the largest field of view in any direction that is not obscured by edge pixels within its neighborhood.

We define a neighborhood of a pixel as a square region of the image of size $(2N+1) \times (2N+1)$ pixels centered on that pixel. We define the set of pixels associated with a neighborhood of a given size N by C_N . We may characterize the location of a pixel in the neighborhood C_N by its Cartesian coordinates, but for our purposes, it is much more convenient to use polar coordinates in the image plane (ρ, θ) . We define these polar coordinates in a square-based region rather than a circular region. The radius ρ is measured outwards from the central pixel; therefore, if a pixel $P \in C_N$ and has coordinates (ρ, θ) , then $\rho \leq N$. It is also follows that if pixel P has a radius ρ , then $P \in C_\rho$, but $P \notin C_{\rho-1}$. The angle θ is also an integer measured around the square region from the top left corner ($\theta = 0$) in a counter-clockwise sense. For neighborhood pixels of a fixed ρ , $0 \leq \theta \leq 8\rho$. These square-based polar coordinates are defined in Fig. 3 (a). Fig. 3 (b) shows the (ρ, θ) coordinates of neighboring pixels.

It is sometimes convenient to relate (ρ, θ) coordinates to (x, y) offsets from the central pixel. We can compute the square polar coordinates from the offsets by the relations

$$\rho = \max(|x|, |y|); \theta = 2\rho + x + y \quad \text{if } x \leq y; \theta = 6\rho - x - y \quad \text{otherwise}$$

The inverse formulae are a little more involved but straightforward to compute.

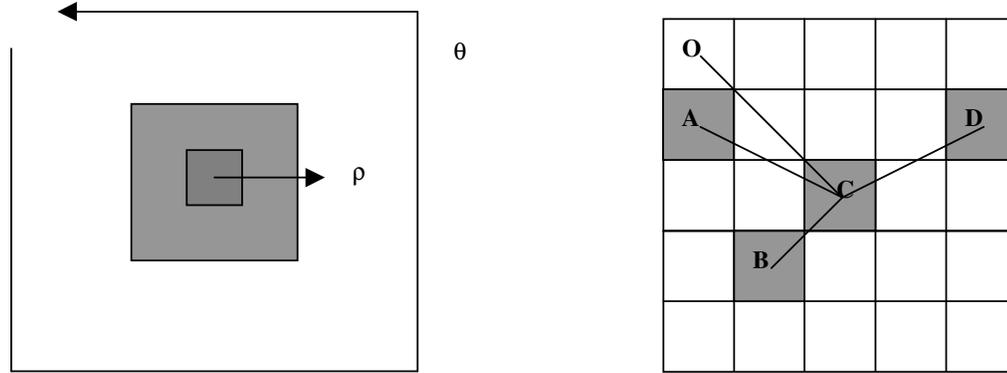


Fig.2.17. (a) Neighborhood around an edge pixel and the square polar coordinates (ρ, θ) that define pixel locations within this neighborhood. As ρ increases so the shading becomes lighter in color. (b) ρ and θ coordinates of boundary pixels in the $N = 2$ neighborhood of pixel C. This pixel has three neighboring boundary pixels. There are three free angles - ABC, BCD and DCA.

Consider now the computation of the free angles. In Fig. 2.17 (b), we show the neighborhood of an edge pixel. It is clear from our definitions above that the free angle is simply the difference in the scaled angles θ/ρ associated with two successive neighbors when the neighbors are ranked in order of increasing θ . It is clear that there are three free angles associated with the pixel C in Fig. 2.17 (b). The angles are ABC, BCD and DCA. Simple arithmetic shows that these free angles are 1.5, 3.5 and 3.0, respectively. Therefore, the maximum free angle is BCD.

From Fig. 2.17 (b), it is evident that edge pixels on the boundary of a textured region will have a large maximum free angle in the direction looking away from the texture. Edge pixels inside the textured region, however, will all have a small free angle, and therefore, they can be removed by thresholding the maximum free angle. We note, however, that the size of the neighborhood plays a crucial role in determining the maximum free angle.

2.3.3 Filter Bank 1D wavelet implementation

One dimensional discrete wavelet transform can be described in terms of a filter bank as shown in Fig.4. An input signal $x(n)$ is input to the analysis low-pass filter $h_0(n)$ and the analysis high-pass filter $h_1(n)$. The odd samples of the outputs of these filters are discarded, corresponding to decimation by a factor two. The decimated outputs of these filters constitute the reference signal $r_1(n)$ and the detail signal $d_1(n)$ for a one level decomposition. For reconstruction, interpolation by a factor two is performed, followed by filtering using the low-pass and high-pass synthesis filters $g_0(n)$ and $g_1(n)$ as shown. Provided that the system satisfies the perfect reconstruction property, the sum of the outputs synthesis filters will give $y(n) = Ax(n - n_d)$ where A is a gain factor and n_d an odd delay.

For a multilevel decomposition, the reference signal $r_1(n)$ serves as the input to a filter bank whose analysis stage is identical to that of Fig.4. This process, which is repeated iteratively as shown in Fig.5 provides, after L levels, a reference signal $r_L(n)$ with resolution reduced by factor 2^L with respect to the original input $x(n)$ as well as the detail signals $d_L(n), d_{L-1}(n), \dots, d_1(n)$. Each detail signal $d_i(n)$ contains precisely the information that,

together with the reference signal $r_i(n)$, enables reconstruction of $r_{i-1}(n)$, which is the reference signal at the next higher resolution.

One can associate a continuous scaling function and a wavelet with a multilevel analysis filter bank. A generally different scaling function and wavelet are associated with the synthesis stage. References [23] [24] contain a through treatment of the relationship between the filter coefficients and scaling functions.

Signal compression can be obtained by wavelet transformation of integer input data followed by quantification and coding. As the quantification is usually lossy, the whole compression scheme is lossy too. The wavelet transform allows implementation of multiresolution sub-band compression schemes, in which the decompressed data are gradually refined, retaining the option of perfect reconstruction [15].

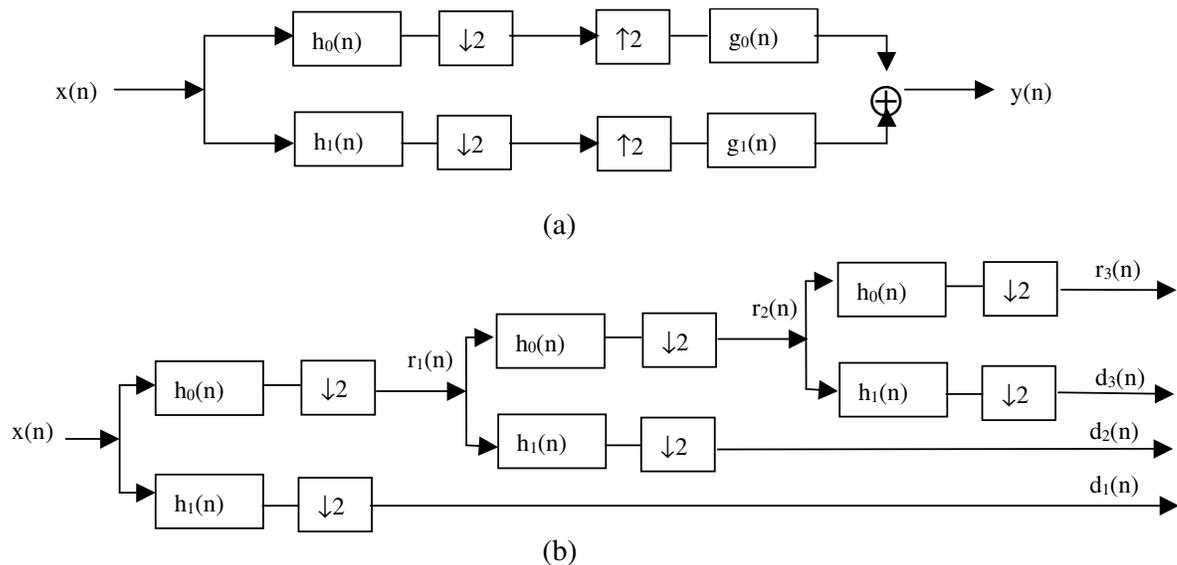


Figure 2.18. (a) Basic filter bank for wavelet transform (b) Tree structure showing three levels decomposition

2.3.4 Outline of the encoding algorithm

The textured regions of the image are first identified using the texture boundary detection algorithm described in [12]. This algorithm works on the basis of the free angle of a detected edge, and it does not use any other information concerning the textured region apart from the fact that textured regions are characterized by high density of edges. The free angle simply measures the maximum angle of the field of view an edge has through its neighboring edges. Edges that form texture boundaries have higher values of the free angle than edges interior to the texture. Some post-processing described in [12] allows one to extract only the closed boundaries of the textured regions, and not edgels along boundaries of non-textured regions. The result of the application of the above mentioned detection algorithm to the two test images is shown in Fig.2.19.

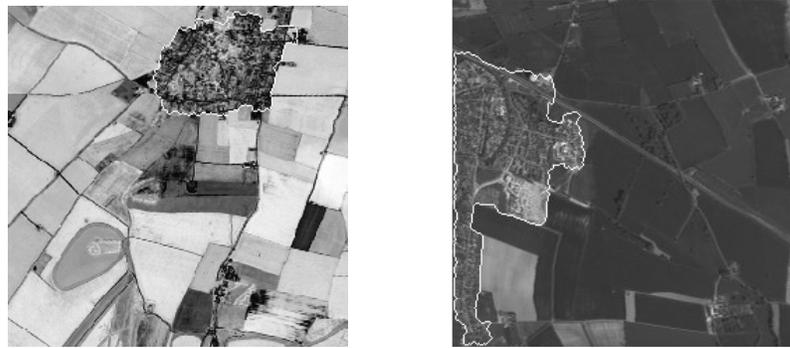


Fig. 2.19. Textured region extraction.

The extracted textured regions are subsequently treated individually; for each region the minimum enclosing box is identified, and the pixels of the region are flagged, using a simple region filling algorithm. The minimum enclosing box is then scanned in a raster fashion and all flagged pixels are read sequentially to form a string of numbers that has either to be carefully encoded or given a flat value, depending on whether we are interested in the textured regions or the background. In the examples presented below, the pixels of the textured regions are given the mean gray value of the region, and the background is encoded carefully. The whole process is shown in Fig.2.20.

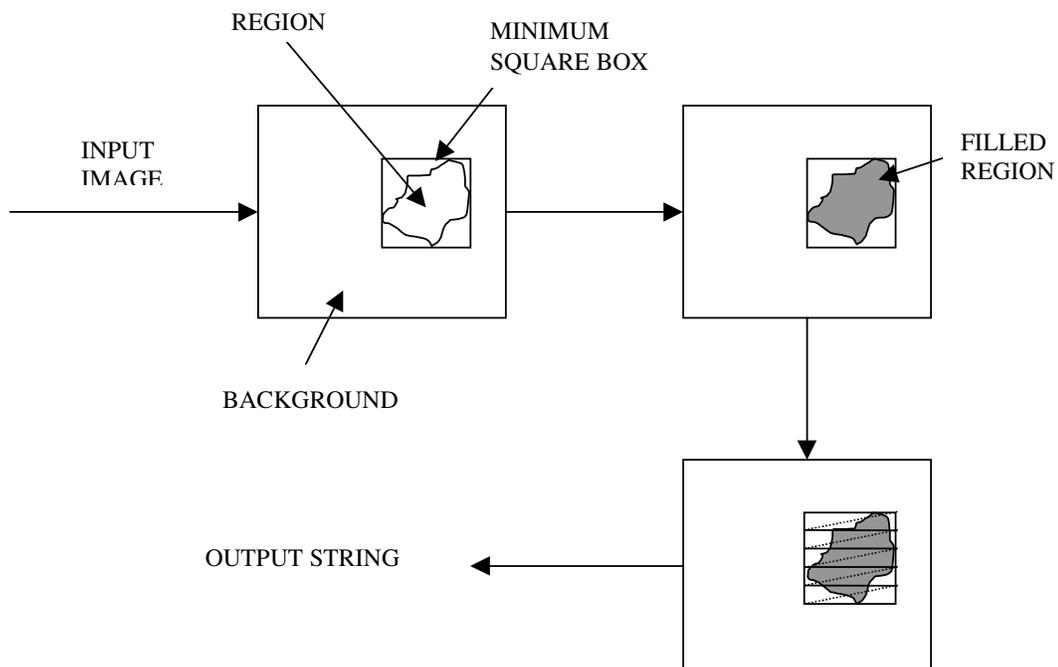


Fig. 2.20. Outline of the processing steps.

The outlines of the textured regions are encoded using Freeman's chain code and transmitted separately to the receiving station. The background pixels are then read sequentially to form a string of data that is encoded using 1D wavelet expansion using Mallat filters [14]. Three levels of expansion and sub-sampling are used. The final low pass component is assumed to

contain most of the information, so it is encoded in a lossless way, using DPCM and the Huffman codes. The successive high pass components are encoded in a lossy way [13]: For each level of resolution the histogram of the wavelet coefficients is created. This histogram is quantized as follows: The weakest 95% of the coefficients are given the mean which is near to 0 value. All other coefficients are encoded by means of a VQ scheme: an alphabet of 16 vectors is created by applying the LBG algorithm [21] to a set of remotely sensed gray level images.

2.3.5 Active zone classification

The statistical distribution of wavelet coefficients in each subband is a nearly Gaussian distribution with zero mean and small variance. This generic property of wavelet coefficients allows the distinction, in each subband, of a zone characterized by greater energetic and informative contents (i.e. an *active zone*). For this zone the coding process must be more accurate in order to guarantee for a better reconstructed image.

After an extensive experimental analysis on different images, a heuristic algorithm for identifying these zones has been designed. In each subband an accurate analysis of the quantized wavelet coefficients histogram is performed in order to apply a thresholding process to the histogram. The histogram is first regularized, to avoid falling into a local minimum during the thresholds searching, then its maximum value is found. In fact, since the histogram can be modeled as unimodal, with the maximum usually located around the zero value, two thresholds can be found proceeding into the two directions starting from this maximum, so that they include a histogram area around the 95% of the total. The search of these thresholds is made adaptively, in the sense that at each step the threshold, whose movement minimizes the variation of the included area, is shifted.

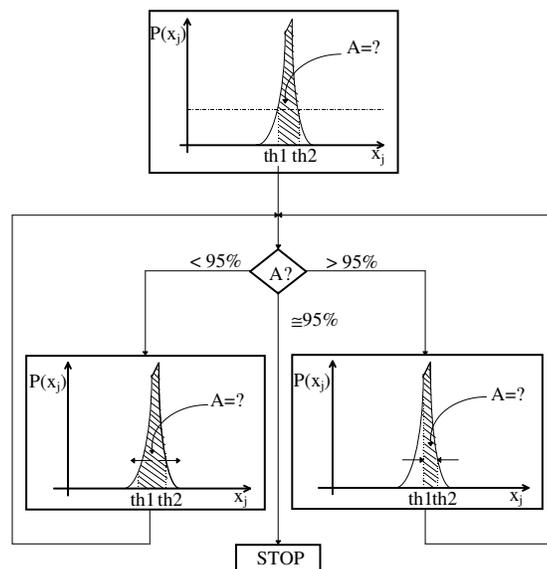


Fig.2.21 The threshold searching algorithm

At the end of this process the distinction between active zone and not, is made in this way: pixels which have values included between the two thresholds found are classified as belonging to the non-active zone, because their values are near the mean value and they constitute an enough uniform zone of the image, while pixels whose values are not included

between the two thresholds are classified as belonging to the active zone, because their values constitute the distribution tails and are quite different from the mean value.

2.3.6 Mask extraction

After the classification process, for each subband a binary-value “mask”, which contains the information relative to the position of the active zone, is extracted. The mask is then scaled down and logically summed with the one obtained with the same procedure at the lower resolution and at the same orientation. At the end, three masks, each of which contains the information regarding the active zones of the subbands at the same orientation, are extracted. In alternative, a unique mask can be extracted by the logic sum of these three ones.

The final mask (or the final masks) is encoded by quadtree or optimized run length, according to convenience.

2.3.7 Wavelet coefficient encoding

For each subband the coefficients belonging to the non-active zone are not coded because their informative content is very low (they constitute the background of the subband). The coefficients of the active zone, on the other hand, have the main part of the energetic and informative content of the subband, so their encoding must be accurate.

These coefficients are coded by vector quantization with parameters chosen from energetic considerations. In fact, as the resolution reduces, the energetic content of a subband increases and thus quantization must be more accurate.

The parameters of the vector quantization (codebooks and codevectors dimensions) are chosen accordingly to the variation of MSE value during the LBG algorithm. Codebooks are specific for each subband (at each resolution) and are generated from the active zones of the subbands of the images belonging to the training set.

2.3.8 Experimental results

Figs. 2.22 and 2.23 show the reconstruction of the image when the whole image is encoded using 2D wavelet decomposition, and the reconstruction of the background when the textured regions are isolated and encoded separately. Both images are encoded at the same compression ratio.

The MSE calculated when the whole image is encoded is, for the image in Fig.2.22, 25.19 and for the image in Fig.2.23, 68.1. The MSE on the other hand when the textured regions are removed is 16.00 and 12.87 respectively.

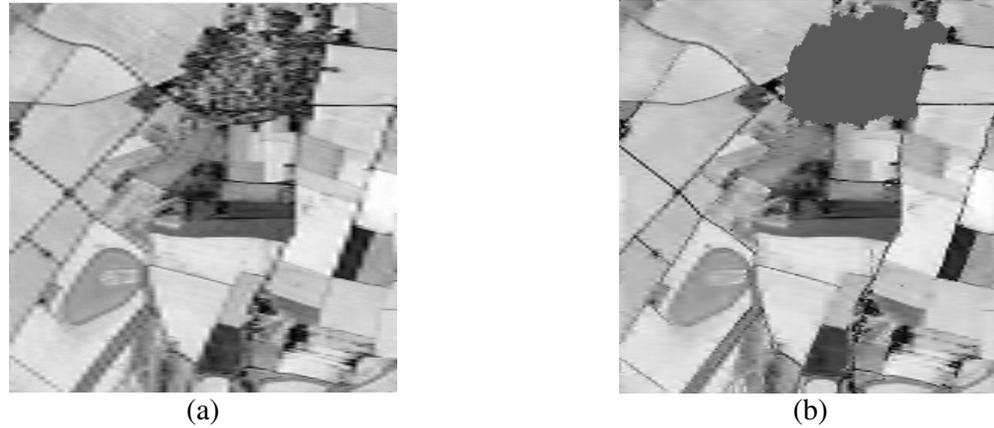


Fig.2.22. Images compressed at CR=20.



Fig.2.23. Images compressed for CR=22.

2.3.9 Conclusions

We propose here a scheme for the region-based encoding of remote sensing images. The scheme is based on the observation that textured regions require a large number of bits to be encoded and they are often of no interest to a large number of applications. The scheme relies on the use of a texture boundary detection algorithm which can identify reliably the regions of a textured region irrespective of the class of the texture [12]. The outline of the textured region can then be encoded and transmitted separately and the rest of the image can be encoded on its own. Because the encoded image of interest will have irregular (non-rectangular) shape, and in order to preserve the detail with which the textured regions have been extracted, we use 1D coding of the string that is created by reading the pixels of the regions of interest sequentially in a raster format. We showed that the 1D wavelet transform can be used to encode the images with improved compression ratio. The scheme is very simple and can be implemented in a completely automatic way, so that it is appropriate for on-board satellite implementation.

References

- [1] P.H.Westerink, J.Biemon, D.E.Boekee, J.W.Woods, "Subband coding of images using vector quantization," IEEE Trans. Comm., Vol.36, N.6, pp.713-719, 1988.
- [2] M.Antonini, M.Barlaud, I.Daubechies, P.Mathieu, "Image coding using vector quantization in the wavelet transform domain," Proc. ICASSP, pp.2297-2300, 1990.
- [3] M.Ancis, D.D.Giusto, "Image data compression by adaptive vector quantization of classified wavelet coefficients," Proc. IEEE PACRIM Conference, pp. 330-333, Victoria, Canada, 1997
- [4] A.N.Netravali, B.G.Haskell, "Digital picture: representation, compression and standards," Plenum Press, 1995.
- [5] H. Samet "A top-down quadtree traversal algorithm," IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-7, 1984.
- [6] R.M. Gray, "Vector quantization," IEEE ASSSP Magazine, pp.4-28, 1984.
- [7] Y. Linde, A. Buzo, R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Comm., Vol. 28, No. 1, pp. 84-95, 1980.
- [8] F.G.B.DeNatale, G.S.Desoli, D.D.Giusto, "Adaptive sampling and interpolation for image data compression," Journal of Visual Communication and Image Representation, Vol.5, no.4, pp.388-402, 1994.
- [9] F.G.B.DeNatale, G.S.Desoli, D.D.Giusto, G.Vernazza, "Polynomial approximation and vector quantization: A region-based integration," IEEE Trans. Communications, Vol.43, pp.198-206, 1995.
- [10] M.Kunt, A.Ikonomopoulos, M.Kocher, "Second-generation image coding techniques," Proc. IEEE, Vol.73, pp.549-575, 1985.
- [11] M.Kunt, M.Benard, R.Leonardi, "Recent results in high compression image coding," IEEE Trans. Circuits and Systems, Vol.34, pp.1306-1336, 1987.
- [12] P.L.Palmer, M.Petrou, "Locating boundaries of textured regions," IEEE Trans. Geoscience and Remote Sensing, Vol.35, pp.1367-1371, 1997.
- [13] S.Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Analysis and Machine Intelligence, 1989.
- [14] S.Dewitte, J.Cornelis, "Lossless integer wavelets transform," IEEE Signal Processing Letters, 1997.

- [15] R.D.Chaney, E.J.VanAllen, D.E.Dudgeon, "Reduction of communication requirements for wide-area surveillance systems," Proc. SPIE, Vol.2755 (Signal Processing, Sensor Fusion, and Target Recognition V, I.Kadar, V.Libby, Eds.), 1996.
- [16] M.Antonini, M.Barlaud, P.Mathieu, I.Daubechies, "Image coding using wavelet transform," IEEE Trans. Image Processing, Vol.1, pp.205-220, 1992.
- [17] A.Said, W.A.Pearlman, "A new fast and efficient image code based on a set partitioning in hierarchical trees," IEEE Trans. Circuits and Systems for Video Technology, Vol.6l, pp.243-250, 1996.
- [18] J.M.Shapiro, "Embadded image coding using zerotrees of wavelet coefficients," IEEE Trans. Signal Processing, Vol.41, pp.3445-3462, 1993.
- [19] P.Sriram, M.W.Marcellin, "Image coding using wavelet transform and entropy-constrained trellis-coded quantization," IEEE Trans. Image Processing, Vol.4, pp.725-733, 1995.
- [20] P.H.Westerink, "Subband coding of images," Ph.D. Thesis, Delft University of Technology, 1989.
- [21] M.Vetterli, J.Kovacevic, Wavelets and subband coding, Prentice Hall, 1995.
- [22] A.Cohen, I.Daubechies, J.C.Feaveau, "Biorthogonal bases of compactly supported wavelets," AT&T Bell Labs Tech. Rep., 1990.
- [23] B.S.Everitt, G.Dunn, "Applied multivariate data analysis," Hodden and Stoghton, 1991.
- [24] A.Grossman, J.Morlet, "Decomposition of handly functions into square integrable wavelets of constant shape," *J.Math.Phys.*, Vol.26, pp.2473-2479, 1985.
- [25] I.Daubechies, "Ten lectures on wavelets," CBMs-NFS, Regional Conferences Series in Applied Math., 1992.
- [26] I.Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE Trans. on Information Theory, Vol.36, no.5, pp.961-1005, 1990.

Chapter 3

WAVELETS AND DATA ZOOMING

5.1 Introduction

Nowadays, the surprising diffusion of multimedia applications (from scientific to commercial, from informative to recreational) has caused an increasing joint development of both hardware and software technologies. Furthermore, this sudden and fast spreading process has generated a large interest of the scientific community in the signal processing field, due to the necessity of dealing with large amounts of data in machines with limited computational and storage capacities. Mainly, the problem arises with multimedia-based content signals (i.e. videos). The common standards developed for digital image and video applications (JPEG, MPEG) seem to have given a solution to the problem, achieving, with added computational complexity, a good trade-off between quality and payload.

The present work proposes an alternative approach to the video database management by using a novel and promising approach exploiting iterated function systems (I.F.S.), usually called *fractals*. Still relatively immature, this technique has been applied as a compression technique for image coding. The aim of this paper is to show how the fractal coding technique can be successfully extended to the video processing. Actually, it has been developed a technique performing a zoom of the video sequence both in the spatial and in the temporal domain, in order to reconstruct the entire sequence by starting from some selected Key-Frames. These are a sample of frames representative of the whole video, which are extracted by a temporal sub-sampling process and selected by an opportune criterion. A three-dimensional fractal coding is applied to the key frames and, during the decoding phase, both a spatial zoom, with overlapped range blocks, and an adaptive temporal zoom are realized in order to reconstruct a sequence classified as an approximation of the original.

In order to reduce the computational complexity of the fractal encoder, the wavelet image representation has been employed: the video sequence is decomposed into several packets with frame dimensions lower than the original and depending, as well as the number of frames deployed, on the level of wavelet decomposition. The fractal encoder, therefore, does not deal anymore with video sequences at the original sizes, but with thumbnails of lower dimensions. Thus, the searching space for the domain blocks is drastically reduced, with optimal consequences from the computational point of view. Moreover, by discarding video packets in the high frequencies bands, it is possible to obtain a strong reduction of the number of pixels elaborated by the encoder, with a relatively low loss of information. As a matter of fact, the main advantage of the use of the wavelet representation is exactly due to its similarity to the human visual system (H.V.S.), by representing images at different levels of detail in different frequency bands. The most of the coding techniques based on this representation

exploits exactly this property by discarding or hardly quantizing the high frequencies information, since the resulting visual quality is only slightly affected by this quantization error. Such a feature has made possible, a decreasing up to the 25% of the number of pixels to be processed. Another advantage that comes from the wavelet decomposition of the packets is that it allows a parallelization of the whole process thus fully exploiting the CODEC potentiality, in case a parallel computer is available. On the whole, a large decrease of the coding/decoding time is obtained with a low loss of visual quality.

In order to reduce the unpleasant blocking effect due to the use of separate range blocks in images with low spatial correlation between adjacent blocks, *overlapped range blocks* (O.R.B.) have been used during the coding and an opportune processing with *ordered square overlapping* (O.S.O.) during the decoding phase. A partial overlapping between blocks and an opportune processing of the shared parts ensure the absence of blocking effect.

Finally, a technique to detect a mask of “active zones” has been performed. These zones locate the background of the video sequence and separate it from the regions that are not static from a temporal point of view. The background of the scene is encoded once for each group of frames (G.O.P.), relieving the encoder of the computational burden deriving from the encoding of these regions frame by frame. So, since the sequence of the images containing the active zones does not contain the background, the actual number of pixels to process is hardly reduced proportionally to the dimensions of the background and to the number of frames to encode. Finally, in order to obtain the complete image, the background, opportunely adapted according to the spatial zoom factor, is inserted into each frame for each GOP. The whole process is done only if the area of the regions constituting the background is large enough, otherwise it is avoided.

Several tests have been done with different sequences and the performance of the proposed method has been estimated from objective and subjective points of view.

3.2 The fractal coding

Coding techniques based on the fractal theory are new and promising approaches for digital image compression.

The general aim in the digital image compression field is to find coding algorithms that generate the required compression ratio and the required image quality with an acceptable encoding and decoding time. The most popular approach consists in the transformation of the image in a different domain (frequency domain), in order to discard the higher frequency coefficients and to use only the low frequency coefficients to describe the original image. The approach based on the fractal theory is radically different and is very promising even if still at a first development.

The idea derives from a work by Barnsley [1], who first worked with IFS. Then Jacquin [11, 12] proposed a fractal image encoder. The basic idea of this approach is to exploit the redundancy given by the self-similarity always contained in all natural images. The purpose is to find a rule to reconstruct an image by approximating the original one.

Many compression techniques are based on the exploitation of the correlation between pixels neighboring in the spatial or the time domain. The fractal compression approach exploits the correlation between not adjacent image parts. The fractal image can be seen as a collage constituted by copies of the original image that have been transformed through opportune geometric and luminance transformations. The novel approach by Jacquin [11] permits to see the image as composed by copies of parts of the image itself. The process consists in the decomposition of the image into regular distinct blocks and in the search, for each block, of another part which matches the block itself. The CODEC is based on the realization of a

specific transformation (*fractal code*) that, once iteratively applied on an arbitrary initial image, produces a sequence of images that converge more and more to an approximation of the original image. Moreover, the use of IFS allows not only the data compression, but also the down- or over-sampling of the data themselves, by exploiting the properties of the fractals.

The reduction of redundancy is obtained by representing the image through contract parts of itself: the image to be processed is partitioned into non-overlapping regular regions, named *range blocks*. The task of the encoder is to find for each range block a larger block in the same image, named *domain block*, such that, after an opportune transformation, this constitutes a good approximation of the range block. This transformation is a combination of geometric and luminance transformations. In the decoding phase, the process is really very fast: starting from an arbitrary image, the convergence to the original image is obtained through an iterative process with a velocity that could assure real time applications. The main problem of this new approach is the long encoding time, due to the complexity of the basis algorithm. This complexity has caused in the past the poor use of fractal encoders for the lower computation potentialities of the past computer generations. Several trials have been made in order to reduce the complexity of the fractal encoder, and thus the encoding time. Further attempts have been done in order to apply the fractal encoding technique in other fields, for example in musical signals, color images and image sequences processing [8, 15].

The mathematical foundation of this technique is the general theory of contractive iterated transformations, based on the work by Barnsley [1, 3, 10]. Fractal coding of an image consists in building a particular transformation τ (a code) such that, if μ_{orig} is the original image, then $\mu_{\text{orig}} \approx \tau(\mu_{\text{orig}})$, that is, μ_{orig} is approximately self-transforming under τ . If τ is a contractive transformation, μ_{orig} is approximately the attractor of τ , that is $\mu_{\text{orig}} \approx \lim_{k \rightarrow \infty} \tau^k(\mu_0)$ for some initial image μ_0 . The code τ is built on a partition of the original image. Each block of this partition R_i (*range block*) is coded independently of the others by a matching with another block D_i (*domain block*) of the image.

In [11] a fractal code $\tau(\mu_{\text{orig}}; R, n, p_x, p_y, \alpha_m)$ is obtained by using the following parameters:

- R is the range block size (squared range blocks) and D the domain block size;
- n is the scale factor for the local self-similarity search, so $D = nR$;
- p_x and p_y are respectively the horizontal and vertical step which define the set of all the domain blocks for an exhaustive research. The coordinates of each domain block's upper left corner are then expressed by (ip_x, jp_y) , where i and j are integers.
- α_m is the upper boundary on the scales. If $\tau_i: D_i \rightarrow R_i$ and $\tau_i = M_i \circ I_i \circ r_{i,n}$, with $M_i(x) = a_i x + b_i$ an affine operator with scale a_i and shift b_i on the luminance of the pixels, I_i a transformation selected from eight discrete isometries and $r_{i,n}$ a reduction by a factor n

using an averaging, than $\tau = \bigcup_{i=1}^N \tau_i$. So $0 \leq \alpha_i < \alpha_m \forall i \in \{1, 2, \dots, N\}$ during the coding stage.

So, $r_{i,2}$ operates the rescaling: thus, for example, a 2×2 pixels block is obtained from a 4×4 one by averaging its values. With I_i one among eight different isometries is applied (identity, horizontal and vertical flipping, reflection of 180° around the principal and the secondary diagonal, clockwise and anti-clockwise rotation of 90° , rotation of 180°). Finally, M_i locates opportunely the original range block. τ_i is the global transformation ($r_{i,2} + I_i + M_i$) that minimizes the error:

$$E(D, R) = \min \|R - (aD + B)\|$$

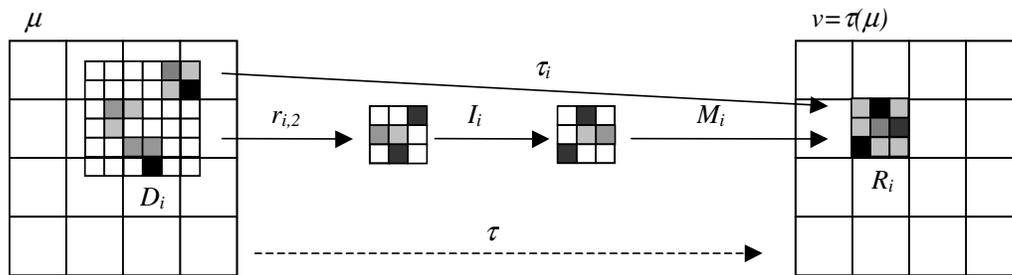


Figure 1 Fractal coding stage

Since the domain block could be located wherever in the image, a long encoding time can be necessary. In order to reduce the complexity of this searching process, a characteristic space, named *feature space* [18], is considered.

Each *domain* and *range* block is pre-classified according to a set of *invariant features* and the range blocks are compared only with the near domain blocks in a n-dimensional feature space, as indicated in figure 2.

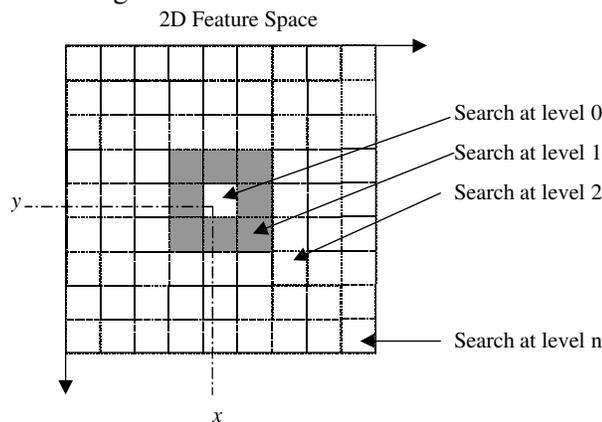


Figure 2 Research and construction in a bidimensional feature space: given a range block identified by two coordinates (x,y), the corresponding domain block is searched in a limited region, which depends on the level of the research, between 0 and n.

Jaquin used a classification scheme derived from the work by Ramamurthi e Gersho [20]. The domain blocks are classified according to their figurative characteristics. Three classes of domain blocks are considered: shade blocks, edge blocks (simple and mixed) and midrange blocks. The classification of the blocks can ever be done through a measure of their variance. Then, since range blocks that can be classified as shade blocks can be well approximated by uniform blocks, it is no more necessary for them to search for a domain block, thus reducing the search only for the other two classes, with a notable decrease of the computational complexity.

A more complex classification technique has been proposed by Fisher and Jacobs [11]. A domain or range block is divided into four parts (upper left, upper right, lower left and lower right), in each of which the mean A_i and the variance V_i are calculated. Every block can be oriented in such a way that the means can be in one of the following configurations: $A_1 \geq A_2 \geq A_3 \geq A_4$, $A_1 \geq A_2 \geq A_4 \geq A_3$ or $A_1 \geq A_4 \geq A_2 \geq A_3$. Once the orientation of the block has been determined, there are 24 possible ways to order the variance. So, for each range block, the research is done in 2 subclasses among 72. Even if this approach has been successful, it is not very satisfying since it does not say anything about the vicinity

among the classes, and if the research in a class does not give a good matching, the extension to the near classes is not very simple.

A solution to this problem has been given by Caso, Obrador and Kuo [3], who substitute a vector of variances to the rigid variances ordering. In this method the variances are rigorously quantized and a set of classes is obtained so that each class has a near one where the research can be easily extended.

3.3 Classical fractal decoder

The classical fractal decoding stage consists simply on an iterated process starting from an initial image μ_0 (for example constituted by black pixels). In fact, if τ is a contractive transformation, $\tau^\infty(\mu_0)$ gives an approximation of the original image μ_{orig} .

By denoting ε_c the coding error, the reconstruction error is bounded, according to the Collage theorem, by the following expression:

$$\varepsilon_r \leq \frac{\varepsilon_c}{1-s} \quad (1)$$

where $\varepsilon_c = d_2(\mu_{\text{orig}}, \tau(\mu_{\text{orig}}))$, $\varepsilon_r = d_2(\mu_{\text{orig}}, \tau^\infty(\mu_0))$ and s the contractivity of the transformation τ . In this context the contractivity constraint is expressed as follows: for all images u and v , $d_2(\tau(u), \tau(v)) \leq s d_2(u, v)$, with $s \in [0, 1[$ and d_2 the Euclidean metric.

It has been noticed that 8-10 iterations are sufficient to find a good approximation and this number represents a good compromise between computational cost and perceptual quality and this does not depend on the original image μ_0 .

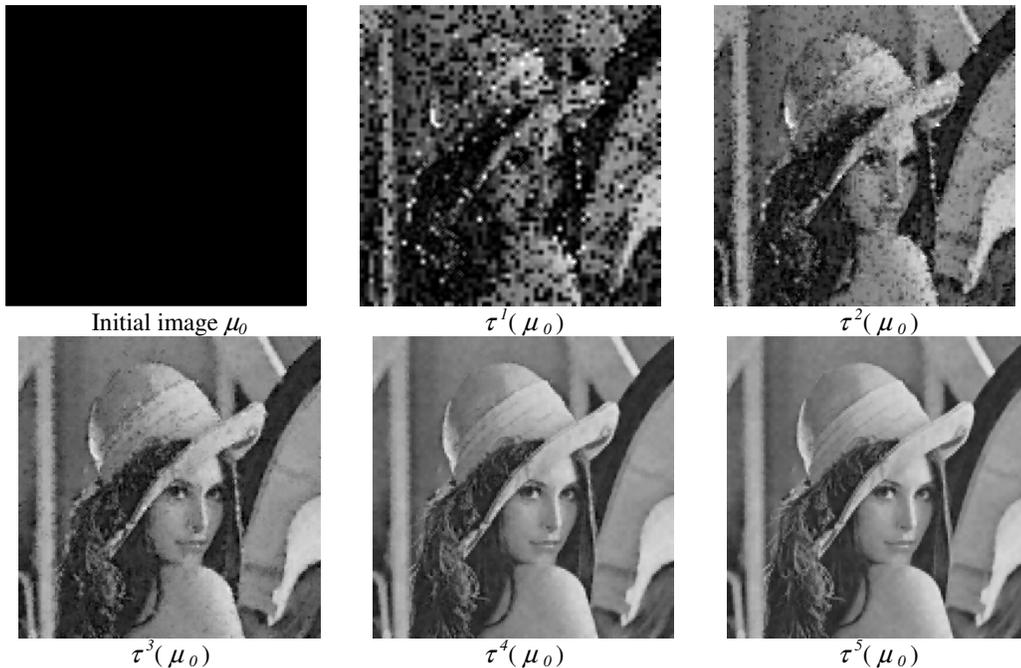




Figure 3 The first 8 iterations of the reconstruction sequence of *Lena* image by a fractal decoder

It. No	1	2	3	4	5	6	7	8	9	10
PSNR	15.9	19.5	22.8	25.2	26.4	27.0	27.3	27.5	27.6	27.6

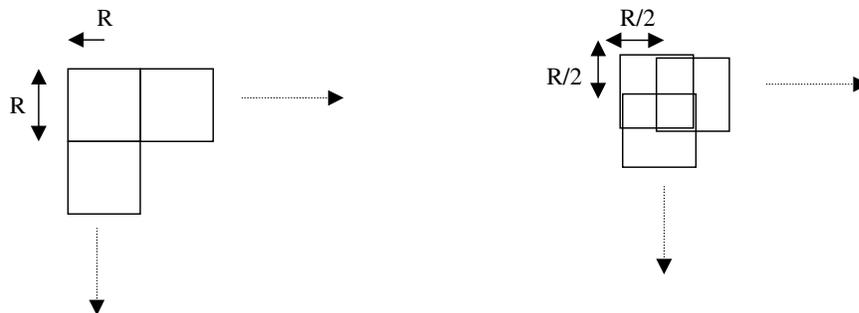
Table 1 PSNR values of the images obtained by the first 10 iterations of the fractal decoding

The fractal code can also be used in order to find an approximation of the original image at a resolution which is not the original one, thus obtaining a zoom of the original image. The main idea of the fractal zoom is based on the following property of fractals: by iterating a determinist transformation on some initial image, a deterministic fractal image is obtained, that is, if fractal coding is really a fractal process, the fractal code's attractor is a fractal object. Then, the fact that the fractal code is a transformation that can be applied on any images enables to zoom, just increasing the range and the domain blocks sizes, but letting the fractal code τ be unchanged.

The assumption at the base of fractal zoom is that self-similarities (i.e. matchings between areas with different sizes in the original image) are scale-independent. The main problem with this technique is an important block effect, due to some important discontinuities along the range blocks sides.

3.3.1 Variants at the classical fractal decoder

In [22] a variant at the classical fractal zoom is described to avoid blockness distortion and perform good visual quality: overlapped range blocks (O.R.B.) are used instead of a partition (see figure 4) and an averaging of the two or four common areas of the range blocks is made.



a) b)

Figure 4 (a) Classical partition and (b) Overlapped range blocks

This case is equivalent to take four partitions: one of the original image and the others of a part of the image with the same cells size, as shown in figure 5 (a)-(d). The image is then divided into four areas categories, as shown in figure 6.

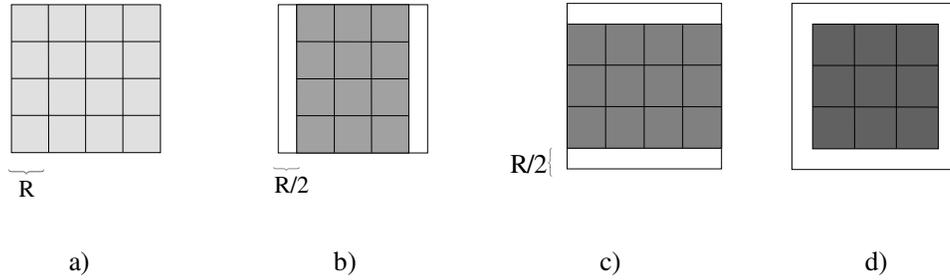


Figure 5 Different areas considered with the partition in figure 4 b

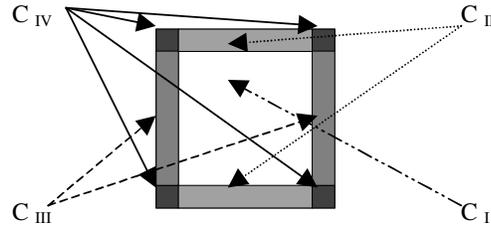


Figure 6 The four areas categories considered with the partition in figure 4 b

To encode the original image is then necessary to know the code $\tau^{(i)}$ of each part (see figure 5 (a)-(d)) of the original image independently.

The fractal code of the original image is now defined by the following expression:

$$\tau(\cdot) = \psi \left[\prod_{j=1}^4 \tau^{(j)}(\cdot) \right] \tag{2}$$

where ψ is a transformation used to stick back together the different codes $\tau^{(i)}$ according to the original image, as shown in figure 7.

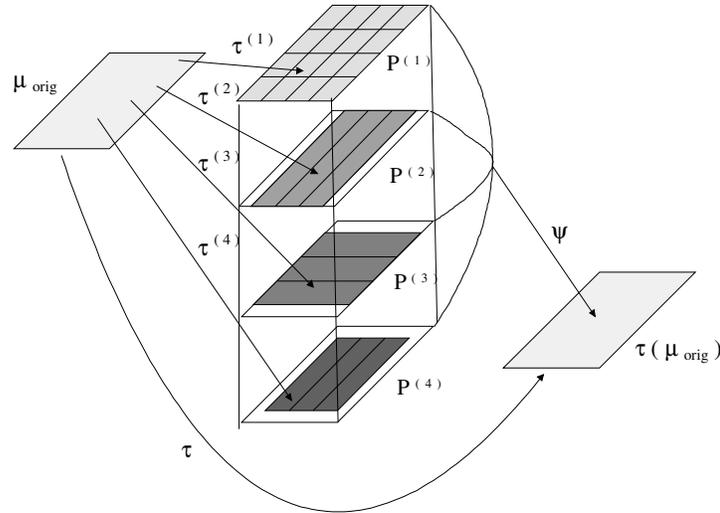


Figure 7 Behaviour of the new code τ and the transformation ψ

Let p be a pixel, (k,l) its coordinates and $V(p)$ the function which gives the value of the pixel p . Let v and $v(j)$, with the partition $P^{(j)}$ for $1 \leq j \leq 4$, be 5 images with the same size. Let $C: U \rightarrow \mathfrak{R}^2: p \mapsto (k,l)$ be the function that gives the coordinates, in the corresponding image, of

a pixel belonging to $U = \prod_{j=1}^4 v^{(j)}$. Now (see figure 8):

- $(k,l) \in C_I \Rightarrow C^{-1}(k,l) = \{p_1, p_2, p_3, p_4\}$
- $(k,l) \in C_{II} \cup C_{III} \Rightarrow C^{-1}(k,l) = \{p_1', p_2'\}$
- $(k,l) \in C_{IV} \Rightarrow C^{-1}(k,l) = \{p_0\}$

So:

$$v = \psi(U) = \bigcup_{p \in v^{(1)}} \xi[C^{-1}(C(p))] \quad (?)$$

and ψ is given by the function ξ , which depends on the method used.

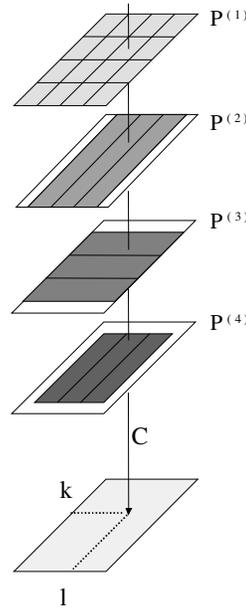


Figure 8 Behavior of the function C

To know the global code it's sufficient to take $v^{(j)} = \tau^{(j)}(\mu)$ for same image μ and then, using (1) and (2), $v = \tau(v)$.

In order to find the value of each transformed pixel, one can, for example, average the four values obtained with the four different partitions of the original image (fractal zoom with total averaging), thus taking the redundancy of the coding easily into account (in fact all pixels belonging to the area C_I are coded four times). In the areas C_{II} and C_{III} the averaging is only done on two values (obtained with the codes of the corresponding partitions) while in the area C_{IV} the only value obtained is from the partition of the original image. Hence, ξ is defined by:

$$V[\xi(\{p_n\})] = \frac{1}{r} \cdot \sum_{p \in \{p_n\}} V(p) \quad (3)$$

for r pixels $\{p_1, p_2, \dots, p_r\}$ ($r \in \{1, 2, 4\}$) and $\xi(\{p_n\})$ is a pixel belonging to v such that its coordinates are equal to $C(p_1) = C(p_2) = \dots = C(p_r)$.

The main problem with this technique is that a total averaging smoothes the image too much. In [19] a fractal zoom with adaptive averaging is described. It is based on this consideration: if the fractal coding was perfect, the four independent codes would give same results. In fact, it is reasonable to suppose that the fractal coding has a rather stable behavior and then to assume that the best values (i.e. closest to the real values) of pixels belonging to the area C_I are given by the two closest founded ones and to average only these two values, while the values belonging to the other areas remaining the same as before. Hence, ξ is defined by:

$$\begin{aligned} - (k, l) \in C_{IV} &\Rightarrow V[\xi(p_0)] = V(p_0) \\ - (k, l) \in C_{II} \cup C_{III} &\Rightarrow V[\xi(p_1', p_2')] = \frac{[V(p_1') + V(p_2')]}{2} \end{aligned}$$

Thus ξ is the same function as before if the coordinates belong to $C_{II} \cup C_{III} \cup C_{IV}$.

If the coordinates belong to C_I , let be $x_j = V(p_j)$ the value of the pixel p_j for $1 \leq j \leq 4$ and $d_{n,m} = |x_n - x_m|$. Therefore, the computation of 6 distances ($d_{1,2}$, $d_{1,3}$, $d_{1,4}$, $d_{2,3}$, $d_{2,4}$ and $d_{3,4}$) is required.

Then there exist at least one couple (n_0, m_0) such that:

$$d_{n_0, m_0} = \min_{n < m} \{d_{n, m}\} \quad (4)$$

Hence:

$$- (k, l) \in C_I \Rightarrow V[\xi(p_1, p_2, p_3, p_4)] = \frac{x_{n_0} + x_{m_0}}{2}$$

In this case ψ is an adaptive averaging which realizes the average computation from only two values, instead of four.

To increase the performances of this technique, we applied this procedure on a 3×3 mask to estimate the value of each pixel.

For each of the 3×3 neighbours and for the central pixel, 4 values are founded according to the four partition of Figure 8, and then the median value of these 36 is an estimation of the central pixel (O.S.O.).

This value constitutes a better estimation than that in [19] because the probable correlation between near pixels is reasonably taken under consideration.

3.3.2 Extension to 3D fractal encoders

The extension from image encoding to video encoding has followed two different ways:

- frame by frame [8, 21, 16] with a possible use of the motion information technique (similarly to MPEG) in order to limitate the time for the research of the affine transformations;
- by extending the concept of the range block to the range cube, that is by using 3D range and domain blocks [13, 4, 5].

The second approach is more complex than the first but it conserves the properties of the fractals that allow the decoding at every resolution by under or oversampling.

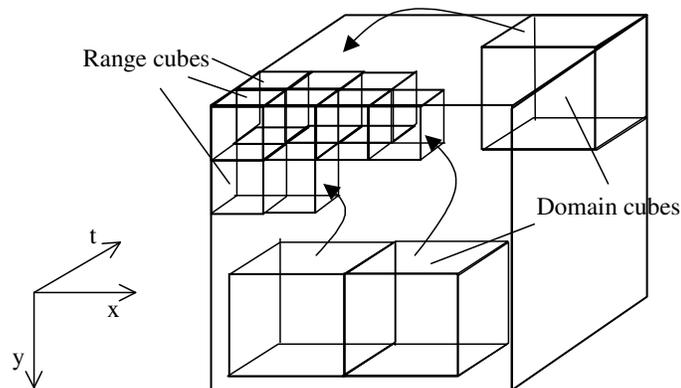


Figure 9 3D video fractal coding

The computational cost of the process, already high, will increase with another dimension, but a higher correlation along the time (Z) axis than along the spatial (X) and (Y) axes can be exploited in order to have a good coding without unacceptable delays.

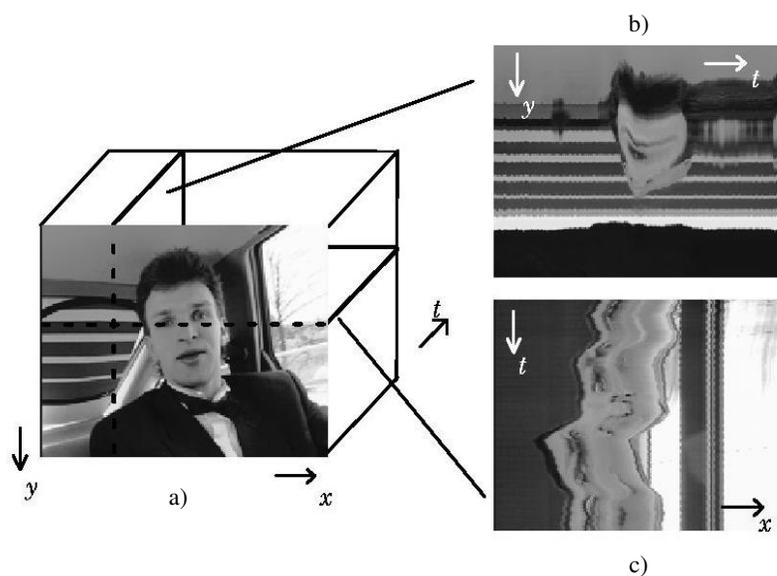


Figure 10 Two different views of a video sequence (a) a video packet of the “Carphone” sequence and two slices along the temporal direction: (b) vertical section and (c) horizontal section. In the first it is possible to note the appearing/disappearing of the head at the center of the sequence, while in the second it is possible to follow the movement of the head itself.

Many approaches to video coding use a technique of motion compensation named “motion estimation”, encoding the frames one at a time and considering the temporal changes between consecutive frames. With this new 3D approach it is possible to encode video sequences simultaneously. Figures 10 b) and c) show a video sequence in the spatial-temporal domain: the edges are orthogonal to the temporal direction and are formed by the movement of objects along the spatial direction of the cut. Many of the patterns of the gray levels in these cut visions are either homogeneous regions or borders. In general, this kind of structure can be efficiently encoded through fractal encoders. The borders generated by the movement can be approximated by contract parts of the near volumes.

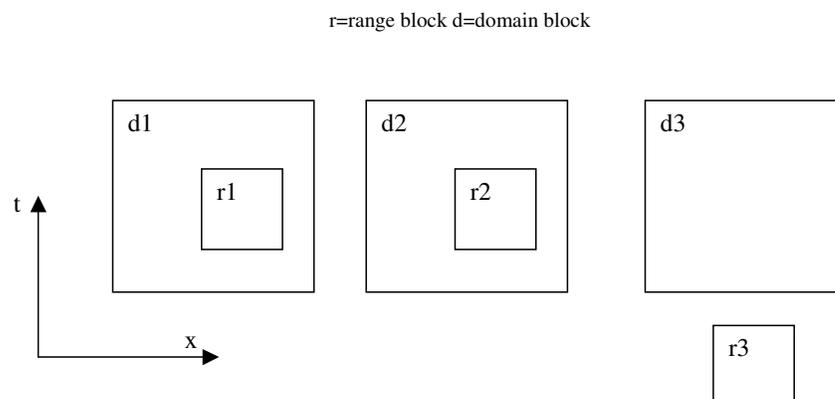


Figure 11 3D fractal coding applied to different kinds of movement: linear shifting, shifting with acceleration and spatial-temporal zoom

So the proposed technique does not use an explicit motion compensation technique. Due to the long required encoding time, this new 3-D approach can be efficiently used to store video data but it is not very convenient for real time applications.

In order to generate the range blocks, an intelligent partition method has been ideated: instead of using a fixed grid, a variable adaptive partition, named *quadtree* (cfr. fig. 12) is used.

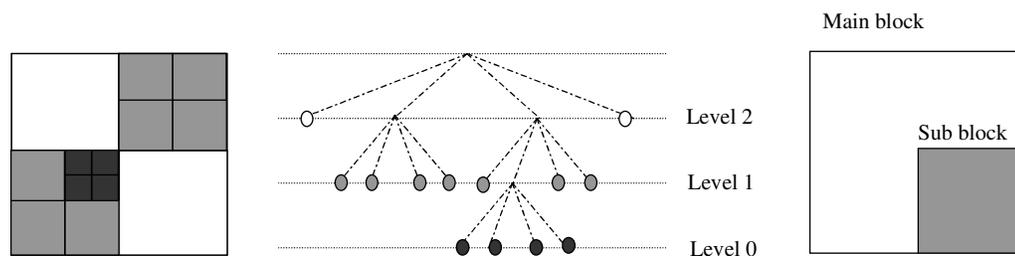


Figure 12 Partition of an image with a quadtree structure (2 levels partition).

This adaptive partition allows the encoding of an image by using larger range blocks in homogeneous regions and smaller ones in regions that are rich of small details.

In a 2D case, the principal $n \times n$ block is encoded and a distortion measure is computed for each of the $n/2 \times n/2$ sub-blocks. If the distortion of a sub-block is higher than a prefixed threshold, this block is divided into other four sub-blocks. The process is carried on in a recursive way until in the most distortion blocks sub-blocks the minimum allowed level is reached. Experimentally, it has been found that 3 levels are enough in most cases.

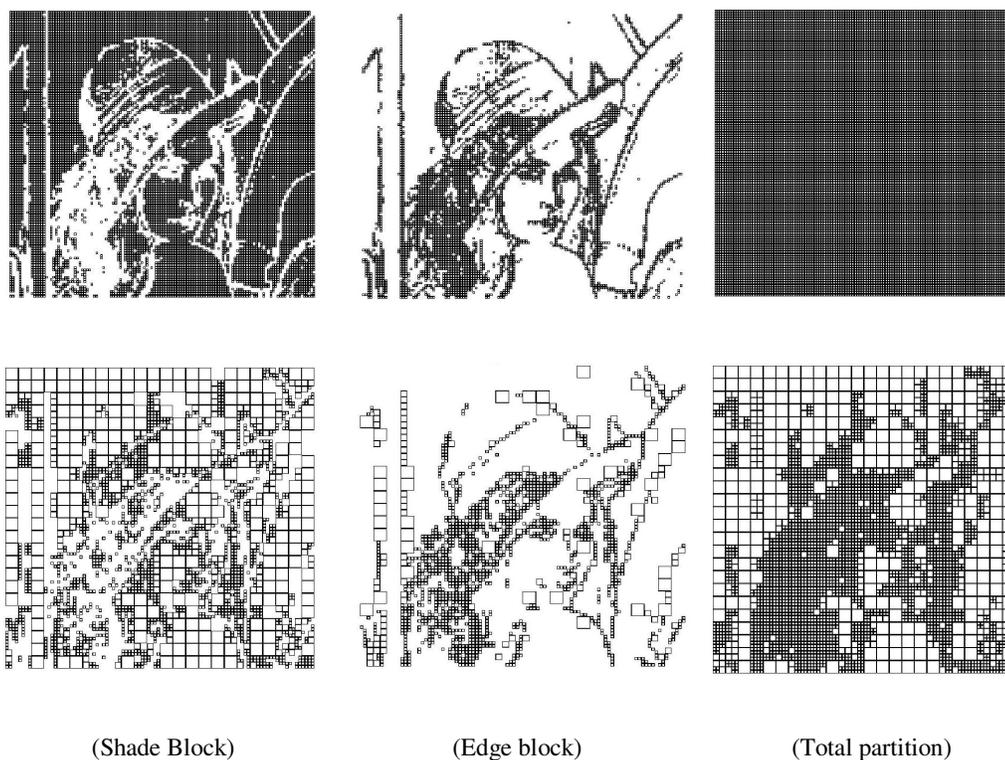


Figure 13 Partition of the image “Lenna”: up, the fixed grid partition and, at the bottom, the 2 levels quadtree

The extension of the quadtree to 3D is called *octree*. A partition of this kind in two levels is shown in figure 14.

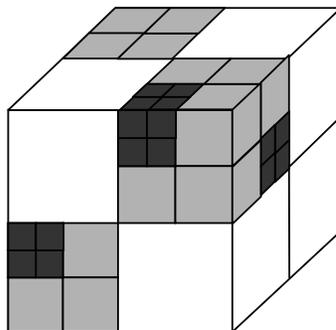


Figure 15 A 2 levels octree.

The use of this encoding technique for video sequences allows to obtain good results as regards compression ratio, quality and encoding time, compared to other fractal techniques (MA QUALI???), so that the computational complexity can be well justified.

3.4 The wavelet transform

The wavelet representation of an image is composed by the approximation of the signal at low resolution and the set of details at several resolutions. The image at low resolution is a low-pass version of the original one, while the details contain the information at high frequencies.

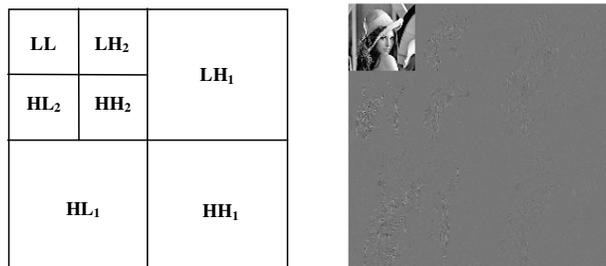


Figure 16 Wavelet representation of “Lenna” image

The signal into each subband is found through an iterative algorithm which decomposes the original signal into four ones, each of which contains the information regarding the original one in a frequency band and at a particular orientation.

The reconstruction algorithm is similar to the decomposition one. The complete signal is found again through a pyramidal algorithm from the low-pass one and the set of details.

The used wavelet representation was studied by Lemarie [14] and Battle [2]. It corresponds to a multiresolution approximation constructed beginning from cubic splines.

Let $N_1(\omega)$ and $N_2(\omega)$ be the following functions:

$$N_1(\omega) = 5 + 30\cos^2(\omega/2) + 30\sin^2(\omega/2)\cos^2(\omega/2)$$

$$N_2(\omega) = 2\sin^4(\omega/2)\cos^2(\omega/2) + 70\cos^4(\omega/2) + 2/3\sin^6(\omega/2)$$

The scaling function associated to the adopted representation can be write as:

$$\hat{\phi}(\omega) = \frac{1}{\omega^4 \sqrt{\Sigma_8(\omega)}}$$

where $\Sigma_8(\omega)$ is defined by:

$$\Sigma_8(\omega) = \frac{N_1(\omega) + N_2(\omega)}{105 \cdot \sin^8(\omega/2)}$$

The QMF used for the wavelet decomposition has then the following expression:

$$H(\omega) = \sqrt{\frac{\Sigma_8(\omega)}{2^8 \Sigma_8(2\omega)}}$$

The corresponding orthonormal wavelet is then given by the following expression:

$$\hat{\psi}(\omega) = e^{-i(\omega/2)} \overline{H\left(\frac{\omega}{2} + \pi\right)} \hat{\phi}\left(\frac{\omega}{2}\right) = \frac{e^{-i(\omega/2)} \sqrt{\Sigma_8\left(\frac{\omega}{2} + \pi\right)}}{\omega^n \sqrt{\Sigma_8(\omega) \Sigma_8\left(\frac{\omega}{2}\right)}}$$

and the other QMF can be found from the following relation:

$$G(\omega) = e^{-j\omega} \overline{H(\omega + \pi)}$$

The image in the LL subband is a low-pass version of the original image. Errors in the encoding process of this image could propagate in the reconstruction phase, making the final image worse. So the encoding of this image must be made with a lossless technique.

The monodimensional DPCM technique does not take advantage of the correlation between near pixels in all the directions (horizontal, vertical and diagonal) so the bidimensional DPCM technique defined in [17] has been used.

First, a 1-D DPCM is applied to the first row and to the first column of the image. Then each pixel value is predicted with a linear combination of its three near pixels values and the prediction error is coded:

$$\begin{aligned} \hat{x}_{i,j} &= 0.75x_{i-1,j} - 0.50x_{i-1,j-1} + 0.75x_{i,j-1} \\ \varepsilon_{x_{i,j}} &= \hat{x}_{i,j} - x_{i,j} \end{aligned}$$

3.4.1 Statistical properties of the wavelet coefficients

The statistical distribution of wavelet coefficients at a fixed resolution and orientation is a symmetric distribution with a nearly zero mean and small variance.

This distribution is often modeled as a Laplacian distribution, but it falls off really more rapidly, and it is better approximated by the generalized Gaussian distribution [24, 23]:

$$p_{2^j}^k = a_{2^j}^k \cdot \exp\left(-|b_{2^j}^k x|^{r_{2^j}^k}\right)$$

$$\text{with } a_{2^j}^k = \frac{b_{2^j}^k r_{2^j}^k}{2\Gamma(1/r_{2^j}^k)} \text{ and } b_{2^j}^k = \frac{1}{\sigma_{2^j}^k} \frac{\Gamma(3/r_{2^j}^k)^{1/2}}{\Gamma(1/r_{2^j}^k)^{1/2}}$$

where $\sigma_{2^j}^k$ is the standard deviation of the subband distribution at orientation K and resolution 2^j and $\Gamma(\cdot)$ is the Gamma function.

This formula contains the Gaussian and the Laplacian PDF as special cases:

- for $r_{2^j}^k = 2$ it is the Gaussian PDF;
- for $r_{2^j}^k = 1$ it is the Laplacian PDF.

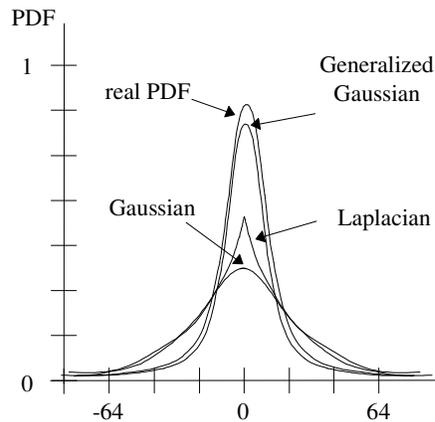


Figure 17 Approximations of the statistical distribution of wavelet coefficients

3.4.1 The classification of the active zones

The statistical distribution of wavelet coefficients in each subband is a nearly Gaussian distribution with zero mean and small variance. This generic property of wavelet coefficients allows the distinction, in each subband, of a zone characterized by greater energetic and informative contents (i.e. an *active zone*). For this zone the coding process must be more accurate in order to guarantee for a better reconstructed image.

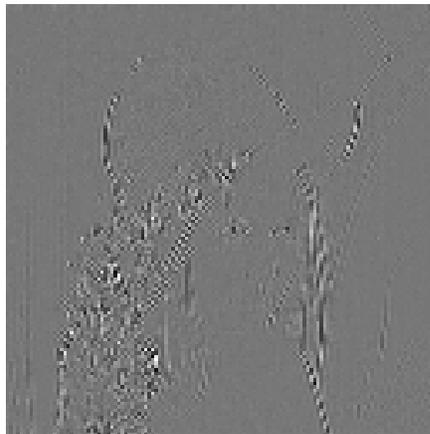


Figure 18 LH₂ subband of wavelet decomposition for “Lenna” image

After an extensive experimental analysis on different images, a heuristic algorithm for identifying these zones has been designed.

In each subband an accurate analysis of the quantized wavelet coefficients histogram is performed in order to apply a thresholding process to the histogram. The histogram is first regularized, to avoid falling into a local minimum during the thresholds searching, then its maximum value is found. In fact, since the histogram can be modeled as unimodal, with the maximum usually located around the zero value, two thresholds can be found proceeding into the two directions starting from this maximum, so that they include a histogram area around the 95% of the total. The search of these thresholds is made adaptively, in the sense that at

each step the threshold, whose movement minimizes the variation of the included area, is shifted.

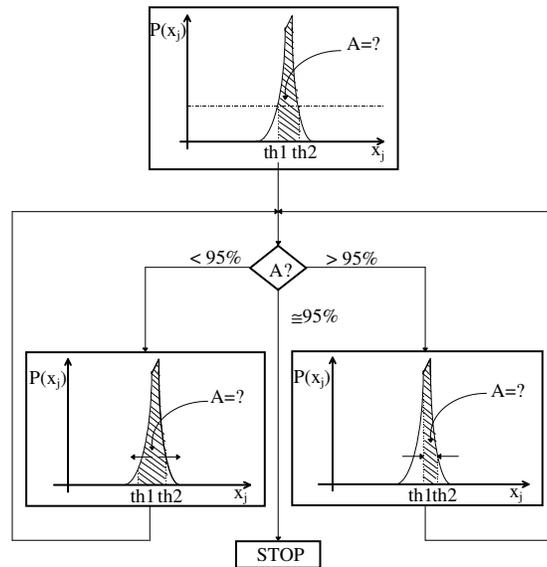


Figure 19 The thresholds searching algorithm

At the end of this process the distinction between active zone and not, is made in this way: pixels which have values included between the two thresholds found are classified as belonging to the non-active zone, because their values are near the mean value and they constitute an enough uniform zone of the image, while pixels whose values are not included between the two thresholds are classified as belonging to the active zone, because their values constitute the distribution tails and are quite different from the mean value.

3.4.3. The extraction of the mask

After the classification process, for each subband a binary-value “mask”, which contains the information relative to the position and the shape of the active zone, is extracted. The mask is then scaled down and logically summed with the one obtained with the same procedure at the lower resolution and at the same orientation.

At the end three masks, each of which contains the informations regarding the active zones of the subbands at the same orientation, are extracted.



a)



b)



c)

Figure 20 The masks of the active zones of “Lenna” image for a) LH orientation, b) HL orientation and c) HH orientation

The final masks are encoded by quadtree or optimized run length, according to the convenience.

3.4.4 The encoding of the wavelet coefficients

In each subband the coefficients belonging to the non-active zone are not coded because their informative content is very low (they constitute the background of the subband). The coefficients of the active zone, on the other hand, have the main part of the energetic and informative content of the subband, so their encoding must be accurate.

These coefficients are coded by vector quantization with parameters chosen from energetic considerations. In fact, the HH subbands are in general poorer of informative content than LH and HL ones. Moreover, a subband at low resolution is more energetic than subbands at the same orientation but at higher resolutions. In fact, as the resolution reduces, the energetic content of a subband increases and thus quantization must be more accurate.

The parameters of the vector quantization (codebooks and codevectors dimensions) are chosen accordingly to the variation of MSE value during the LBG algorithm. Codebooks are specific for each subband (at each resolution and orientation) and are generated from the active zones of the subbands of the images belonging to the training set.

3.4.5 The wavelet representation for video sequences

The main advantage of the use of the wavelet transform for video encoding is in the strong reduction of the computational complexity of the algorithm and, thus, in the reduction of the encoding time. As a matter of fact, the use of the wavelet representation allows the encoding of the video packets at lower resolutions with parallel processes. Furthermore, the properties of the wavelets permit to encode differentially the subbands according to their informative content, allowing eventually to discard some subbands from the encoding process (as, for example, for the higher frequency band, HH). Besides, the research of the matching between range and domain blocks is carried out in a smaller region, reducing thus the encoding time, that, as already noticed, is due essentially to this search. It should also be noticed that the computational burden introduced by the wavelet filtering is not very appreciable.

The best compromise between quality and compression ratio can be found by applying the described O.S.O. technique only to the LL subband (which is usually the most informative).

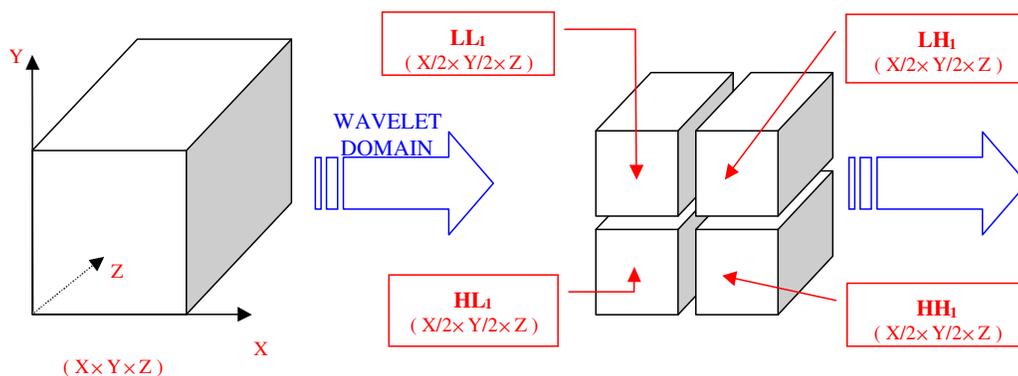


Figure 21 1 level wavelet decomposition of a video packet with dimensions XxYxZ

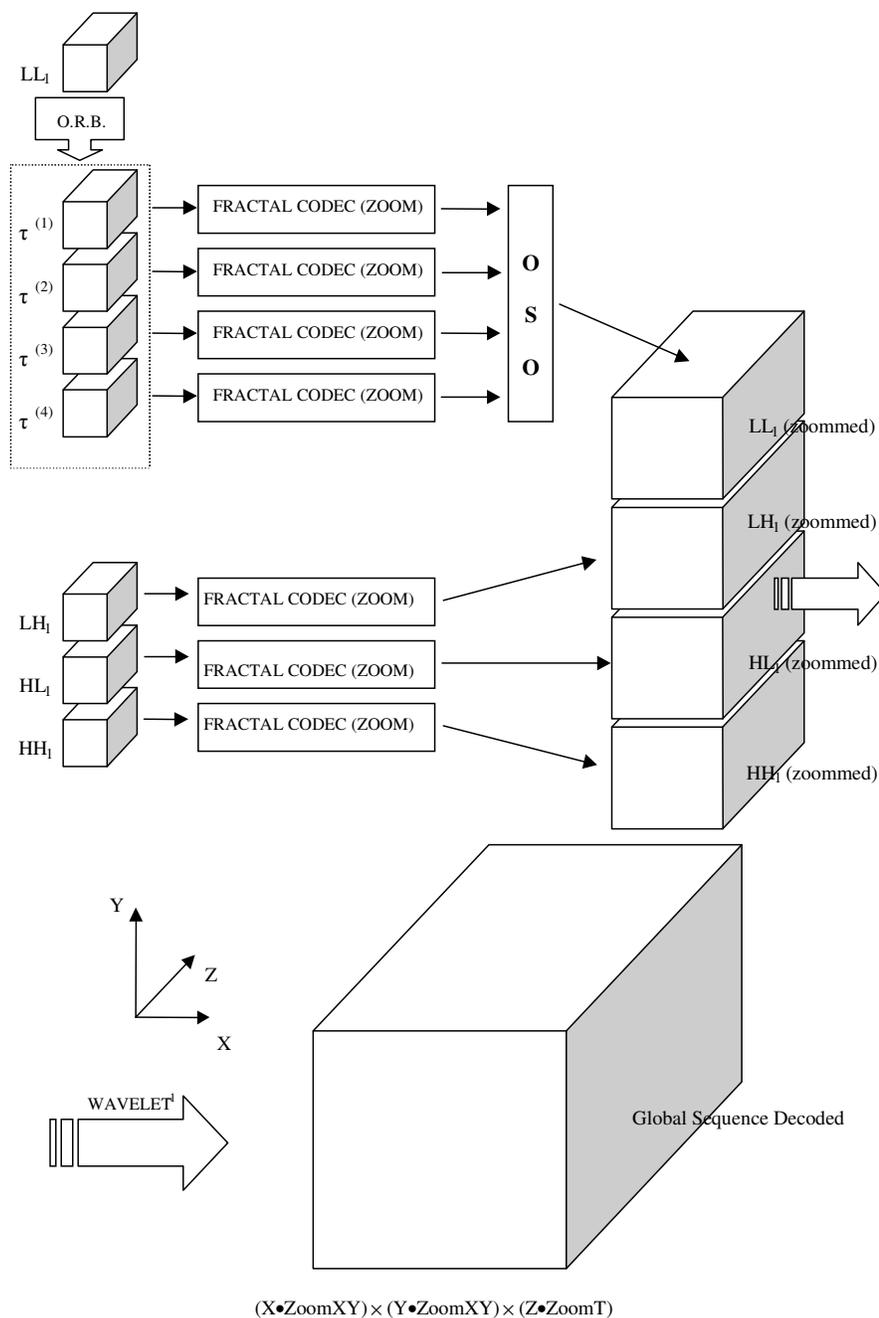


Figure 22 Parallelization of the CODEC process with 3D fractal zoom in the wavelet domain

3.5 Temporal subsampling with *key frames* extraction and use

In order to reduce the memory amount necessary to store the video sequence, a subsampling of the sequence along the temporal axis can be done by extracting only selected frames. In the decoding phase the sequence will be reconstructed by the fractal interpolation. For the selection of the key frames, three different approaches have been tested:

- Fixed grid

The key frames are chosen according to a fixed step. If N is the original number of frames and P the step of the fixed grid, the number of the selected frames is K :

$$K = \left\lceil \frac{N-1}{P} \right\rceil$$

The main drawback of this approach is that frames rich of informative content (i.e. the scene is changing from one frame to the next) could be discarded, while frames without information (i.e. static frames, no change from the previous frame) can be selected.

- MSE method

In order to overcome the drawback deriving from the use of a fixed grid, the frames are chosen according to their difference from the other selected frames. So, if the difference (measured by the MSE) between the current frame and the previous one is over a fixed threshold, this means that an important changing in the scene is happening and the frame should be selected, otherwise it would be discarded. So the reduction of data is obtained through the analysis and the reduction of the correlation between consecutive frames.

The main drawback of this variant is essentially in the fact that to a change of scene does not always corresponds a large MSE and thus the selection depends on the type of sequence. This approach cannot be applied a priori to every sequence, since the absence of a normalization of the MSE could cause the disregarding of important scenes.

- Adaptive with frames difference method

The basic idea of this variant is to take into account the differences between two frames by considering a frame formed by the difference itself.

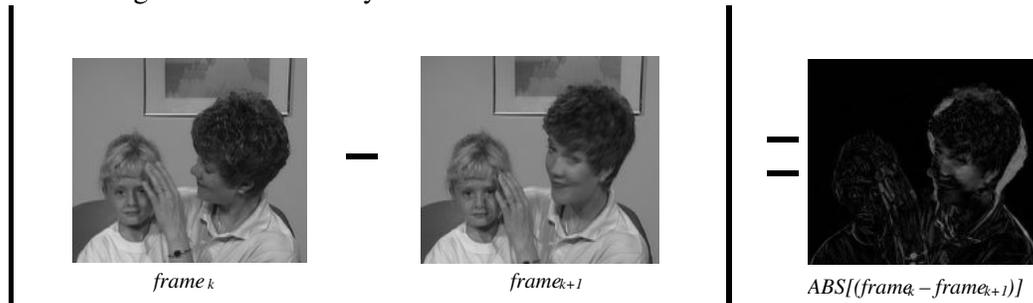


Figure 23 Extraction of the difference frame. It is possible to note the regions of the image corresponding to movements of the head and the body.

The percentage of the number of the pixels that are different from zero is compared to a prefixed threshold and determines the selection of the frame as key frames. The main problem is that a threshold should be selected for each sequence, and the choice of this threshold does influence strongly the number of the selected frames, which could vary largely. In order to perform an adaptive method for the selection of the frame, the analysis of the histogram has been performed and an adaptive threshold has been chosen according to the entity of the peaks in the histograms.

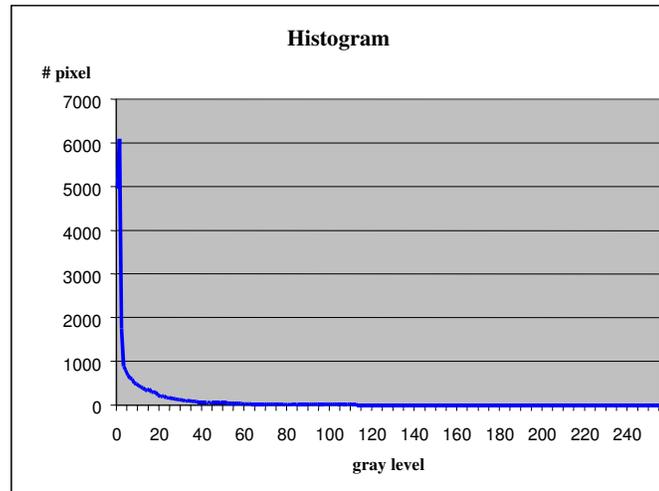


Figure 24 Histogram of the difference image in fig. 23

The ideated algorithm consists in the calculation of the maximum peak in the histogram (which corresponds to the black background) and in the choice of the threshold as the gray level corresponding to the first peak that is larger or equal to the 5% of the maximum. Then the selection proceeds as indicated by the flow diagram in fig.25

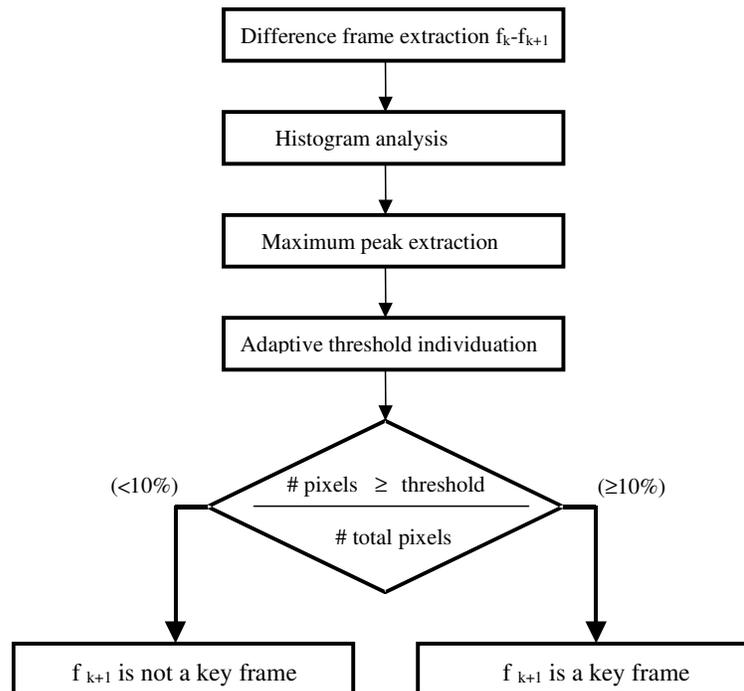


Figure 25 Algorithm for the extraction of the key frames

3.6 Selection of the active scenario

In order to reduce the computational complexity of the entire process, the background of the video sequence, which is almost static, can be encoded once for all, and the most of the bit-

stream can be dedicated to the encoding of the active part of the sequence. In order to select the active regions from the background, a binary mask is extracted: the resolution is proportional to the dimension of the range blocks and a mask is extracted for each selected key frame.

Each key frame is decomposed into blocks, which dimensions are equal or multiple of the range blocks', and for each of these blocks a parallelepiped is formed by considering all the corresponding blocks in the other key frames, as indicated in figure. ..

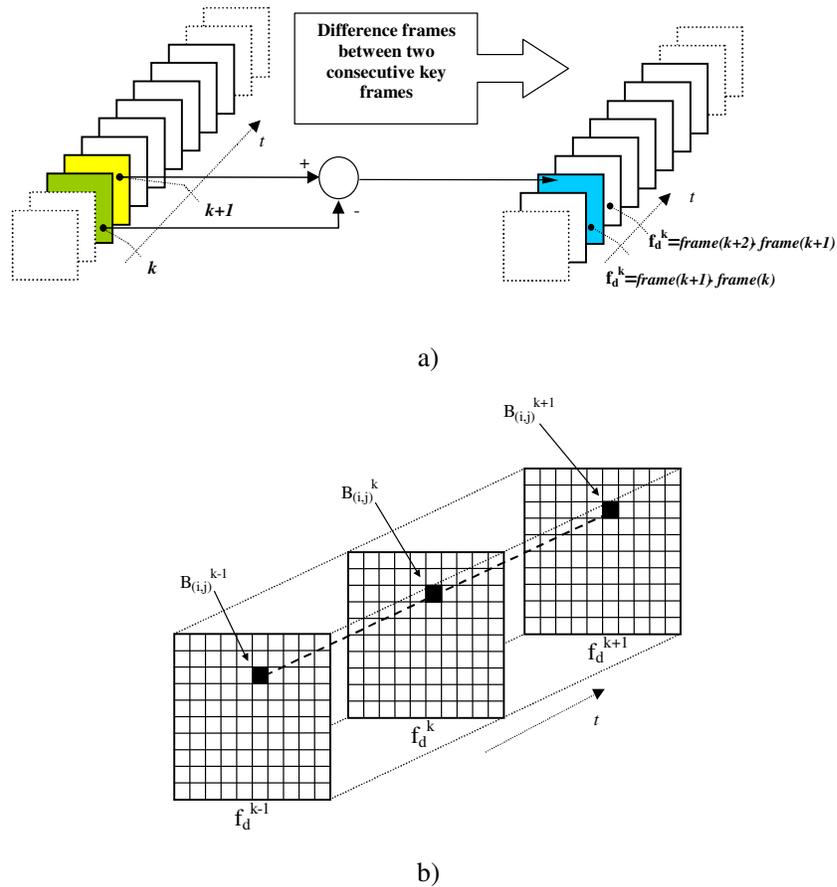


Figure 26 a) Individuation of the parallelepiped $P_{i,j} = \bigcup_{k=1}^{Z-1} B_{i,j}^k$; b) Zoomed view of the difference frames

In order to classify each region and build the binary mask, two different approaches have been taken:

- the parallelepiped $P_{i,j}$ is classified as an active region (white) if the mean value of a block $B_{i,j}^k$ is over a prefixed threshold;
- the parallelepiped $P_{i,j}$ is classified as an active region (white) if at least one of its pixels is over a prefixed threshold.

Tests have shown the equivalence of the two proposed approaches. The second has been chosen for its greater velocity, since it is not necessary to calculate any mean.

In order to avoid the risk that a region, classified as background, is encoded with a luminance which has changed gradually in the scene, and to have in this way a background with a mean

luminance different from the rest of the frame, the background is refreshed once in a GOP (which has been chosen of 8 frames) (see fig.27).

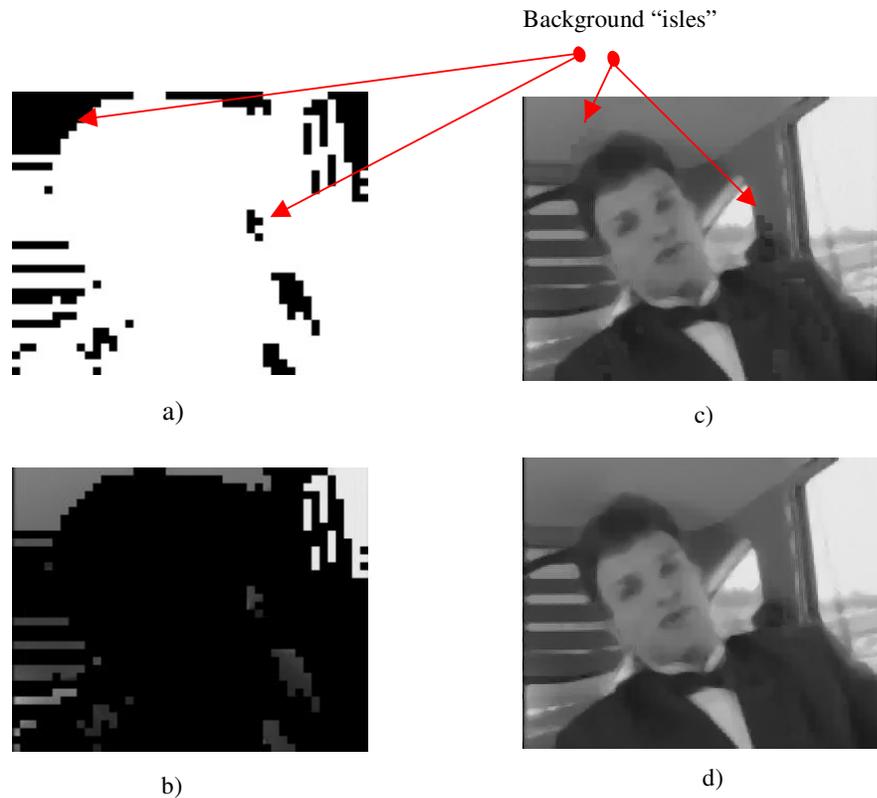


Figure 27 Frame 190 extracted from the video sequence “Carphone”: a) Extracted mask; b) Background obtained with the mask on the original frame; c) Decoded frame without refresh of the background; d) Decoded frame with refresh of the background.

Since the encoding of a background with a small area would not increase the compression ratio, this process is done only if the area of the mask extracted is at least equal to the 30% of the whole area. Otherwise the mask is not used and the frame is entirely encoded.

During the decoding, it is not necessary to use the mask, since the background itself can be used for this purpose. A new problem arises with the use of the ORB, since background and borders of the active zones could not correspond (see fig. 28). In order to overcome to this problem, the borders of the active zones and the background should be overlapping, in such a way to have a right reconstruction. A dilation operator is applied on the mask in such a way that also the regions containing the borders of the active zones are encoded with the active zones (see fig. 29).

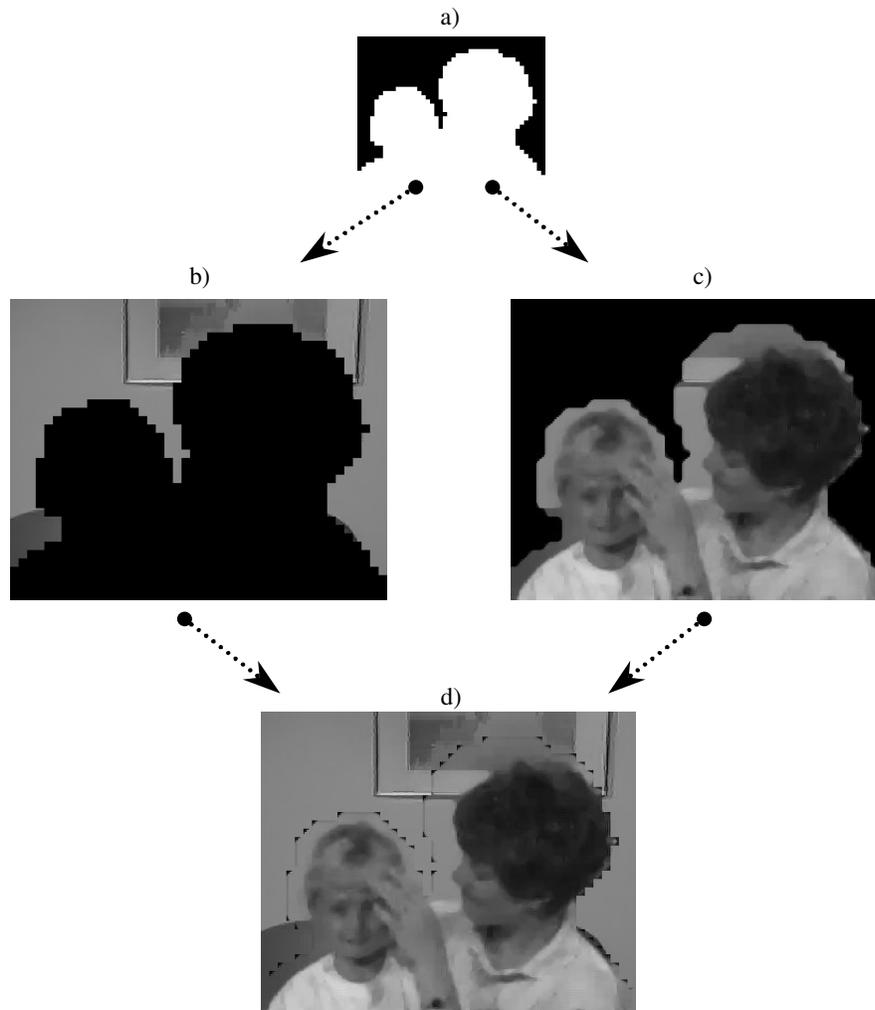


Figure 28 Bad reconstruction of the zones between background and moving image: a) Mask, b) and c) Background and active zones frame decoded by spatial zooming, d) Reconstructed frame. Note the presence of black zones in the region between b) and c)

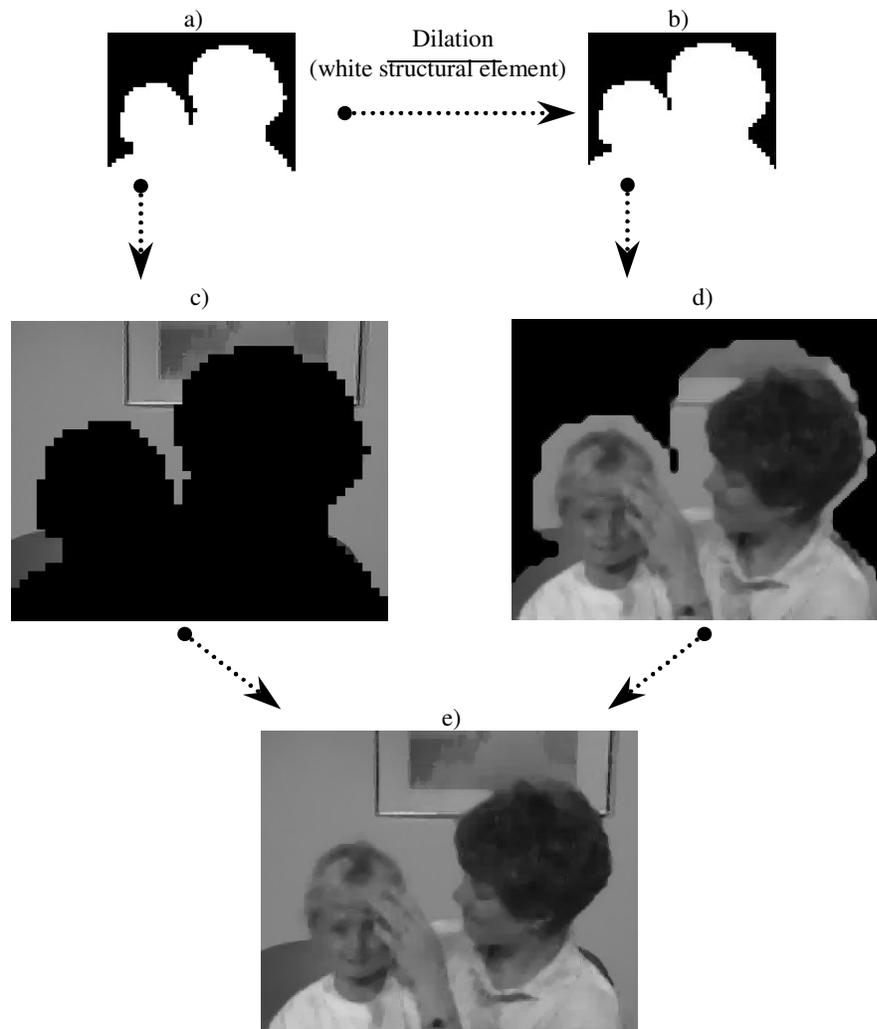
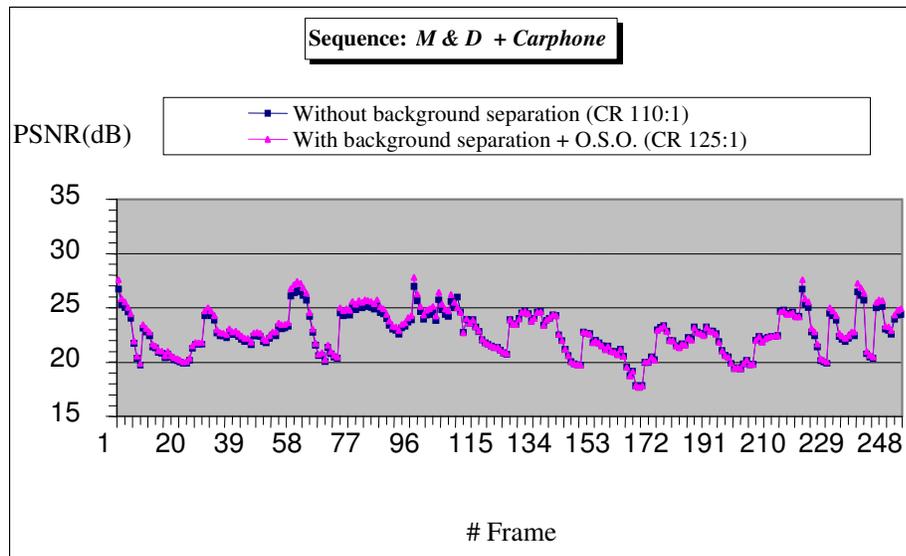


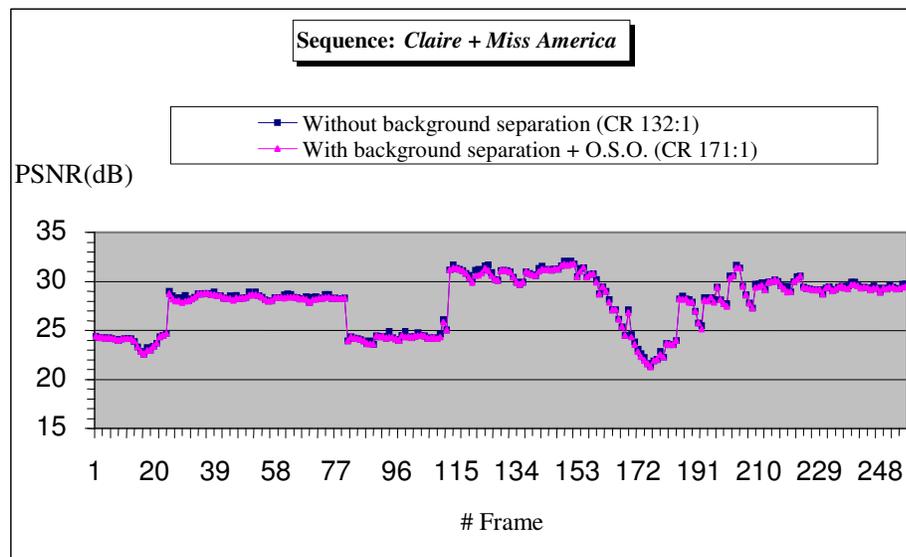
Figure 29 Correct reconstruction of the zones between background and moving image: a) Mask for the background, b) dilated mask for the moving image, c) Background with spatial zooming, d) moving image with spatial zooming, e) Reconstructed frame. Note the absence of black zones in the region between c) and d)

3.7 Tests and results

Several tests have been done on well-known sequences, with different quality factors and compression ratios. The behavior of the PSNR value for the two sequences “Mother and daughter + Carphone” and “Claire + Miss America” are reported in fig. 30.



a)



b)

Figure 30 PSNR values for the two test sequences: a) “Mother & Daughter” and “Carphone” respectively composed by 128+128 frames in CIF format (352×288), b) “Claire” and “Miss America” respectively composed by 128+128 frames in CIF format (352×288)

The decoding time has been analyzed in the two different situations (with and without separation of the background) and the results for the two sequences are reported in tab.

DECODING				
Sequence	# frame	Format	Time (s)	
			With O.S.O.	With O.S.O. & background sep.
<i>M&D - Car</i>	256	CIF (352×288)	5902	4356 (-26%)
<i>Cla - Mis</i>	256	CIF	5895	2160 (-63%)

Table 2 Decoding time for two test sequences

No variation has been noticed in the behavior of the PSNR (see fig. 30) between video sequences encoded with and without the separated encoding of the background. As regards the decoding times, they decrease if the background is encoded separately, since it does not need the use of the OSO filter, even if a longer encoding time should be taken into account, due to the selection of the active zones.

In particular, for the sequence “Mother and daughter + Carphone” it has been computed a decrease of the decoding time equal to the 26% when the selection of the background is applied. However, for the sequence “Claire + Miss America” the decrease of the decoding time is of about the 63%. This reflects the fact that the advantages of the selection of the background, as regards the decrease of the decoding times, is strictly related to the characteristics of the sequence to be encoded: as a matter of fact, the latter sequence is composed by frames with a strong interframe temporal correlation (low velocity in the movements of the objects) and a large background area during the entire sequence (small area occupied by the moving objects).

3.8 Conclusions and further developments

A novel and promising approach to the zoom of video sequences with the joined use of fractals and wavelets has been described and discussed. The development of the presented method has been done with increasing refinements in order to find the best compromise between quality and compression. The use of these two powerful mathematical tools (fractals and wavelets) combines the advantages of both characteristics in order to exploit the redundancy of the video signals, both in the spatial and temporal domains. The originality of the work has birth in the idea of extending the fractal zooming properties to 3D signals and combining them with the selectivity of the wavelet representation, which allow the selection and discarding of redundancy in the frequency domain. The individuation and use of the key frames and of the active zones inside them light the amount of data without discarding information by using adaptive and intelligent algorithms.

A strong reduction of the amount of data with limited decoding time has been observed. Further improvements of the method could render it very feasible for applications related to video demand, database search and videoconferencing.

References

- [1] BARNESLEY M.F., DEMKO S., “*Iterated function systems and the global construction of fractals*”, Proc. Royal Soc. London, vc A399, pp. 243-275, 1985.
- [2] BARNESLEY M.F., “*Fractals Everywhere*”, Academic Press, 1988.
- [3] BARNESLEY M.F., ELTON J.H., HARDIN D.P., “*Recurrent iterated function systems*”, Constructive Approximation, pp. 3-31, Berlin, 1989.
- [4] BARTHEL K.U., VOYE T., “*Three-Dimensional FractalVideo Coding*“, Proc. ICIP-95 IEEE International Conference on Image Processing, Washington, D.C., 1995.
- [5] BARTHEL K.U., VOYE T. RUHL G., “*Combining Wavelet and Fractal Coding for 3-D Video Coding*”, Proc. ICIP-96 IEEE International Conference on Image Processing, Lausanne, 1996.
- [6] BATTLE G., “*A block spin construction of ondelettes. Part I: Lemarie functions*”, Commun. Math. Ohys, Vol. 110, pp. 601-615, 1987.
- [7] CASO G., OBRADOR P., KUO C.J., “*Fast methods for fractal image encoding*”, Proc. IST/SPIE, Symposium on Electronic Imaging: Science & Technology, Vol. 2501, pp. 582-594, 1995.
- [8] FISHER Y., ROGOVIN D., SHEN T.P., “*Fractal (Self-VQ) Encoding of Video Sequences*”, Proc. SPIE, “*Visual Communications and Image Processing*”, 1994.
- [9] GONZALES R.C., WOODS R.E., “*Digital Image Processing*”, Addison Wesley, 1992.
- [10] JACQUIN A.E., “*A fractal theory of iterated Markov operators with application to digital image coding*”, Ph. D. dissertation, Georgia Tech, 1989.
- [11] JACOBS E.W., FISHER Y., BOSS R.D., “*Image compression: A study of the Iterated Transform method*”, Signal processing n. 29, pp. 251-263, 1992.
- [12] JACQUIN A.E., “*Fractal Image Coding : A review*”, Proc. IEEE, pp. 1451-1465, 1993.
- [13] LAZAR M.S., BRUTON L.T., “*Fractal Block Coding of Digital Video*”, IEEE Trans. On Circuits and Systems for Video technology 4, pp. 297-308, 1994.
- [14] LEMARIE P.G., « *Ondelettes a localization exponentielles* », Journal Math. Pures et Applications, Vol. 67, pp. 227-236, 1998.
- [15] LI H., NOVAK M., FORCHHEIMER R., “*Fractal-Based Image Sequence Compression Scheme*”, Optical Engineering, pp. 1588-1595, 1993.
- [16] MONRO D.M., NICHOLLS J., “*Real Time Fractal Video for personal Communications*“, Fractals, Montreal, Canada, Vol. 2, pp. 391-394, 1994.
- [17] NETRAVALI A.N., HASKELL B.G., “*Digital Picture: representation, compression and*

- standards*”, Plenum Press, 1995.
- [18] NOVAK M., “*Attractor Coding of Images*”, Department of Electrical Engineering, Lindkûping University, Sweden, 1992.
- [19] POLIDORI E., DUGELAY J., “Zooming using iterated function systems,” *Fractals*, vol.5, pp.111-123, Apr. 1997.
- [20] RAMAMURTHI B., GERSHO A., “*Classified vector quantization of images*”, *IEEE trans. Comm.*, COM 34, 1986.
- [21] REUSENS E., “*Sequence Coding based on the Fractal Theory of Iterated Transformations Systems*”, SPIE, Visual Communications and Image Processing, Boston, Vol. 2094, pp. 132-140, 1993.
- [22] REUSENS E., “*Overlapped Adaptive Partitioning for Image Coding Based on the theory of Iterated Function Systems*”, International Conference on Acoustic, Speech and Signal Processing, 1994.
- [23] VETTERLI M., KOVACEVIC J., “*Wavelets and subband coding*”, Prentice Hall, U.S.A. 1995.
- [24] WESTERINK P.H., “*Subband coding of images*”, Ph. D. Thesis, Delft University of Technology, 1989.

Chapter 4

Performance Analysis of Fractal Modulation Transmission over Fast-Fading Wireless Channels

4.1 Introduction

Prevalent communication systems for time-varying channels are basically designed to achieve the required performance under the worst-case channel conditions. Usually, to compensate channel variations, large margins are then taken into account during designing. A major problem is that such a large fading margins planning does not take maximum advantage of the available channel capacity, especially for high-variable channels. In fact, the process of designing to worst-case channel capacity results in periods of very good performance and periods of performance at, or close to the maximum error rate, in effect yielding variable performance with a fixed data rate. It follows that a worst-case approach to the communications system design will not be able to use the full channel capacity and it will be thus a sub-optimal design.

In those particular transmitting conditions where the channel availability and bandwidth are stochastically time-varying, a complex adaptive scheme, able to evaluate the channel characteristics changes and accordingly to adjust the system, can be alternatively, adopted. However this approach achieves a high-complex structure and depends on getting a prompt feedback from the receiver regarding channel conditions. This makes it unsuitable for application like broadcast communications over rapidly-varying channels.

A viable solution to this problem requires a transmission strategy where the data to be transmitted can be found at more frequency bands in order to allow an efficient reception also with channel condition variations. In such a case the transmitter is not required to change the transmission configuration, but is the receiver is the one which makes the necessary changes according to the channel conditions variations. This is the basic concept of fractal modulation, where the transmission spectral efficiency is maintained over a broad range of rate-bandwidth combinations using a fixed transmitter configuration. A rather natural strategy of this type arises out of the concept of embedding the data to be transmitted into a homogeneous signal [1].

The fractal modulation paradigm has been essentially proposed as an interesting potential application of dy-homogeneous signals. Aside from a scale factor, this class of signals remains invariant under scaling of the time axis. In [2], Wornell and Oppenheim have shown that this kind of signal can carry information distributed over multiple time scales and frequency bands and that they appear well-suited for transmission over noisy channels of

simultaneously unknown duration and bandwidth. Additionally, they have the advantage of being efficiently represented using orthonormal wavelet basis sets.

The most interesting application of this technique is represented by the wireless communication, especially in urban areas, where the network distribution allows a mobile user to be reached by signals coming from different based stations. Many random signals that propagate through different signal paths from the transmitter are superposed at the receiver to produce standing waves. When a receiver and/or a transmitter moves in the standing waves, the receiver experiences random variation in the signal level and in the phase, and also a Doppler shift. As a consequence of the standing wave characteristics, the minimum distance between signal level drops is one half of the wavelength; this microscopic deviation of the received signal is called *fast fading* [3].

In this paper, we present a performance analysis of fractal modulation transmission over a AWGN fast-fading channel. A quadrature transmission scheme is simulated and compared with frequently used transmission systems achieving better results in terms of error robustness and low complexity.

Previous works on the analysis of the performances of such a modulation technique have the main limit of not considering time-varying channels [4][5]. Instead, these are very important since they represent the main configuration of interest for such modulation technique. The originality of our work relies on the comparison with the QAM transmitter, core of the OFDM modulation system, which is extensively employed in broadcasting technology, over a test-bed simulation environment including typical problems occurring in wireless transmission like fast-fading and additive noise. Furthermore, we performed experiments, using different wavelet families. The results of the experiments show the effectiveness of the fractal modulation paradigm and confirm its effective utilization in data broadcasting.

The concept of fractal modulation is based on the properties and characteristics of homogeneous signals, and its implementation relies on the wavelet theory. Based on that, in the following two sections of the paper the characteristics of homogeneous signals and their wavelet representation are summarized: excellent introductions to these concepts can be found in [6] and [7]. In section 4, the structure of fractal modulation transmitter and receiver are described. Finally, sections 5 and 6 are devoted to the simulation process and its results.

4.2 Homogeneous signals and wavelet representation

A dy-homogeneous signal $x(t)$ is a dyadic self-similar signal satisfying the deterministic scale-invariance property:

$$x(t) = 2^{-kH} x(2^k t), \quad (1)$$

for all integers k . This is the class of signal of interest for fractal modulation applications.

Homogeneous signals are inherently well suited as modulating waveforms for use on the channels previously described. In fact, as a consequence of their intrinsic self-similarity, these waveforms have the property that an arbitrarily short duration time-segment is sufficient to recover the entire waveform, and hence the embedded information, given adequate bandwidth. Likewise, an arbitrarily low-bandwidth approximation of the waveform is sufficient to recover the undistorted waveform, and again the embedded information, given an adequate time duration [2]. Additionally, homogeneous signals can be efficiently represented through a wavelet basis, that can be exploited in the development of practical systems.

The expansion of an arbitrary signal $x(t)$ in a orthonormal wavelet basis takes the following form:

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x_n^m(t) \psi_n^m(t), \quad (2.a)$$

$$\text{where } x_n^m = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt. \quad (2.b)$$

The orthonormal wavelet basis functions are related as usual, according to:

$$\psi_n^m(t) = 2^{m/2} \psi(2^m t - n), \quad (2.c)$$

where $\psi(t)$ represents the basic wavelet function, and where m and n are the dilation and translation indexes, respectively [8], [9], [10], [11],[12], [13]

When $x(t)$ is a homogeneous signal, it follows from (2.b) that the wavelet coefficients take the form:

$$x_n^m = \beta^{-m/2} x_n^0, \quad (3)$$

where $\beta = 2^{2H+1} = 2^\gamma$. Denoting x_n^0 with $q[n]$, then (2.a) becomes:

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \beta^{-m/2} q[n] \psi_n^m(t). \quad (4)$$

From (4) it arises that $x(t)$ is completely specified in terms of $q[n]$, that is referred to as the *generating sequence* for the homogeneous signal $x(t)$. The associated time-frequency portrait of a homogeneous signal, expressed in terms of generating sequence, is depicted in Fig. 1 (for purposes of illustration, H has been set to $-1/2$). The partitioning in such time frequency portraits is of course idealized; in general, there is both spectral and temporal overlap between cells.

As Fig. 1 shows, the synthesis of homogeneous signals can be accomplished by replicating a generating sequence $q[n]$ at each scale in the representation (4) via an expansion in terms of an orthonormal wavelet basis [14],[15]. According to this expansion, the detail signals in the associated multiresolution representation are simply time-dilated versions of one another, to within an amplitude factor:

$$D_m \{x(t)\} = \beta^{-m/2} \sum_{n=-\infty}^{\infty} q[n] \psi_n^m(t). \quad (5)$$

With reference to this wavelet representation, the corresponding resolution 2^m approximation of the homogeneous signal is then obtained as the projection of $x(t)$ onto the subspace \mathbf{V}_m of the signal space \mathbf{V} , is as follows:

$$A_m \{x(t)\} = \sum_{n=-\infty}^{\infty} a_n^m \phi_n^m(t), \quad (6)$$

where $\phi_n^m(t)$ are the orthonormal basis of the considered subspace and expressed in terms of the associated scaling function $\phi(t)$ according to:

$$\phi_n^m(t) = 2^{m/2} \phi(2^m t - n). \quad (7)$$

Since the coefficients a_n^m are obtained by projecting $x(t)$ onto the $\phi_n^m(t)$ basis function, the homogeneity of $x(t)$ implies that they are identical at all scales to within an amplitude factor:

$$a_n^m = \beta^{-m/2} a_n^0. \quad (8)$$

From this, it results that the sequence a_n^0 is an additional characterization of $x(t)$ since its knowledge is sufficient to reconstruct the signal with arbitrary accuracy. Such a sequence is

referred to as the *characteristic sequence* and represented with $p[n]$ notation. Hence, $p[n]$ characterizes arbitrarily fine approximations of $x(t)$, and, in turn, $x(t)$ itself.

frequency

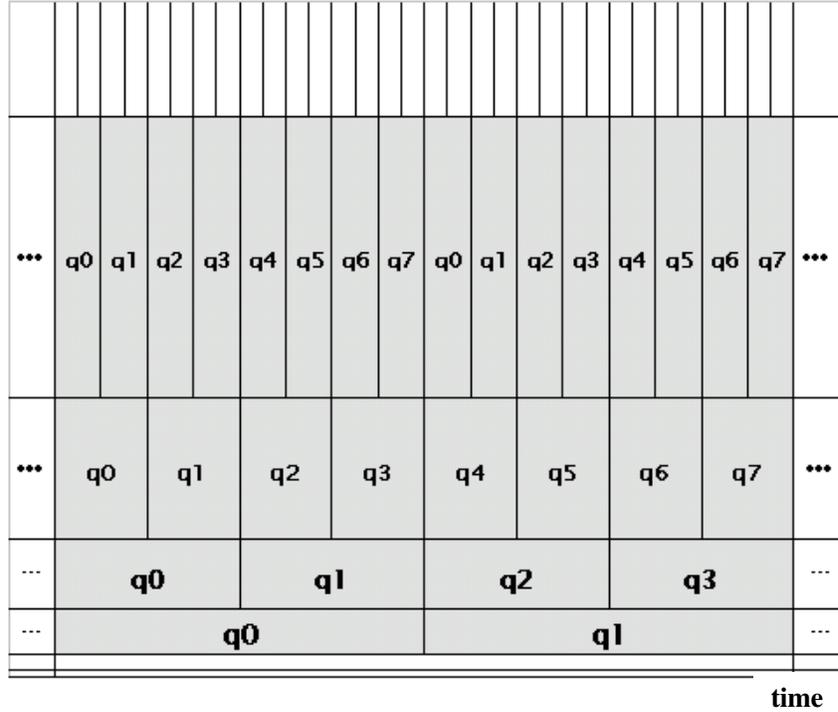


Fig.1 The time-frequency portrait of a homogeneous signal.

4.2.1 Iterative synthesis of homogeneous signals

There are interesting relations between the two sequences $p[n]$ and $q[n]$ that can be exploited in a discrete synthesis of the homogeneous signal $x(t)$, based on the use of the quadrature mirror filter (QMF) pair associated with the wavelet basis [9]. Let refer with $H_0[n]$ and $H_1[n]$ to the filters pair, so the wavelet coefficients a_n^m are:

$$a_n^m = \sum_l H_0[l - 2n]a_l^{m+1}, \text{ with } x_n^m = \sum_l H_1[l - 2n]a_l^{m+1}. \quad (9)$$

From this expression it is straightforward to obtain the generating sequence $q[n]$ from the characteristic sequence $p[n]$ as follows:

$$\beta^{1/2}q[n] = \sum_k H_1[k - 2n]p[k] \quad (10)$$

while the inverse procedure is more difficult and requires the use of an iterative algorithm for constructing $p[n]$ from $q[n]$:

$$p^{[0]}[n] = 0$$

$$p^{[l+1]}[n] = \beta^{1/2} \sum_{k=-\infty}^{\infty} \{H_0[n - 2k]p^{[l]}[k] + H_1[n - 2k]q[k]\} \quad (11)$$

This recursive upsample-filter-merge algorithm, depicted in Fig.2, can be viewed as repeatedly modulating $q[n]$ with the appropriate gain into successively lower octave band of the frequency interval. Note that the quantity $q_*[n]$ is the sequence $q[n]$ modulated into essentially the upper half band of frequencies.

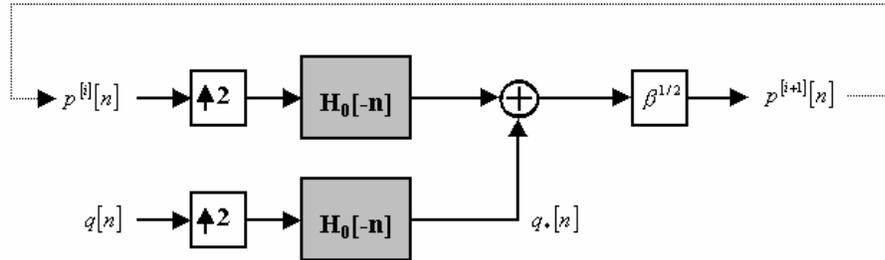


Fig. 2 Iterative synthesis of the characteristic sequence $p[n]$ from the generating sequence $q[n]$

4.3 Fractal modulation transmitter and receiver

The results of the preceding section suggest an efficient way for embedding a symbol stream $q[n]$ into a homogeneous waveform $x(t)$ by only using such a sequence as the generating one for the signal $x(t)$. This synthesis is the essence of the fractal modulation, which yields $q[n]$ to be modulated into a finite number of contiguous octave-width frequency bands.

The fractal modulation transmitter can be implemented in a computationally efficient manner using the discrete-time algorithm of the previous section. In particular, after obtaining $p^{[M]}[n]$ from $q[n]$ using M iterations of the synthesis algorithm (11), the result is mapped into the associated continuous-time waveform by modulating with the appropriate scaling function as follows:

$$x(t) = \sum_n p^{[M]}[n] 2^M \phi(2^M t - n). \quad (12)$$

Finite length messages are accommodated by modulating their periodic extensions $q[n \bmod L]$, thereby generating a transmitted waveform:

$$x(t) = \sum_{n=-\infty}^{\infty} q[n \bmod L] \theta_n^H(t), \quad (13)$$

$$\text{where } \theta_n^H(t) = \sum_{m=-\infty}^{\infty} \beta^{-m/2} \psi_n^m(t).$$

Denoting with $\mathbf{q} = [q[0] \dots q[L-1]]$ the data vector, the time-frequency portrait associated with this signal is shown in Fig. 3. This naturally leads to a strategy for data transmission on block-by-block basis.

The parameter H in fractal modulation controls the relative power distribution among frequency bands and, hence, the overall transmitted power spectrum.

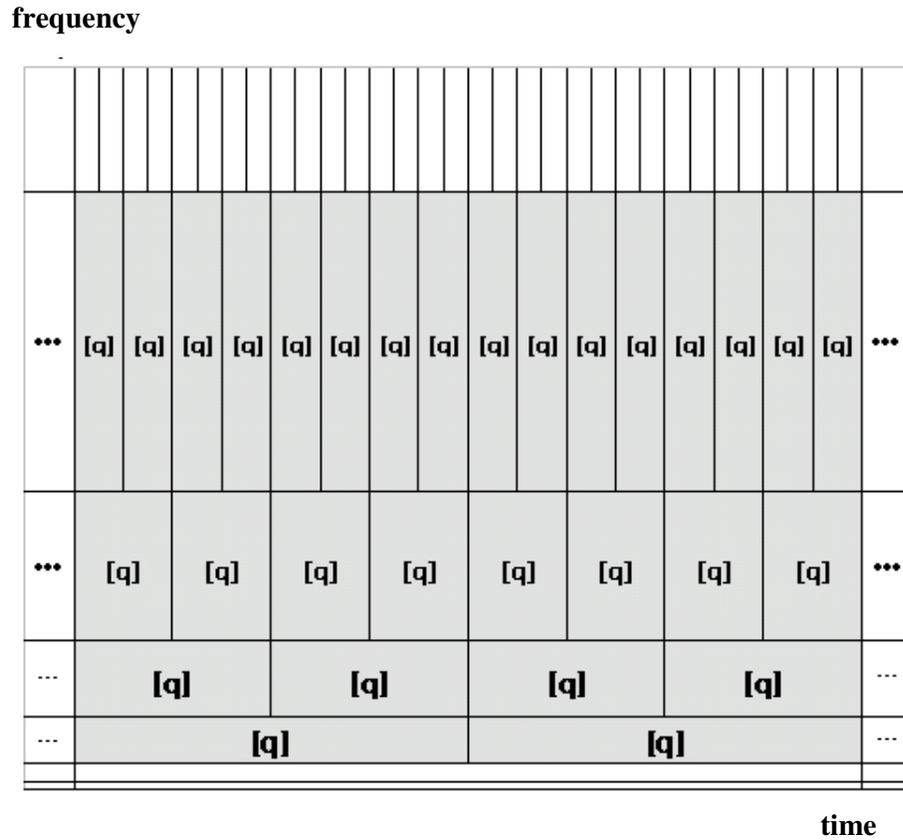


Fig. 3. A portion of the time-frequency portrait of the transmitted signal for fractal modulation of a finite length data vector q

For transmission of finite-length messages composed of M -ary symbols in the presence of Gaussian noise, efficient maximum likelihood (minimum probability of error) receivers can be developed [2]. Such receivers exploit processing in the wavelet coefficient domain. Accordingly, the first stage receiver extracts the wavelet coefficients r_n^m of the received waveform $r(t)$ using the DWT. These coefficients take the form:

$$r_n^m = \beta^{-m/2} q[n \bmod L] + z_n^m, \quad (14)$$

where z_n^m are the wavelet coefficients of the noise process $z(t)$.

The duration-bandwidth characteristics of the channel in general affect which observation coefficients r_n^m may be accessed and, hence, the available redundancy. If the channel is band-limited to 2^{T_U} Hz for some integer T_U , this precludes access to the coefficients at scales corresponding to $m > T_U$. Simultaneously, the duration-constraint in the channel results in a minimum allowable decoding rate of 2^{T_L} symbols/sec for some integer T_L , which precludes access to the coefficients at scales corresponding to $m < T_L$. As a result, the available number of noisy measurements of the message is:

$$K = \sum_{m=T_L}^{T_U} 2^{m-T_L} = 2^{T_U-T_L+1} - 1. \quad (15)$$

Under the hypothesis of z_n^m being a Gaussian variable with zero mean and variance $\text{var } z_n^m = \sigma_z^2 \beta^{-m}$, the resulting bit-error probability that characterizes such a fractal demodulator can be written as:

$$P_e = Q\left(\sqrt{\sigma_c^2 \left[\frac{2\eta_F}{R/W} - 1\right]}\right), \quad (16)$$

where σ_c^2 represents the signal to noise ratio.

4.4 Experiments set-up

The performance of a high-order constellation fractal modulation scheme has been investigated with the aid of the *Ptolemy* tool, that is a heterogeneous simulation environment supporting different computational models developed at the Berkeley University [16]. Aim of our experiments is to investigate the robustness of fractal modulation schemes for digital data transmission over AWGN channels affected by multi-path fading.

Initially, the orthonormal wavelets firstly described by Daubechies was used in the experiments, due to the compactness of the associated discrete wavelet transform filters [17]-[18]. The order used for these wavelets has been selected taking into account the complexity in the realization of the basic function and the fact that they are not time-limiting, that is, the wavelets of the same dilation overlap each other. The number of wavelets overlapped by a given wavelet is $2N - 2$, where N is the order of the wavelet family. As N increases the number of data symbols simultaneously being modulated within one-frequency band increases as well. The implementation of high-order system would require heavy increase of the system complexity, due to the additional buffers, modulators, and other components to be added. Then we selected $N = 4$ as a reasonable order to use for our simulation, according also with the strategy already adopted by Ptasinaky and Fellman [4]. The wavelet have been generated by means of the MacWavelets [19].

When digitizing wavelets, we must use a sufficient number of samples to maintain their desired characteristics. Since the wavelets are not ideally band-limited, any sampling introduces some distortion. The main characteristic of interest is the orthogonality of wavelets at the different dilations. We selected a 1024 points representation for the prototype wavelet, and a number of 4 dilations obtaining the following signal representation:

$$x(t) = \sum_{m=-3}^0 \beta^{-m/2} \sum_{n=-\infty}^{\infty} q[n] \psi_n^m(t), \quad (17)$$

according to the representation of the dilation with the longest time duration to be the $m = 0$ dilation, and that with the shortest duration the $m = -3$ one. It can be shown that this strategy for Daubechies wavelet of order 10 lead to a 40 dB separation between dilations that reduces to 24 dB increasing the number of dilation to 6 [5]. For simplicity, we will not consider the sign in the dilation indexes and we shall indicate each signal dilation by means of the indexes m_0, m_1, m_2 and m_3 decreasing the order of resolution.

Other than the Daubechies families of wavelt of order 8, 18, 20, the Beylkin wavelets of order 18, the Coifman families of order 30, and the Vaidyanathan wavelets of order 24 have been considered. We refer to these with D8, D18, D20, B18, C30 and V24 respectively. The same considerations made for the Daubechies wavelets can be extended for all the considered families. All the wavelets have normalized energy of 1 W/Hz, and the scaling factor in the frequency band modulators was set $\beta = 1$ corresponding to a Hausdorf dimension $H = -1/2$ optimizing the system for AWGN channels. We chose 1024 points for m_0 , and this

strategy lead to a 2048, 4096 and 8192 points representation for m_1 , m_2 and m_3 respectively. The energy associated to each dilation reduces to half within each decreasing of resolution. We implemented a digital quadrature modulation scheme, coding the inphase and quadrature component with 4 bits performing a 16 level constellation.

4.4.1 Channel Model

Noise is a random process affecting the decoding of the transmitted data, making it uncertain. Generally the effect of the noise is represented in simulation test-bed by an additive source at the front of the receiver. In this work, we considered the transmission of digital data over a AWGN channel, with zero mean and spectral power density N_0 affected by fast-fading effect. Fading effect is typically due to a particular terrain geometry or to a particular atmospheric conditions. We characterize the phenomena by defining the *fading-rate*, the number of variation of the signal level per time unit, and by the *fading intensity*, generally measured in dB that can be higher than 20 dB.

The fast-fading effect is simulated by adding some signals with different amplitudes and frequencies, distributed in a Doppler frequency range [20]. The rate is proportional to the frequency carrier (f_c) and to the speed of the mobile receiver (V_R). For instance, for $f_c = 900\text{MHz}$ and $V_R = 100\text{Km/h}$, the maximum Doppler frequency is about 90 Hz. Fig. 4. shows the corresponding *Ptolemy*-based implementation used in the simulations.

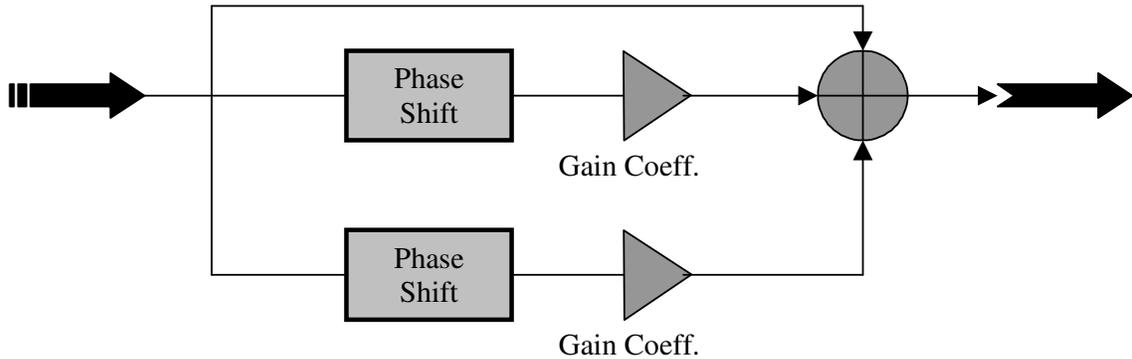


Fig.4 Fast-fading simulator scheme

4.4.2 Quadrature scheme transmitter/receiver

The scheme in Fig. 5 implements for each dilation m_i the following operation:

$$\sum_n q[n \bmod L] \psi_n^m (2^m t - n), \quad (18)$$

where $q[n]$ is a complex number representing the position of the corresponding symbol in the constellation.

The sum of four dilations will produce the signal to be transmitted over the channel.

We implemented the receiver by means of matched filters and samplers. The inphase and quadrature components are quantized separately and following added. Fig 6 and 7 show the decoding procedures.

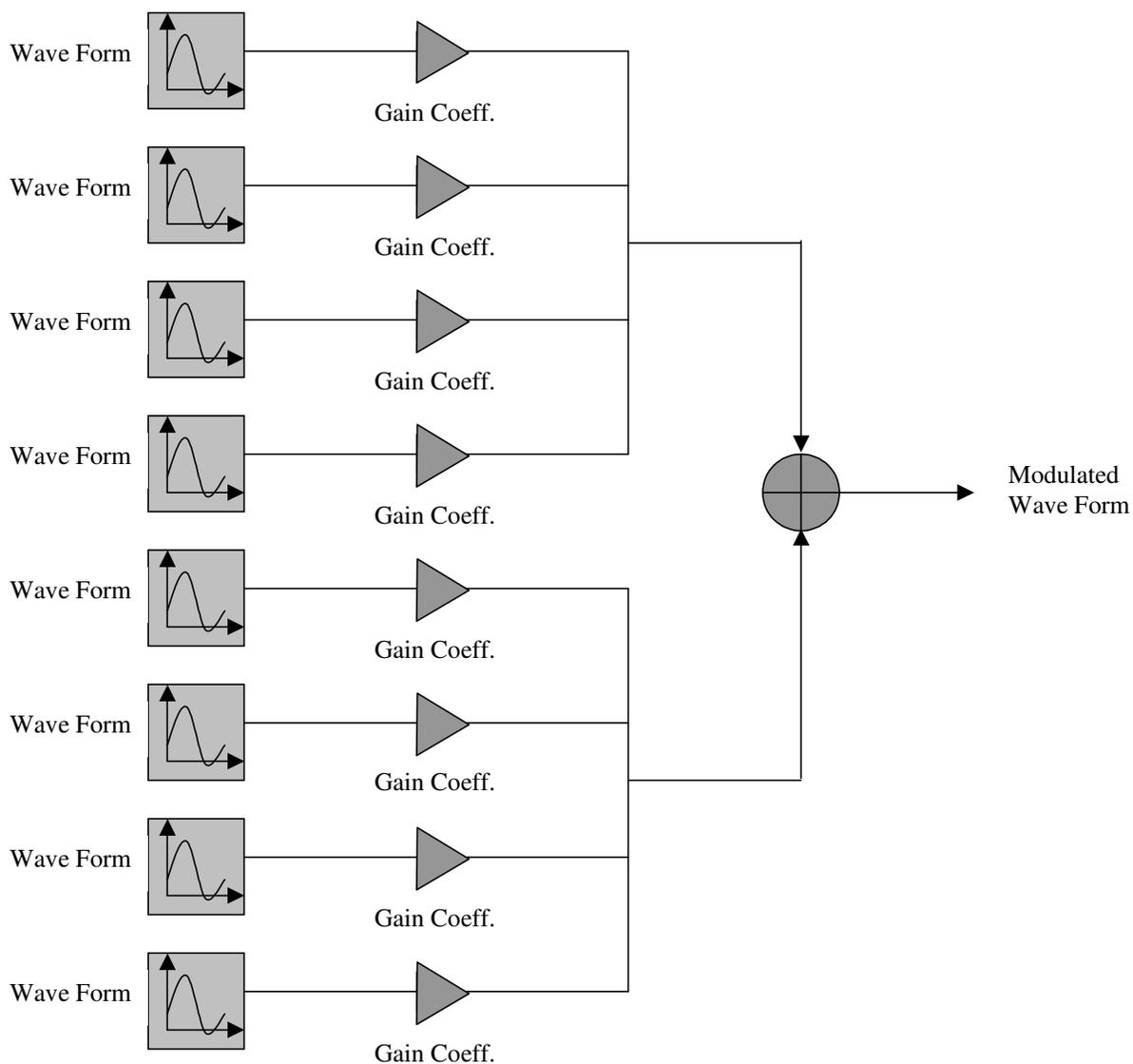


Fig. 5 Quadrature fractal modulation transmitter scheme (number of dilation $N=4$)

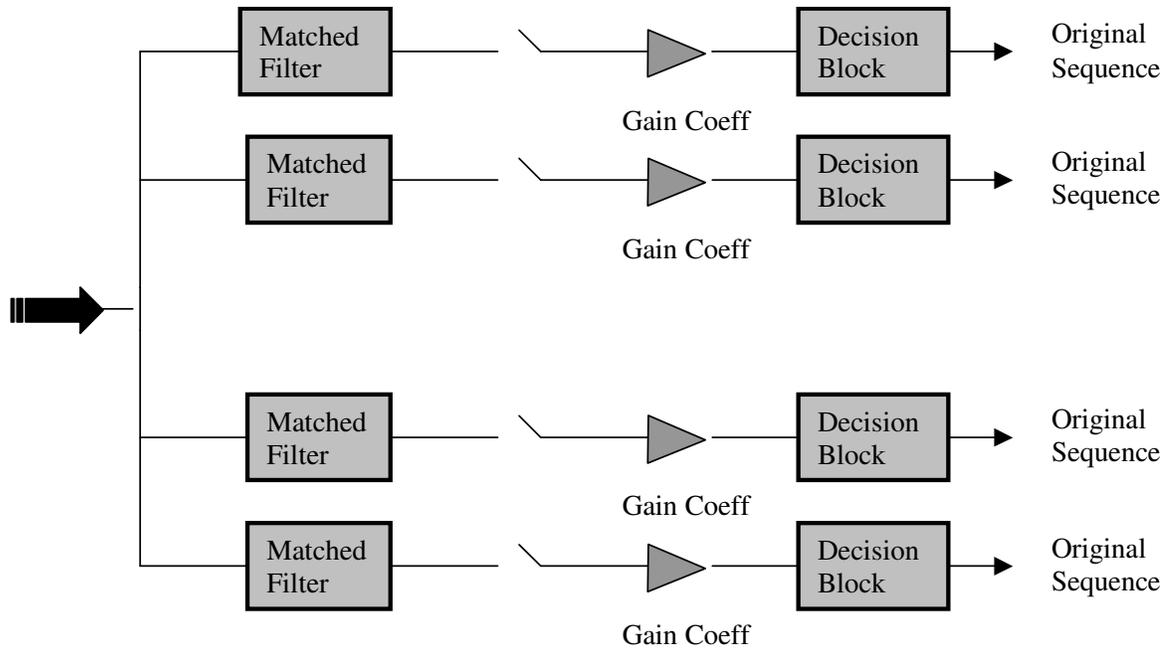


Fig. 6. Quadrature receiver scheme.

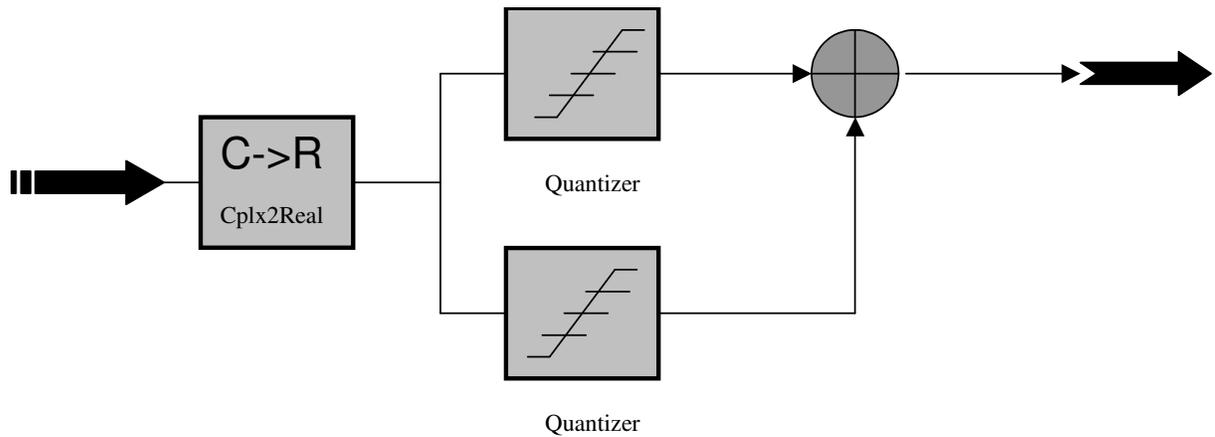


Fig. 7. Decision block for quadrature transmission

4.5 Performance analysis

In order to compare the performance of different modulation schemes, there is a need of exploiting a method not dependent to the bit-rate, received power and noise at the front receiver. Accordingly, often one refers to the energy per bit- noise ratio E_b / N_0 .

The signal transmitted over the channel is composed by wave with different energy, therefore to make consistent comparison, in our experimentation we collected the performance relevant to the simulated transmission schemes in terms of bit error rate at varying noise spectral power density N_0 .

In order to analyze the performance of the quadrature scheme, we simulated a classic 16-QAM modulation system. Tests have been accomplished by fixing the energy per bit E_b and varying N_0 . We used as test sequence to be transmitted the binary pattern 1101001101001001, corresponding to the four position (1,3), (-3,3), (-1, -3), (1, -1) in the constellation. The used mean energy per bit has been set to 2.5 W/Hz. Accordingly, in the dilation m_0 for example a noise spectral power density of 0.01 W/Hz yields to a σ_c^2 of about 24 dB.

The "fast-fading" channel has been simulated by adding at the front receiver two identical signals attenuated to 1/10 respect to the original power. Furthermore, these signals have been translated at frequency and phase of 75 Hz and 25 degrees, and 50 Hz and 35 degrees, respectively.

We present the performance results for all the different wavelet families previously mentioned: we dedicated special attention to the Daubechies's wavelets, extensively investigated in previous works [4][5][2][1], presenting the corresponding results in separate graphics at varying order. The results obtained with this wavelets family are presented in Figs. 8-11 for the different dilations that are compared with the QAM error rate. It can be observed that better results have been obtained increasing the resolution order. In particular, it results that the Daubechies wavelets allow to obtain significant performance improvement at the highest dilation compared to the standard QAM. In fact, the latter provides a bit-error rate more than 18 times higher respect to all the family order. At lower resolutions, the fractal modulation performance decreased achieving results comparable with that of the QAM in the case of the m_3 dilation.

A similar behavior has been found even for the other wavelet families, whose results are depicted in Figs. 12-15. Except the Coifman's wavelet families, that showed their relevant sensitivity to the fading-effect, all other considered families achieved good results.

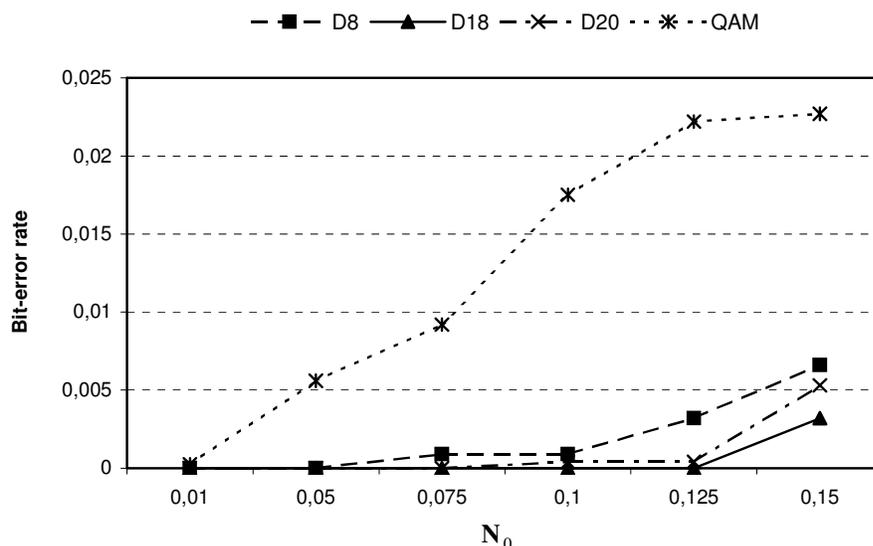


Fig. 8. Daubechies wavelet performance at m_0 dilation.

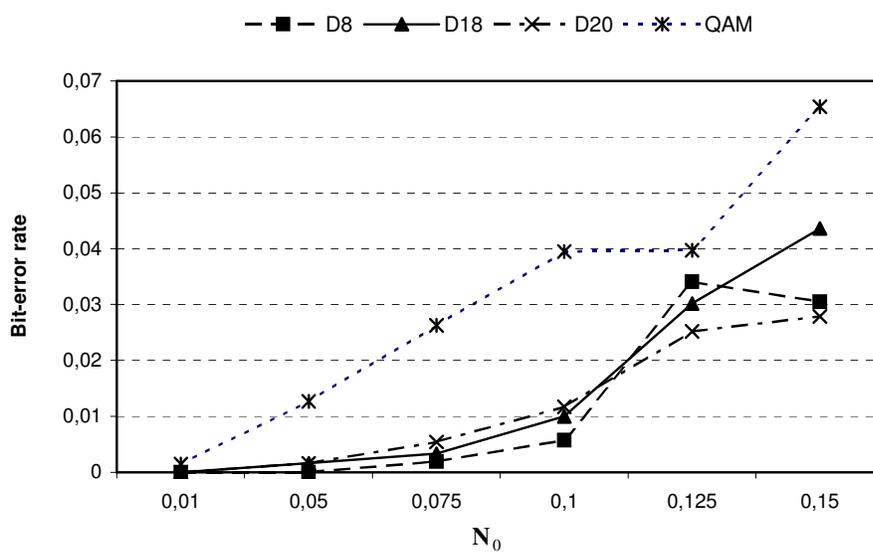


Fig. 9. Daubechies wavelet performance at m_1 dilation.

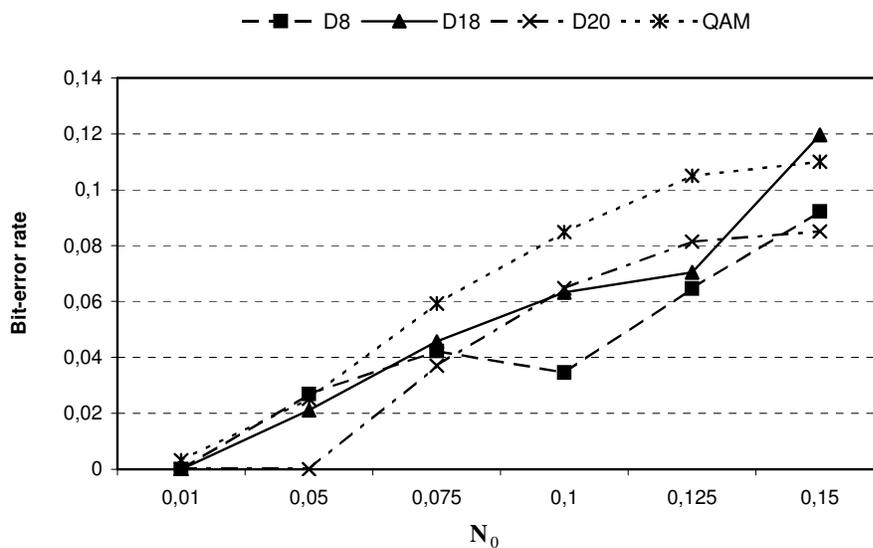


Fig. 10. Daubechies wavelet performance at m_2 dilation.

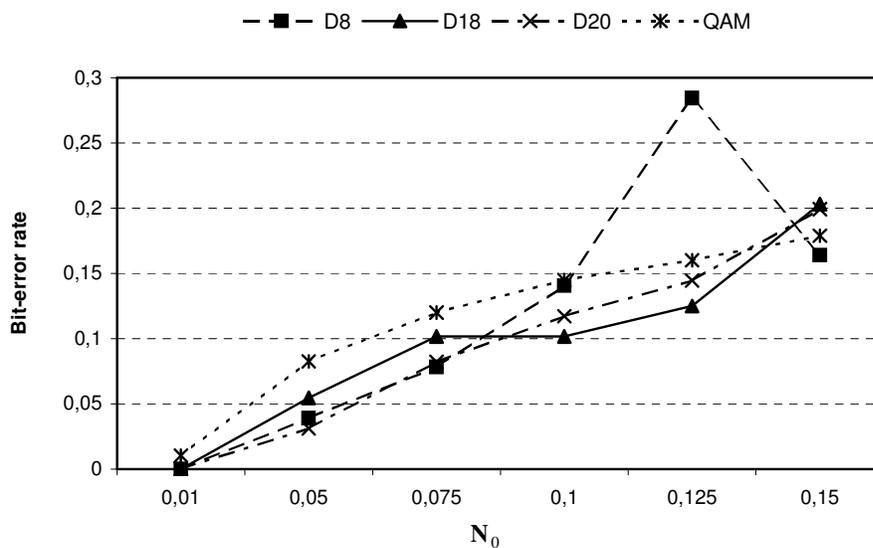


Fig. 11. Daubechies wavelet performance at m_3 dilation.

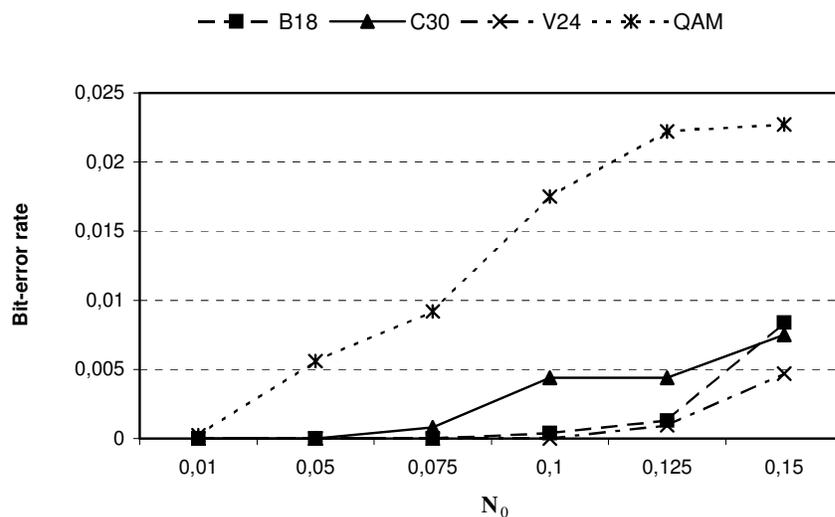


Fig. 12. Beylkin, Coifman and Vaidyanathan wavelet performance at m_0 dilation.

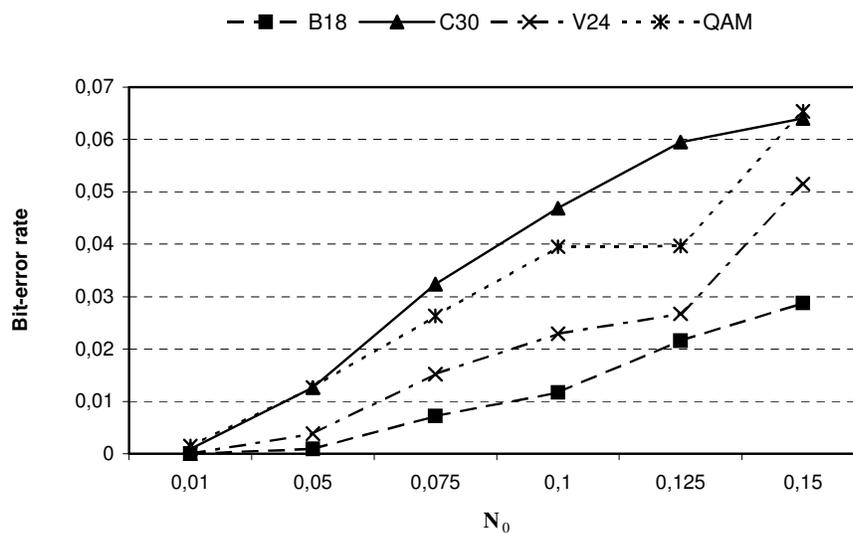


Fig. 13. Beylkin, Coifman and Vaidyanathan wavelet performance at m_1 dilation.

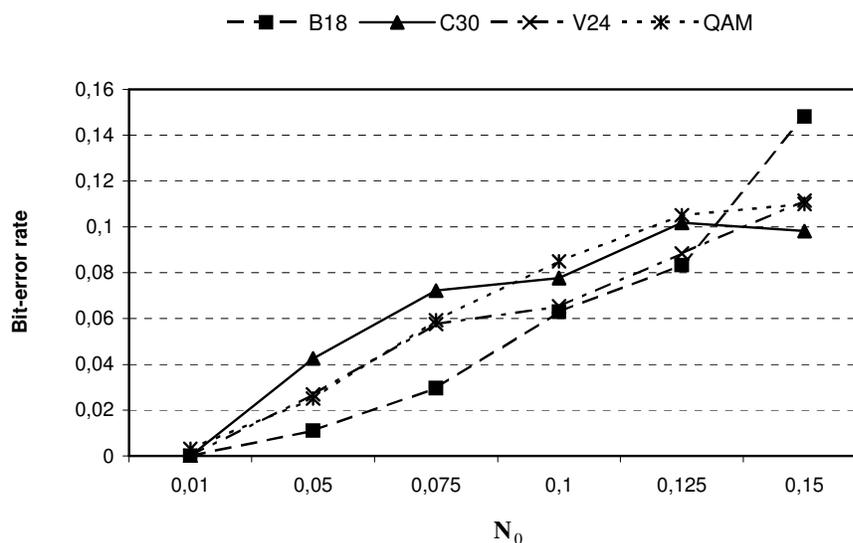


Fig. 14. Beylkin, Coifman and Vaidyanathan wavelet performance at m_2 dilation.

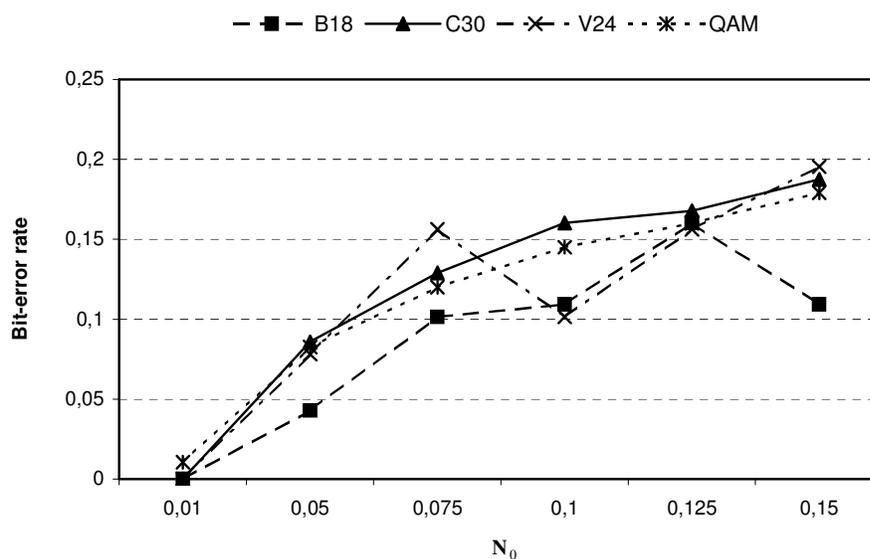


Fig. 15. Beylkin, Coifman and Vaidyanathan wavelet performance at m_3 dilation.

4.6 Conclusion

In this paper we investigated the problem of transmitting digital data over a fast-fading affected AWGN channel, by means of a fractal modulation system. We performed experiments using different wavelet families and we tested the effectiveness of the system for different orders.

We implemented a 16 level quadrature scheme and we compared it with a classic QAM system. Tests showed as the fractal implementation outperform of the latter in terms of achieved bit error rate and that that Beylkin's and Daubechies's wavelet of order 8, 18 and 20 represent the more reliable and efficient waves in terms of error robustness and bandwidth. By the light of the achieved results, further developments of the algorithm could concern a comparison between the proposed scheme and an OFDM transmission system for broadcasting applications.

References.

- [1] Wornell G. W., "Signal Processing with Fractals: A Wavelet Based Approach", Upper Saddle River, NJ: Prentice-Hall, 1995.
- [2] Wornell G. W. and Oppenheim A. V., "Wavelet-Based Representations for a class of Self-Similar Signals with Application to Fractal Modulation", *IEEE Trans. Inform. Theory*, Vol. 38, pp. 785-800, March 1992.
- [3] Akaiwa Y., "Introduction to Digital Mobile Communication", John Wiley & Sons Inc., 1997.
- [4] Ptasiński H. S. and Fellman R. D., "Implementation and Simulation of a Fractal Modulation Communication System", *IEEE Intern. Conf. on Communications*, Vol. 3, pp. 1551-1555, 1994.
- [5] Ptasiński H. S. and Fellman R. D., "Performance Analysis of a Fractal Modulation Communication System", *SPIE Proceedings Vol. 2242-Wavelet Applications*, pp. 78-86, 1994.
- [6] Vetterli M. and Kovacevic J., "Wavelets and Subband Coding", Prentice Hall, Signal Processing Series, Englewood Cliffs, NJ, 1995.
- [7] Daubechies I., "Ten Lectures on Wavelets", Philadelphia, PA, SIAM, 1992.
- [8] Grossman A. and Morlet J., "Decomposition of Hardy functions into Square Integrable Wavelets of Constant shape", *SIAM Journal of Mathematical Analysis*, vol. 15, pp. 723-736, 1984.
- [9] Haar A., "Zur Theorie der orthogonalen Funktionen-Systeme", *Math. Ann.*, Vol. 69, pp. 331-371, 1910.
- [10] Meyer Y., "Ondelettes et fonctions splines", in *Sem. Equations aux Derivées Partielles*, Ecole Polytechnique, Paris, France, Dec. 1986.
- [11] Strang G., "Wavelets", *American Scientist*, Vol. 82, pp. 250-255, April 1994.
- [12] Strang G., "Wavelets and Dilation Equations: A Brief Introduction", *SIAM Rev.*, Vol. 31, pp. 614-627, Dec. 1989.

- [13] Strang G. and Nguyen T., “Wavelets and Filter Banks”, Wellesley-Cambridge Press, 1996.
- [14] Mallat S., “Multiresolution approximations and wavelet orthonormal bases of $L^2\mathbb{R}$ ”, *Trans. Amer. Math. Soc.*, Vol. 315, pp. 69-87, Sep. 1989.
- [15] Mallat S., “A Theory for Multiresolutions Signal Decompositions: The Wavelet Representation”, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-11, pp. 674-693, July 1989.
- [16] <http://ptolemy.eecs.berkeley.edu>.
- [17] Daubechies I., “The Wavelet Transform, Time-Frequency Localization and Signal Analysis”, *IEEE Trans. Inform. Theory*, Vol. IT-36, pp. 961-1005, Sep1990.
- [18] Daubechies I., “Orthonormal Bases of Compactly Supported Wavelets”, *Commun. Pure Appl. Math.*, Vol. 41, pp. 909-996, Nov. 1988.
- [19] <http://www.intergalact.com/macwavelets/macwavelets.html>.
- [20] Freeman R. L., “Radio System Design for Telecommunications” 2nd ed., John Wiley & Sons Inc., 1997.

Appendix A

SOME WAVELET FILTERS

A.1 Haar filters

The simplest wavelet filters have been found by Haar in 1910 [1]. These filters are of length 2 and the low-pass filter is an averaging operator, while the high-pass is a difference operator. So, the two filters have the following expression:

$$h(n) = \left\{ \frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}} \right\} \quad (\text{A.1})$$

$$g(n) = \left\{ \frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}} \right\} \quad (\text{A.2})$$

and derive from the following scaling function and wavelet:

$$\phi(n) = \left\{ \frac{1}{2} \quad \frac{1}{2} \right\} \quad (\text{A.3})$$

$$\psi(n) = \left\{ \frac{1}{2} \quad -\frac{1}{2} \right\} \quad (\text{A.4})$$

A.2 Daubechies filters

Daubechies' family of wavelets is built in order to produce maximally flat filters [2]. The design procedure consists of finding orthogonal low-pass filters with a large number of zeros (that is, maximum flatness) at $\omega=\pi$. In the particular case of two zeros, this construction leads to the D_2 filter, that is a length-4 orthogonal filter. The length-2 Daubechies filter coincides with the Haar filter.

Daubechies has calculated sets of orthonormal compactly supported wavelets and has proved that there is no closed formula, apart from Haar wavelets, for this kind of wavelets.

Since the wavelet and scaling functions are quadrature mirror filters and their dilation and translation can span two complement spaces, a trigonometric polynomial $m_0(f)$ can be defined [3] in order to satisfy the following conditions:

$$|m_0(f)|^2 + |m_0(f + \pi)|^2 = 1 \quad (\text{A.5})$$

$$m_0(0) = 1 \quad (\text{A.6})$$

The scaling function ϕ and the corresponding wavelet ψ can be then defined by the following:

$$\hat{\phi}(f) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}f) \quad (\text{A.7})$$

$$\hat{\psi}(f) = -e^{-if/2} m_0\left(\frac{f}{2} + \pi\right) \hat{\phi}\left(\frac{f}{2}\right) \quad (\text{A.8})$$

Daubechies proposed the following trigonometric polynomial:

$$|m_0(f)|^2 = \left(\frac{1+e^{-if}}{2}\right)^2 N \sum_{k=0}^{N-1} \binom{N-1+k}{k} \left(\sin^2 \frac{f}{2}\right)^k + \left(\sin^2 \frac{f}{2}\right)^N R\left(\frac{1}{2} - \sin^2 \frac{f}{2}\right) \quad (\text{A.9})$$

where R is an odd polynomial, chosen so that the right hand side of the equation becomes positive.

Daubechies introduced also other wavelets with more vanishing moments for both the scaling function and the wavelet. Thus, they should satisfy also the following conditions:

$$\int \phi(x) dx = 1 \quad (\text{A.10})$$

$$\int x^l \phi(x) dx = 0 \quad \text{for } l = 1, 2, \dots, L-1 \quad (\text{A.11})$$

$$\int x^l \psi(x) dx = 0 \quad \text{for } l = 1, 2, \dots, L-1 \quad (\text{A.12})$$

These new constraints mean that the inner product of the scaling function with a smooth function f only depends on $f(2^j k)$ and its derivatives of order greater or equal to L . Imposing these vanishing moments forces symmetry to $\phi(x)$ as well. The wavelets derived from these criteria are called ‘‘Coiflets of order L ’’, as Coifman requested these filters for some signal processing applications [4].

The coefficients of the 4-length filter are reported in table A.1.

n	h(n)
0	0.48296
1	0.83652
2	0.22414
3	-0.12941

Table A.1 Coefficients of 4-length Daubechies filter (D_4)

A.3 Lemariè-Battle filters

Lemariè [5] and Battle [6] studied a class of multiresolution approximation of $L^2(\mathbb{R})$ constructed starting from polynomial splines of order $2p+1$.

Let V_1 be the vectorial space of all continuous and derivable functions in $L^2(\mathbb{R})$ that coincide for p times with polynomials of order $2p+1$ into each interval $[k, k+1]$, for each $k \in \mathbb{Z}$. A multiresolution approximation of $L^2(\mathbb{R})$ is composed by V_1 and by all V_{2^j} obtained by V_1 through causality properties (see 3. in 1.2.1).

Lemariè has shown that the scaling function associated with this multiresolution approximation can be written in the following way:

$$\hat{\phi}(\omega) = \frac{1}{\omega^n \sqrt{\sum_{2n}(\omega)}} \quad (\text{A.13})$$

with $n = 2 + 2p$ and the function $\sum_n(\omega)$ defined as:

$$\sum_n(\omega) = \sum_{k=-\infty}^{+\infty} \frac{1}{(\omega + 2k\pi)^n} \quad (\text{A.14})$$

An explicit expression of $\sum_n(\omega)$ can be obtained from the calculation of the derivative of order $n - 2$ of the following expression:

$$\sum_2(\omega) = \frac{1}{4 \operatorname{sen}^2 \frac{\omega}{2}} \quad (\text{A.15})$$

For theorem 1.2, the function $\hat{\phi}(\omega)$ is related to the transfer function $\mathbf{H}(\omega)$ of a QMF filter by the following:

$$\hat{\phi}(2\omega) = \mathbf{H}(\omega) \cdot \hat{\phi}(\omega) \quad (\text{A.16})$$

And for the (A.13) it holds:

$$\mathbf{H}(\omega) = \sqrt{\frac{\sum_{2n}(\omega)}{2^{2n} \sum_{2n}(2\omega)}} \quad (\text{A.17})$$

The Fourier transform of the corresponding orthonormal wavelet can be obtained from the (1.22) of theorem 1.3:

$$\hat{\psi}(\omega) = e^{-i\frac{\omega}{2}} \cdot \overline{\mathbf{H}}\left(\frac{\omega}{2} + \pi\right) \hat{\phi}\left(\frac{\omega}{2}\right) = \frac{e^{-i\frac{\omega}{2}}}{\omega^n} \cdot \frac{\sqrt{\sum_{2n}\left(\frac{\omega}{2} + \pi\right)}}{\sqrt{\sum_{2n}(\omega) \sum_{2n}\left(\frac{\omega}{2}\right)}} \quad (\text{A.18})$$

The wavelet defined by (A.18) is a function that decreases in an exponential manner. The choice of $p = 1$ and then of $n = 4$ corresponds to a multiresolution approximation built starting from cubic splines.

The function $\sum_8(\omega)$ can be expressed in the following way:

$$\sum_8(\omega) = \frac{N_1(\omega) + N_2(\omega)}{105 \cdot \operatorname{sen}^8\left(\frac{\omega}{2}\right)} \quad (\text{A.19})$$

$$\text{with } N_1(\omega) = 5 + 30 \cdot \cos^2\left(\frac{\omega}{2}\right) + 30 \cdot \operatorname{sen}^2\left(\frac{\omega}{2}\right) \cdot \cos^2\left(\frac{\omega}{2}\right) \quad (\text{A.20})$$

$$\text{and } N_2(\omega) = 2 \cdot \text{sen}^4\left(\frac{\omega}{2}\right) \cdot \text{cos}^2\left(\frac{\omega}{2}\right) + 70 \cdot \text{cos}^4\left(\frac{\omega}{2}\right) + \frac{2}{3} \cdot \text{sen}^6\left(\frac{\omega}{2}\right). \quad (\text{A.21})$$

For such multiresolution approximation the Fourier transforms of the scaling function and of the corresponding orthonormal wavelet can be derived from the (A.13) and (A.18) with $n = 4$.

The transfer function of the filter $\mathbf{H}(\omega)$ is given by the (A.17) and the one of the mirror filter $\mathbf{G}(\omega)$ by the (1.22).

The filter $h(n)$ is symmetrical with respect to the origin. The first 25 coefficients of the impulsive response of the filter are reported in table A.2.

n	h(n)
±12	0.000
±11	-0.002
±10	-0.003
±9	0.006
±8	0.006
±7	-0.013
±6	-0.012
±5	0.030
±4	0.023
±3	-0.078
±2	-0.035
±1	0.307
0	0.542

Table A.2 Coefficients of length-25 Lemarié-Battle filter

Appendix B

STATISTICAL PROPERTIES OF THE WAVELET TRANSFORM

The statistical distribution of wavelet coefficients at a fixed resolution and orientation is a symmetric distribution with a nearly zero mean and small variance (figure B.1).

This distribution is often modeled as a Laplacian distribution, but it falls off really more rapidly, and it is better approximated by the *generalized Gaussian distribution* [7]:

$$p_{2^j}^k = a_{2^j}^k \cdot \exp\left(-|b_{2^j}^k x|^{r_{2^j}^k}\right) \text{ with } a_{2^j}^k = \frac{b_{2^j}^k r_{2^j}^k}{2\Gamma(1/r_{2^j}^k)} \text{ and } b_{2^j}^k = \frac{1}{\sigma_{2^j}^k} \frac{\Gamma(3/r_{2^j}^k)^{1/2}}{\Gamma(1/r_{2^j}^k)^{1/2}} \quad (\text{B.1})$$

where $\sigma_{2^j}^k$ is the standard deviation of the subband distribution at orientation K and resolution 2^j and $\Gamma(\cdot)$ is the Gamma function.

This formula contains the Gaussian and the Laplacian PDF as special cases:

- for $r_{2^j}^k = 2$ it is the Gaussian PDF;
- for $r_{2^j}^k = 1$ it is the Laplacian PDF.

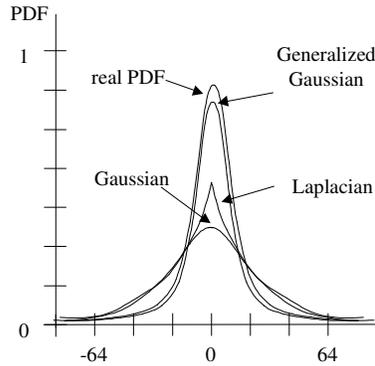


Figure B.1
Approximations of the statistical distribution of wavelet coefficients

Some data regarding the distribution of the wavelet coefficients of the images used in (2.2.4) and in (5.4.2) are reported in the following. All original images are composed by 512x512 pixels and quantized with 256 gray levels.

Image	Orientation	Maximum value	Minimum value	Mean value	Variance
Agave	Resolution j=-1				
	Horizontal	46.977	-49.433	0.026	14.473
	Vertical	53.846	-55.426	-0.026	16.095
	Diagonal	25.112	-24.503	-0.001	3.043
	Resolution j=-2				
	Horizontal	59.087	-43.655	-0.007	30.870
	Vertical	57.265	-55.168	0.032	28.427
	Diagonal	29.198	-35.783	0.009	11.643
	Resolution j=-3				
Horizontal	64.468	-44.537	-0.021	50.081	
Vertical	44.074	-42.388	0.088	33.202	
Diagonal	37.789	-36.058	0.037	17.607	
Airplane	Resolution j=-1				
	Horizontal	11.746	-8.262	0.002	1.050
	Vertical	44.363	-54.093	-0.002	6.271
	Diagonal	6.231	-3.953	0.001	0.295
	Resolution j=-2				
	Horizontal	39.200	-40.917	0.018	16.243
	Vertical	41.899	-47.995	0.001	24.189
	Diagonal	15.257	-14.367	0.002	2.827
	Resolution j=-3				
Horizontal	58.900	-55.204	0.027	54.335	
Vertical	45.378	-49.223	0.107	58.040	
Diagonal	25.259	-22.646	0.027	13.679	
Baboon	Resolution j=-1				
	Horizontal	23.365	-21.617	0.002	12.612
	Vertical	46.708	-49.384	-0.005	35.583
	Diagonal	15.953	-14.458	-0.004	6.304
	Resolution j=-2				
	Horizontal	37.391	-34.196	-0.018	44.683
	Vertical	49.000	-46.803	-0.083	54.912
	Diagonal	23.911	-39.020	-0.019	16.963
	Resolution j=-3				
Horizontal	40.381	-38.972	-0.108	47.675	
Vertical	41.075	-41.771	-0.037	54.621	
Diagonal	27.476	-34.579	-0.009	25.669	
Carmen	Resolution j=-1				
	Horizontal	67.672	-61.027	-0.000	41.679
	Vertical	78.366	-70.164	0.026	38.523
	Diagonal	30.885	-30.527	-0.000	4.006
	Resolution j=-2				
	Horizontal	70.156	-81.172	0.014	52.523
	Vertical	62.198	-101.056	-0.092	42.684
	Diagonal	41.935	-41.031	-0.024	14.223
	Resolution j=-3				
Horizontal	82.395	-70.729	0.033	111.376	
Vertical	74.758	-68.106	-0.064	100.582	
Diagonal	48.886	-39.041	-0.050	40.080	

Image	Orientation	Maximum value	Minimum value	Mean value	Variance
<i>Cat</i>	Resolution j=-1				
	Horizontal	53.957	-45.495	-0.003	10.314
	Vertical	60.520	-68.022	-0.041	15.941
	Diagonal	41.274	-41.060	0.003	4.563
	Resolution j=-2				
	Horizontal	47.924	-53.713	-0.0045	20.272
	Vertical	52.418	-55.870	0.070	24.044
	Diagonal	40.257	-25.781	0.033	7.717
	Resolution j=-3				
	Horizontal	47.862	-68.457	-0.006	41.098
	Vertical	50.724	-56.765	0.118	52.834
	Diagonal	44.543	-33.540	0.018	16.917
<i>Lena</i>	Resolution j=-1				
	Horizontal	20.788	-24.561	0.003	3.625
	Vertical	34.932	-34.234	0.031	10.663
	Diagonal	13.415	-13.694	0.001	2.345
	Resolution j=-2				
	Horizontal	51.086	-38.512	0.000	10.290
	Vertical	69.465	-65.162	0.025	26.210
	Diagonal	24.427	-29.562	0.015	6.970
	Resolution j=-3				
	Horizontal	37.329	-32.903	-0.009	17.239
	Vertical	41.591	-41.315	-0.100	52.683
	Diagonal	26.947	-29.629	-0.034	16.324
<i>Masquerade</i>	Resolution j=-1				
	Horizontal	75.575	-60.380	-0.014	32.560
	Vertical	94.231	-78.869	-0.024	47.285
	Diagonal	80.298	-85.806	-0.002	17.193
	Resolution j=-2				
	Horizontal	71.393	-58.058	-0.046	44.073
	Vertical	91.993	-63.802	0.008	63.595
	Diagonal	34.432	-34.028	-0.014	20.063
	Resolution j=-3				
	Horizontal	49.614	-54.072	0.055	64.650
	Vertical	52.685	-73.564	-0.243	91.682
	Diagonal	31.134	-31.237	0.018	27.035
<i>Model</i>	Resolution j=-1				
	Horizontal	49.911	-46.912	0.022	15.224
	Vertical	51.322	-56.583	-0.003	20.051
	Diagonal	24.364	-25.812	-0.003	2.096
	Resolution j=-2				
	Horizontal	67.136	-60.309	-0.036	18.781
	Vertical	43.262	-46.637	0.031	26.261
	Diagonal	26.636	-29.915	-0.019	5.433
	Resolution j=-3				
	Horizontal	44.426	-40.109	-0.120	30.509
	Vertical	47.578	-52.157	-0.003	50.070
	Diagonal	30.913	-25.789	0.029	12.632

Image	Orientation	Maximum value	Minimum value	Mean value	Variance
<i>Peppers</i>	Resolution j=-1				
	Horizontal	36.705	-42.713	0.001	2.636
	Vertical	46.311	-39.925	-0.021	6.099
	Diagonal	20.895	-15.738	0.002	1.937
	Resolution j=-2				
	Horizontal	27.341	-23.143	0.022	7.735
	Vertical	34.695	-39.420	0.040	14.329
	Diagonal	16.979	-38.834	-0.006	1.884
	Resolution j=-3				
	Horizontal	47.368	-41.769	0.028	27.218
	Vertical	36.480	-48.707	-0.145	39.237
	Diagonal	26.009	-24.779	-0.019	6.924
<i>Tiffany</i>	Resolution j=-1				
	Horizontal	28.780	-23.527	0.006	3.284
	Vertical	35.931	-47.363	0.014	8.174
	Diagonal	22.113	-21.304	0.002	2.661
	Resolution j=-2				
	Horizontal	22.868	-32.275	-0.011	4.815
	Vertical	39.879	-46.565	-0.031	12.868
	Diagonal	21.837	-15.930	0.003	3.449
	Resolution j=-3				
	Horizontal	30.636	-31.758	0.029	9.803
	Vertical	34.533	-45.096	-0.018	15.485
	Diagonal	14.149	-18.241	-0.031	5.109
<i>Residual Baboon</i>	Resolution j=-1				
	Horizontal	22.977	-21.865	0.002	13.130
	Vertical	45.752	-49.898	-0.001	36.078
	Diagonal	15.424	-15.223	-0.006	6.739
	Resolution j=-2				
	Horizontal	37.566	-33.719	0.009	45.190
	Vertical	48.196	-47.061	-0.098	55.517
	Diagonal	24.855	-39.242	-0.015	17.242
	Resolution j=-3				
	Horizontal	39.857	-36.675	-0.011	60.309
	Vertical	47.983	-46.052	0.069	73.102
	Diagonal	28.332	-35.476	-0.063	25.977
<i>Residual Lena</i>	Resolution j=-1				
	Horizontal	20.139	-23.153	0.001	4.141
	Vertical	35.959	-34.041	0.051	11.192
	Diagonal	13.229	-13.022	-0.002	2.768
	Resolution j=-2				
	Horizontal	51.059	-38.230	-0.004	10.680
	Vertical	69.089	-65.152	0.004	26.568
	Diagonal	24.260	-28.489	0.018	7.174
	Resolution j=-3				
	Horizontal	38.360	-38.513	0.020	26.456
	Vertical	51.180	-49.385	0.003	68.278
	Diagonal	27.755	-31.451	-0.018	17.722

Image	Orientation	Maximum value	Minimum value	Mean value	Variance
Residual <i>Masquerade</i>	Resolution j=-1				
	Horizontal	75.868	-60.560	-0.019	32.650
	Vertical	94.068	-78.492	-0.024	47.304
	Diagonal	80.452	-86.051	-0.002	17.208
	Resolution j=-2				
	Horizontal	71.290	-58.372	-0.062	44.620
	Vertical	91.424	-64.221	0.006	63.711
	Diagonal	35.030	-33.967	-0.014	20.051
	Resolution j=-3				
	Horizontal	61.259	-56.766	-0.008	89.203
	Vertical	55.678	-77.320	-0.254	116.173
	Diagonal	34.021	-35.067	0.019	28.115

References

- [1] A. Haar, "Zur Theorie der Orthogonalen Functionensysteme," Math. Annal., Vol. 69, pp. 331-371, 1910
- [2] I. Daubechies, "Orthonormal bases of compactly supported wavelets," Comm. on Pure and Applied Mathematics, Vol. XLI, pp. 909-996, 1988
- [3] I. Daubechies, "Ten lectures on wavelets," CBMS-NFS, Regional Conf. Series in Applied Mathematics, 1992
- [4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE trans. On Info. Theory, Vol. 36, No. 5, pp. 961-1005, 1990
- [5] P.G.Lemariè, "Ondelettes à localisation exponentielles," Journ. Math. Pures et Appl., Vol.67, pp.227-236, 1988
- [6] G.Battle, "A block spin construction of ondelettes. Part I: Lemariè functions," Commun. Math. Phys, Vol.110, pp.601-615, 1987
- [7] P.H. Westerink, "Subband coding of images," Ph.D. Thesis, Delft University of Technology, The Netherlands, 1989