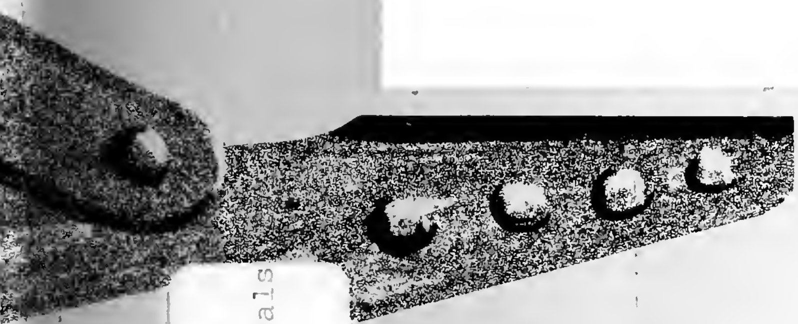
Robotics Research Technical Report

Characterization of Signals from Multiscale Edges

by

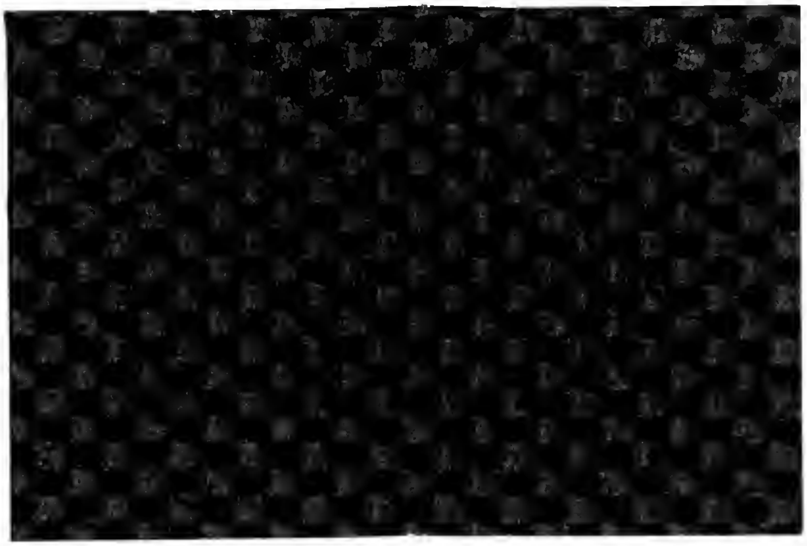
Stephane Mallat and Sifen Zhong

Technical Report No. 592
Robotics Report No. 258
November, 1991



lew York University
Cour titute of Mathematical Sciences

omputer Science Division
Mercer Street New York, N.Y. 10012



Characterization of Signals from Multiscale Edges

by

Stephane Mallat and Sifen Zhong

Technical Report No. 592
Robotics Report No. 258
November, 1991

New York University
Dept. of Computer Science
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, New York 10012

Work on this paper has been supported by NSF grant IRI-890331, AFOSR grant #90-0040 and ONR grant N00014-91-J-1967.

CHARACTERIZATION OF SIGNALS FROM MULTISCALE EDGES

Stephane Mallat and Sifen Zhong

Courant Institute, New York University
251 Mercer Street, New York, NY 10012

Abstract

A multiscale Canny edge detection is equivalent to finding the local maxima of a wavelet transform. We study the properties of multiscale edges through the wavelet formalism. For pattern recognition, one often needs to discriminate different types of edges. We show that the evolution of wavelet local maxima across scales characterize the local shape of irregular structures. Numerical descriptors of edge types are derived. The completeness of a multiscale edge representation is also studied. We describe an algorithm that reconstructs a close approximation of one and two-dimensional signals from their multiscale edges. For images, the reconstruction errors are below our visual sensitivity. As an application, we implement a compact image coding algorithm that selects important edges and compresses the image data by factors over 30.

This research was supported by the NSF grant IRI-890331, AFOSR grant AFOSR-90-0040 and ONR grant N00014-91-J-1967.

1. Introduction

Points of sharp variations are often among the most important features for analyzing the properties of transient signals or images. In images, they are generally located at the boundaries of important image structures. In order to detect the contours of small structures as well as the boundaries of larger objects, several researchers in computer vision have introduced the concept of multiscale edge detection [24,29,31]. The scale defines the size of the neighborhood where the signal changes are computed. The wavelet transform is closely related to multiscale edge detection and can provide a deeper understanding of these algorithms. We concentrate on the Canny edge detector [3], which is equivalent to finding the local maxima of a wavelet transform modulus.

There are many different types of sharp variation points in images. Edges created by occlusions, shadows, highlights, roofs, textures... have very different local intensity profiles. To label more precisely an edge that has been detected, it is necessary to analyze its local properties. In mathematics, singularities are generally characterized by their Lipschitz exponents. The wavelet theory proves that these Lipschitz exponents can be computed from the evolution across scales of the wavelet transform modulus maxima. We derive a numerical procedure to measure these exponents. If an edge is smooth, we can also estimate how smooth it is from the decay of the wavelet transform maxima across scales. Lipschitz exponents and smoothing factors are numerical descriptors that allow us to discriminate the intensity profiles of different types of edges.

An important open problem in computer vision is to understand how much information is carried by multiscale edges and how stable is a multiscale edge representation. This issue is important in pattern recognition where one needs to know whether some interesting information is lost when representing a pattern with edges. We study the reconstruction of one and two-dimensional signals from multiscale edges detected by the wavelet transform modulus maxima. It has been conjectured [22,24] that multiscale edges characterize uniquely one and two-dimensional signals, but recently Meyer [27] has found counter-examples to these conjectures. In spite of these counter-examples, we show that one can reconstruct a close approximation of the original signal from multiscale edges. The reconstruction algorithm is based on alternate projections. We prove its convergence and derive a lower bound for the convergence rate. Numerical results are given both for one and two-dimensional signals. The differences between the original and reconstructed images are not visible on a high quality video monitor.

The ability to reconstruct images from multiscale edges has many applications in signal processing. It allows us to process the image information with edge based algorithms. We describe a compact image coding algorithm that keeps only the "important" edges. The image that is recovered from these main features has lost some small details but is visually of good quality. Examples with compression ratio over 30 are shown. Another application to the removal of noises from signals is described in [23].

The article is organized as follow. Section 2 relates multiscale edge detection to the wavelet transform. It shows that a Canny edge detector is equivalent to finding the local maxima of a wavelet transform modulus. Until Section 6, we concentrate on one-dimensional signals. Section 3.1 reviews the wavelet transform properties that are important for understanding multiscale edges. The wavelet transform is first defined over functions of continuous variables and Section 3.2 explains how to discretize this model. The numerical implementation of fast wavelet transform algorithms is given in Appendix 2. Section 4 explains how to characterize different types of sharp signal variations, from the evolution across scales of the wavelet transform maxima. Section 5 studies the reconstruction of signals from multiscale edges. We review some previous results and explain how to formalize the reconstruction problem within the wavelet framework. The reconstruction algorithm is described in Section 5.2 and numerical results are presented in Section 5.3. A two-dimensional extension of the wavelet transform is given in Section 6.1 and its discrete version is explained in Section 6.2. Fast two-dimensional wavelet algorithms are given in Appendix 4. Section 7 differentiates the edges of an image from the evolution across scales of the wavelet modulus maxima. The reconstruction of images from multiscale edges is explained in Section 8.1 and numerical examples are shown in Section 8.2. Section 9 describes an application to compact image coding.

Notation

$\mathbf{L}^2(\mathbf{R})$ denotes the Hilbert space of measurable, square-integrable one-dimensional functions $f(x)$. For $f(x) \in \mathbf{L}^2(\mathbf{R})$ and $g(x) \in \mathbf{L}^2(\mathbf{R})$, the inner product of $f(x)$ with $g(x)$ is written:

$$< g(x), f(x) > = \int_{-\infty}^{+\infty} g(x) f(x)\, dx.$$

The norm (energy) of $f(x) \in \mathbf{L}^2(\mathbf{R})$ is given by

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(x)|^2\, dx.$$

We denote the convolution of two functions $f(x) \in \mathbf{L}^2(\mathbf{R})$ and $g(x) \in \mathbf{L}^2(\mathbf{R})$ by

$$f * g(x) = \int_{-\infty}^{+\infty} f(u)\, g(x-u)\, du.$$

The Fourier transform of $f(x) \in \mathbf{L}^2(\mathbf{R})$ is written $\hat{f}(\omega)$ and is defined by

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)\, e^{-i\omega x}\, dx.$$

$\mathbf{L}^2(\mathbf{R}^2)$ is the Hilbert space of measurable, square-integrable two dimensional functions $f(x,y)$. The norm of $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ is given by

$$\|f\|^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x,y)|^2\, dx\, dy \quad .$$

The Fourier transform of $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ is written $\hat{f}(\omega_x, \omega_y)$ and is defined by

$$\hat{f}(\omega_x, \omega_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\, e^{-i(\omega_x x + \omega_y y)}\, dx\, dy\ .$$

2. Multiscale Edge Detection

Most multiscale edge detectors smooth the signal at various scales and detect sharp variation points from their first or second order derivative. The extrema of the first derivative correspond to the zero-crossings of the second derivative and to the inflection points of the smoothed signal (see Fig. 1). This section explains how these multiscale edge detection algorithms are related to the wavelet transform.

We call a *smoothing function* any function $\theta(x)$ whose integral is equal to 1 and which converges to 0 at infinity. For example, one can choose $\theta(x)$ equal to a Gaussian. We suppose that $\theta(x)$ is twice differentiable and define respectively $\psi^a(x)$ and $\psi^b(x)$ as the first and second order derivative of $\theta(x)$

$$\psi^a(x) = \frac{d\theta(x)}{dx} \quad \text{and} \quad \psi^b(x) = \frac{d^2\theta(x)}{dx^2} \quad . \tag{1}$$

By definition, the functions $\psi^a(x)$ and $\psi^b(x)$ can be considered as wavelets because their integral is equal to 0

$$\int_{-\infty}^{+\infty} \psi^a(x)\,dx = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} \psi^b(x)\,dx = 0 \ .$$

In this article, we denote

$$\xi_s(x) = \frac{1}{s}\,\xi(\frac{x}{s})$$

the dilation by a scaling factor s of any function $\xi(x)$. A wavelet transform is computed by convolving the signal with a dilated wavelet. The wavelet transform of $f(x)$ at the scale s and position x, computed with respect to the wavelet $\psi^a(x)$, is defined by

$$W_s^a f(x) = f * \psi_s^a(x) \ . \tag{2}$$

The wavelet transform of $f(x)$ with respect to $\psi^b(x)$ is

$$W_s^b f(x) = f * \psi_s^b(x) \ . \tag{3}$$

We derive that

$$W_s^a f(x) = f * (s\,\frac{d\theta_s}{dx})(x) = s\,\frac{d}{dx}(f * \theta_s)(x) \quad \text{and} \tag{4}$$

$$W_s^b f(x) = f * (s^2\,\frac{d^2\theta_s}{dx^2})(x) = s^2\,\frac{d^2}{dx^2}(f * \theta_s)(x) \ . \tag{5}$$

The wavelet transforms $W_s^a f(x)$ and $W_s^b f(x)$ are respectively the first and second derivative of the signal smoothed at the scale s. The local extrema of $W_s^a f(x)$ thus correspond to the zero-crossings of $W_s^b f(x)$ and to the inflection points of $f * \theta_s(x)$, as illustrated in Fig. 1. In the particular case where $\theta(x)$ is a Gaussian, the zero-crossing detection is equivalent to a Marr-Hildreth [25] edge detection where as the extrema detection corresponds to a Canny [3] edge detection. When the scale s is small, the smoothing of $f(x)$ by $\theta_s(x)$ is negligible and hence these edge detections provide the locations of most sharper variations of $f(x)$. When s is large, the convolution with $\theta_s(x)$ removes small signal fluctuations so we only detect the sharp variations of larger structures.

Detecting zero-crossings or local extrema are similar procedures but the local extrema approach has some important advantages. An inflection point of $f * \theta_s(x)$ can either be a maximum or a minimum of the absolute value of its first derivative. The maxima of the absolute value of the first derivative are sharp variation points of $f * \theta_s(x)$, whereas the minima correspond to slow variations (see Fig. 1). With a second derivative operator it is difficult to distinguish these two types of zero-crossings. On the contrary, with a first order derivative, we easily select the sharp variation points by detecting only the local maxima of $|W_s^a f(x)|$. Also, zero-crossings give a position information but do not differentiate small amplitude fluctuations from important discontinuities. When detecting local maxima, we can also record the values of $W_s^a f(x)$ at the maxima locations, which measure the derivative at the inflection points. Section 4 explains how to characterize different types of sharp variation points from the evolution across scales of $W_s^a f(x)$ at the modulus maxima locations, defined as follow.

Definition 1

We call a *modulus maximum* of $g(x)$ any abscissa x_0 where $|g(x)|$ is locally maximum and is strictly maximum either over the left or the right neighborhood of x_0.

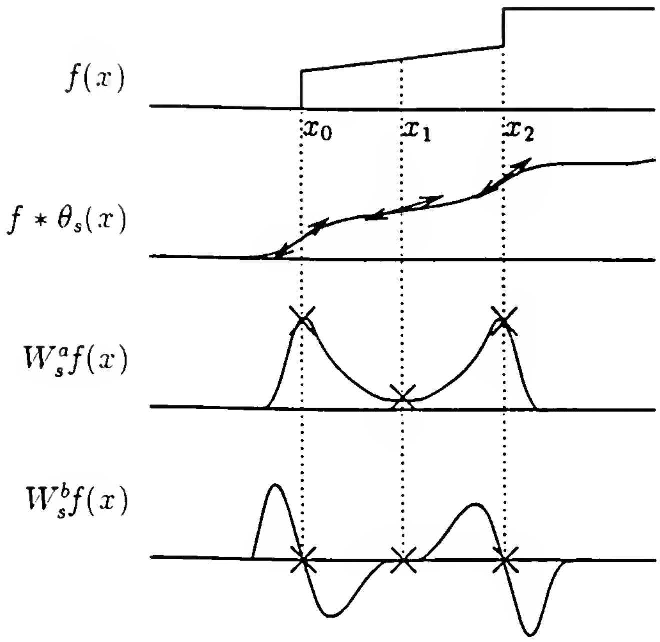


Fig. 1: *The extrema of* $W_s^a f(x)$ *and the zero-crossings of* $W_s^b f(x)$ *appear at the inflection points of* $f * \theta_s(x)$. *We only record the points of abscissa* x_0 *and* x_2 *where* $|W_s^a f(x)|$ *is locally maximum, because they locate the sharper variation points of* $f(x)$ *smoothed at the scale* s. *The local minima of* $|W_s^a f(x)|$ *at* x_1 *is also a zero-crossing of* $W_s^b f(x)$ *but corresponds to a slow variation point.*

The Canny edge detector is easily extended in two-dimensions. We denote by

$$\xi_s(x,y) = \frac{1}{s^2}\,\xi(\frac{x}{s},\frac{y}{s}) ,$$

the dilation by s of any two-dimensional function $\xi(x,y)$. We call two-dimensional smoothing function any function $\theta(x,y)$ whose integral over x and y is equal to 1 and which converges to 0 at infinity. The image $f(x,y)$ is smoothed at different scales s by a convolution with $\theta_s(x,y)$. We then compute the gradient vector $\vec{\nabla}(f * \theta_s)(x,y)$. The direction of the gradient vector at a point (x_0,y_0) indicates the direction in the image plane (x,y) along which the directional derivative of $f(x,y)$ has the largest absolute value. Edges are defined as points (x_0,y_0) where the modulus of the gradient vector is maximum in the direction where the gradient vector point to in the image plane. Edge points are inflection points of the surface $f * \theta_s(x,y)$. Let us relate this edge detection to a two-dimensional wavelet transform. We define two wavelet functions $\psi^1(x,y)$ and $\psi^2(x,y)$ such that

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \quad \text{and} \quad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} . \tag{6}$$

Let $\psi_s^1(x,y) = \frac{1}{s^2}\,\psi^1(\frac{x}{s},\frac{y}{s})$ and $\psi_s^2(x,y) = \frac{1}{s^2}\,\psi^2(\frac{x}{s},\frac{y}{s})$. Let $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$. The wavelet transform of $f(x,y)$ at the scale s has two components defined by

$$W_s^1 f(x,y) = f * \psi_s^1(x,y) \quad \text{and} \quad W_s^2 f(x,y) = f * \psi_s^2(x,y) \ . \tag{7}$$

Similarly to equation (4), one can easily prove that

$$\begin{pmatrix} W_s^1 f(x,y) \\ W_s^2 f(x,y) \end{pmatrix} = s \begin{pmatrix} \frac{\partial}{\partial x}(f * \theta_s)(x,y) \\ \frac{\partial}{\partial y}(f * \theta_s)(x,y) \end{pmatrix} = s\,\vec{\nabla}(f * \theta_s)(x,y) \ . \tag{8}$$

Hence, edge points can be located from the two components $W_s^1 f(x,y)$ and $W_s^2 f(x,y)$ of the wavelet transform.

In this article, we explain how to characterize different types of edges from the evolution of the wavelet transform maxima across scales. We also study how much information is embedded in multiscale edges and how to reconstruct a close approximation of the original signal. Up to Section 6, we concentrate on one-dimensional signals. The next section reviews the main properties of a wavelet transform that are needed to understand the behavior of multiscale edges. For most purposes, the wavelet model does not require to keep a continuum of scales s. To allow fast numerical implementations, we impose that the scale parameter varies only along the dyadic sequence $\left(2^j\right)_{j\in Z}$.

3. Dyadic Wavelet Transform in One Dimension

3.1. General Properties

We review the main properties of a dyadic wavelet transform and explain under what condition it is complete and stable. For thorough presentations of the wavelet transform, the reader is referred to the mathematical books of Meyer [26] and Daubechies [8] or to signal processing oriented reviews [20,28]. The wavelet model has first been formalized by Grossmann and Morlet [13]. A wavelet is a function $\psi(x)$ whose average is zero. We denote by $\psi_{2^j}(x)$ the dilation of $\psi(x)$ by a factor 2^j

$$\psi_{2^j}(x) = \frac{1}{2^j}\,\psi(\frac{x}{2^j}) \ .$$

The wavelet transform of $f(x)$ at the scale 2^j and at the position x is defined by the convolution product

$$W_{2^j} f(x) = f * \psi_{2^j}(x) \ . \tag{9}$$

We call the *dyadic wavelet transform* the sequence of functions

$$\mathbf{W}f = \left[W_{2^j} f(x) \right]_{j \in \mathbf{Z}} ,$$

and $\mathbf{W}$ is the dyadic wavelet transform operator.

Let us study the completeness and stability of a dyadic wavelet transform. The Fourier transform of $W_{2^j} f(x)$ is

$$\hat{W}_{2^j} f(\omega) = \hat{f}(\omega)\, \hat{\psi}(2^j \omega) \ . \tag{10}$$

By imposing that there exists two strictly positive constants A_1 and B_1 such that

$$\forall \omega \in \mathbf{R} \quad A_1 \leq \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2 \leq B_1 \ , \tag{11}$$

we insure that the whole frequency axis is covered by dilations of $\hat{\psi}(\omega)$ by $\left[2^j \right]_{j \in \mathbf{Z}}$, so that no information about $\hat{f}(\omega)$ is lost. We call reconstructing wavelet $\chi(x)$ any function whose Fourier transform satisfies

$$\sum_{j=-\infty}^{+\infty} \hat{\psi}(2^j \omega)\, \hat{\chi}(2^j \omega) = 1 \ . \tag{12}$$

The function $f(x)$ is recovered from its dyadic wavelet transform with the summation

$$f(x) = \sum_{j=-\infty}^{+\infty} W_{2^j} f * \chi_{2^j}(x) \ . \tag{13}$$

This equation is proved by computing its Fourier transform and inserting equations (10) and (12). There exists an infinite number of functions $\hat{\chi}(\omega)$ that satisfy equation (12). One example is

$$\hat{\chi}(\omega) = \frac{\bar{\hat{\psi}}(\omega)}{\sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2} \ , \tag{14}$$

where $\bar{\hat{\psi}}(\omega)$ denotes the complex conjugate of $\hat{\psi}(\omega)$. With the Parseval theorem, we derive from equations (10) and (11) an energy equivalence equation

$$A_1 \ \|f\|^2 \leq \sum_{j=-\infty}^{+\infty} \|W_{2^j} f(x)\|^2 \leq B_1 \ \|f\|^2 \ . \tag{15}$$

This proves that the dyadic wavelet transform is not only complete but also stable. The closer $\frac{B_1}{A_1}$ to 1 the more stable it is.

A dyadic wavelet transform is more than complete, it is redundant. Any sequence $\left[g_j(x) \right]_{j \in \mathbf{Z}}$, with $g_j(x) \in \mathbf{L}^2(\mathbf{R})$, is not necessarily the dyadic wavelet transform of some function in $\mathbf{L}^2(\mathbf{R})$. We denote by $\mathbf{W}^{-1}$ the operator defined by

$$\mathbf{W}^{-1}\left(g_j(x)\right)_{j\in\mathbf{Z}} = \sum_{j=-\infty}^{+\infty} g_j * \chi_{2^j}(x) \ . \tag{16}$$

The reconstruction formula (13) shows that $\left(g_j(x)\right)_{j\in\mathbf{Z}}$ is the dyadic wavelet transform of some function in $\mathbf{L}^2(\mathbf{R})$ if and only if

$$\mathbf{W}\left(\mathbf{W}^{-1}\left(g_l(x)\right)_{l\in\mathbf{Z}}\right) = \left(g_j(x)\right)_{j\in\mathbf{Z}} \ . \tag{17}$$

If we replace the operators $\mathbf{W}$ and $\mathbf{W}^{-1}$ by their expression given in equations (9) and (16), we obtain

$$\forall j \in \mathbf{Z} \quad \sum_{l=-\infty}^{+\infty} g_l * K_{l,j}(x) = g_j(x) \ , \ \text{with} \tag{18}$$

$$K_{l,j}(x) = \chi_{2^l} * \psi_{2^j}(x) \ . \tag{19}$$

These equations are called reproducing kernel equations. The energy of the kernel $K_{l,j}(x)$ measures the redundancy of the wavelet transform at the scales 2^j and 2^l.

For numerical applications, it is necessary to discretize the parameter x while keeping and complete and stable representation. One approach is to build a frame of $\mathbf{L}^2(\mathbf{R})$ by sampling uniformly the parameter x at different scales [7]. The value of $W_{2^j}f(x)$ at x_0 can be written as an inner product in $\mathbf{L}^2(\mathbf{R})$. Indeed

$$W_{2^j}f(x_0) = f * \psi_{2^j}(x_0) = \int_{-\infty}^{+\infty} f(x)\,\psi_{2^j}(x_0-x)\,dx \ , \ \text{thus}$$

$$W_{2^j}f(x_0) = \langle f(x) \,,\, \psi_{2^j}(x_0-x) \rangle \ . \tag{20}$$

By imposing a weak condition on $\psi(x)$, Daubechies [7] proved that if the sampling interval a is small enough, the family of functions $\left(\psi_{2^j}(a2^jn-x)\right)_{(n,j)\in\mathbf{Z}^2}$ is a frame of $\mathbf{L}^2(\mathbf{R})$. This means that for any $f(x) \in \mathbf{L}^2(\mathbf{R})$, the inner products $\left(W_{2^j}f(a2^jn) = \langle f(x) \,,\, \psi_{2^j}(a2^jn-x) \rangle\right)_{(n,j)\in\mathbf{Z}^2}$ provide a complete and stable characterization of $f(x)$. The function $f(x)$ is thus characterized by sampling uniformly $W_{2^j}f(x)$ at intervals $a2^j$, at each scale 2^j. Wavelet orthonormal bases [20, 26] are particular example of frames with $a = 1$. The Laplacian pyramid of Burt [2] as well as the DOLP transform of Crowley [5] are other examples of multiresolution frames. A major inconvenience of frames is that the signal descriptors $\left(W_{2^j}f(a2^jn)\right)_{(n,j)\in\mathbf{Z}^2}$ are considerably modified when the signal is translated. This is a problem in pattern recognition where one does not know a priori the position of the patterns we want to analyze. Let us explain this behavior with translation. Let $f_\tau(x) = f(x-\tau)$ be a translation of $f(x)$ by τ. Since a wavelet transform at a scale 2^j is given by a convolution product (equation (9)), it is clear that $W_{2^j} f_\tau(x) = W_{2^j} f(x-\tau)$. However, the samples $W_{2^j} f_\tau(a 2^j n)$ and $W_{2^j} f(a 2^j n)$ may be totally different, unless $\tau = ka2^j$, with $k \in \mathbf{Z}$ (see Fig. 2). Hence, when a signal is translated its wavelet frame coefficients are not translated but modified. Distortions through translation are maximum for wavelet orthogonal bases [20].

We saw in Section 2 that if the wavelet $\psi(x)$ is the first derivative of a smoothing function, then a Canny edge detection is equivalent to an adaptive sampling of $W_{2^j} f(x)$ at the locations where $|W_{2^j} f(x)|$ is locally maximum. Since $W_{2^j} f(x)$ is defined by the convolution of two functions in $\mathbf{L}^2(\mathbf{R})$, it is a continuous function and hence this adaptive sampling is well defined. When $f(x)$ is translated by τ, $W_{2^j} f(x)$ is also translated by τ so its modulus maxima are translated as well. The amplitude of the wavelet transform modulus maxima are therefore not modified by translations.

Fig. 3(a) is a quadratic spline wavelet of compact support which is further defined in Appendix 1. It is the derivative of the smoothing function $\theta(x)$ shown in Fig. 3(b). Fig. 4(a) is the plot of a discrete signal of 256 samples. Fig. 4(b) shows its discrete dyadic wavelet transform computed on 9 scales. At each scale 2^j, we compute a uniform sampling of the dyadic wavelet transform that we denote $W^d_{2^j} f$. The next section explains how to discretize the continuous wavelet model and solve border problems. Fast algorithms to compute the wavelet and the inverse wavelet transform are described in Appendix 2. The reader not interested by numerical issues might want to skip Section 3.2. Fig. 4(c) gives the locations and values of the modulus maxima of the dyadic wavelet transform. At each scale 2^j, each modulus maximum is represented by a Dirac which has the same location and whose amplitude is equal to the value of $W_{2^j} f(x)$. The modulus maxima detection is an adaptive sampling that depends upon the local signal regularity. The first part of the signal in Fig. 4(a) has few sharp variations and thus creates few wavelet modulus maxima in Fig. 4(c). The second part is an irregular texture which produces many modulus maxima.

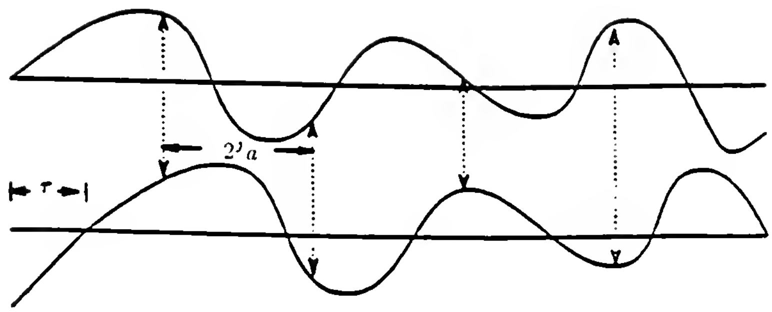
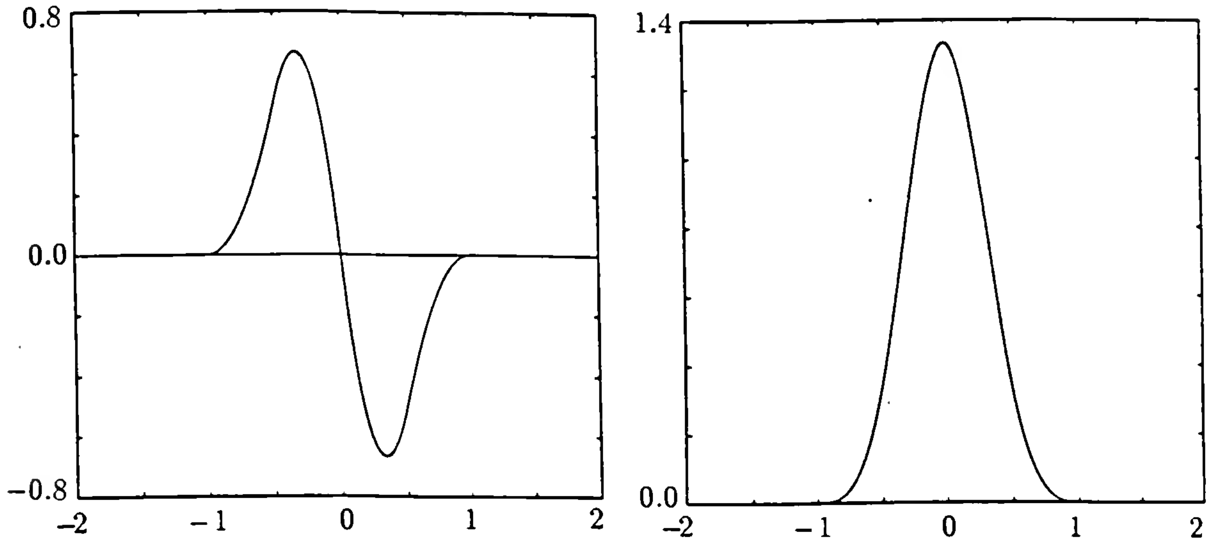


Fig. 2: *This drawing shows that the sample values of a wavelet transform (given by the arrows) are modified by translating the signal. If* $f_\tau(x) = f(x-\tau)$, *then* $W_{2^j} f_\tau(x) = W_{2^j} f(x-\tau)$, *but the samples do not translate if* τ *is not proportional to the sampling interval* $a2^j$.



(a) *(b)*

Fig. 3: *(a): This wavelet is a quadratic spline of compact support which is is continuously differentiable. It is defined in Appendix 1 and it is the derivative of the cubic spline function* $\theta(x)$ *shown in (b).*

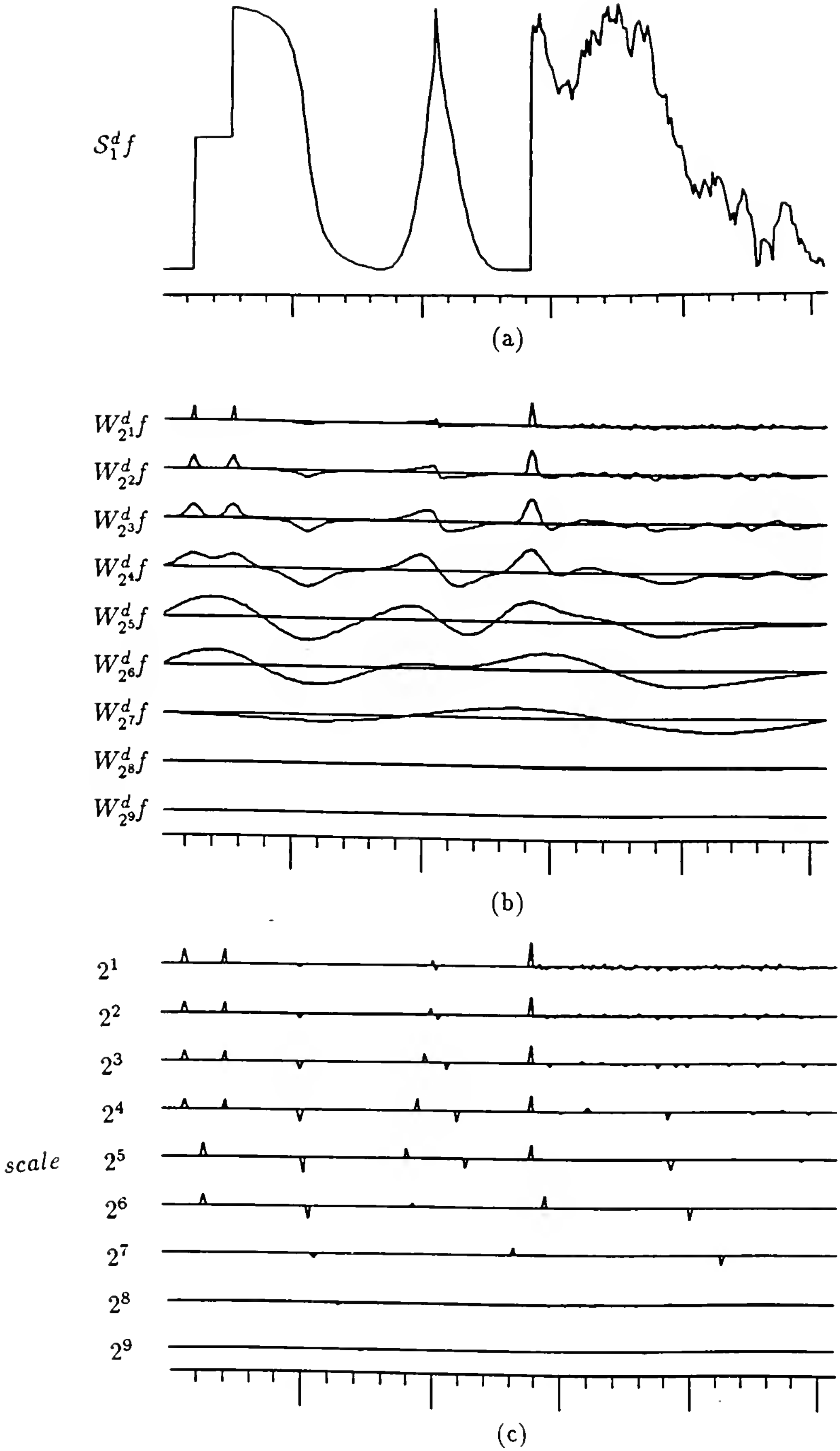


Fig. 4: *(a): Signal of 256 samples.*

(b): Discrete dyadic wavelet transform of signal (a) computed on 9 scales. At each scale 2^j, we plot the signal $W^d_{2^j} f$ which also has 256 samples.

(c): Modulus maxima of the dyadic wavelet transform shown in (b). Each Dirac indicates the position and amplitude of a modulus maximum.

3.2. Discrete Wavelet Transform

A proper implementation of a discrete dyadic wavelet transform raises several important questions. The input signal is generally measured with a finite resolution which imposes a finer scale when computing the wavelet transform. We only know the signal on a finite domain which limits the coarser scale. We introduce a wavelet transform over a finite range of scales and give efficient discrete algorithms to compute the wavelet transform and its inverse.

Since the input signal is measured at a finite resolution, we cannot compute the wavelet transform at an arbitrary fine scale. Let us normalize the finest scale to 1. In order to model this scale limitation, we introduce a real function $\phi(x)$ whose Fourier transform is an aggregation of $\hat{\psi}(2^j\omega)$ and $\hat{\chi}(2^j\omega)$ at scales 2^j larger than 1

$$|\hat{\phi}(\omega)|^2 = \sum_{j=1}^{+\infty} \hat{\psi}(2^j\omega)\,\hat{\chi}(2^j\omega) \quad . \tag{21}$$

We suppose here that the reconstructing wavelet $\chi(\omega)$ has been chosen so that

$$\sum_{j=1}^{+\infty} \hat{\psi}(2^j\omega)\,\hat{\chi}(2^j\omega)$$

is a positive real function. As an example, the reconstruction wavelet $\hat{\chi}(\omega)$ defined by equation (14) satisfies this property. As a consequence of equation (12), one can derive that

$$\lim_{\omega \to 0} |\hat{\phi}(\omega)| = 1 \quad .$$

If $\phi(x)$ is real, this implies that the integral of $\phi(x)$ is equal to 1 and hence that it is a smoothing function. Let S_{2^j} be the smoothing operator defined by

$$S_{2^j} f(x) = f * \phi_{2^j}(x) \quad \text{with} \quad \phi_{2^j}(x) = \frac{1}{2^j}\phi(\frac{x}{2^j}) \quad . \tag{22}$$

The larger the scale 2^j, the more details of $f(x)$ are removed by the smoothing operator S_{2^j}. In between the scales 1 and 2^J, equation (21) yields

$$|\hat{\phi}(\omega)|^2 - |\hat{\phi}(2^J\omega)| = \sum_{j=1}^{J} \hat{\psi}(2^j\omega)\,\hat{\chi}(2^j\omega) \quad . \tag{23}$$

One can derive from this equation that the higher frequencies of $S_1 f(x)$ which have disappeared in $S_{2^J} f(x)$ can be recovered from the dyadic wavelet transform $\left(W_{2^j} f(x)\right)_{1 \le j \le J}$, between the scales 2^1 and 2^J. We call *finite-scale wavelet transform* of $S_1 f(x)$, the sequence of functions

$$\left\{ S_{2^J} f(x) \, , \, \left(W_{2^j} f(x)\right)_{1 \le j \le J} \right\} \quad .$$

The discrete data that we process are often obtained by filtering and sampling uniformly an analog signal. We suppose that there exists two constants $C_1 > 0$ and $C_2 > 0$ such that $\hat{\phi}(\omega)$ satisfies

$$\forall \omega \in \mathbf{R} \quad , \quad C_1 \le \sum_{n=-\infty}^{+\infty} |\hat{\phi}(\omega+2n\pi)|^2 \le C_2 \ . \tag{24}$$

It has been proved [22] that for any discrete signal $D = \left(d_n\right)_{n \in \mathbf{Z}}$ of finite energy, there exists a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ (not unique) such that

$$\forall n \in \mathbf{Z} \ , \ S_1 f(n) = d_n \ . \tag{25}$$

The discrete signal D can thus be rewritten $D = \left(S_1 f(n)\right)_{n \in \mathbf{Z}}$. Any finite energy discrete signal is therefore interpreted as the uniform sampling a function in $\mathbf{L}^2(\mathbf{R})$ that is smoothed at the scale 1. For a particular class of wavelets $\psi(x)$ described in Appendix 1, the samples $\left(S_1 f(n)\right)_{n \in \mathbf{Z}}$ are sufficient to compute a uniform sampling of the finite scale wavelet transform of $S_1 f(x)$:

$$\left\{ \left(S_{2^J} f(n+w)\right)_{n \in \mathbf{Z}} , \left(\left(W_{2^j} f(n+w)\right)_{n \in \mathbf{Z}} \right)_{1 \le j \le J} \right\} .$$

The sampling shift w depends upon the wavelet $\psi(x)$. Let us denote

$$W^d_{2^j} f = \left(W_{2^j} f(n+w)\right)_{n \in \mathbf{Z}} \quad \text{and} \quad S^d_{2^j} f = \left(S_{2^j} f(n+w)\right)_{n \in \mathbf{Z}} \ . \tag{26}$$

The sequence of discrete signals $\left\{ S^d_{2^J} f , \left(W^d_{2^j} f\right)_{1 \le j \le J} \right\}$ is called a *discrete dyadic wavelet transform* of $D = \left(S_1 f(n)\right)_{n \in \mathbf{Z}}$.

In practice, the original discrete signal D has a finite number N of non-zero values: $D = \left(d_n\right)_{1 \le n \le N}$. To solve border problems, we use the same periodization technique as in a cosine transform. We define the signal $\tilde{D} = \left(\tilde{d}_n\right)_{n \in \mathbf{Z}}$ which has a period of $2N$ samples and such that

$$\tilde{d}_n = \begin{cases} d_n & \text{if } \ 1 \le n \le N \\ d_{2N+1-n} & \text{if } \ N < n \le 2N \end{cases} . \tag{27}$$

By periodizing the signal with a symmetry, we avoid to create a discontinuity at the borders. Since $\tilde{D}$ has a period of $2N$ samples, the corresponding discrete wavelet signals $W^d_{2^j} f$ have also a period of $2N$ samples. If the wavelet is antisymmetrical with respect to 0, like in fig. 3(a), then the values $\left[W_{2^j} f(n+w) \right]_{n \in Z}$ are antisymmetrical with respect to the abscissa 1/2 and $N+1/2$. Each discrete signal $W^d_{2^j} f$ is thus characterized by the N samples whose abscissa are between 1/2 and $N+1/2$. Similarly, the discrete signals $S^d_{2^j} f$ also have a period of $2N$ samples. For the wavelets defined in Appendix 1, if $2^j \geq 2N$, the signal $S^d_{2^j} f$ is constant and equal to the mean value of the original signal D whereas $W^d_{2^j} f$ is equal to 0. The signal D is thus completely characterized by the discrete wavelet transform signals $W^d_{2^j} f$ over $\log_2(N)+1$ scales plus its mean value. Appendix 2 describes a fast discrete wavelet transform algorithm that requires $O(N \log(N))$ operations. The fast inverse wavelet transform also requires $O(N \log(N))$ operations.

From the discrete wavelet transform, at each scale 2^j, we detect the modulus maxima by finding the points where $|W_{2^j} f(n+w)|$ is larger than its two closest neighbor values and strictly larger than at least one of them. We record the abscissa $n+w$ and the value $W_{2^j} f(n+w)$ at the corresponding locations.

4. Analysis of the Multiscale Information

One signal sharp variation produces modulus maxima at different scales 2^j. We know that the value of a modulus maximum at a scale 2^j measures the derivative of the signal smoothed at the scale 2^j, but it is not clear how to combine these different values to characterize the signal variation. The wavelet theory gives an answer to this question by showing that the evolution across scales of the wavelet transform depends upon the local Lipschitz regularity of the signal. This section explains what is a Lipschitz exponent and how this exponent is computed from the behavior across scales of the wavelet transform maxima. A more detailed mathematical and numerical analysis of this issue can be found in [23]. When the signal is not singular, we show that one can still measure how smooth the signal is by estimating the decay of the wavelet maxima across scales.

Definition 2

Let $0 \leq \alpha \leq 1$. A function $f(x)$ is Lipschitz α at x_0 if and only if there exists a constant K such that for all x in a neighborhood of x_0, we have

$$|f(x) - f(x_0)| \leq K\, |x - x_0|^{\alpha} \; . \tag{28}$$

The function $f(x)$ is uniformly Lipschitz α over an interval $]a,b[$ if and only if there exists a constant K such that equation (28) holds for any $(x, x_0) \in \;]a,b[^2$.

If $f(x)$ is differentiable at x_0, then it is Lipschitz $\alpha = 1$. We call Lipschitz regularity of $f(x)$ at x_0, the upper bound α_0 of all α such that $f(x)$ is Lipschitz α at x_0. The larger the Lipschitz regularity α_0, the more "regular" the singularity at x_0. If $f(x)$ is discontinuous but bounded at x_0, its Lipschitz regularity at x_0 is 0. We also define the uniform Lipschitz regularity of $f(x)$ over an interval $]a,b[$ as the upper bound of all α such that $f(x)$ is uniformly Lipschitz α over $]a,b[$. Theorem 1 proves that the Lipschitz exponent of a function can be measured from the evolution across scales of the absolute value of the wavelet transform. We suppose that the wavelet $\psi(x)$ is continuously differentiable and has a decay at infinity which is $O(\frac{1}{1+x^2})$.

Theorem 1

Let $0 < \alpha < 1$. A function $f(x)$ is uniformly Lipschitz α over $]a,b[$ if and only if there exists a constant $K > 0$ such that for all $x \in]a,b[$, the wavelet transform satisfies

$$|W_{2^j} f(x)| \leq K (2^j)^{\alpha} \quad . \tag{29}$$

The proof of this theorem can be found in [26]. From equation (29) we derive that

$$\log_2 |W_{2^j} f(x)| \leq \log_2(K) + \alpha j \quad . \tag{30}$$

Theorem 1 characterizes uniform Lipschitz exponents over intervals but not pointwise Lipschitz exponents. Jaffard [16] as well as Holschneider and Tchamitchian [14] have shown that the Lipschitz regularity of a function at a point x_0 can also be characterized by the decay of the wavelet transform across scales, but this theorem is more technical and is not useful for the scope of this article. A tutorial review of the characterization of Lipschitz exponents from the wavelet transform is in [23]. To study isolated singularities, Theorem 1 is sufficient. We shall say that a function has an isolated singularity at x_0 if there exists a neighborhood $]a,b[$ of x_0, where the worst singularity is at x_0. In other words, the uniform Lipschitz regularity of the signal over $]a,b[$ is equal to the pointwise Lipschitz regularity at x_0.

If the Lipschitz regularity is positive, equation (29) implies that the amplitude of the wavelet transform modulus maxima should decrease when the scale decreases. This is not the case for the modulus maxima corresponding to the singularity at the abscissa 3 of Fig. 5(b), where the wavelet transform maxima increases when the scale decreases. Such singularities can be described with negative Lipschitz exponents which means that they are more singular than discontinuities. The signal is viewed as a tempered distribution and at the abscissa 3 of Fig. 5(b) this distribution is locally equal to a Dirac. The reader might want to consult the book of Folland [10] for a quick presentation of the mathematical theory of distributions. The wavelet transform of tempered distributions of order k is well defined if the wavelet is k times continuously differentiable. For example, a Dirac $\delta(x)$ is a tempered distribution of order 0 and

$$W_{2^j} \delta(x) = \delta * \psi_{2^j}(x) = \psi_{2^j}(x),$$

if $\psi(x)$ is continuous. Definition 3 enables us to extend Lipschitz exponents to negative values for tempered distributions such as Diracs.

Definition 3

Let $f(x)$ be a tempered distribution of finite order and $\alpha < 0$. The distribution $f(x)$ is said to have a uniform Lipschitz regularity equal to α on $]a,b[$, if and only if, its primitive has a uniform Lipschitz regularity equal to $\alpha+1$ on $]a,b[$.

For example, the primitive of a Dirac centered at x_0 is a function which is bounded and has a discontinuity at x_0 (step edge). The uniform Lipschitz regularity of the primitive of this Dirac is thus equal to 0 in the neighborhood of x_0. Definition 3 implies that a Dirac centered at x_0 has a uniformly Lipschitz regularity equal to -1 in the neighborhood of x_0. One can prove that Theorem 1 is also valid for negative Lipschitz exponents. Let $\alpha_0 < 1$ be a real number that may be negative. A tempered distribution $f(x)$ has a uniform Lipschitz regularity equal to α_0 over $]a,b[$ if and only if α_0 is the upper bound of the α such that there exists K that satisfy

$$|W_{2^j} f(x)| \leq K\,(2^j)^{\alpha} \ . \tag{31}$$

Since the Lipschitz regularity of a Dirac is -1, this result implies that the maxima values of $|W_{2^j}\delta(x)|$ increase proportionally to the scale 2^j. This can indeed be verified in Fig. 5(b).

In practice, we can only process discrete signals that approximate the original function at a finite resolution, that we normalize to 1. Strictly speaking, it is not meaningful to speak about singularities, discontinuities or Diracs. In fact we can not compute the wavelet transform at scales finer than 1 and thus can not verify equation (31) at scales smaller than 1. Even though we are limited by the resolution of measurements, we can still use the mathematical tools that differentiate singularities. Suppose that the approximation of $f(x)$ at the resolution 1 is given by a set of samples $\left(f_n\right)_{n \in Z}$, with $f_n = 0$ for $n < n_0$ and $f_n = 1$ for $n \geq n_0$, like at the abscissa 2 of Fig. 5(a). At the resolution 1, $f(x)$ behaves as if it has a discontinuity at $n = n_0$, although $f(x)$ might be continuous at n_0 with a continuous sharp transition at that point which is not visible at this resolution. The characterization of singularities from the decay of the wavelet transform gives a precise meaning to this "discontinuity at the resolution 1". We measure the decay of the wavelet transform up to the finer scale available and the Lipschitz regularity is computed by finding the coefficient α_0 such that $K\,(2^j)^{\alpha_0}$ approximates at best the decay of $|W_{2^j} f(x)|$ over a given range of scales larger than 1. In Fig. 5(b), in the neighborhood of the $x = 2$, the maxima values of $|W_{2^j} f(x)|$ remain constant over a large range of scales. Equation (31) implies that the Lipschitz regularity α_0 is equal to 0 at that point which means that this singularity is a discontinuity. In the edge detection procedure described in Section 2, we only keep the wavelet transform modulus maxima. One can prove that any singularity creates modulus maxima in the wavelet transform [23]. If there is no modulus maxima at fine scales over a given interval, then the function is uniformly Lipschitz 1 in this interval [23]. If there are some modulus maxima, the decay of the wavelet transform is bounded by the decay of these modulus maxima and we thus measure the uniform Lipschitz regularity from this decay.

A signal is often not singular in the neighborhood of local sharp variations. An example is the smooth edge at the abscissa 1 of fig. 5(a). It is generally important to estimate how smooth is the signal variation in such cases. We model a smooth variation at x_0, as a singularity convolved with a Gaussian of variance σ^2. Since the Gaussian is the Green's function of the heat equation, one can prove that σ is proportional to the time it would take to create a singularity at the point x_0 if we apply a backwards heat equation to the signal. Let us explain how to measure the smoothing component σ as well as the Lipschitz regularity of the underlined singularity. We suppose that locally the signal $f(x)$ is equal to the convolution of a function $h(x)$, which has a singularity at x_0, with a Gaussian of variance σ^2

$$f(x) = h * g_\sigma(x) \quad \text{with} \quad g_\sigma(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp(-\frac{x^2}{2\sigma^2}) \ . \tag{32}$$

We also suppose that $h(x)$ has a uniform Lipschitz regularity equal to α_0 in a neighborhood of x_0. If the wavelet $\psi(x)$ is the derivative of a smoothing function $\theta(x)$, equation (4) proves that the wavelet transform of $f(x)$ can be written

$$W_{2^j} f(x) = 2^j \frac{d}{dx} (f * \theta_{2^j})(x) = 2^j \frac{d}{dx} (h * g_\sigma * \theta_{2^j})(x) \ . \tag{33}$$

Let us suppose that the function $\theta(x)$ is close to a Gaussian function in the sense that

$$\theta_{2^j} * g_\sigma(x) \simeq \theta_{s_0}(x) \quad \text{with} \quad s_0 = \sqrt{2^{2j} + \sigma^2} \ . \tag{34}$$

Equation (33) can thus be rewritten

$$W_{2^j} f(x) = 2^j \frac{d}{dx} (h * \theta_{s_0})(x) = \frac{2^j}{s_0} W_{s_0} h(x) \ , \tag{35}$$

where $W_{s_0} h(x)$ is the wavelet transform of $h(x)$ at the scale s_0

$$W_{s_0} h(x) = h * \psi_{s_0}(x) \ .$$

This equation proves that the wavelet transform at the scale 2^j of a singularity smoothed by a Gaussian of variance σ^2, is equal to the wavelet transform of the non-smoothed singularity $h(x)$ at the scale $s_0 = \sqrt{2^{2j} + \sigma^2}$. Equation (29) of Theorem 1 proves that the Lipschitz regularity is the upper bound of the set of α that satisfy

$$|W_{2^j} h(x)| \le K \, (2^j)^\alpha \ .$$

This result is valid for any scale $s > 0$. Hence, α_0 is the upper bound of the set of α such that there exists K that satisfy

$$|W_s h(x)| \leq K\, s^{\alpha} \ ,$$

for any scale $s > 0$ and any x in the corresponding neighborhood of x_0. By inserting this inequality in equation (35), we obtain

$$|W_{2^j} f(x)| \leq K\, 2^j\, s_0^{\alpha - 1} \ , \ \text{with} \ \ s_0 = \sqrt{2^{2j} + \sigma^2} \ . \qquad (37)$$

This equation is satisfied at all points x, if and only if, it is satisfied at the locations of all the local maxima of $|W_{2^j} f(x)|$. Fig. 5 gives the examples of a step edge and a Dirac smoothed by Gaussians of different variances. The decay of the maxima are clearly affected by the different Lipschitz exponents as well as the variance of the Gaussian smoothing.

Let us explain how to compute numerically the Lipschitz regularity α_0 and the smoothing scale σ, from the evolution of the wavelet transform modulus maxima across scales. If we detect the modulus maxima at all scales s, instead of just dyadic scales 2^j, their position would define smooth curve in the scale-space plane (s,x). These curves have been called "finger prints" by Witkin [31]. We say that a modulus maxima at the scale 2^j *propagates* to a maxima at the coarser scale 2^{j+1}, if and only if, both maxima belong to the same maxima curve in the scale-space plane (s,x). In Fig. 5, there is one sequence of maxima which belongs to the same maxima curve and converges to the position of the discontinuity at $x = 2$. For the Dirac at the abscissa 3, there are two such sequences. Each one gives an information respectively on the left and the right part of the Dirac singularity. In order to find which maxima propagate to the next scale, one should compute the wavelet transform on a dense sequence of scales. However, with a simple ad-hoc algorithm, one can still try to estimate which maxima propagate to the next scale, by looking at their value and position with respect to other maxima at the next scale. The propagation algorithm supposes that a modulus maximum propagates from a scale 2^j to a coarser scale 2^{j+1}, if it has a large amplitude and if its position is close to a maximum at the scale 2^{j+1} that has the same sign. This ad-hoc algorithm is not exact but saves computations since we do not need to compute the wavelet transform at any other scale. The Lipschitz regularity as well as the smoothing variance σ^2 of a sharp variation point are then computed from the evolution of the modulus maxima that propagate across scales. Let us suppose that we have a sequence of modulus maxima that propagate from the scale 2^l up to the scale 2^1 and converge to the abscissa x_0. Let a_j be the value of the wavelet transform at the maximum location at the scale 2^j and let us also suppose that in a given neighborhood of x_0 the wavelet transform modulus is smaller than a_j. This means that the signal change at x_0 is the sharpest variation in this neighborhood. We compute the three values K, σ and α_0 so that the inequality of equation (37) is as close as possible to an equality for each maximum a_j. These values are obtained by minimizing

$$\sum_{j=1}^{I}\left[\log_2|a_j| - \log_2(K) - j - \frac{\alpha_0-1}{2}\log_2(\sigma^2+2^{2j})\right]^2 \quad . \tag{38}$$

This is done with a steepest gradient descent algorithm. The value K gives the amplitude of the sharp variation. If the signal is multiplied by a constant λ then K is also multiplied by λ, but σ and α_0 are not affected. On the contrary, if the signal is smoothed by a Gaussian of variance σ_0^2 (and integral 1), then K and α_0 are not affected but σ^2 becomes $\sigma^2 + \sigma_0^2$. This shows clearly that the parameters α, σ and K describe different properties of the sharp variation that occurs at x_0. When computing the values of σ and α from the evolution of the maxima across scales in Fig. 8, we have a numerical error of less than 10%, which is mainly due to the fact that the wavelet we use is not the derivative of a Gaussian but only an approximation. In this case, $\theta(x)$ is the cubic spline shown in Fig. 3(b). When the variance σ^2 increases, the measurement of α_0 becomes more unstable because the smoothing removes the fine scale components that characterize reliably α_0. For singularities of fractal textures such as in the right part of Fig. 4(a), this analysis is not valid because singularities are not isolated and none of the singularities dominate the others in a given neighborhood. The behavior of the wavelet transform modulus maxima of non-isolated singularities is studied in more detail in [23].

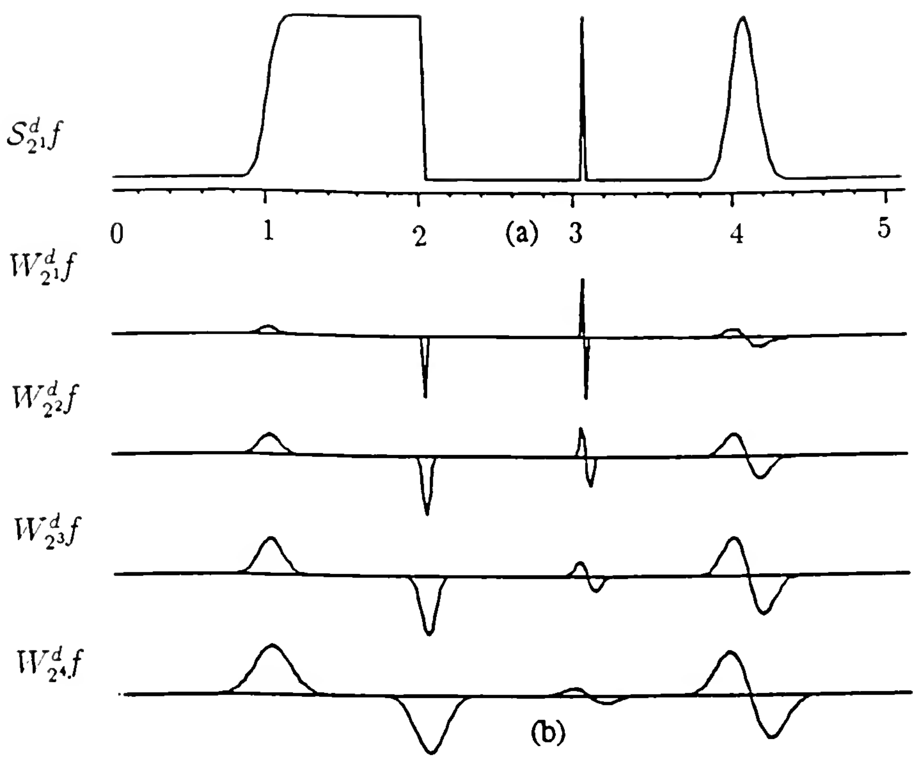


Fig. 5: *(a): The four sharp variation points of this signal have different Lipschitz regularity* α_0 *and smoothing variance* σ^2. *These values are respectively given by* $(\alpha_0 = 0, \sigma = 3)$, $(\alpha_0 = 0, \sigma = 0)$, $(\alpha_0 = -1, \sigma = 0)$ *and* $(\alpha_0 = -1, \sigma = 4)$.

(b): The behavior of the modulus maxima across scales depends upon the Lipschitz regularity α_0 *and the smoothing factor* σ.

5. Signal Reconstruction from Multiscale Edges

Section 4 shows that one can get a precise description of the signal sharp variation points from the evolution of the wavelet transform modulus maxima across scales. An important question is to understand whether the whole signal information is embedded into these modulus maxima. Is it possible to have a stable signal reconstruction only from the modulus maxima information at the dyadic scales 2^j ? The next section reviews briefly some results on the reconstruction of signals from zero-crossings and multiscale edges. Section 5.2 describes an algorithm that reconstructs a close approximation of the original signal from the wavelet transform modulus maxima. Numerical results are presented in Section 5.3.

5.1. Previous Results

The reconstruction of signals from multiscale edges has mainly been studied in the zero-crossing framework. We saw in Section 2 that if the wavelet is given by $\psi^b(x) = \frac{d^2\theta(x)}{dx^2}$, multiscale edges are detected from the zero-crossings of the wavelet transform $W_s^b f(x)$. The most basic result concerning the reconstruction of signals from zero-crossing is the Logan theorem [19]. However, as it is explained in [22], the hypothesis of the Logan theorem are not appropriate to study the reconstruction of signals from multiscale edges. The Logan theorem has been generalized by several authors [6, 30, 34] and the reader is referred to a review by Hummel and Moniot [15] for more details.

If the smoothing function $\theta(x)$ is a Gaussian, the properties of the wavelet transform zero-crossings are more easily understood because $W_s^b f(x)$ can be interpreted as the solution of a heat diffusion process, at time $t = s^2$ [17]. With this approach, Hummel and Moniot [15] as well as Yuille and Poggio [33] have proved some completeness properties under restrictive conditions, like supposing that $f(x)$ is a polynomial [35]. In general, there are know counter-examples which prove that the positions of the zero-crossings of $W_s^b f(x)$ do not characterize uniquely the function $f(x)$. For example, the wavelet transforms of $\sin(x)$ and $\sin(x) + \frac{1}{5}\sin(2x)$ have the same zero-crossings at all scales $s > 0$. Meyer [27] found a large family of such counter-examples.

To obtain a complete and stable zero-crossing representation, one of us [22] conjectured that it is sufficient to record the zero-crossing positions of $W_{2^j}^b f(x)$ at all dyadic scales $\left(2^j\right)_{j\in\mathbf{Z}}$, as well the integral values

$$e_n = \int_{z_n}^{z_{n+1}} W_{2^j}^b f(u)\, du \quad , \tag{39}$$

between any pair of consecutive zero-crossings (z_n, z_{n+1}). This conjecture was motivated by a reconstruction algorithm that is able to reconstruct a close approximation of the original signal, from these zero-crossings and integral values [22]. We proved in Section 2 that the zero-crossings of $W^b_{2^j} f(x)$ occur at the extrema points of the wavelet transform $W^a_{2^j} f(x)$, defined with respect to the wavelet $\psi^a(x) = \frac{d\theta(x)}{dx}$. From equations (4), (5) and (39), we derive that

$$e_n = W^a_{2^j} f(z_{n+1}) - W^a_{2^j} f(z_n) . \tag{40}$$

To record the zero-crossing positions and integral values of $W^b_{2^j} f(x)$ is therefore equivalent to record the positions where $W^a_{2^j} f(x)$ has local extrema and the value of $W^a_{2^j} f(x)$ at the corresponding locations. Meyer [27] proved that the completeness of this representation depends upon the choice of the smoothing function $\theta(x)$ but that the conjecture is not valid in general. Let

$$f_0(x) = \begin{cases} \frac{1}{2\pi}(1+\cos(x)) \text{ if } |x| < \pi \\ 0 \text{ otherwise} \end{cases} . \tag{41}$$

For the wavelet shown in Fig. 3(a), Meyer [27] found a non-countable family of functions

$$f_\varepsilon(x) = f_0(x) + \chi_\varepsilon(x) \ ,$$

such that at all scales 2^j, $W_{2^j} f_\varepsilon(x)$ and $W_{2^j} f_0(x)$ have the same extrema (positions and values). The functions $\chi_\varepsilon(x)$ are small high frequency perturbations, implicitly defined by constraint equations that guaranty that the local extrema of $W_{2^j} f_0(x)$ are not modified. It seems that in order to maintain the local extrema of $W_{2^j} f_0(x)$ unchanged, the perturbations $\chi_\varepsilon(x)$ must remain small, which would explain the quality of the signal reconstructions obtained in [22], but this has not been proved. For another wavelet defined by $\psi^a(x) = \frac{d\theta(x)}{dx}$, with $\theta(x) = f_0(x)$, Meyer proved that any function of compact support is uniquely characterized by the zero-crossings and integral values of its dyadic wavelet transform. This characterization is however not stable at high frequencies. The numerical precision of reconstructions is thus not improved with this other wavelet. A discrete analysis of the completeness conjecture was done independently by Berman [1] who found numerical examples which contradict the completeness conjecture.

We explained in Section 2 that for a wavelet equal to the first derivative of a smoothing function, the local minima of the wavelet transform modulus correspond to slow variation points of the signal. Hence, among all the wavelet transform extrema, we detect only the points where the wavelet transform modulus is locally maximum. For the quadratic wavelet of Fig. 3(a), since the wavelet transform local extrema do not provide a complete signal representation, the subset of modulus maxima is certainly not complete either. The next section describes an algorithm that still recovers a precise approximation of the original signals from these modulus maxima.

5.2. Reconstruction Algorithm

Let $f(x) \in \mathbf{L}^2(\mathbf{R})$ and $\left(W_{2^j}f(x)\right)_{j\in \mathbf{Z}}$ be its dyadic wavelet transform. We describe an algorithm that reconstructs an approximation of $\left(W_{2^j}f(x)\right)_{j\in \mathbf{Z}}$, given the positions of the local maxima of $|W_{2^j}f(x)|$ and the values of $W_{2^j}f(x)$ at these locations. For this purpose, we characterize the set of functions $h(x)$ such that at each scale 2^j, the modulus maxima of $W_{2^j}h(x)$ are the same as the modulus maxima of $W_{2^j}f(x)$. We suppose that the wavelet $\psi(x)$ is differentiable in the sense of Sobolev. Since $W_{2^j}f(x)$ is obtained through a convolution with $\psi_{2^j}(x)$, it is also differentiable in the sense of Sobolev and it has at most a countable number of modulus maxima. Let $(x_n^j)_{n\in \mathbf{Z}}$ be the abscissa where $|W_{2^j}f(x)|$ is locally maximum. The maxima constraints on $W_{2^j}h(x)$ can be decomposed in two conditions.

(a) At each scale 2^j, for each local maximum located at x_n^j, $W_{2^j}h(x_n^j) = W_{2^j}f(x_n^j)$.

(b) At each scale 2^j, the local maxima of $|W_{2^j}h(x)|$ are located at the abscissa $(x_n^j)_{n\in \mathbf{Z}}$.

Let us first analyze the condition (a). We saw in equation (20) that a wavelet transform can be rewritten as an inner product

$$W_{2^j}f(x_0) = \langle f(u), \psi_{2^j}(x_0-u)\rangle \ . \tag{42}$$

The condition (a) is thus equivalent to

$$\langle f(u), \psi_{2^j}(x_n^j-u)\rangle = \langle h(u), \psi_{2^j}(x_n^j-u)\rangle \ . \tag{43}$$

Let $\mathbf{U}$ be the closure in $\mathbf{L}^2(\mathbf{R})$ of the space of functions that are linear combinations of functions in the family

$$\left\{\psi_{2^j}(x_n^j-x)\right\}_{(j,n)\in \mathbf{Z}^2} \ . \tag{44}$$

One can easily prove that the functions $h(x)$ that satisfy equations (43) for all abscissa $(x_n^j)_{(n,j)\in \mathbf{Z}^2}$, are the functions whose orthogonal projection on $\mathbf{U}$ is equal to the orthogonal projection of $f(x)$ on $\mathbf{U}$. Let $\mathbf{O}$ be the orthogonal complement of $\mathbf{U}$ in $\mathbf{L}^2(\mathbf{R})$, which means that the space $\mathbf{O}$ is orthogonal to $\mathbf{U}$ and that

$$\mathbf{O} \oplus \mathbf{U} = \mathbf{L}^2(\mathbf{R}) \ . \tag{45}$$

The functions that satisfy equations (43) for all abscissa $(x_n^j)_{(n,j)\in \mathbf{Z}^2}$ can therefore be written

$$h(x) = f(x) + g(x) \quad \text{with} \quad g(x) \in \mathbf{O} \ . \tag{46}$$

This defines an affine space that we denote $f + \mathbf{O}$. If $\mathbf{U} = \mathbf{L}^2(\mathbf{R})$, then $\mathbf{O} = \{\ 0\ \}$ which implies that $h(x)$ must be equal to $f(x)$. In general this is not the case so the equations (43) do not characterize uniquely $f(x)$.

The condition (b) is more difficult to analyze because it is not convex. One can generally find two functions $h_1(x)$ and $h_2(x)$ that satisfy (b) and α, β with $\alpha+\beta=1$, such that $\alpha\, h_1(x)+\beta\, h_2(x)$ does not satisfy the condition (b). In order to solve this problem numerically, we approximate the condition (b) with a convex constraint. The condition (a) defines the value of the wavelet transform at the points $(x_n^j)_{(n,j)\in Z^2}$. Instead of imposing that the local maxima of $W_{2^j}h(x)$ are located at these points, we impose that $|W_{2^j}h(x)|^2$ is as small as possible on average. This generally creates local modulus maxima at the positions $(x_n^j)_{(n,j)\in Z^2}$. The number of modulus maxima of $W_{2^j}f(x)$ depends upon how much this function oscillates. To have as few modulus maxima as possible outside the abscissa $(x_n^j)_{(n,j)\in Z^2}$, we also minimize the energy of the derivative of $W_{2^j}h(x)$. Since these conditions must be imposed at all scales 2^j, we minimize globally

$$|||h|||^2 = \left| \left(W_{2^j}h(x) \right)_{j\in Z} \right|^2 = \sum_{j=-\infty}^{+\infty} \left(\| W_{2^j}h \|^2 + 2^{2j} \left\| \frac{dW_{2^j}h}{dx} \right\|^2 \right) . \tag{47}$$

The weight 2^{2j} expresses that the relative smoothness of $W_{2^j}f(x)$ increases with the scale 2^j. Let $\psi^1(x)$ be the derivative of $\psi(x)$. If there exists two constants $A_2>0$ and B_2 such that for all $\omega\in \mathbf{R}$

$$A_2 \le \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j\omega)|^2 + \sum_{j=-\infty}^{+\infty} |\hat{\psi}^1(2^j\omega)|^2 \le B_2 , \tag{48}$$

then for any $h(x)\in \mathbf{L}^2(\mathbf{R})$

$$A_2 \, \|h\|^2 \le |||h|||^2 \le B_2 \, \|h\|^2 . \tag{49}$$

Hence, $|||\ \ |||$ is a norm over $\mathbf{L}^2(\mathbf{R})$ which is equivalent to the classical $\mathbf{L}^2(\mathbf{R})$ norm. We prove that (48) implies (49) by observing that

$$2^j \frac{dW_{2^j}h(x)}{dx} = f * \psi^1_{2^j}(x) .$$

Like for the energy equivalence equation (15), we then prove the implication by applying the Parseval theorem to each $\mathbf{L}^2(\mathbf{R})$ norm component of the norm defined by equation (47). Equation (48) is valid for any dyadic wavelet $\psi(x)$ that satisfies equation (11) and which is smooth enough. For example, there exists two such constant A_2 and B_2 for the wavelet shown in Fig. 3(a). By replacing the condition (b) by the minimization of $|||h|||$, we define a problem that has a unique solution. Indeed, the condition (a) imposes that $h(x)$ must belong to the closed affine space $f+\mathbf{O}$ and the minimization of a norm over such a closed convex has a unique solution.

Let us study a simple particular example that explains why a multiscale edge detection is an efficient strategy to obtain an adaptive signal characterization. If the two constants A_2 and B_2 of equation (48) are equal, the norm $|||\ \ |||$ is proportional to the classical $\mathbf{L}^2(\mathbf{R})$ norm. One can then easily prove that the solution of the minimization problem is the orthogonal projection of $f(x)$ over $\mathbf{U}$. The reconstruction error is then given by

$$\|f - \mathbf{P_U} f\|^2 = \|f\|^2 - \|\mathbf{P_U} f\|^2 \ . \tag{50}$$

This error is decreased by increasing $\|\mathbf{P_U} f\|$. The space $\mathbf{U}$ is defined from the wavelets $\psi_{2^j}(x_n^j - u)$ such that $|\langle f(u), \psi_{2^j}(x_n^j - u)\rangle|$ are local maxima. By chosing wavelets whose inner products with $f(x)$ are large, we guarantee that the projection of $f(x)$ on $\mathbf{U}$ is large. The local maxima detection can thus be interpreted as a procedure to define an adaptive space $\mathbf{U}$ where the projection of $f(x)$ has a large energy, in order to minimize the reconstruction error.

Although there exists a unique element of $f + \mathbf{O}$ whose norm $|||\ |||$ is minimum, the computation of this function might not be stable. If the two constants A_2 and B_2 of equations (48) are equal, we saw that the solution is equal to $\mathbf{P_U} f(x)$. The frame theory proves [9] that one can make a stable computation of $\mathbf{P_U} f(x)$ from the inner products $\left(\langle f(x), \psi_{2^j}(x_n^j - x)\rangle\right)_{(n,j)\in \mathbf{Z}^2}$, if and only if, the family of functions $\left(\sqrt{2^j}\ \psi_{2^j}(x_n^j - x)\right)_{(n,j)\in \mathbf{Z}^2}$ is a frame of $\mathbf{U}$. The factor $\sqrt{2^j}$ normalizes the $\mathbf{L}^2(\mathbf{R})$ norm of each function in the family. By definition, such a family is a frame of $\mathbf{U}$ if and only if there exists two constants $A_3 > 0$ and B_3 such that for any function $g \in \mathbf{U}$

$$A_3 \|g\|^2 \le \sum_{(n,j)\in \mathbf{Z}^2} 2^j\, |\langle g(u), \psi_{2^j}(x_n^j - u)\rangle|^2 \le B_3 \|g\|^2 \ . \tag{51}$$

When the two constants A_2 and B_2 of equation (48) are different, the norm $|||\ |||$ is not equal but equivalent to the classical $\mathbf{L}^2(\mathbf{R})$ norm, so one can prove that the stability also depends whether the family of wavelets is a frame of $\mathbf{U}$. The closer to 0 the value of $\frac{B_3 - A_3}{B_3 + A_3}$, the more stable the computations. Outside a few particular cases, it is difficult to prove analytically whether a given family of wavelets $\left(\sqrt{2^j}\ \psi_{2^j}(x_n^j - x)\right)_{(n,j)\in \mathbf{Z}^2}$ is or is not a frame of the space $\mathbf{U}$ that it generates because the points x_n^j are not uniformly distributed.

Let us now describe an algorithm that computes the solution of our minimization problem. Instead of computing the solution itself, we reconstruct its wavelet transform with an alternate projection algorithm. Let $\mathbf{K}$ be the space of all sequences of functions $\left(g_j(x)\right)_{j\in \mathbf{Z}}$ such that

$$\left|\left(g_j(x)\right)_{j\in \mathbf{Z}}\right|^2 = \sum_{j=-\infty}^{+\infty} \left(\|g_j\|^2 + 2^{2j}\, \|\frac{dg_j}{dx}\|^2\right) < +\infty \ . \tag{52}$$

The norm $|\ |$ defines a Hilbert structure over $\mathbf{K}$. Let $\mathbf{V}$ be the space of all dyadic wavelet transforms of functions in $\mathbf{L}^2(\mathbf{R})$. Equation (49) proves that $\mathbf{V}$ is included in $\mathbf{K}$. Let Γ be the affine space of sequences of functions $\left(g_j(x)\right)_{j\in \mathbf{Z}} \in \mathbf{K}$ such that for any index j and all maxima positions x_n^j

$$g_j(x_n^j) = W_{2^j} f(x_n^j) \ .$$

One can prove that Γ is closed in $\mathbf{K}$. The dyadic wavelet transforms that satisfy the condition (a) are the sequences of functions that belong to

$$\Lambda = \mathbf{V} \cap \Gamma \ .$$

We must therefore find the element of Λ whose norm $\| \ \|$ is minimum. This is done by alternating projections on $\mathbf{V}$ and Γ.

Equation (17) shows that any dyadic wavelet transform is invariant under the operator

$$\mathrm{P_V} = \mathrm{W} \circ \mathrm{W}^{-1} \ . \tag{53}$$

For any sequence $Y = \left[g_j(x) \right]_{j \in \mathbf{Z}} \in \mathbf{K}$, it is clear that $\mathrm{P_V} Y \in \mathbf{V}$, so $\mathrm{P_V}$ is a projector on $\mathbf{V}$. We saw in equation (19) that this operator is characterized by the kernels

$$K_{l,j}(x) = \chi_{2^l} * \psi_{2^j}(x) \ .$$

One can easily prove that the projector $\mathrm{P_V}$ is self-adjoint and therefore orthogonal if and only if the kernels $K_{l,j}(x)$ are symmetrical functions. This is the case if the wavelet $\psi(x)$ is symmetrical or antisymmetrical. For the wavelet shown in Fig. 3(a), the orthogonal projection on the space $\mathbf{V}$ is thus implemented by applying the operator W^{-1} followed by the operator W. The fast discrete implementation of these operators is given in Appendix 2. Appendix 5 characterizes the projection on the affine set Γ which is orthogonal with respect to the norm $\| \ \|$. We prove that this operator P_Γ is implemented by adding piece-wise exponential curves to each function of the sequence that we project on Γ. Let $\mathrm{P} = \mathrm{P_V} \circ \mathrm{P}_\Gamma$ be alternate projections on both spaces. Let $\mathrm{P}^{(n)}$ be n iterations over the operator P. Since Γ is an affine space and $\mathbf{V}$ a Hilbert space, a classical result on alternate projections [32] proves that for any sequence of functions $X = \left[g_j(x) \right]_{j \in \mathbf{Z}} \in \mathbf{K}$

$$\lim_{n \to +\infty} \mathrm{P}^{(n)} X = \mathrm{P}_\Lambda X \ . \tag{54}$$

Alternate projections on Γ and $\mathbf{V}$ converge strongly to the orthogonal projection on Λ. If X is the zero element of $\mathbf{K}$ which means that $g_j(x) = 0$ for all $j \in \mathbf{Z}$, the alternate projections converge to the element of Λ which is the closest to zero and thus whose norm $\| \ \|$ is minimum. This is illustrated by Fig. 6. This iterative algorithm can be related to techniques based on frame operators for reconstructing signals from irregular samplings [12].

If the minimization problem is unstable, the convergence of the alternate projections is extremely slow. We saw that the numerical stability depends whether $\left[\sqrt{2^j}\ \psi_{2^j}(x_n^j - x)\right]_{(n,j)\in \mathbf{Z}^2}$ is a frame of $\mathbf{U}$. Appendix 6 proves that if $\left[\sqrt{2^j}\ \psi_{2^j}(x_n^j - x)\right]_{(n,j)\in \mathbf{Z}^2}$ is a frame of $\mathbf{U}$, and if there exists a constant $0 < D \leq 1$ such that at all scales 2^j the distances between any two consecutive maxima satisfy

$$|x_n^j - x_{n-1}^j| \geq D\, 2^j ,$$

then the convergence is exponential. Moreover, there exists a constant R such that for any $X \in \mathbf{K}$

$$\| \mathbf{P}^{(n)}X - \mathbf{P}_{\Lambda}X \| \leq R\,(1 - \frac{D\,A_3}{2B_2})^{n/2} , \qquad (55)$$

where the A_3 is the frame bound defined in equation (51) and B_2 the norm equivalence bound defined in equation (48). This equation gives a lower bound for the convergence rate and shows how it decreases when the frame bound A_3 goes to zero.

When the original wavelet transform $W_{2^j}f(x)$ has an abrupt transition, the minimization of $\| \; \|$ can yield a smoother solution $W_{2^j}h(x)$ which oscillates slightly at the location where $W_{2^j}f(x)$ has this sharp change. These oscillations are similar to a Gibbs phenomenon. Appendix 5 explains how to modify the alternate projections in order to suppress these oscillations. Numerical experiments show that this oscillation removal does not perturbate the convergence of the algorithm.

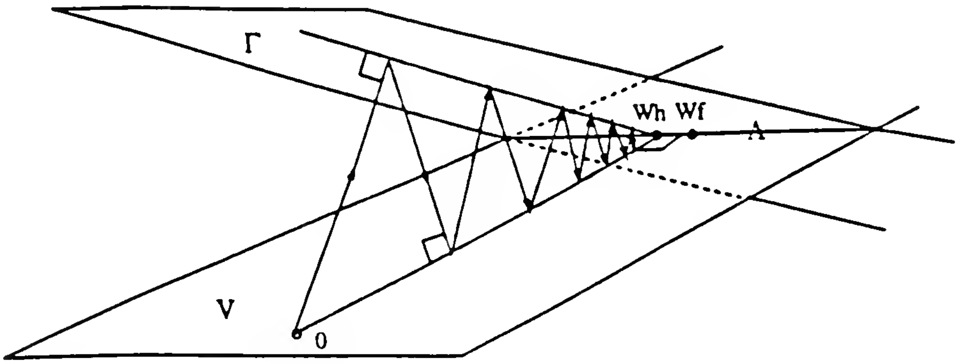


Fig. 6: *An approximation of the wavelet transform of $f(x)$ is reconstructed by alternating orthogonal projections on an affine space Γ and on the space $\mathbf{V}$ of all dyadic wavelet transforms. The projections begin from the zero element and converges to its orthogonal projection on $\Gamma \cap \mathbf{V}$.*

5.3. Numerical Reconstruction of One-Dimensional Signals from Local Maxima

There are several open issues behind the reconstruction algorithm that we described. From the results of Meyer's work [27], we know that in general we can not reconstruct exactly a function from the modulus maxima of its wavelet transform. Our algorithm approximates this inverse problem by replacing the maxima constraint by the minimization of a norm and thus has a unique solution. We thus do not converge towards the wavelet transform of the original signal but towards some other wavelet transform that we hope to be close to the original one. We also explained that the computation of the solution might be unstable, in which case the alternate projections converge very slowly. It therefore important to measure how far we are from the convergence point after a given number of iterations.

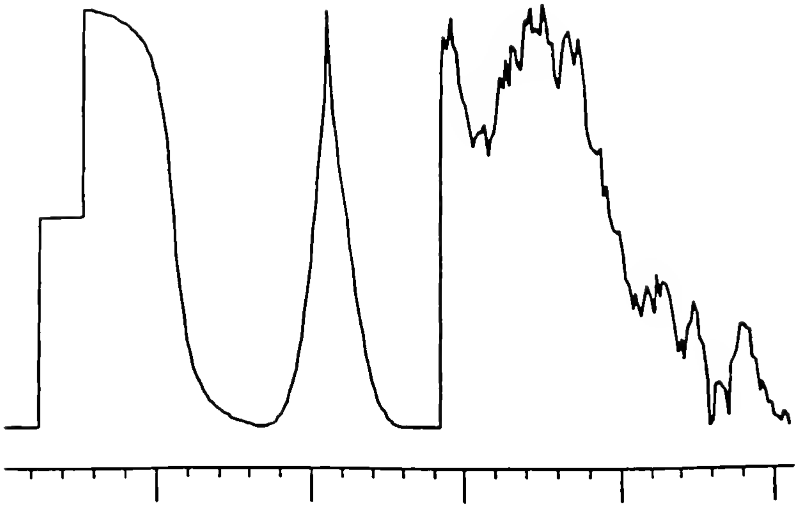
If the original signal has N samples, we record the positions and values of the modulus maxima at all scales 2^j, for $1 \leq j \leq \log_2(N)+1$. We also keep the average value of the original discrete signal, which characterizes $S_{2^J}^d f$ for $J = \log_2(N)+1$, as explained in Section 3.2. Equation (53) proves that we can compute the projection $\mathbf{P_V}$ by implementing $\mathbf{W}^{-1}$ followed by $\mathbf{W}$. With the fast algorithms described in Appendix 2, this requires a total of $O(N \log_2(N))$ operations. Appendix 5 proves that the implementation of $\mathbf{P}_\Gamma$ also requires $O(N \log_2(N))$ operations. The projection operator that suppresses the wavelet transform oscillations is computed with the same complexity. Hence, each iteration on $\mathbf{P}$ involves $O(N \log_2(N))$ operations.

The signal to noise ratio of the reconstructions is measured in db. Let σ_g^2 be the variance of the function $g(x)$ that we want to reconstruct and σ_e^2 be the variance of the reconstruction error $e(x)$. The SNR of $g(x) + e(x)$ with respect to $g(x)$ is

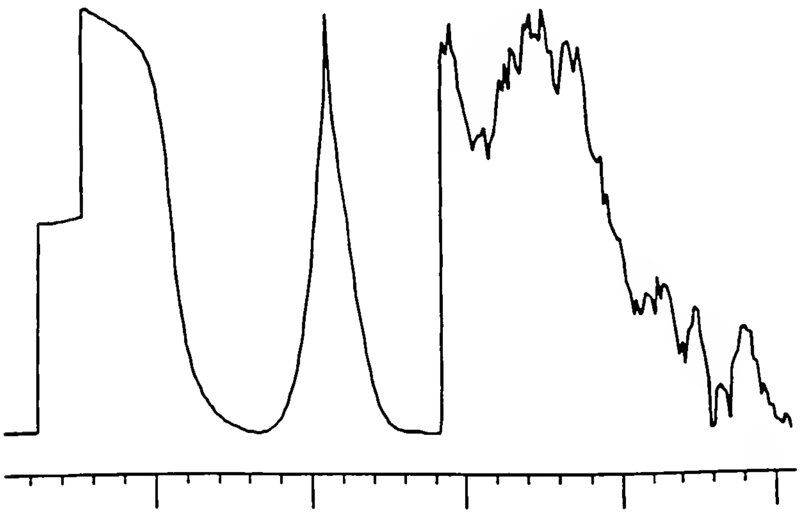
$$SNR = 20 \log_{10}(\frac{\sigma_g}{\sigma_e}) .$$

At the scale 2^j, for $1 \leq j \leq 6$, Fig. 8(a) gives the value of the SNR for the reconstruction of $W_{2^j}^d f$, after n iterations on the operator $\mathbf{P}$, with $1 \leq n \leq 100$. At all scales, the error decreases quickly during the first 20 iterations and then decays much more slowly. For a fixed number of iterations on $\mathbf{P}$, the SNR increases when the scale increases. This proves that the remaining error is rather concentrated at fine scales, like in the counter-examples of Meyer [27]. After n iterations, we can reconstruct a signal by applying the inverse wavelet transform operator on the reconstructed wavelet transform. Fig. 8(b) shows the increase of the SNR, computed with respect to the original signal. This SNR is an aggregation of the wavelet transform SNR at all scales. The signal in Fig. 7(b) is reconstructed by applying the inverse wavelet operator on the reconstructed wavelet transform after 20 iterations. In this case, the SNR is 34.6 db. The remaining error after n iterations has two components. The first one is the distance between the reconstructed wavelet transform and the wavelet transform we converge to. The other one is the distance between the wavelet transform we converge to and the wavelet transform of the original signal. We saw that the convergence rate of the algorithm is related to the frame properties of the family of wavelets defined by the maxima positions. In numerical computations, there is a finite number of maxima so the family of wavelets that generates $\mathbf{U}$ is finite. A finite family of vectors is always a frame, but the frame bound A_3 can be very small so that the lower bound of the convergence rate given by equation (55) is also very small. Fig. 9 is the SNR of the reconstructed signal computed with respect to the signal we converge to, instead of the original signal. After 30 iterations, the slope of the SNR curve is constant, which proves that the convergence is exponential, but the convergence rate is slow. In Fig. 8(b) and Fig. 9, the increase of the SNR slows down at the same point which corresponds approximatively to 30 iterations. One can verify that at this point, the distance between the reconstructed signal and the signal we converge to is of the same order as the distance between the original signal and the signal we converge to. Increasing the number of iterations slowly reduces the distance with respect to the point we converge to, but does not decrease much the distance with respect to the original signal. This is why SNR in Fig. 9 continues to increase slowly while the SNR in Fig. 8(b) reaches a maximum which is of the order of 38 db.

We made extensive numerical tests including reconstructions of special functions such as sinusoidal waves, Gaussians, step edges, Diracs, fractals, and the counter example of Meyer given by equation (41). In all these examples, the SNR has the same type of behavior as in Fig. 8 and 9. In some cases, the convergence rate is faster than in Fig. 9, but this does not influence much the numerical precision of reconstructions since we are limited by the distance between the point we converge to and the original signal. In most cases, after 30 iterations, the relative increase of precision that is obtained by increasing the number of iterations is negligible. Since each iteration requires $O(N\ log(N))$ operations, these reconstructions do not require extensive computations and can be done in real time. The reconstructed functions are not equal to the original signal but are numerically close. They have no spurious oscillations and the same type of sharp variations. Qualitatively, the reconstructed signals are thus very similar to the original one and the errors are hardly noticeable by comparing the graphs, as shown by Fig. 7. We have no upper bound on the error due to the distance between the signal we converge to and the original signal. This is an open mathematical problem, but the numerical precision of this reconstruction algorithm is sufficient for many signal processing applications.



(a)



(b)

Fig. 7: *(a): Original signal. (b): Signal reconstructed with 20 iterations from the modulus maxima shown in Fig. 4(c).*

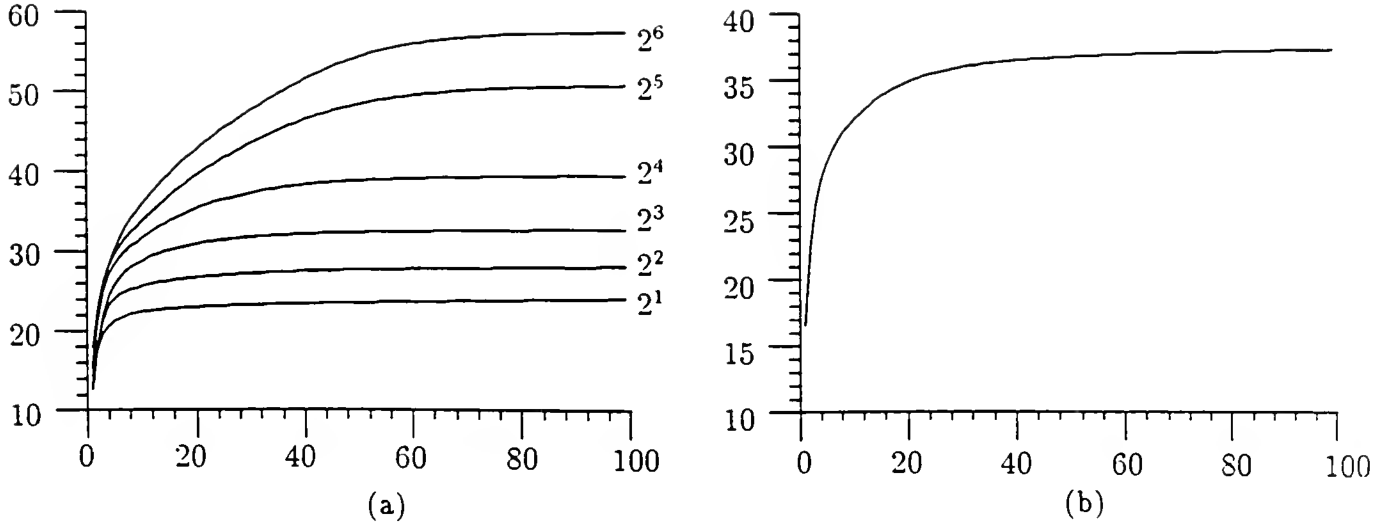


Fig. 8: *(a): Signal to Noise Ratio for the reconstruction of the wavelet transform* $W^d_{2^j}f$*, as a function of the number of iterations on the operator* **P**. *Each curve is labeled by the scale* 2^j*, for* $1 \le j \le 6$.

(b): SNR of the reconstructed signal computed with respect to the original signal, as a function of the number of iterations on the operator **P**.

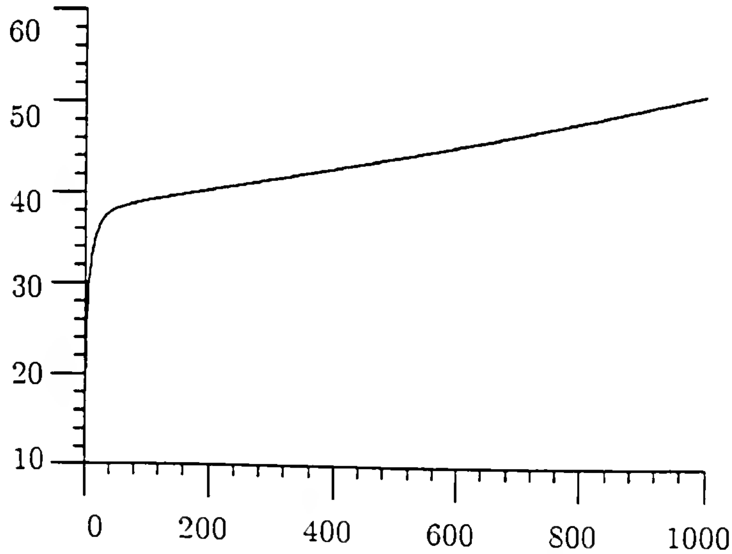


Fig. 9: *SNR of the reconstructed signal computed with respect to the signal we converge to, as a function of the number of iterations on the operator* **P**.

6. Wavelet Transform of Images

We explained in Section 2 that in two dimensions, a multiscale edge detection can be reformalized through a wavelet transform defined with respect to two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$. In the second part of this article, we extend our one-dimensional results for image processing applications.

6.1. General Properties

We denote $\psi^1_{2^j}(x,y) = \frac{1}{2^{2j}}\,\psi^1(\frac{x}{2^j},\frac{y}{2^j})$ and $\psi^2_{2^j}(x,y) = \frac{1}{2^{2j}}\,\psi^2(\frac{x}{2^j},\frac{y}{2^j})$. The wavelet transform of a function $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$, at the scale 2^j, has two components defined by

$$W^1_{2^j} f(x,y) = f * \psi^1_{2^j}(x,y) \quad \text{and} \quad W^2_{2^j} f(x,y) = f * \psi^2_{2^j}(x,y) \ . \tag{56}$$

We call the *two-dimensional dyadic wavelet transform* of $f(x,y)$, the set of functions

$$\mathbf{W}f = \left(W^1_{2^j} f(x,y)\,,\, W^2_{2^j} f(x,y) \right)_{j \in \mathbf{Z}} \ . \tag{57}$$

Let $\hat{\psi}^1(\omega_x,\omega_y)$ and $\hat{\psi}^2(\omega_x,\omega_y)$ be the Fourier transforms of $\psi^1(x,y)$ and $\psi^2(x,y)$. The Fourier transforms of $W^1_{2^j} f(x,y)$ and $W^2_{2^j} f(x,y)$ are respectively given by

$$\hat{W}^1_{2^j} f(\omega_x,\omega_y) = \hat{f}(\omega_x,\omega_y)\,\hat{\psi}^1(2^j\omega_x,2^j\omega_y) \ , \tag{58}$$

$$\hat{W}^2_{2^j} f(\omega_x,\omega_y) = \hat{f}(\omega_x,\omega_y)\,\hat{\psi}^2(2^j\omega_x,2^j\omega_y) \ . \tag{59}$$

To insure that a dyadic wavelet transform is a complete and stable representation of $f(x,y)$, we impose that the two-dimensional Fourier plane is covered by the dyadic dilations of $\hat{\psi}^1(\omega_x,\omega_y)$ and $\hat{\psi}^2(\omega_x,\omega_y)$. This means that there exists two strictly positive constants A_4 and B_4 such that

$$\forall(\omega_x,\omega_y) \in \mathbf{R}^2 \quad , \quad A_4 \le \sum_{j=-\infty}^{+\infty} \left(|\hat{\psi}^1(2^j\omega_x,2^j\omega_y)|^2 + |\hat{\psi}^2(2^j\omega_x,2^j\omega_y)|^2 \right) \le B_4 \ . \tag{60}$$

Let $\chi^1(x,y)$ and $\chi^2(x,y)$ be two functions whose Fourier transform satisfy

$$\sum_{j=-\infty}^{+\infty} \left(\hat{\psi}^1(2^j\omega_x,2^j\omega_y)\,\hat{\chi}^1(2^j\omega_x,2^j\omega_y) + \hat{\psi}^2(2^j\omega_x,2^j\omega_y)\,\hat{\chi}^2(2^j\omega_x,2^j\omega_y) \right) = 1 \ . \tag{61}$$

There is an infinite number of choices for $\chi^1(x,y)$ and $\chi^2(x,y)$. We can derive from equations (58), (59) and (61) that $f(x,y)$ is reconstructed from its dyadic wavelet transform with

$$f(x,y) = \sum_{j=-\infty}^{+\infty} \left(W^1_{2^j} f * \chi^1_{2^j}(x,y) + W^2_{2^j} f * \chi^2_{2^j}(x,y) \right) \ . \tag{62}$$

A two-dimensional dyadic wavelet transform is more than complete, it is redundant. Any sequence of functions $\left(g^1_j(x,y), g^2_j(x,y) \right)_{j \in \mathbf{Z}}$ is not necessarily the dyadic wavelet transform of some functions in $\mathbf{L}^2(\mathbf{R}^2)$. We denote by $\mathbf{W}^{-1}$ the operator defined by

$$\mathbf{W}^{-1}\left[g_j^1(x,y),g_j^2(x,y)\right]_{j\in\mathbf{Z}} = \sum_{j=-\infty}^{+\infty}\left(g_j^1 * \chi^1_{2^j}(x,y) + g_j^2 * \chi^2_{2^j}(x,y)\right) . \tag{63}$$

The sequence $\left[g_j^1(x,y),g_j^2(x,y)\right]_{j\in\mathbf{Z}}$ is a dyadic wavelet transform if and only if

$$\mathbf{W}\left[\mathbf{W}^{-1}\left[g_j^1(x,y),g_j^2(x,y)\right]_{j\in\mathbf{Z}}\right] = \left[g_j^1(x,y),g_j^2(x,y)\right]_{j\in\mathbf{Z}} . \tag{64}$$

In Section 2, we explained that multiscale sharp variation points can be obtained from a dyadic wavelet transform if

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \quad \text{and} \quad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} . \tag{65}$$

Equation (8) proves that the wavelet transform can be rewritten

$$\begin{bmatrix} W^1_{2^j}f(x,y) \\ W^2_{2^j}f(x,y) \end{bmatrix} = 2^j \begin{bmatrix} \frac{\partial}{\partial x}(f * \theta_{2^j})(x,y) \\ \frac{\partial}{\partial y}(f * \theta_{2^j})(x,y) \end{bmatrix} = 2^j\, \vec{\nabla}(f * \theta_{2^j})(x,y) . \tag{66}$$

The two components of the wavelet transform are proportional to the two components of the gradient vector $\vec{\nabla}(f * \theta_{2^j})(x,y)$. This appears clearly in Fig. 10 that shows the two-dimensional wavelet transform of the image of a circle. At each scale 2^j, the modulus of the gradient vector is proportional to

$$M_{2^j}f(x,y) = \sqrt{|W^1_{2^j}f(x,y)|^2 + |W^2_{2^j}f(x,y)|^2} . \tag{67}$$

The angle of the gradient vector with the horizontal direction is given by

$$A_{2^j}f(x,y) = argument(\, W^1_{2^j}f(x,y) + i\, W^2_{2^j}f(x,y)\,) . \tag{68}$$

Like in the Canny algorithm [3], the sharp variation points of $f * \theta_{2^j}(x,y)$ are the points (x,y) where the modulus $M_{2^j}f(x,y)$ has a local maxima in the direction of the gradient given by $A_{2^j}f(x,y)$. We record the position of each of these modulus maxima as well the values of the modulus $M_{2^j}f(x,y)$ and the angle $A_{2^j}f(x,y)$ at the corresponding locations.

The circle image at the top of Fig. 10 has 128 by 128 pixels. The first two columns of Fig. 10 gives the discrete wavelet transform $W^{1,d}_{2^j}f$ and $W^{2,d}_{2^j}f$, for $1 \le j \le 8$. The next section explains how to define such a discrete dyadic wavelet transform and how to solve border problems. The reader not interested by numerical implementations can skip Section 6.2. The discrete modulus images $M^d_{2^j}f$ and angle images $A^d_{2^j}f$ are shown along the next two columns. Along the border of the circle, the angle value turns from 0 to 2π and the modulus has a maximum amplitude. When the scale 2^j is larger than 2^6, we see that the circle is deformed due to the image periodization that we use for border computations. The position of the modulus maxima at all scales are given in the last column on the right. The original Lena image is shown at the top left of Fig. 12 and has 256 by 256 pixels. The third column of Fig. 11 displays its discrete modulus images $M^d_{2^j}f$ and the fifth column gives the position of the modulus maxima, for $1 \le j \le 9$. At fine scales, there are many maxima created by the image noise. At these locations, the modulus value has a small amplitude. The last column displays the maxima whose modulus are larger than a given threshold, at all scales. The edge points with a high modulus value correspond to the sharper intensity variations of the image. At coarse scales, the modulus maxima have different positions than at fine scales. This is due to the smoothing of the image by $\theta_{2^j}(x,y)$.

Fig. 10: *see caption page 39.*

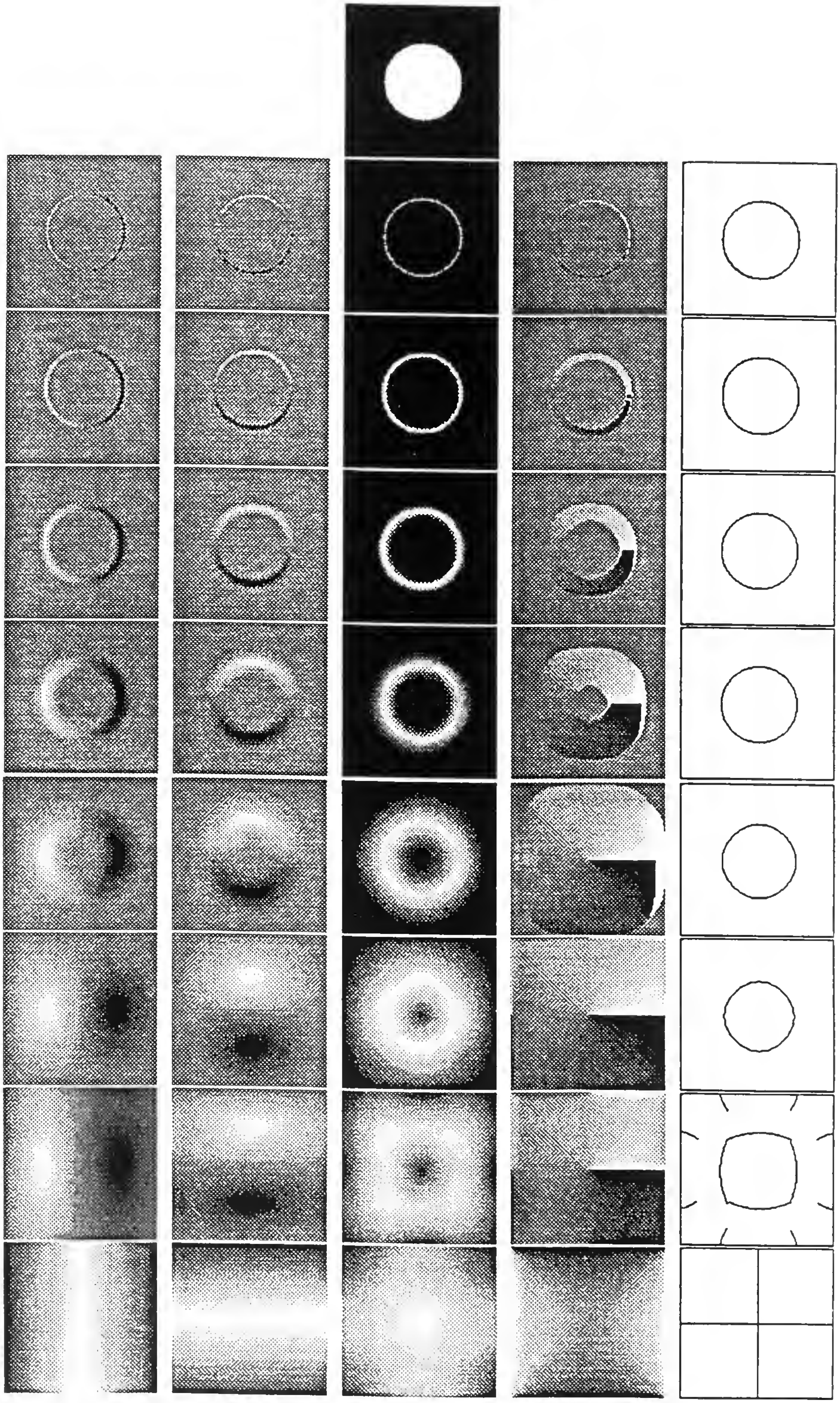


Fig. 11: *see caption page 39.*

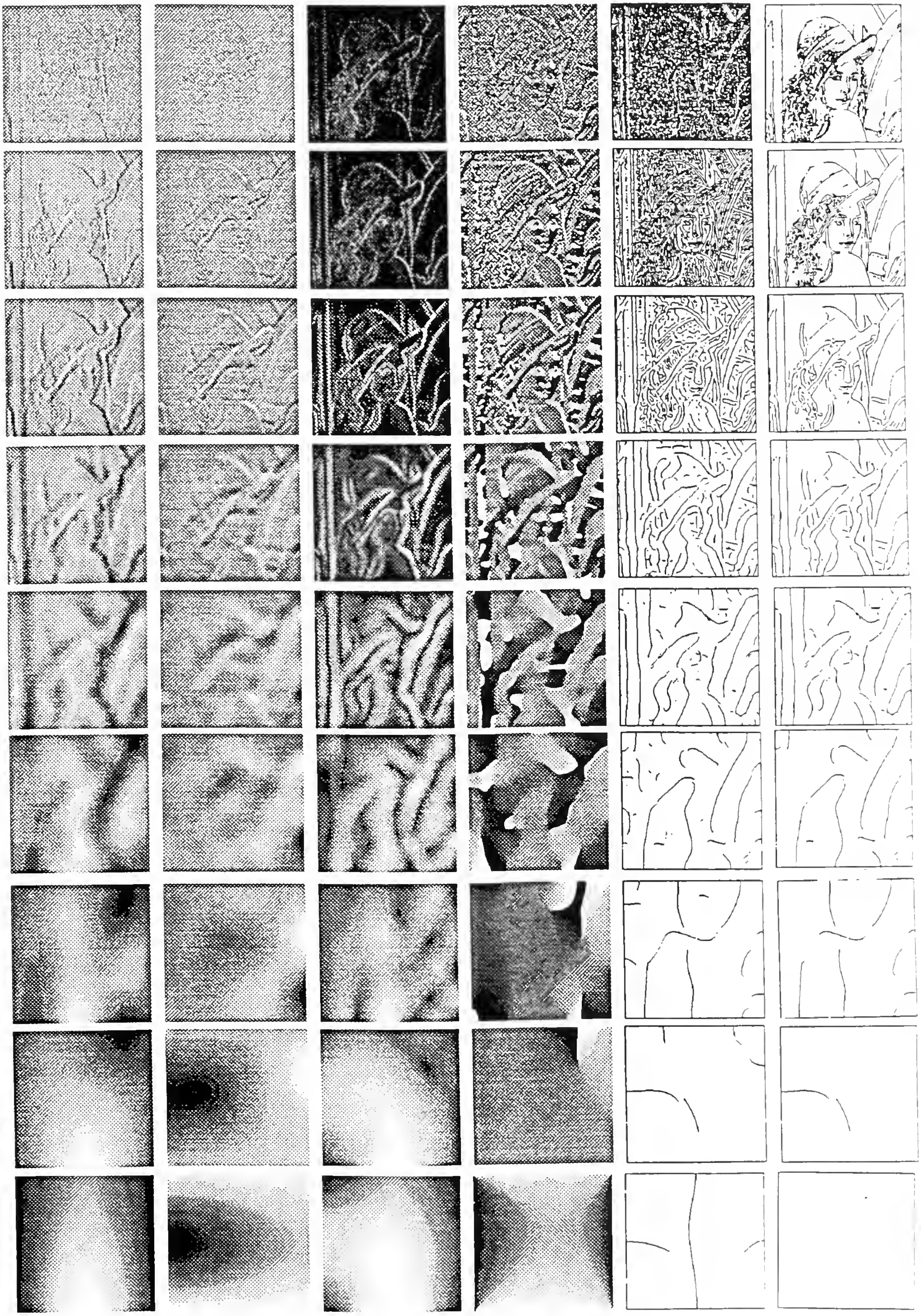


Fig. 10: *The original image at the top has 128 by 128 pixels. The first two columns from the left show respectively* $\left(W^1_{2^j}{}^d f)\right)_{1 \le j \le 8}$ *and* $\left(W^2_{2^j}{}^d f\right)_{1 \le j \le 8}$. *The scale increases from top to bottom. Black, grey and white pixels indicate respectively negative, zero and positive pixel values. The third column displays the modulus images* $\left(M^d_{2^j} f\right)_{1 \le j \le 8}$. *Black pixels indicate zero values whereas white ones correspond to highest values. The fourth column shows the angle images* $\left(A^d_{2^j} f\right)_{1 \le j \le 8}$. *The angle value turns from* 0 *(white) to* 2π *(black) along the circle contour. The images of the last column display in black the points where* $M^d_{2^j} f$ *has local maxima in the direction indicated* $A^d_{2^j} f$.

Fig. 11: *The original Lena image at the top left of Fig. 12 and has 256 by 256 pixels. The first two columns from the left show respectively* $\left(W^1_{2^j}{}^d f\right)_{1 \le j \le 9}$ *and* $\left(W^2_{2^j}{}^d f\right)_{1 \le j \le 9}$. *The third column displays the modulus images* $\left(M^d_{2^j} f\right)_{1 \le j \le 9}$. *The fourth column shows the angle images* $\left(A^d_{2^j} f\right)_{1 \le j \le 9}$. *The fifth column displays the position of the local maxima of* $M^d_{2^j} f$, *for* $1 \le j < 9$. *The last column gives the positions of local maxima where the modulus value is larger than a given threshold. Local maxima that correspond to light texture variations are removed by the thresholding.*

6.2. Discrete Wavelet Transform of Images

The discretization of the two-dimensional wavelet transform raises the same problems as in one-dimension. Images are measured at a finite resolution so we cannot compute the wavelet transform at scales below the limit set by this resolution. As in one dimension, in order to model the limitation of resolution, we introduce a smoothing function $\phi(x,y)$ whose Fourier transform satisfies

$$|\hat{\phi}(\omega_x,\omega_y)|^2 = \sum_{j=1}^{+\infty} \left[\hat{\psi}^1(2^j\omega_x,2^j\omega_y)\,\hat{\chi}^1(2^j\omega_x,2^j\omega_y) + \hat{\psi}^2(2^j\omega_x,2^j\omega_y)\,\hat{\chi}^2(2^j\omega_x,2^j\omega_y)\right] \quad (69)$$

As a consequence of the admissibility condition (61), one can derive that

$$\lim_{(\omega_x,\omega_y)\to(0,0)} |\hat{\phi}(\omega_x,\omega_y)| = 1 \ .$$

We also impose that $\phi(x,y)$ is real. This limit implies that the integral of $\phi(x,y)$ is equal to 1 which means that it is a smoothing function. We define the smoothing operator S_{2^j} by

$$S_{2^j}f(x,y) = f * \phi_{2^j}(x,y) \quad \text{with} \quad \phi_{2^j}(x,y) = \frac{1}{2^j}\,\phi\left(\frac{x}{2^j},\frac{y}{2^j}\right) \ . \qquad (70)$$

Like in one dimension, one can prove that the wavelet transform between the scales 1 and 2^J, $\left[W^1_{2^j}f(x,y)\ ,\ W^2_{2^j}f(x,y)\right]_{1\le j\le J}$, provides the details available in $S_1f(x,y)$ but that have disappeared in $S_{2^J}f(x,y)$. Any image $S_1f(x,y)$ is thus characterized by the *finite dyadic wavelet transform*

$$\left\{\left[W^1_{2^j}f(x,y)\ ,\ W^2_{2^j}f(x,y)\right]_{1\le j\le J}\ ,\ S_{2^J}f(x,y)\right\} \ .$$

In order to compute the wavelet transform with a minimum amount of operations, we choose two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ that can be written as separable products of functions of the x and y variables. Given a one-dimensional dyadic wavelet $\psi(x)$ such that $\psi(x) = \frac{d\theta(x)}{dx}$, we build the two wavelets

$$\psi^1(x,y) = \psi(x)\,\xi(y) \quad \text{and} \quad \psi^2(x,y) = \xi(x)\,\psi(y) \ . \qquad (71)$$

The function $\xi(x)$ is a one-dimensional smoothing function that is defined in Appendix 3, so that the two-dimensional wavelet transform can be implemented with a fast pyramidal algorithm. Since $\psi(x) = \frac{d\theta(x)}{dx}$, these two wavelets can be rewritten

$$\psi^1(x,y) = \frac{\partial\theta^1(x,y)}{\partial x} \quad \text{and} \quad \psi^2(x,y) = \frac{\partial\theta^2(x,y)}{\partial y} \ , \quad \text{with} \qquad (72)$$

$$\theta^1(x,y) = \theta(x)\,\xi(y) \quad \text{and} \quad \theta^2(x,y) = \xi(x)\,\theta(y) \ .$$

The fast discrete wavelet algorithm does not allow us to have $\xi(x) = \theta(x)$. Hence, $\theta^1(x,y) \neq \theta^2(x,y)$. Although the two functions $\theta^1(x,y)$ and $\theta^2(x,y)$ are different, we choose $\xi(x)$ so that they are close enough and can be considered as equal to a single function $\theta(x,y)$, in a first approximation.

At the output of a camera digitizer, an image is a finite energy two-dimensional discrete signal $D = \left[d_{n,m} \right]_{(n,m) \in \mathbf{Z}^2}$. As in one dimension, one can prove that there exists a function $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ (not unique) such that

$$\forall (n,m) \in \mathbf{Z}^2 \quad , \quad S_1 f(n,m) \;=\; d_{n,m} \; . \tag{73}$$

The discrete image can thus be rewritten $D = \left[S_1 f(n,m) \right]_{(n,m) \in \mathbf{Z}^2}$. For the class of wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ given in Appendix 3, from the sample values $\left[S_1 f(n,m) \right]_{(n,m) \in \mathbf{Z}^2}$, we can compute a uniform sampling of the finite-scale dyadic wavelet transform of $S_1 f(x,y)$. For any scale 2^j, we denote

$$W_{2^j}^{1\,d} f \;=\; \left[W_{2^j}^1 f(n+w, m+w) \right]_{(n,m) \in \mathbf{Z}^2} , \quad W_{2^j}^{2\,d} f \;=\; \left[W_{2^j}^2 f(n+w, m+w) \right]_{(n,m) \in \mathbf{Z}^2} \quad \text{and}$$

$$S_{2^j}^d f \;=\; \left[S_{2^j} f(n+w, m+w) \right]_{(n,m) \in \mathbf{Z}^2} \; .$$

The sampling shift w depends upon the choice of wavelets. Given an image D, Appendix 4 describes a fast algorithm that computes the *discrete dyadic wavelet transform*

$$\left\{ S_{2^J}^d f \, , \; \left[W_{2^j}^{1\,d} f \right]_{1 \le j \le J} , \; \left[W_{2^j}^{2\,d} f \right]_{1 \le j \le J} \right\} \; . \tag{74}$$

Images are finite two-dimensional discrete signals $D = \left[d_{n,m} \right]_{1 \le n,m \le N}$ of N by N pixels. We solve border problems like in a two-dimensional cosine transform. We define an infinite two-dimensional discrete signal $\tilde{D} = \left[\tilde{d}_{n,m} \right]_{(n,m) \in \mathbf{Z}^2}$ of period $2N$ by $2N$, by extending the image D into an image of $2N$ by $2N$ pixels with a symmetry with respect to its left and right borders, and periodizing the resulting signal. The signal $\tilde{D}$ is given by

$$\tilde{d}_{n,m} \;=\; \begin{cases} d_{n,m} & \text{if} \;\; 1 \le n \le N \;\; \text{and} \;\; 1 \le m \le N \\ d_{2N+1-n,m} & \text{if} \;\; N+1 \le n \le 2N \;\; \text{and} \;\; 1 \le m \le N \\ d_{n,\,2N+1-m} & \text{if} \;\; 1 \le n \le N \;\; \text{and} \;\; N+1 \le m \le 2N \\ d_{2N+1-n,\,2N+1-m} & \text{if} \;\; N+1 \le n \le 2N \;\; \text{and} \;\; N+1 \le m \le 2N \end{cases} \; .$$

The discrete wavelet images $W_{2^j}^{1\,d} f$ and $W_{2^j}^{2\,d} f$ are computed from $\tilde{D}$ by a sequence of discrete convolutions that are described in Appendix 4. Since $\tilde{D}$ has a period of $2N$ by $2N$ pixels, $W_{2^j}^{1\,d} f$

and $W_{2^J}^{2\,d} f$ have also have a period of $2N$ by $2N$ samples. The two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ are defined from the one-dimensional functions $\psi(x)$ and $\xi(x)$, that are respectively antisymmetrical and symmetrical with respect to $x = 0$. Hence, each row of $W_{2^J}^{1\,d}$ is antisymmetrical with respect to the points of abscissa $1/2$ and $N + 1/2$ and each column is symmetrical with respect to $1/2$ and $N + 1/2$. The transposed result is valid for $W_{2^J}^{2\,d}$. We thus only need N^2 samples to characterize $W_{2^J}^{1\,d}$ as well as $W_{2^J}^{2\,d}$. The smooth signals $S_{2^J}^d f$ has also a period of $2N$ by $2N$ samples. For $J = \log_2(N) + 1$, one can prove that $S_{2^J}^d f$ is constant and equal to the average of the original image D. We thus decompose images of N^2 pixels over $\log_2(N)$+1 scales. Fig. 10 and 11 are two examples. The numerical complexity of the fast discrete wavelet transform is $O(N^2 \log(N))$. The reconstruction the original image from its discrete wavelet transform is also performed with $O(N^2 \log(N))$ operations. The discrete modulus images $M_{2^j}^d f$ and angle images $A_{2^j}^d f$ are computed with equations (67) and (68). The modulus maxima are the points of the modulus images $M_{2^j}^d f$ that are larger than the two neighbors whose position are in the direction indicated by the angle value of $A_{2^j}^d f$, at the corresponding location.

7. Characterization of Image Edges

Sharp variations of two dimensional signals are often not isolated but belong to curves in the image plane. Along these curves, the image intensity can be singular in one direction while varying smoothly in the perpendicular direction. It is well known that such curves are more meaningful than edge points by themselves because they generally are the boundaries of the image structures. For discrete images, we reorganize the maxima representation into chains of local maxima to recover these edge curves. Like in one dimension, we then characterize the properties of edges from the modulus maxima evolution across scales.

At a scale 2^j, the wavelet modulus maxima detect the sharp variation points of $f * \theta_{2^j}(x,y)$. Some of these modulus maxima define smooth curves in the image plane, along which the profile of the image intensity varies smoothly. At any point along maxima curve, $\vec{\nabla}(f * \theta_{2^j})(x,y)$ is perpendicular to the tangent of the edge curve. We thus chain two adjacent local maxima if their respective position is perpendicular to the direction indicated by $A_{2^j} f\,(x,y)$. Since we want to recover edge curves along which the image profile varies smoothly, we only chain together maxima points where the modulus $M_{2^j} f\,(x,y)$ has close values. This chaining procedure defines an image representation that is a set of maxima chains. Image edges might correspond to very different types of sharp variations. Like in one dimension, we discriminate different types of singularities by measuring their local Lipschitz regularity.

Definition 4

Let $0 < \alpha < 1$. A function $f(x)$ is said to be Lipschitz α at (x_0, y_0) if and only if there exists a constant K such that for all (x,y) in a neighborhood of (x_0, y_0),

$$|f(x,y) - f(x_0,y_0)| \leq K \, |(x-x_0)^2 + (y-y_0)^2|^{\alpha/2} \ . \tag{75}$$

The function $f(x)$ is uniformly Lipschitz α over an open set of $\mathbf{R}^2$ if and only if there exists a constant K such that equation (75) is satisfied for all (x,y) and (x_0,y_0) in this open set. The uniform Lipschitz regularity of $f(x)$ over this open set is the superior bound of all α such that $f(x)$ is uniformly Lipschitz α.

In two-dimensions, the Lipschitz regularity is characterized by the decay across scales of both $|W^1_{2^j}f(s,x)|$ and $|W^2_{2^j}f(s,x)|$. The decay of these two-components is bounded by the decay of $M_{2^j}f(x,y)$. Let us suppose that the two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ are continuously differentiable and that their decay at infinity is $O(\frac{1}{(1+x^2)\,(1+y^2)})$.

Theorem 2

Let $0 < \alpha < 1$. A function $f(x,y)$ is uniformly Lipschitz α over an open set of $\mathbf{R}^2$, if and only if, there exists a constant K such that for all points (x,y) of this open set

$$M_{2^j}f(x,y) \leq K\,(2^j)^{\alpha} \ . \tag{76}$$

This theorem is the two-dimensional extension of Theorem 1 and its proof is essentially the same [26]. The logarithm of equation (76) yields

$$\log_2(M_{2^j}f(x,y)) \leq \log_2(K) + \alpha\, j \ .$$

Uniform Lipschitz exponents can thus be measured from the evolution across scales of $\log_2(M_{2^j}f(x,y))$. This result enables us to discriminate different type of singularities.

When the signal variations are smooth, we can measure how smooth they are with the same approach as in one dimension. Locally, we model the smooth variation of $f(x,y)$ at (x_0,y_0), as the convolution of a function $h(x,y)$ that has a singularity at (x_0,y_0), with a two dimensional rotationally symmetric Gaussian of variance σ^2

$$f(x,y) = h * g_\sigma(x,y) \;\; \text{with} \;\; g_\sigma(x,y) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2}) \ . \tag{77}$$

We suppose that the uniform Lipschitz regularity of $h(x,y)$ in a neighborhood of (x_0,y_0) is α_0. If the two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ are the partial derivatives of a smoothing function $\theta(x,y)$ which closely approximates a rotationally symmetrical Gaussian, then we can estimate the variance σ^2. The wavelet transform modulus of $f(x,y)$ is defined at any scale s by

$$M_s f(x,y) = \sqrt{|W_s^1 f(x,y)|^2 + |W_s^2 f(x,y)|^2}\ . \tag{78}$$

With the same derivations as for equation (35), we prove that

$$M_{2^j} f(x,y) = \frac{2^j}{s_0} M_{s_0} h(x,y) \quad \text{with} \quad s_0 = \sqrt{2^{2j} + \sigma^2}\ . \tag{79}$$

Equation (76) of Theorem 2 is valid not only at dyadic scales 2^j, but at all scales $s > 0$. For $\alpha < \alpha_0$, $h(x,y)$ is uniformly Lipschitz α in a neighborhood of (x_0,y_0). Hence, there exists $K > 0$ such that for any points (x,y) in this neighborhood

$$M_{s_0} h(x,y) \leq K\, s_0^{\alpha}\ .$$

We thus derive from equation (79) that

$$M_{2^j} f(x,y) \leq K\, 2^j\, s_0^{\alpha-1} \quad \text{with} \quad s_0 = \sqrt{2^{2j} + \sigma^2}\ . \tag{80}$$

Along a maxima chain, the singularity type varies smoothly so the parameters K, α_0 and σ^2 do not change much. We thus estimate these values for portions of chains, by looking at the evolution of the modulus values across scales. Let us suppose that we have a portion of maxima chain that propagates between the scales 2^1 and 2^l. We also suppose that in a given neighborhood, at each scale 2^j, the value of $M_{2^j} f(x,y)$ is bounded by its values along this maxima chain. This means that the maxima chain corresponds to the sharpest image variation in the neighborhood. Since $M_{2^j} f(x,y)$ is bounded by the maxima values, we estimate the parameters α and σ that satisfy equation (80), from the evolution across scales of these modulus maxima values. In theory, this should be done by using the absolute maximum of $M_{2^j} f(x,y)$ along the maxima chain, at each scale 2^j. It is often better to regularize these computations by averaging the maxima modulus value along the corresponding portion of chain. This is justified since we suppose that the singularity type does not vary much along this portion of chain. Let a_j be the average value of $M_{2^j} f(x,y)$. Like in one dimension, we estimate the smoothing factor σ and the Lipschitz regularity α_0 by computing the values that minimize

$$\sum_{j=1}^{l} \left[\log_2 |a_j| - \log_2(K) - j - \frac{\alpha_0 - 1}{2} \log_2(\sigma^2 + 2^{2j}) \right]^2\ . \tag{81}$$

This algorithm associates to each portion of maxima chain, three constants K, α_0 and σ that describe the intensity profile of the image sharp variation, along the chain. Such a characterization of edge types is important for pattern recognition. For example, we can discriminate occlusions from shadows by looking whether the image intensity is discontinuous or is smoothly varying. For the circle image of Fig. 10, the wavelet transform modulus along the boundary remains constant across scales, which means that $\alpha_0 = 0$ and $\sigma = 0$. Indeed, the image intensity is discontinuous along the border and the constant K gives the amplitude of the discontinuity. In general, we believe that an edge detection should not be viewed as a binary process that labels the image pixels as edge points or non edge points but as a procedure that characterizes precisely the different types of image sharp variations.

8. Reconstruction of Images from Multiscale Edges

8.1. Reconstruction Algorithm

The algorithm that reconstructs images from the local maxima of their wavelet transform modulus is an extension of the one-dimensional algorithm described in Section 5.2. Let $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ and $\left[W^1_{2^j} f(x,y) \,,\, W^2_{2^j} f(x,y) \right]_{j \in \mathbf{Z}}$ be its dyadic wavelet transform. For each scale 2^j, we detect the local maxima of $M_{2^j} f(x,y)$ along the direction given by the angle image $A_{2^j} f(x,y)$. We record the positions of the modulus maxima, $\left[(x^j_v, y^j_v) \right]_{v \in \mathbf{R}}$, as well as $\left[M_{2^j} f(x^j_v, y^j_v) \,,\, A_{2^j} f(x^j_v, y^j_v) \right]_{v \in \mathbf{R}}$. In two dimensions, the number of modulus maxima is not countable anymore. From $M_{2^j} f(x^j_v, y^j_v)$ and $A_{2^j} f(x^j_v, y^j_v)$, we can compute $W^1_{2^j} f(x^j_v, y^j_v)$ and $W^2_{2^j} f(x^j_v, y^j_v)$, and vice-versa. The inverse problem consists in finding the set of functions $h(x,y)$ that satisfy the following two constraints.

(a) At each scale 2^j and for each modulus maxima location (x^j_v, y^j_v), we have $W^1_{2^j} h(x^j_v, y^j_v) = W^1_{2^j} f(x^j_v, y^j_v)$ and $W^2_{2^j} h(x^j_v, y^j_v) = W^2_{2^j} f(x^j_v, y^j_v)$.

(b) At each scale 2^j, the modulus maxima obtained from $W^1_{2^j} h(x,y)$ and $W^2_{2^j} h(x,y)$, are located at the abscissa $\left[(x^j_v, y^j_v) \right]_{v \in \mathbf{R}}$.

Let us analyze the property (a). At any point (x_0, y_0) the wavelet transform can be rewritten as inner products

$$W^1_{2^j} h(x_0, y_0) = \langle f(x,y) \,,\, \psi^1_{2^j}(x_0 - x, y_0 - y) \rangle \,, \tag{82}$$
$$W^2_{2^j} h(x_0, y_0) = \langle f(x,y) \,,\, \psi^2_{2^j}(x_0 - x, y_0 - y) \rangle \,.$$

Let $\mathbf{U}$ be the closure of the set of functions that are linear combinations of any function of the family

$$\left\{ 2^j \, \psi^1_{2^j}(x^j_v - x, y^j_v - y) \,,\, 2^j \, \psi^2_{2^j}(x^j_v - x, y^j_v - y) \right\}_{(j,v) \in \mathbf{Z} \times \mathbf{R}} \,. \tag{83}$$

The factor 2^j normalizes the $\mathbf{L}^2(\mathbf{R}^2)$ norm of each function. One can prove that the set of functions $h(x,y)$ whose wavelet transform satisfy the condition (a) are the functions whose orthogonal projection on $\mathbf{U}$ is equal to the orthogonal projection of $f(x,y)$ on $\mathbf{U}$. Let $\mathbf{O}$ be the orthogonal complement of $\mathbf{U}$ in $\mathbf{L}^2(\mathbf{R}^2)$. This set is therefore the affine space $f + \mathbf{O}$ of functions that can be written

$$h(x,y) = f(x,y) + g(x,y) \quad \text{with} \quad g(x,y) \in \mathbf{O} \ . \tag{84}$$

The condition (b) is not convex. We replace it by convex constraint which has a similar effect, in order to solve the problem numerically. We do not impose that the points $\left[(x_v^j, y_v^j)\right]_{(j,v)\in \mathbf{Z}\times\mathbf{R}}$ are the only modulus maxima of the wavelet transform, but minimize a Sobolev norm defined by

$$|||h|||^2 = \left| \left[W_{2^j}h^1(x,y), W_{2^j}h^2(x,y) \right]_{j\in\mathbf{Z}} \right|^2 =$$

$$\sum_{j=-\infty}^{+\infty} \left[\| W_{2^j}^1 h \|^2 + \| W_{2^j}^2 h \|^2 + 2^{2j} \left(\| \frac{\partial W_{2^j}^1 h}{\partial x} \|^2 + \| \frac{\partial W_{2^j}^2 h}{\partial y} \|^2 \right) \right] . \tag{85}$$

The minimization of this norm creates a wavelet transform whose horizontal and vertical components have an $\mathbf{L}^2(\mathbf{R}^2)$ norm as small as possible. In conjunction with the condition (a), this has a tendency to create modulus maxima at the positions (x_v^j, y_v^j). The partial derivative components are added in order to create a wavelet transform with as few spurious oscillations as possible. Since $W_{2^j}^1 h(x,y)$ is computed by smoothing the signal and taking the partial derivative along x, it oscillates mostly along the x direction and we use a partial derivative along x in (85) to minimize these oscillations. The transpose result is valid for $W_{2^j}^2 h(x,y)$. The weight on the derivative components is proportional to the scale 2^j because the smoothness $W_{2^j}^1 h(x,y)$ and $W_{2^j}^2 h(x,y)$ increases with the scale 2^j.

Let $\psi^3(x,y) = \dfrac{\partial \psi^1(x,y)}{\partial x}$ and $\psi^4(x,y) = \dfrac{\partial \psi^2(x,y)}{\partial y}$. If there exists two constants $A_5 > 0$ and $B_5 > 0$ such that for all $(\omega_x, \omega_y) \in \mathbf{R}^2$,

$$\begin{aligned} A_5 \le \sum_{j=-\infty}^{+\infty} &\left[|\hat{\psi}^1(2^j\omega_x, 2^j\omega_y)|^2 + |\hat{\psi}^2(2^j\omega_x, 2^j\omega_y)|^2 \right] + \\ \sum_{j=-\infty}^{+\infty} &\left[|\hat{\psi}^3(2^j\omega_x, 2^j\omega_y)|^2 + |\hat{\psi}^4(2^j\omega_x, 2^j\omega_y)|^2 \right] \le B_5 \end{aligned} \tag{86}$$

then for any function $h(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$, the norm defined in (85) is equivalent to the $\mathbf{L}^2(\mathbf{R}^2)$ norm

$$A_5 \, \|h\|^2 \le |||h|||^2 \le B_5 \, \|h\|^2 . \tag{87}$$

Similarly to equation (49), we prove this implication by applying the Parseval theorem on each $\mathbf{L}^2(\mathbf{R}^2)$ norm component of the norm defined in (85). We saw that the set of functions $h(x,y)$ whose wavelet transform satisfy the condition (a) is the closed affine space $f + \mathbf{O}$. This minimization of the norm $|||\ |||$ over this closed convex has a unique solution, whose computation might however not be stable. Like in one dimension, we can prove that the computation of this minimum is stable, if and only if, the family of functions

$$\left\{ 2^j\ \psi^1_{2^j}(x^j_\nu - x, y^j_\nu - y)\ ,\ 2^j\ \psi^2_{2^j}(x^j_\nu - x, y^j_\nu - y) \right\}_{(j,\nu)\in \mathbf{Z}\times\mathbf{R}} \quad . \tag{88}$$

is a frame of the space $\mathbf{U}$ that they generate. The factor 2^j normalizes the $\mathbf{L}^2(\mathbf{R}^2)$ norm of the functions in this family. The frame condition expresses the equivalence of the $\mathbf{L}^2(\mathbf{R}^2)$ norm of any function in $\mathbf{U}$ and the sum square of the inner products of this function with each function of the family (88).

To compute the solution of our minimization problem, we use an alternate projection algorithm, like in one dimension. Let $\mathbf{K}$ the space of all sequences of function $\left[g^1_j(x,y), g^2_j(x,y) \right]_{j\in \mathbf{Z}}$ such that

$$\left\| \left[g^1_j(x,y), g^2_j(x,y) \right]_{j\in \mathbf{Z}} \right\| < +\infty\ ,$$

where the norm $\| \ \|$ is defined by the expression (85). We define the set Γ of all sequences of functions $\left[g^1_j(x,y), g^2_j(x,y) \right]_{j\in \mathbf{Z}} \in \mathbf{K}$, such that for any index j and all maxima position (x^j_ν, y^j_ν)

$$g^1_j(x^j_\nu, y^j_\nu) = W^1_{2^j} f\,(x^j_\nu, y^j_\nu) \ \text{ and } \ g^2_j(x^j_\nu, y^j_\nu) = W^2_{2^j} f\,(x^j_\nu, y^j_\nu)\ . \tag{89}$$

The set Γ is an affine space which is closed in $\mathbf{K}$. Let $\mathbf{V}$ be the space of dyadic wavelet transforms of all functions in $\mathbf{L}^2(\mathbf{R}^2)$. Equation (87) proves that $\mathbf{V} \subset \mathbf{K}$. The sequence of functions that satisfy the condition (a) are the dyadic wavelet transforms that belong to Γ. These are the elements of $\mathbf{K}$ that belong to

$$\Lambda = \mathbf{V} \cap \Gamma\,.$$

To reconstruct the element of $\Gamma \cap \mathbf{V}$ that minimizes the norm $\| \ \|$, we alternate projections on Γ and $\mathbf{V}$, that are orthogonal with respect to the norm $\| \ \|$. Like in one dimension, one can prove that the orthogonal projection on $\mathbf{V}$ is the operator

$$\mathbf{P_V} = \mathbf{W} \circ \mathbf{W}^{-1}\ ,$$

that was defined in equation (64). The orthogonal projections $\mathbf{P}_\Gamma$ on Γ is defined in Appendix 5. For a discrete image of N^2 pixels, the implementations of both $\mathbf{P_V}$ and $\mathbf{P}_\Gamma$ require $O(N^2\log_2(N))$ operations. Let $\mathbf{P} = \mathbf{P_V} \circ \mathbf{P}_\Gamma$ be the alternate projection on both sets. Since Γ is an affine space and $\mathbf{V}$ a vector space, it has been proved [32] that for any initial sequence $X = \left[g^1_j(x,y), g^2_j(x,y) \right]_{j\in \mathbf{Z}}$, $\lim_{n\to+\infty} \mathbf{P}^{(n)}X$ converges strongly to the orthogonal projection of X onto

$\Lambda = \Gamma \cap \mathbf{V}$. Hence, if we begin the iteration from the zero element of $\mathbf{K}$, the algorithm converges strongly to the element of Λ whose norm $\| \; \|$ is minimum.

8.2. Numerical Reconstruction of Images from Multiscale Edges

We study the error of the reconstruction algorithm as a function of the number of iterations on the operator **P**. At each scale 2^j, the SNR integrates the error on the horizontal and the vertical components of the wavelet transform. Fig. 12(a) gives the evolution of the SNR when reconstructing the wavelet transform of the Lena image, from the modulus maxima shown in the Fig. 11. After n iterations, we reconstruct an image by applying the inverse wavelet transform operator on the reconstructed wavelet transform. Fig. 12(b) is the SNR of the reconstructed images, computed with respect to the original one, as a function of the number of iterations on the operator **P**. The graphs of Fig. 12 are very similar to the graphs of Fig. 8 that show the reconstruction SNR for a one-dimensional signal. The increase is fast during the first 20 iterations and then slows down. After a given number of iterations, Fig. 12(a) shows that the error is mostly concentrated at fine scales. This error has two components. The first one is the distance to the wavelet transform we converge to and the other one is the distance between the point we converge to and the wavelet transform of the original image. Like in one dimension, the convergence is exponential but the convergence rate is very slow. After 20 iterations, the distance between the reconstructed image and the image we converge to is of the same order as the distance between the original image and the image we converge to. Increasing the number of iterations thus do not increase much the SNR. The top right image in Fig. 13 is reconstructed with 10 iterations. The SNR is 28 db. The reconstructed image has no visual difference with the original image shown at the top left of Fig. 13 which means that the errors are below our visual sensitivity. Qualitatively the original image is well reproduced because the reconstruction has no spurious oscillation, the important singularities are well reproduced and the errors are mostly concentrated at fine scales where our visual sensitivity is not so acute.

The reconstruction algorithm has been tested for a large collection of images including special two-dimensional functions such as Diracs, sinusoidal waves, step edges, Brownian noises... For all these experiments, the SNR behaves similarly to Fig. 12. The visual quality of reconstructed images, with 10 iterations, is as good as in Fig. 13. For image processing applications, the numerical precision of this reconstruction algorithm is sufficient, even if we limit the number of iterations below 10. Since each iterations requires $O(N^2 \log_2(N))$ computations, this reconstruction can be implemented in hardware for real time applications. The reconstruction algorithm is stable for precisions of the order of 30 db. We can therefore slightly perturbate the wavelet transform modulus maxima and reconstruct a close image. The lower left image in Fig. 13 is reconstructed from the modulus maxima shown in the third column of Fig. 11. By thresholding the wavelet transform modulus maxima based on their modulus values, we suppressed the modulus maxima produced by the image noise and the light textures. As expected, these textures have disappeared in the reconstructed image but the sharp variations are not affected. In the lady's shoulder, the thresholding removes the maxima created by the image noise and the reconstructed image reproduces a skin quality which is much smoother, while the boundaries of the shoulder are kept sharp. This thresholding can be viewed as a non-linear noise removal technique. Hwang and one of us [23] have developed a more sophisticated procedure to suppress white noises from images, which removes the maxima produced by the noise through an analysis of their behavior across scales.

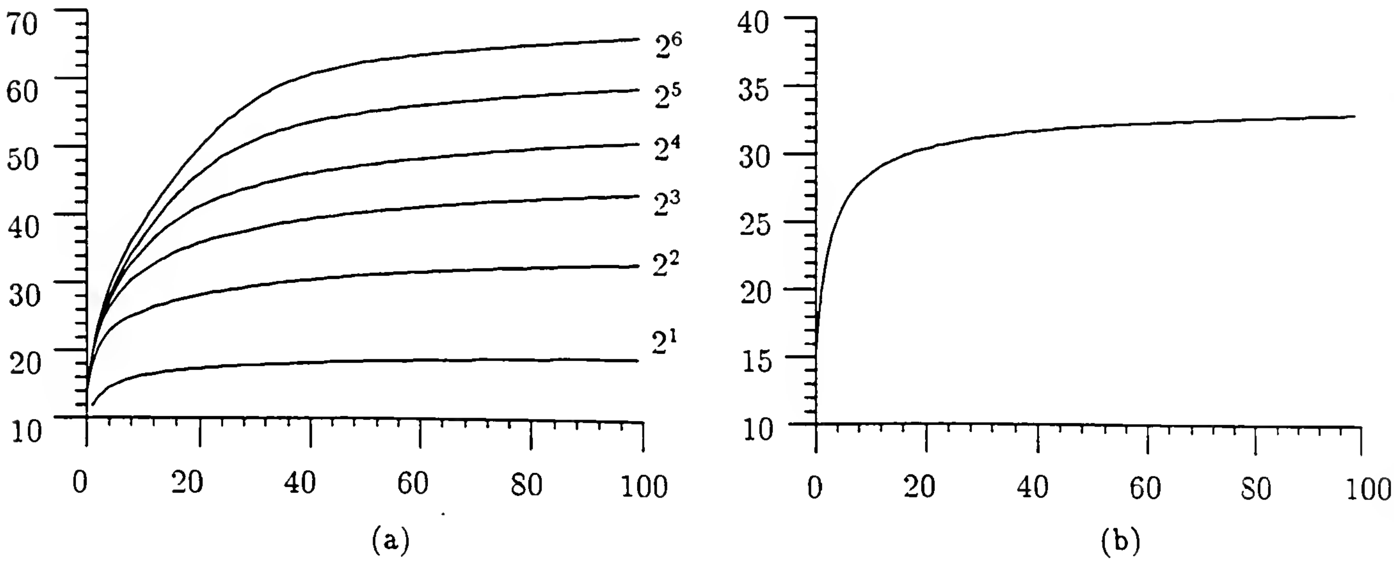


Fig. 12: *(a): Signal to Noise Ratio when reconstructing the wavelet transform components* $W_{2^j}^{1\,d}f$ *and* $W_{2^j}^{2\,d}f$ *from the modulus maxima of the Lena image, shown in Fig. 11. The abscissa gives the number of iterations on the operator* P. *Each curve is labeled by the scale* 2^j, *for* $1 \le j \le 6$.
(b): SNR of the reconstructed Lena images computed with respect to the original image, as a function of the number of iterations on the operator P.

Fig. 13: *Top left: original image. Top right: image reconstructed from the maxima representation shown in the second column of Fig. 11. This reconstruction is performed with 10 iterations and the SNR is 28 db. Lower left: image reconstructed from the thresholded modulus maxima shown in the third column of Fig. 11, with 10 iterations on the operator* **P**.

9. Compact Image Coding from Multiscale Edges

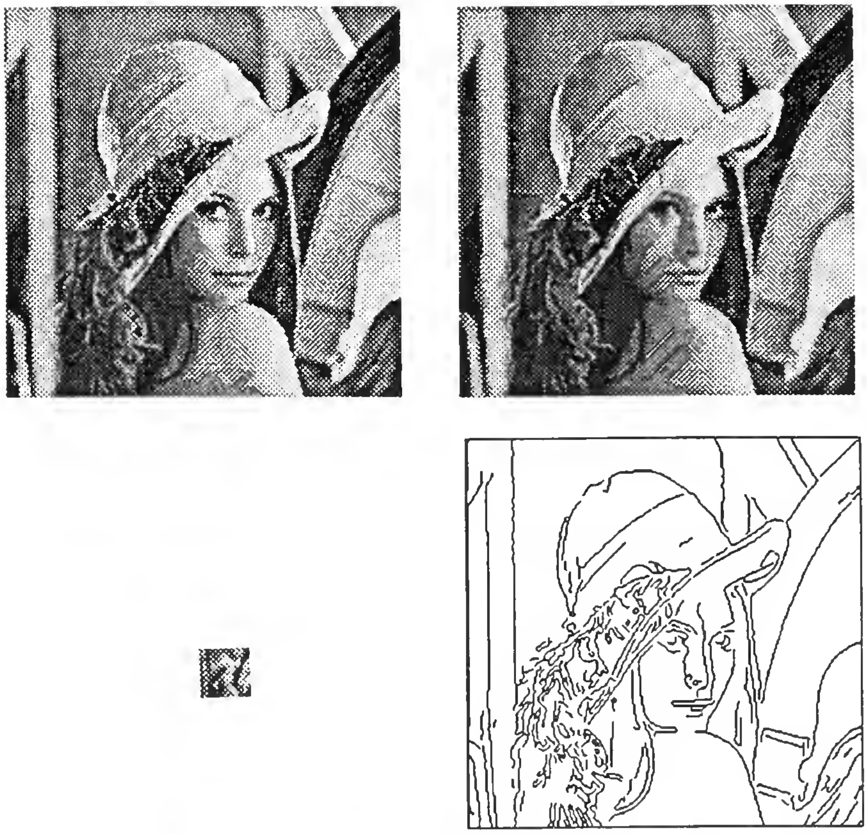
An important problem in image processing is to code images with a minimum number of bits for transmission or storage. To obtain high compression rates in image coding, we cannot afford to code all the information available in the image. It is necessary to remove part of the image components that are not important for the visualization. A major problem is to identify the "important" information that we need to keep. From this point of view, the problems encountered in compact image coding are similar to computer vision tasks, where one also wants to extract the "important" information for recognition purposes. Since edges provide meaningful features for image interpretation, it is natural to represent the image information with an edge based representation, in order to select the information to be coded. Previous edge coding algorithms have already been developed by Carlsson [4] and Kunt et. al. [18], but at a single scale. This section describes a compact coding algorithm based on the wavelet transform modulus maxima. The coding algorithm involves two steps. First we select the edge points that we consider important for the visual image quality. This preprocessing is identical to the feature extraction stage of a pattern recognition algorithm. We then make an efficient coding of this edge information. To select the "most important" edge curves can require sophisticated algorithms, if we take into account the image context. For example, in the Lena image, it is important to introduce no distortion around the eyes because these are highly visible for a human observer. In the following, we do not introduce such context information in the selection.

To code efficiently the edge information, we need to take advantage of the similarities between edges obtained at different scales. As it can be observed in Fig. 11, the edges of the main image structures have similar positions at the three finer scales 2^1, 2^2 and 2^3. These three finer scales also carry more than 90% of the image frequency band-width and thus covers most of the image information. We build our edge encoding from these scales only. The coarse scale information, corresponding to the wavelet transform at scales $2^j > 2^3$, is kept as a low frequency image $S^d_{2^3} f$ defined in Section 6.2. The edge selection is first performed at the scale 2^2 because at the finer scale 2^1 edges are too much contaminated by high frequency noises. The boundaries of the important coherent structures often generate long edge curves. We thus remove any edge curve whose length is smaller than a given length threshold. Among the remaining curves, we select the ones that correspond to the sharpest image variations. This is done by removing the edge curves along which the average value of the wavelet transform modulus is smaller than a given amplitude threshold.

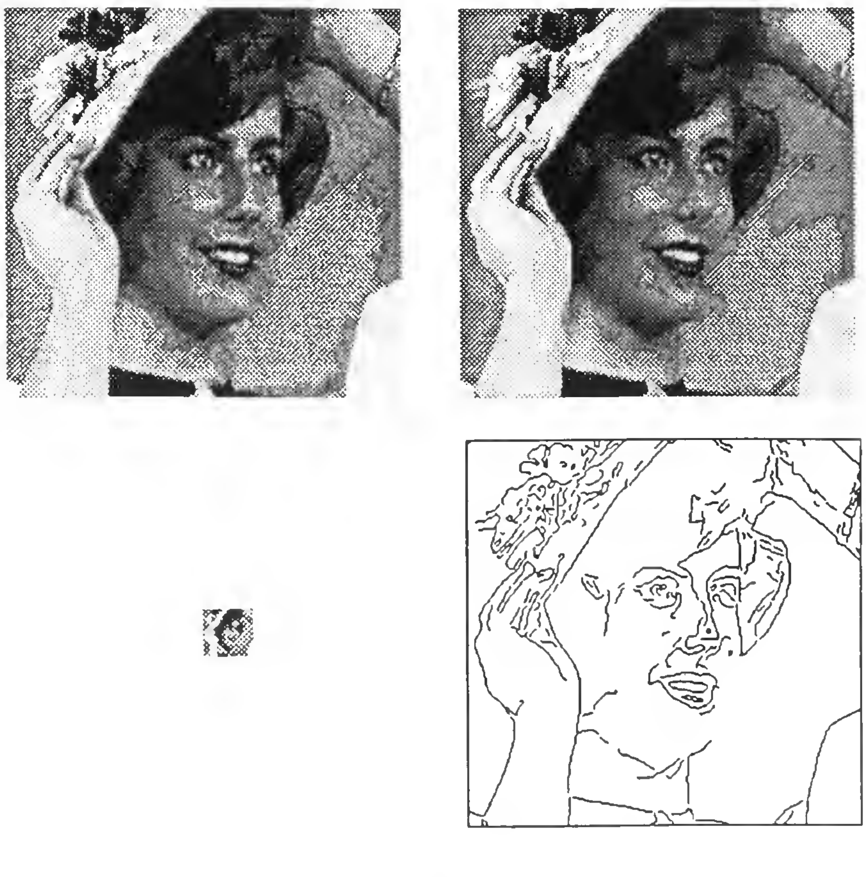
Once the selection is done, we must code efficiently the remaining information. This requires coding the position, modulus value and angle value of each modulus maximum along the maxima curves, at the scales 2^1, 2^2 and 2^3, plus the low frequency image $S^d_{2^3} f$. The geometry of the edge curves is coded only at the scale 2^2 because we consider that the maxima positions are the same at the scales 2^1 and 2^3. Maxima chains are coded by recording the position of the first point of each chain, and then coding the increment between the position of one edge point to the next one, along the chain. Carlsson [4] showed that this requires on average 10 bits for the first point and 1.3 bits per point along the chain, with an entropy coding. At each scale, the direction of the gradient image intensity at the edge locations is approximately orthogonal to the tangent of the edge curves. We thus do not code the angle values, but approximate each of them by the orthogonal direction of the edge tangent at the corresponding location. The values of the modulus along the edge curves at the scales 2^1, 2^2 and 2^3, are recorded with a simple predictive coding using a coarse quantization of the prediction values. In the frequency domain, the image $S^d_{2^3}f$ has an energy mostly concentrated in a domain that is $(2^3)^2$ times smaller than the frequency support of the original image. We thus subsample this image along its rows and columns by a factor 2^3, and its grey level values are coded on 6 bits.

We give in Fig. 14(a), 14(b) and 14(c) three examples of images coded with this algorithm. The same length and amplitude thresholds were used for each of these images to select the edge chains at the scale 2^2. For each example, we display at the top left the original image, at the top right the image reconstructed from the coded representation, at the bottom right the edge map at the scale 2^2 that is encoded and at the bottom left the subsampling of the low-frequency image $S^d_{2^3}f$. Each original image has 256 by 256 pixels. The total amount of data to code the reconstructed images are: 0.30 bits per pixels for Fig. 14(a), 0.24 bits per pixel for Fig. 14(b) and 0.19 bits per pixel for Fig. 14(c). The compression rate varies with the number of edge points that remain after the selection operation. This type of coding removes the image textures, however it does not produce distortions such as Gibbs phenomena. For the Lena image, errors are particularly visible around the eyes because too much edge points have been removed in this region by our simple selection algorithm. Although a lot of details have been removed in the coded images, they remain sharp and most of the information is kept.

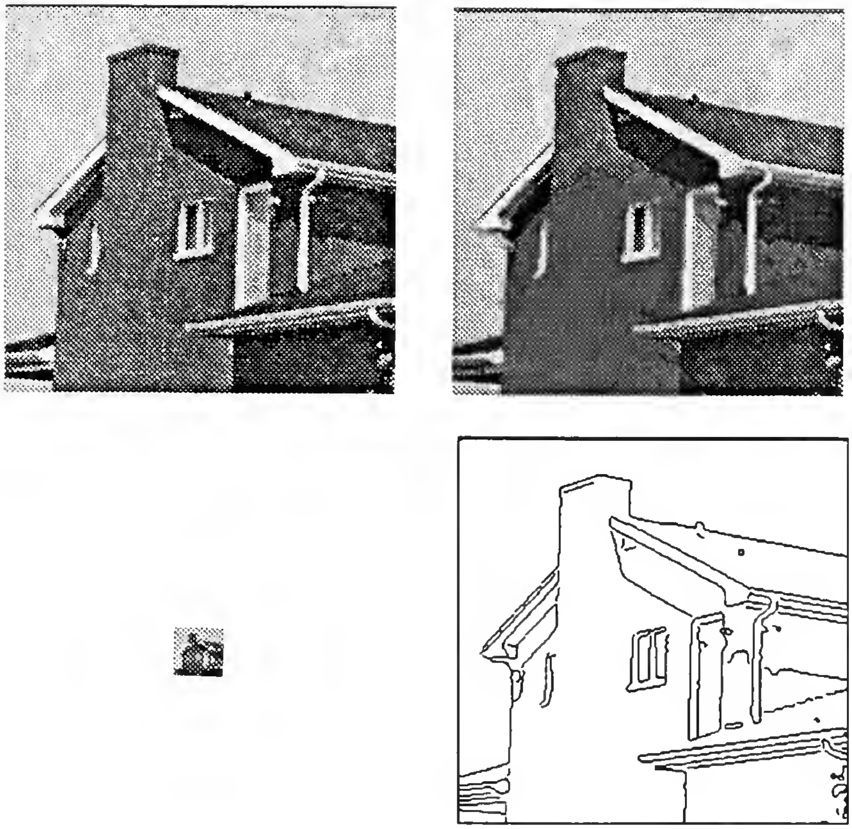
This compact coding algorithm is a feasibility study and it can certainly be improved both at the selection and the coding stages. For applications to images where textures are important, Froment and one of us [11] have extended the method by developing an algorithm that makes a specific coding of textures after this edge based coding. Distortions of textures are generally much less visible than distortion of edges and a separate coding of these two types of features can be adapted to the specificity of the visual perception. High compression rates are also obtained without a complete removal of textures contrarily to the coded images in Fig. 14.



(a)



(b)



(c)

Fig. 14: *Top left: original image of 256 by 256 pixels. Top right: reconstructed image from the coded multiscale edge representation. Image (a) requires 0.30 bits per pixel, image (b) 0.24 bits per pixel and image (c) 0.19 bits per pixel. Bottom left: image* $S^d_{2^3}f$ *subsampled by a factor* 2^3 *along its rows and columns. Bottom right: position of the modulus maxima selected and encoded at the scales* 2^1, 2^2 *and* 2^3.

10. Conclusion

We showed that multiscale edges can be detected and characterized from the local maxima of a wavelet transform. One can estimate the Lipschitz regularity as well as the smoothing component of sharp variation points from the evolution of the wavelet maxima across scales. We believe that this complement of information is important for pattern recognition algorithms based on edges.

The reconstruction algorithms that are described in one and two dimensions recover a close approximation of the original signals. For images, the reconstruction errors are below our visual sensitivity and can thus be neglected in image processing or computer vision applications. To reconstruct such signals requires few iterations which can be implemented in real time on a pipeline hardware architecture. The stability and approximation range of the reconstruction algorithm are open mathematical problems. As an application, we described a compact image coding procedure that selects the important visual information before coding. The compression rates are between 30 and 40 in the examples that are shown, but most of the light image textures are not coded. A double layer coding based on multiscale edges and textures has recently been developed by Froment and one of us [11].

Appendix 1

A Particular Class of One-Dimensional Wavelets

This appendix defines a class of wavelets that can be used for a fast implementation of discrete algorithms. To compute the wavelet transform with cascade of discrete convolutions, we define a wavelet $\psi(x)$ which can be factorized into convolutions of discrete filters. The same idea was used to build orthogonal wavelets [21]. The fast discrete wavelet transform algorithm is described in Appendix 2. We first define the smoothing function $\phi(x)$ introduced in Section 3.2 then we build the wavelet $\psi(x)$ and the reconstructing function $\chi(x)$ that satisfies equation (12).

We impose that the Fourier transform of the smoothing function $\phi(x)$ defined by equation (21) can be written as an infinite product

$$\hat{\phi}(\omega) = e^{-iw\omega} \prod_{p=1}^{+\infty} H(2^{-p}\omega) \ , \tag{90}$$

where $H(\omega)$ is a 2π periodic function such that

$$|H(\omega)|^2 + |H(\omega+\pi)|^2 \leq 1 \quad \text{and} \quad |H(0)| = 1 \ . \tag{91}$$

One can prove [21] that equation (90) defines the Fourier transform of a function $\phi(x)$ which is in $\mathbf{L}^2(\mathbf{R})$. The parameter w is the sampling shift that was introduced in Section 3.2. It is adjusted in order to obtain a smoothing function $\phi(x)$ which is symmetrical with respect to 0. Equation (90) implies that

$$\hat{\phi}(2\omega) = e^{-iw\omega} H(\omega)\, \hat{\phi}(\omega) \ . \tag{92}$$

We define a wavelet $\psi(x)$ whose Fourier transform $\hat{\psi}(\omega)$ is given by

$$\hat{\psi}(2\omega) = e^{-iw\omega} G(\omega)\, \hat{\phi}(\omega) \ , \tag{93}$$

where $G(\omega)$ is a 2π periodic function. Let $\hat{\chi}(\omega)$ be a function that satisfies

$$\hat{\psi}(2\omega)\, \hat{\chi}(2\omega) = |\hat{\phi}(\omega)|^2 - |\hat{\phi}(2\omega)|^2 \tag{94}$$

One can prove that if $|\hat{\phi}(\omega)|$ converges to 0 at infinity and to 1 when ω goes to 0, then equation (12) is also valid which means that $\chi(x)$ is a reconstructing wavelet with respect to $\psi(x)$. If $H(\omega)$ is differentiable at $\omega = 0$ then $|\hat{\phi}(\omega)|$ does converge to 1 when ω goes to 0 [21]. Let us suppose that $\hat{\chi}(\omega)$ can be written

$$\hat{\chi}(2\omega) = e^{iw\omega} K(\omega)\, \hat{\phi}(\omega) \ , \tag{95}$$

where $K(\omega)$ is a 2π periodic function. Substituting equations (92), (93) and (95) in equation (94) yields

$$|H(\omega)|^2 + G(\omega)\, K(\omega) = 1 \ . \tag{96}$$

This equation on $K(\omega)$ is necessary and sufficient so that the function $\hat{\chi}(\omega)$ defined by equation (95) is a reconstructing wavelet.

We want a wavelet $\psi(x)$ equal to the first order derivative of a smoothing function $\theta(x)$. This implies that $\hat{\psi}(\omega)$ must have a zero of order 1 at $\omega = 0$. Since $|\hat{\phi}(0)| = 1$, equation (93) yields that $G(\omega)$ must have a zero of order 1 at $\omega = 0$. We choose $H(\omega)$ in order to obtain a wavelet $\psi(x)$ which is antisymmetrical, as regular as possible and with a small compact support. A family of 2π periodic functions $H(\omega)$, $G(\omega)$ and $K(\omega)$ that satisfy these constraints is given by

$$H(\omega) = e^{i\omega/2}\left[\cos(\omega/2)\right]^{2n+1} , \tag{97}$$

$$G(\omega) = 4\, i\, e^{i\omega/2} \sin(\omega/2) , \tag{98}$$

$$K(\omega) = \frac{1 - |H(\omega)|^2}{G(\omega)} . \tag{99}$$

From equation (90) and (93) we derive

$$\hat{\phi}(\omega) = \left[\frac{\sin(\omega/2)}{\omega/2}\right]^{2n+1} , \tag{100}$$

$$\hat{\psi}(\omega) = i\omega \left[\frac{\sin(\omega/4)}{\omega/4}\right]^{2n+2} . \tag{101}$$

The Fourier transform $\hat{\theta}(\omega)$ of the primitive is therefore

$$\hat{\theta}(\omega) = \left[\frac{\sin(\omega/4)}{\omega/4}\right]^{2n+2} . \tag{102}$$

In the example of Fig. 3, we chose $2n+1 = 3$. In order to have a wavelet antisymmetrical with respect to 0 and $\phi(x)$ symmetrical with respect to 0, the shifting constant w of equation (92) is equal to $1/2$. Equations (101) proves that $\psi(x)$ is a quadratic spline with compact support where as $\theta(x)$ is a cubic spline whose integral is equal to 1. The 2π periodic function $H(\omega)$, $G(\omega)$ and $K(\omega)$ can be viewed as the transfer function of discrete filters with finite impulse response. The corresponding impulse responses are given by Table 1. These filters are used in fast wavelet transform computations.

n	H	G	K
-3			0.0078125
-2			0.054685
-1	0.125		0.171875
0	0.375	-2.0	-0.171875
1	0.375	2.0	-0.054685
2	0.125		-0.0078125

Table 1: *Finite impulse response of the filters H, G and K corresponding to the quadratic spline wavelet of Fig. 3(a).*

Appendix 2

Fast Wavelet Algorithms for One-Dimensional Signals

This appendix describes an algorithm for computing a discrete wavelet transform and the inverse algorithm that reconstructs the original signal from its wavelet transform. We suppose that the wavelet $\psi(x)$ is characterized by the three discrete filters H, G and K described in Appendix 1. We denote H_p, G_p and K_p the discrete filters obtained by putting $2^p - 1$ zeros between each coefficients of the filters H, G and K. The transfer function of these filters is respectively $H(2^p\omega)$, $G(2^p\omega)$ and $K(2^p\omega)$. We also denote by $\tilde{H}_p$ the filter whose transfer function is the complex conjugates of $H(2^p\omega)$: $\overline{H(2^p\omega)}$. We denote by $A * B$ the convolution of two discrete signals A and B.

The following algorithm computes the discrete wavelet transform of the discrete signal $S_1^d f$. At each scale 2^j, it decomposes $S_{2^j}^d f$ into $S_{2^{j+1}}^d f$ and $W_{2^{j+1}}^d f$.

j = 0

while (j < J)

$$W_{2^{j+1}}^d f = \frac{1}{\lambda_j} \,.\, S_{2^j}^d f * G_j$$

$$S_{2^{j+1}}^d f = S_{2^j}^d f * H_j$$

$j = j + 1$

end of while

The proof of this algorithm is based on the properties of the wavelet $\psi(x)$ described in Appendix 1. At each scale 2^j, we divide the values of the samples of $S^d_{2^j}f * G_j$ by λ_j to obtain accurate measures of Lipschitz exponents from the wavelet maxima. Due to discretization, the wavelet modulus maxima of a step edge do not have the same amplitude at all scales, as they should in a continuous model. The constants λ_j guarantees that values of the maxima modulus remain constant at all scales, for a step edge. This enhance the accuracy of the measurement of the Lipschitz regularity for all types of singularities. The values of λ_j are given in Table 2 for the quadratic spline wavelet. The border problems are treated by making symmetry and a periodization of $\left(S_1 f(n)\right)_{1 \le n \le N}$, as explained in Section 3.2. The convolutions must take into account this periodization. The complexity of this discrete wavelet transform algorithm is $O(N\ log(N))$, and the complexity constant is proportional to the number of non-zero coefficients in the impulse response of the filters H and G.

The inverse wavelet transform algorithm reconstructs $S^d_1 f$ from the discrete dyadic wavelet transform. At each scale 2^j, it reconstructs $S^d_{2^{j-1}}f$ from $S^d_{2^j}f$ and $W^d_{2^j}f$. The complexity of this reconstruction algorithm is also $O(N\ log(N))$.

j = J
while (j > 0)
$\quad S^d_{2^{j-1}}f = \lambda_j \,.\, W^d_{2^j}f * K_{j-1} + S^d_{2^j}f * \bar{H}_{j-1}$
$\quad$ *j = j - 1*
end of while

j	λ_j
1	1.50
2	1.12
3	1.03
4	1.01
5	1.00

Table 2: *Normalization coefficients λ_j for the quadratic wavelet of Fig. 3(a). For $j > 5$, $\lambda_j = 1$.*

Appendix 3

A Particular Class of Two-Dimensional Dyadic Wavelets

In this appendix we characterize the two-dimensional wavelets used for numerical computations. The two wavelets

$$\psi^1(x,y) = \psi(x)\,\xi(y) \quad \text{and} \quad \psi^2(x,y) = \xi(x)\,\psi(y) \tag{103}$$

are defined from a one-dimensional wavelet $\psi(x)$ that belongs to the class described in Appendix 1 and whose Fourier transform is defined by

$$\hat{\psi}(2\omega) = e^{-iw\omega}\, G(\omega)\, \hat{\phi}_0(\omega) \quad \text{with} \quad \hat{\phi}_0(\omega) = e^{-iw\omega} \prod_{p=1}^{+\infty} H(2^{-p}\omega) \ . \tag{104}$$

We impose that the smoothing function $\phi(x,y)$ of Section 6.2 can be written

$$\phi(x,y) = \phi_0(x)\,\phi_0(y) \ ,$$

where the Fourier transform of $\phi_0(x)$ is given by equation (104). Let us now find two functions $\hat{\chi}^1(\omega_x,\omega_y)$ and $\hat{\chi}^2(\omega_x,\omega_y)$ such that

$$\begin{aligned} &|\hat{\phi}(\omega_x,\omega_y)|^2 - |\hat{\phi}(2\omega_x,2\omega_y)|^2 = \\ &\hat{\psi}^1(2\omega_x,2\omega_y)\,\hat{\chi}^1(2\omega_x,2\omega_y) + \hat{\psi}^2(2\omega_x,2\omega_y)\,\hat{\chi}^2(2\omega_x,2\omega_y) \ . \end{aligned} \tag{105}$$

One can prove that if $|\hat{\phi}(\omega_x,\omega_y)|$ converges to 0 at infinity and to 1 when (ω_x,ω_y) converges to (0,0) then the functions $\chi^1(\omega_x,\omega_y)$ and $\chi^2(\omega_x,\omega_y)$ satisfy the admissibility condition (61) of reconstructing wavelets. Let us impose that the smoothing function $\xi(x)$ that defines the two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ in equation (103) is $\xi(x) = 2\,\phi_0(2x)$. Equations (103) and (104) imply that the Fourier transform of the two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ are given by

$$\begin{aligned} &\psi^1(2\omega_x,2\omega_y) = e^{-iw\omega_x}\, G(\omega_x)\, \hat{\phi}_0(\omega_x)\, \hat{\phi}_0(\omega_y) \ \text{and} \\ &\psi^2(2\omega_x,2\omega_y) = \hat{\phi}_0(\omega_x)\, e^{-iw\omega_y}\, G(\omega_y)\, \hat{\phi}_0(\omega_y) \ . \end{aligned} \tag{106}$$

The functions $\hat{\chi}^1(\omega_x,\omega_y)$ and $\hat{\chi}^2(\omega_x,\omega_y)$ that satisfy equation (105) can be written

$$\hat{\chi}^1(2\omega_x,2\omega_y) = e^{iw\omega_x}\, K(\omega_x)\, L(\omega_y)\, \hat{\phi}(\omega_x)\, \hat{\phi}(\omega_y) \ ,$$

$$\hat{\chi}^2(2\omega_x,2\omega_y) = e^{iw\omega_y}\, K(\omega_y)\, L(\omega_x)\, \hat{\phi}(\omega_x)\, \hat{\phi}(\omega_y) \ .$$

The fonctions $K(\omega)$ and $L(\omega)$ are 2π periodic and satisfy

$$G(\omega)\,K(\omega) + |H(\omega)|^2 = 1 \ , \tag{107}$$

$$L(\omega) = \frac{1 + |H(\omega)|^2}{2} \ . \tag{108}$$

If we could choose $G(\omega) = i\omega$, then $\psi^1(x,y)$ and $\psi^2(x,y)$ would be the partial derivatives along x and y of $\theta(x,y) = 4\ \phi(2x)\ \phi(2y)$. For discrete implementations, we must choose a function $G(\omega)$ which is 2π periodic. Like in Appendix 1, we choose $G(\omega) = 4\ i\ e^{i\omega/2}\ \sin(\omega/2)$ to approximate a derivative. The two-dimensional wavelets used in the computations of this article are derived from the one-dimensional quadratic spline wavelet shown in Fig. 3. The values of the discrete filters H, G and K are given by Table 1. Table 3 gives the finite impulse response corresponding to the transfer function $L(\omega)$.

n	L
-3	0.0078125
-2	0.046875
-1	0.1171875
0	0.65625
1	0.1171875
2	0.046875
3	0.0078125

Table 3: *Finite impulse response of the filter L.*

Appendix 4

Fast Wavelet Algorithms for Two-Dimensional Signals

We describe two fast algorithms to implement the wavelet transform and the inverse wavelet transform in two dimensions. We suppose that the two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ are characterized by the three discrete filters H , G, K and L mentioned in Appendix 3. We use the same notations as in Appendix 2 and L_p is the discrete filter obtained by putting $2^p - 1$ zeros between consecutive coefficients of the filter L. We also denote by D the Dirac filter whose impulse response is equal to 1 at 0 and 0 otherwise. We denote by $A * (H,L)$ the separable convolution respectively of the rows and columns of the image A with the one-dimensional filters H and L.

The following algorithm computes the two-dimensional discrete wavelet transform of an image $S_1^d f$. At each scale 2^j, the algorithm decomposes $S_{2^j}^d f$ into $S_{2^{j+1}}^d f$, $W_{2^{j+1}}^{1,d} f$ and $W_{2^{j+1}}^{2,d} f$.

j = 0
while (j > J)

$$W^{1\,d}_{2^{j+1}} f = \frac{1}{\lambda_j} \,.\, S^d_{2^j} f * (G_j, D)$$

$$W^{2\,d}_{2^{j+1}} f = \frac{1}{\lambda_j} \,.\, S^d_{2^j} f * (D, G_j)$$

$$S^d_{2^{j+1}} f = S^d_{2^j} f * (H_j, H_j)$$

j = j + 1
end of while

The proof of this algorithm is based on the properties of the wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ described in Appendix 3. At each scale 2^j, the division by λ_j has the same normalization purpose as in Appendix 2. The values of λ_j is given in Table 2. If the original image $\left[S_1 f(n,m) \right]_{(n,m) \in Z^2}$ has N^2 non-zero pixels, the complexity of the algorithm is $O(N^2 \log(N))$. As explained in Section 6.2, border problems are solved by making a symmetry of the image with respect to each of its borders and a periodization. The separable convolutions must take into account this border procedure.

Like in the one-dimensional case, the reconstruction algorithm computes $S^d_1 f$ by reconstructing at each scale 2^j the signal $S^d_{2^{j-1}} f$ from $S^d_{2^j} f$, $W^{1\,d}_{2^j} f$ and $W^{2\,d}_{2^j} f$. The complexity of this reconstruction algorithm is also $O(N^2 \log(N))$.

j = J
while (j > 0)

$$S^d_{2^{j-1}} f = \lambda_j \,.\, W^{1\,d}_{2^j} f * (K_{j-1}, L_{j-1}) + \lambda_j \,.\, W^{2\,d}_{2^j} f * (L_{j-1}, K_{j-1}) + S^d_{2^j} f * (\tilde{H}_{j-1}, \tilde{H}_{j-1})$$

j = j - 1
end of while

Appendix 5
Projection Operator on Γ

In this appendix, we characterize the orthogonal projection on Γ in one and two dimensions and explain how to suppress oscillations for one-dimensional reconstructions. We first study the one-dimensional case. The operator $\mathbf{P}_\Gamma$ transforms any sequence $\left[g_j(x) \right]_{j \in Z} \in \mathbf{K}$ into the closest sequence $\left[h_j(x) \right]_{j \in Z} \in \Gamma$ with respect to the norm $\| \; \|$. Let $\varepsilon_j(x) = h_j(x) - g_j(x)$.

Each function $h_j(x)$ is chosen so that

$$\sum_{j=-\infty}^{+\infty} \| \varepsilon_j \|^2 + 2^{2j} \| \frac{d\varepsilon_j}{dx} \|^2 \qquad (109)$$

is minimum. To minimize this sum, we minimize separately each component

$$\| \varepsilon_j \|^2 + 2^{2j} \| \frac{d\varepsilon_j}{dx} \|^2 .$$

Let x_0 and x_1 be the abscissa of two consecutive modulus maxima of $Wf_{2^j}(x)$. Since $\left[h_j(x) \right]_{j \in \mathbf{Z}} \in \Gamma$, we have

$$\begin{cases} \varepsilon_j(x_0) = W_{2^j} f(x_0) - g_j(x_0) \\ \varepsilon_j(x_1) = W_{2^j} f(x_1) - g_j(x_1) . \end{cases} \qquad (110)$$

Between the abscissa x_0 and x_1, the minimization of (109) is equivalent to the minimization of

$$\int_{x_0}^{x_1} \left[|\varepsilon_j(x)|^2 + 2^{2j} \left| \frac{d\varepsilon_j(x)}{dx} \right|^2 \right] dx . \qquad (111)$$

The Euler equation associated with this minimization is

$$\varepsilon_j(x) - 2^{2j} \frac{d^2 \varepsilon_j(x)}{dx^2} = 0 , \qquad (112)$$

for $x \in]x_0, x_1[$. The constraints (110) are the border conditions of this membrane equation. The solution is

$$\varepsilon_j(x) = \alpha \, e^{2^{-j} x} + \beta \, e^{-2^{-j} x} , \qquad (113)$$

where the constants α and β are adjusted to satisfy equations (110).

In numerical computations, $W^d_{2^j} f$ is a uniform sampling of $W_{2^j} f(x)$ at the rate 1 and has a total of N samples. At each scale 2^j, the operator $\mathbf{P}_\Gamma$ modifies a discrete signal $g^d_j = \left[g_j(n) \right]_{1 \le n \le N}$ by adding a discrete signal $\varepsilon^d_j = \left[\varepsilon_j(n) \right]_{1 \le n \le N}$ that is computed from equation (113), between two consecutive modulus maxima. This requires $O(N)$ computations. Since there are at most $\log_2(N)+1$ scales, the total number of computations to implement $\mathbf{P}_\Gamma$ is $O(N \log_2(N))$.

We know that the modulus maxima of the original wavelet transform are only located at the positions x^j_n. We can thus also impose sign constraints in order to suppress any spurious oscillation in the reconstructed wavelet transform. This is done by imposing that the solution belongs to an appropriate convex set $\mathbf{Y}$. Let $sign(x)$ be the sign of the real number x. Let $\mathbf{Y}$ be the set of sequences $\left[g_j(x) \right]_{j \in \mathbf{Z}} \in \mathbf{K}$ such that for any pair of consecutive maxima positions (x^j_n, x^j_{n+1}) and

$x \in [x_n^j, x_{n+1}^j]$,

$$\begin{cases} sign(g_j(x)) = sign(x_n^j) \ \text{ if } \ sign(x_n^j) = sign(x_{n+1}^j) \\ sign(\frac{dg_j(x)}{dx}) = sign(x_{n+1}^j - x_n^j) \ \text{ if } \ sign(x_n^j) \neq sign(x_{n+1}^j) \end{cases}$$

The set Y is a closed convex and $\left(W_{2^j}f\right)_{j \in \mathbf{Z}} \in \mathrm{Y}$. Instead of minimizing $\| \ \|$ over $\Gamma \cap \mathrm{V}$ as explained in Section 5.2, we can minimize it over $\mathrm{Y} \cap \Gamma \cap \mathrm{V}$. We thus alternate projections on Y, Γ and V. To compute the orthogonal projection on the convex Y we need to solve an elastic membrane problem under constraints. This can be done with an iterative algorithm that is computationally intensive. Instead, we implement a simpler projector $\mathbf{P_Y}$ on Y which is not orthogonal with respect to the norm $\| \ \|$. Let $\left(g_j(x)\right)_{j \in \mathbf{Z}} \in \mathbf{K}$ and $\mathbf{P_Y}\left(g_j(x)\right)_{j \in \mathbf{Z}} = \left(h_j(x)\right)_{j \in \mathbf{Z}}$. For each index j, $h_j(x)$ is obtained by clipping the oscillations of $g_j(x)$ as illustrated in Fig. 15. If the original signal has N samples, at each scale 2^j the discrete implementation of this clipping procedure requires $O(N)$ computations. The total number of computations to implement $\mathbf{P_Y}$ is thus $O(N \log_2(N))$. Since this projector $\mathbf{P_Y}$ is not orthogonal, the iteration on the alternate projection operator $\mathbf{P} = \mathbf{P_V} \circ \mathbf{P_\Gamma} \circ \mathbf{P_Y}$ is not guaranteed to converge. Numerical experiments shows that in most cases, after a few iterations we stay inside Y, even after projections on Γ and V. Hence, the operator $\mathbf{P_Y}$ acts as the identity operator and $\mathbf{P}$ can be rewritten $\mathbf{P} = \mathbf{P_V} \circ \mathbf{P_\Gamma}$. The analysis of Section 5.2 proves that we are then guaranteed to converge strongly to an element in $\mathrm{Y} \cap \Gamma \cap \mathrm{V}$.

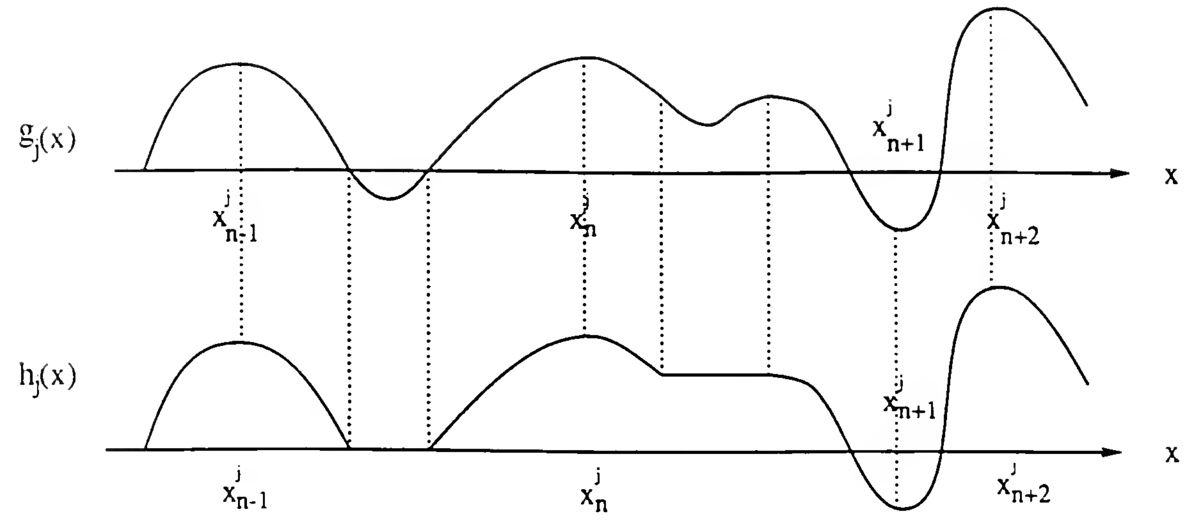


Fig. 15: *The projector* $\mathbf{P_Y}$ *is defined by clipping the function* $g_j(x)$ *so that the resulting function* $h_j(x)$ *has a constant sign between two consecutive modulus maxima or its derivative has a constant sign.*

In two dimensions, the operator $\mathbf{P_\Gamma}$ transforms a sequence $\left[g^1_j(x,y), g^2_j(x,y)\right]_{j\in\mathbf{Z}} \in \mathbf{K}$ into the closest sequence $\left[h^1_j(x,y), h^2_j(x,y)\right]_{j\in\mathbf{Z}} \in \Gamma$. Let $\left[\varepsilon^1_j(x,y), \varepsilon^2_j(x,y)\right]_{j\in\mathbf{Z}}$ be such that for any $j \in \mathbf{Z}$, $\varepsilon^1_j(x,y) = g^1_j(x,y) - h^1_j(x,y)$ and $\varepsilon^2_j(x,y) = g^2_j(x,y) - h^2_j(x,y)$. The sequence $\left[h^1_j(x,y), h^2_j(x,y)\right]_{j\in\mathbf{Z}}$ is chosen so that

$$\sum_{j=-\infty}^{+\infty}\left[\,\|\varepsilon^1_j\|^2 + \|\varepsilon^2_j\|^2 + 2^{2j}\left(\|\frac{\partial\varepsilon^1_j}{\partial x}\|^2 + \|\frac{\partial\varepsilon^2_j}{\partial y}\|^2\right)\right] \quad (114)$$

is minimum. The constraints on $\varepsilon^1_j(x,y)$ and $\varepsilon^2_j(x,y)$ are independent. The minimization of equation (114) is obtained by minimizing each component

$$\|\varepsilon^1_j\|^2 + 2^{2j}\,\|\frac{\partial\varepsilon^1_j}{\partial x}\|^2 \quad (115)$$

and

$$\|\varepsilon^2_j\|^2 + 2^{2j}\,\|\frac{\partial\varepsilon^2_j}{\partial y}\|^2 ,$$

for all integer $j \in \mathbf{Z}$. Let us concentrate on the minimization of equation (115). Let (x_0,y) and

(x_1,y) be two consecutive modulus maxima position at a fixed y. The function $\varepsilon_j^1(x,y)$ must satisfy

$$\begin{cases} \varepsilon_j^1(x_0,y) = W_{2^j}^1 f(x_0,y) - g_j^1(x_0,y) \\ \varepsilon_j^1(x_1,y) = W_{2^j}^1 f(x_1,y) - g_j^1(x_1,y) \end{cases} \quad (116)$$

The minimization of equation (115) subject to these constraints is obtained by minimizing

$$\int_{x_0}^{x_1} \left[|\varepsilon_j^1(x,y)|^2 + 2^{2j} \, |\frac{\partial \varepsilon_j^1(x,y)}{\partial x}|^2 \right] dx \ . \quad (117)$$

For y fixed, we obtain a one-dimensional minimization problem which is identical the minimization of the expression (111). The solution is a sum of two exponentials like in equation (113). This analysis shows that the solution of the two-dimensional minimization problem is obtained by fixing the parameter y for $\varepsilon_j^1(x,y)$ and computing the one-dimensional solution along the x variable, between two consecutive modulus maxima. The same analysis can be performed on the other component $\varepsilon_j^2(x,y)$. The discrete implementation is thus a straight forward extension of the one-dimensional algorithm that is applied along the rows and columns of the images that belong to the sequence that we project on Γ. One can verify that if the original image has N^2 pixels, the implementation of $\mathbf{P}_\Gamma$ requires $O(N^2 \, log(N))$ computations. In two dimensions we do not introduce any sign constraint as it is done in one-dimensional reconstructions.

Appendix 6

Convergence Rate of the Alternate Projection Algorithm

We prove that if $\left(\sqrt{2^j} \ \psi_{2^j}(x_n^j - x) \right)_{(n,j) \in \mathbf{Z}^2}$ is a frame of $\mathbf{U}$ then the alternate projection converges exponentially and we give a lower bound of the convergence rate. We first prove that there exists constant $C > 0$ such that for any element $X = \left(W_{2^j} g(x) \right)_{j \in \mathbf{Z}} \in \mathbf{V}$

$$\| X - \mathbf{P}_\Gamma X \|^2 \geq C \ \| X - \mathbf{P}_\Lambda X \|^2 \ . \quad (118)$$

Let $\varepsilon_j(x_n^j) = W_{2^j} f(x_n^j) - W_{2^j} g(x_n^j)$ be the error at each modulus maxima location. We first prove that there exists two constants $C_1 > 0$ and $C_2 > 0$ such that

$$\| X - \mathbf{P}_\Gamma X \|^2 \geq C_1 \sum_{(n,j) \in \mathbf{Z}^2} 2^j \, |\varepsilon_j(x_n^j)|^2 \ , \quad (119)$$

and

$$\| X - \mathrm{P}_{\Lambda} X \|^2 \le C_2 \sum_{(n,j) \in \mathbf{Z}^2} 2^j \, |\varepsilon_j(x_n^j)|^2 \ . \tag{120}$$

Let us begin with equation (119). Let $\left(h_j(x)\right)_{j \in \mathbf{Z}} = \mathrm{P}_\Gamma \left(W_{2^j} g(x)\right)_{j \in \mathbf{Z}} = \mathrm{P}_\Gamma X$ and $\varepsilon_j(x) = h_j(x) - W_{2^j} g(x)$. By definition

$$\| X - \mathrm{P}_\Gamma X \|^2 = \| \left(\varepsilon_j(x)\right)_{j \in \mathbf{Z}} \|^2 = \sum_{j=-\infty}^{+\infty} \| \varepsilon_j \|^2 + 2^{2j} \, \| \frac{d\varepsilon_j}{dx} \|^2 \ . \tag{121}$$

Hence,

$$\| X - \mathrm{P}_\Gamma X \|^2 = \sum_{(n,j) \in \mathbf{Z}^2} \left(\int_{x_n^j}^{x_{n+1}^j} |\varepsilon_j(x)|^2 \, dx + 2^{2j} \int_{x_n^j}^{x_{n+1}^j} |\frac{d\varepsilon_j(x)}{dx}|^2 \, dx \right) . \tag{122}$$

We saw in Appendix 5 that $\varepsilon_j(x)$ satisfies the differential equation (112), so by integrating by parts we obtain

$$\| X - \mathrm{P}_\Gamma X \|^2 = \sum_{(n,j) \in \mathbf{Z}^2} \left(2^{2j} \, \varepsilon_j(x_{n+1}^j) \, \frac{d\varepsilon_j(x_{n+1}^j)}{dx} - 2^{2j} \, \varepsilon_j(x_n^j) \, \frac{d\varepsilon_j(x_n^j)}{dx} \right) \tag{123}$$

The derivative at x_{n+1}^j is the left derivative whereas the derivative at x_n is the right derivative. We know that the function $\varepsilon_j(x)$ is the sum of two exponentials given by equation (113) between any two consecutive modulus maxima located at x_n^j and x_{n+1}^j. If we replace the constants α and β by their values specified by $\varepsilon_j(x_n^j)$ and $\varepsilon_j(x_{n+1}^j)$, with a few algebraic manipulations we derive that

$$\begin{aligned} & 2^{2j} \, \varepsilon_j(x_{n+1}^j) \, \frac{d\varepsilon_j(x_{n+1}^j)}{dx} - 2^{2j} \, \varepsilon_j(x_n^j) \, \frac{d\varepsilon_j(x_n^j)}{dx} \ge \\ & \frac{2^j}{4} \left(|\varepsilon_j(x_{n+1}^j)|^2 + |\varepsilon_j(x_n^j)|^2 \right) Min(2^{-j} \, (x_{n+1}^j - x_n^j) \, , \, 1.) \, . \end{aligned} \tag{124}$$

Since we suppose that there exists a constant $D > 0$ such that $2^{-j} \, |x_n^j - x_{n-1}^j| \ge D \le 1$, we obtain

$$\| X - \mathrm{P}_\Gamma X \|^2 \ge \frac{D}{2} \sum_{(n,j) \in \mathbf{Z}^2} 2^j \, |\varepsilon_j(x_n^j)|^2 \ , \tag{125}$$

which proves equation (119) for $C_1 = \frac{D}{2}$. Let us now prove equation (120). The element X is the dyadic wavelet transform of the function $g(x)$. Let $\mathbf{U}$ and $\mathbf{O}$ be the spaces defined in Section 5.2. The function $g(x)$ can be decomposed into

$$g(x) = g_1(x) + g_2(x) \, , \tag{126}$$

with $g_1(x) \in \mathbf{U}$ and $g_2(x) \in \mathbf{O}$. The original function $f(x)$ can also be decomposed into

$$f(x) = f_1(x) + f_2(x) \, , \tag{127}$$

with $f_1(x) \in \mathbf{U}$ and $f_2(x) \in \mathbf{O}$. Let us now define the function

$$h(x) = f_1(x) + g_2(x) .$$

Since $h(x) = f(x) + u(x)$, with $u(x) \in \mathbf{O}$, we know from equation (46) that $h(x)$ satisfies the constraint (a) and thus that $\left[W_{2^j} h(x) \right]_{j \in \mathbf{Z}} \in \Gamma$. We also have $h(x) - g(x) = f_1(x) - g_1(x) \in \mathbf{U}$. Since we suppose that $\left[\sqrt{2^j}\ \psi_{2^j}(x_n^j - x) \right]_{(n,j) \in \mathbf{Z}^2}$ is a frame of $\mathbf{U}$, equation (53) implies that

$$\| h(x) - g(x) \|^2 \leq \frac{1}{A_3} \sum_{(n,j) \in \mathbf{Z}^2} 2^j \, |<h(x) - g(x), \psi_{2^j}(x_n^j - x)>|^2 . \tag{128}$$

Since $\left[W_{2^j} h(x) \right]_{j \in \mathbf{Z}} \in \Gamma$, we have $W_{2^j} h(x_n^j) = W_{2^j} f(x_n^j)$ and thus

$$<h(x) - g(x)\,,\, \psi_{2^j}(x_n^j - x)> = W_{2^j} f(x_n^j) - W_{2^j} g(x_n^j) = \varepsilon_j(x_n^j) . \tag{129}$$

From the norm equivalence of equation (49), we can also derive that

$$\left| \left[W_{2^j} h(x) \right]_{j \in \mathbf{Z}} - \left[W_{2^j} g(x) \right]_{j \in \mathbf{Z}} \right|^2 \leq B_2 \, \| h(x) - g(x) \|^2 . \tag{130}$$

Since the projector $\mathbf{P}_\Lambda$ is orthogonal and $\left[W_{2^j} h(x) \right]_{j \in \mathbf{Z}} \in \Lambda$

$$\left| \left[W_{2^j} g(x) \right]_{j \in \mathbf{Z}} - \mathbf{P}_\Lambda \left[W_{2^j} g(x) \right]_{j \in \mathbf{Z}} \right|^2 \leq \left| \left[W_{2^j} g(x) \right]_{j \in \mathbf{Z}} - \left[W_{2^j} h(x) \right]_{j \in \mathbf{Z}} \right|^2 . \tag{131}$$

Equation (128), (129), (130) and (131) imply that

$$\left| \left[W_{2^j} g(x) \right]_{j \in \mathbf{Z}} - \mathbf{P}_\Lambda \left[W_{2^j} g(x) \right]_{j \in \mathbf{Z}} \right|^2 \leq \frac{B_2}{A_3} \sum_{(n,j) \in \mathbf{Z}^2} 2^j \, |\varepsilon_j(x_n^j)|^2 . \tag{132}$$

This proves equation (120) for $C_2 = \frac{B_2}{A_3}$. From equation (125) and (132), we then derive that

$$\left| X - \mathbf{P}_\Gamma X \right|^2 \geq \frac{D\, A_3}{2 B_2} \left| X - \mathbf{P}_\Lambda X \right|^2 . \tag{133}$$

This inequality gives a lower bound for the "angle" between the affine space Γ and the space $\mathbf{V}$. Let $\mathbf{P} = \mathbf{P}_\mathbf{V} \circ \mathbf{P}_\Gamma$ be the alternate projection on both spaces. A classical result on alternate projections [32] enables us to derive that for any element $X \in \mathbf{K}$ there exists a constant R such that

$$\left| \mathbf{P}_\Lambda X - \mathbf{P}^{(n)} X \right| \leq R \left(1 - \frac{D\, A_3}{2 B_2}\right)^{n/2} . \tag{134}$$

This proves that the algorithm converges exponentially with a convergence rate larger than $\left(1 - \frac{D\, A_3}{2 B_2}\right)^{-1/2}$.

References

1. Berman, Z., "The uniqueness question of discrete wavelet maxima representation," *System Research Center, TR 91-48*, University of Maryland, College Park, April 1991.

2. Burt, P. J. and Adelson, E. H., "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communications, vol 31, pp 532-540*, April 1983.

3. Canny, J., "A Computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, 1986.

4. Carlsson, S., "Sketch based coding of grey level images," *Signal Proessing, North Holland*, vol. 15, pp. 57-83, July 1988.

5. Crowley, J., "A representation for visual information," *Tech. Rep. CMU-RI-TR-82-7*, Robotic Inst. Carnegie-Mellon Univ., 1987.

6. Curtis, S., Shitz, S., and Oppenheim, V., "Reconstruction of Non-periodic two-dimensional signals from zero-crossings," *IEEE Trans. Acoustic Speech and Signal Process.*, vol. 35, pp. 890-893, 1987.

7. Daubechies, I., "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. on Information Theory*, vol. 36, pp. 961-1005, Sept. 1990.

8. Daubechies, I., *Ten lectures on wavelets*, CBMS-NSF Series in Appl. Math., SIAM, 1991.

9. Duffin, R. and Schaeffer, A., "A class of nonharmonic Fourier series," *Trans. American Math. Society*, vol. 72, pp. 341-366, 1952.

10. Folland, G., in *Introduction to partial differential equations*, Mathematical Notes, Princeton University Press, New Jersey, 1976.

11. Froment, J. and Mallat, S., "Second generation compact image coding with wavelets," in *Wavelets--A Tutorial in Theory and Applications*, ed. C. Chui, pp. 655-678, Academic Press, Jan. 1992.

12. Grochenig, K., "Sharp results on random sampling of band-limited functions," *Proc. of the NATO ASI workshop on Stochastic Processes*, Kluwer, Italy, August 1991.

13. Grossmann, A. and Morlet, J., "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math.*, vol. 15, pp. 723-736, 1984.

14. Holschneider, M. and Tchamitchian, P., "Regularite locale de la fonction non-differentiable de Riemann," in *Les ondelettes en 1989*, ed. P.G. Lemarie, Lecture notes in Mathematics, Springer-Verlag, 1989.

15. Hummel, R. and Moniot, R., "Reconstruction from zero-crossings in scale-space," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. 37, no. 12, Dec. 1989.

16. Jaffard, S., "Pointwise smoothness, two microlocalisation and wavelet coefficients," *Publicacions Mathematiques*, vol. 35, 1991.

17. Koenderink, J., "The structure of images," *Biological Cybernetics*, Springer Verlag, 1984.

18. Kunt, M., Ikonomopoulos, A., and Kocher, M., "Second generation image coding technics," *Proceed. of the IEEE*, vol. 74, pp. 549-575, April 1985.

19. Logan, B., "Information in the zero-crossings of band pass signals," *Bell Systems Tech. Journ.* , vol. 56, p. 510, 1977.

20. Mallat, S., "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. 37, no. 12, pp. 2091-2110, Dec. 1989.

21. Mallat, S., "Multiresolution approximation and wavelet orthonormal bases of L2," *Trans. of the American Mathematical Society.*, vol. 315, pp. 69-87, Sept. 1989.

22. Mallat, S., "Zero-crossings of a wavelet transform," *IEEE Trans. on Information Theory*, vol. 37, July, 1991.

23. Mallat, S. and Hwang, W. L., *Singularity detection and processing with wavelets*, IEEE Trans. on Information Theory, March 1992.

24. Marr, D., in *Vision*, W.H.Freeman and Company, 1982.

25. Marr, D. and Hildreth, E., "Theory of edge detection," *Proc. of the Royal Society of London*, vol. 207, pp. 187-217, 1980.

26. Meyer, Y., in *Ondelettes et Operateurs*, Hermann, 1990.

27. Meyer, Y., "Un contre-exemple a la conjecture de Marr et a celle de S. Mallat," *Preprint*, 1991.

28. Rioul, O. and Vetterli, M., *Wavelets and Signal Processing*, p. IEEE Signal Processing Magazine, Oct. 1991.

29. Rosenfeld, A. and Thurston, M., "Edge and curve detection for visual scene analysis," *IEEE Trans. on Computers*, vol. C-20, pp. 562-569, 1971 .

30. Sanz, J. and Huang, T., "Theorem and experiments on image reconstruction from zero-crossings," *Research report RJ5460* , IBM, Jan. 1987.

31. Witkin, A., "Scale space filtering," *Proc. Int. Joint Conf. Artificial Intell.* , 1983.

32. Youla, D. and Webb, H., "Image restoration by the method of convex projections," *IEEE Trans. Medical Imaging*, vol. 1, pp. 81-101, Oct. 1982.

33. Yuille, A. and Poggio, T., "Scaling theorems for zero crossings," *IEEE trans on PAMI*, vol. 8, Jan 1986.

34. Zeevi, Y. and Rotem, D., "Image reconstruction from zero-crossings," *IEEE Acoustic Speech and Signal Proc.*, vol. 34, pp. 1269-1277, 1986.

35. Zhong, S., *Edges representation from wavelet transform maxima*, Ph.D. Thesis, New York University, Sept. 1990.
