Characterization of Signals from Multiscale Edges

Stephane Mallat and Sifen Zhong

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 theory. 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 1-D and 2-D 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.

Index Terms—Edge detection, feature extraction, level crossings, multiscale wavelets.

I. 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 [18], [23], [25]. 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 [2], 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, etc. 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 a multiscale edge representation is. 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 1-D and 2-D signals from multiscale edges detected by the wavelet transform modulus maxima. It has been conjectured [16], [18] that multiscale edges characterize uniquely 1-D and 2-D signals, but recently, Meyer [21] has found counterexamples to these conjectures. In spite of these counterexamples, 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 1-D and 2-D 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 [17].

The article is organized as follows. Section II 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 VI, we concentrate on 1-D signals. Section III-A 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 III-B explains how to discretize this model. The numerical implementation of fast wavelet transform algorithms is given in Appendix B. Section IV explains how to characterize different types of sharp signal variations from the evolution across scales of the wavelet transform maxima. Section V 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 V-B, and numerical results are presented in Section V-C. A 2-D extension of the wavelet transform is given in Section VI-A,

and its discrete version is explained in Section VI-B. Fast 2-D wavelet algorithms are given in Appendix D. Section VII 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 VIII-A, and numerical examples are shown in Section VIII-B. Section IX describes an application to compact image coding.

Notation: $L^2(\mathbf{R})$ denotes the Hilbert space of measurable, square-integrable 1-D functions $f(x)$. For $f \in L^2(\mathbf{R})$ and $g \in L^2(\mathbf{R})$, the inner product of $f(x)$ with $g(x)$ is written:

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

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

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

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

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

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

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

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

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

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

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

II. 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. 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 that 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 $\theta(x)$

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

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

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

In this paper, we denote

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

which is 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

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

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

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 $f * \theta_s(x)$. In the particular case where $\theta(x)$ is a Gaussian, the zero-crossing detection is equivalent to a Marr-Hildreth [19] edge detection, whereas the extrema detection corresponds to a Canny [2] edge detection. When the scale s is large, the convolution with $\theta_s(x)$ removes small signal fluctuations; we therefore only detect the sharp variations of large 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. 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 $|W_s^a f(x)|$. In addition, zero crossings give 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 IV 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.

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 2-D function $\xi(x, y)$. We use the term 2-D smoothing function to describe any function $\theta(x, y)$ whose integral over x and y is equal to 1 and 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 towards which the gradient vector points 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 2-D wavelet transform. We define two wavelet functions $\psi^1(x, y)$ and $\psi^2(x, y)$ such that

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

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 $f(x, y) \in L^2(R^2)$. The wavelet transform of $f(x, y)$ at the scale s has two components defined by

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

Similarly to (4), one can easily prove that

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

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.

III. DYADIC WAVELETTRANSFORM IN ONE DIMENSION

A. General Properties

For most purposes, the wavelet model is not required to keep a continuous scale parameter s. To allow fast numerical implementations, we impose that the scale varies only along the dyadic sequence $(2^j)_{j \in Z}$. 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, refer to the mathematical books of Meyer [20] and Daubechies [5] or to signal processing oriented reviews [15], [22]. The wavelet model has first been formalized by Grossmann and Morlet [10]. 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

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

We refer to the *dyadic wavelet transform* as the sequence of functions

$$Wf = (W_{2^j} f(x))_{j \in Z} \quad (10)$$

and 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

$$\hat{W}_{2^j} f(\omega) = \hat{f}(\omega) \hat{\psi}(2^j \omega). \quad (11)$$

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

$$\forall \omega \in R, A_1 \leq \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2 \leq B_1 \quad (12)$$

we ensure that the whole frequency axis is covered by dilations of $\hat{\psi}(\omega)$ by $(2^j)_{j \in Z}$ so that $\hat{f}(\omega)$, and thus, $f(x)$ can be recovered from its dyadic wavelet transform. The reconstructing wavelet $\chi(x)$ is any function whose Fourier transform satisfies

$$\sum_{j=-\infty}^{+\infty} \hat{\psi}(2^j \omega) \hat{\chi}(2^j \omega) = 1. \quad (13)$$

If property (12) is valid, there exists an infinite number of functions $\hat{\chi}(\omega)$ that satisfy (13). The function $f(x)$ is recovered from its dyadic wavelet transform with the summation

$$f(x) = \sum_{j=-\infty}^{+\infty} W_{2^j} f * \chi_{2^j}(x). \quad (14)$$

This equation is proved by computing its Fourier transform and inserting (11) and (13). With the Parseval theorem, we derive from (11) and (12) a norm equivalence relation

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

This proves that the dyadic wavelet transform is not only complete but stable as well. If $\frac{B_1}{A_1}$ is closer to 1, it will be more stable.

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

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

The reconstruction formula (14) shows that $(g_j(x)_{j \in Z})$ is the dyadic wavelet transform of some function in $L^2(R)$, if and only if

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

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

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

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

These equations are known as 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.
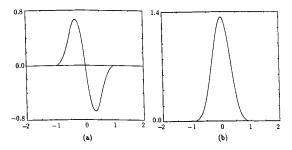
Fig. 1. (a) This wavelet is a quadratic spline of compact support that is continuously differentiable. It is defined in Appendix A, and it is the derivative of the cubic spline function $\theta(x)$ shown in (b).

Fig. 1(a) is a quadratic spline wavelet of compact support, which is further defined in Appendix A. It is the derivative of the smoothing function $\theta(x)$ shown in Fig. 1(b). Fig. 2(a) is the plot of a discrete signal of 256 samples. Fig. 2(b) shows its discrete dyadic wavelet transform computed on nine scales. At each scale 2^j, we compute a uniform sampling of the wavelet transform that we denote $W_{2^j}^d f$. 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 B. The reader not interested in numerical issues might want to skip Section III-B. Since our wavelet is the derivative of a smoothing function, (4) proves that $W_{2^j}^d f$ is proportional to the derivative of the original signal smoothed at the scale 2^j. Fig. 2(c) gives the locations and values of the local maxima of the dyadic wavelet transform modulus, as in a Canny edge detection. At each scale 2^j, each modulus maximum is represented by a Dirac that 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 finds the signal sharp variation points.

B. Discrete Wavelet Transform

In numerical applications, the input signal is measured at a finite resolution; therefore, 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

$$|\hat{\phi}(\omega)|^2 = \sum_{j=1}^{+\infty} \hat{\psi}(2^j \omega) \hat{\chi}(2^j \omega). \qquad (20)$$

We suppose here that the reconstructing wavelet $\chi(\omega)$ is such that $\hat{\psi}(\omega)\hat{\chi}(\omega)$ is a positive, real, even function. One can prove that property (13) 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

$$S_{2^j} f(x) = f * \phi_{2^j}(x) \text{ with } \phi_{2^j}(x) = \frac{1}{2^j}\phi(\frac{x}{2^j}). \qquad (21)$$

If the scale 2^j is larger, the more details of $f(x)$ are removed by S_{2^j}. For any scale $2^J > 1$, (20) yields
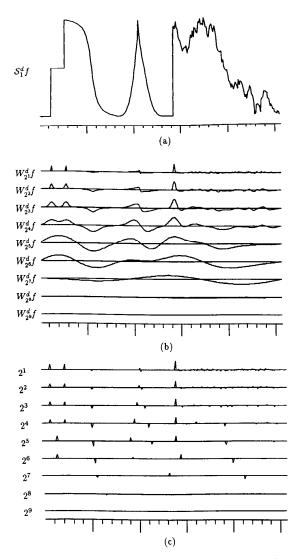

Fig. 2. (a) Signal of 256 samples; (b) discrete dyadic wavelet transform of signal (a) computed on nine scales. At each scale 2^j, we plot the signal $W_{2^j}^d f$, which also has 256 samples; (c) modulus maxima of the dyadic wavelet transform shown in (b). Each Dirac indicates the position and amplitude of a modulus maximum.

$$|\hat{\phi}(\omega)|^2 - |\hat{\phi}(2^J \omega)| = \sum_{j=1}^{J} \hat{\psi}(2^j \omega) \hat{\chi}(2^j \omega). \qquad (22)$$

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 $(W_{2^j} f(x))_{1 \le j \le J}$ between the scales 2^1 and 2^J.

We suppose that the original signal is a discrete sequence $D = (d_n)_{n \in Z}$ of finite energy. If there exists two constants $C_1 > 0$ and $C_2 > 0$ such that $\hat{\phi}(\omega)$ satisfies

$$\forall \omega \in R, \quad C_1 \le \sum_{n=-\infty}^{+\infty} |\hat{\phi}(\omega + 2n\pi)|^2 \le C_2 \qquad (23)$$

then one can prove prove [18] that there exists a function $f(x) \in L^2(R)$ (not unique) such that

$$\forall n \in Z, \quad S_1 f(n) = d_n. \tag{24}$$

The input signal can thus be rewritten $D = (S_1 f(n))_{n \in Z}$. For a particular class of wavelets defined in Appendix A, the discrete signal $D = (S_1 f(n))_{n \in Z}$ allows us to compute a uniform sampling of the wavelet transform of $f(x)$ at any scale larger than 1. Let us denote

$$W^d_{2^j} f = (W_{2^j} f(n + w))_{n \in Z} \text{ and } S^d_{2^j} f = (S_{2^j} f(n + w))_{n \in Z} \tag{25}$$

where w is a sampling shift that depends only on $\psi(x)$. For any coarse scale 2^J, the sequence of discrete signals

$$\left\{ S^d_{2^J} f, \left(W^d_{2^j} f \right)_{1 \le j \le J} \right\} \tag{26}$$

is called the *discrete dyadic wavelet transform* of $D = (S_1 f(n))_{n \in Z}$. The coarse signal $S^d_{2^J}$ provides the signal components below the scale 2^J. A fast discrete wavelet transform algorithm and its inverse are described in Appendix B.

In practice, the original discrete signal D has a finite number N of nonzero values: $D = (d_n)_{1 \le n \le N}$. To solve the border problems, we use the same periodization technique as in a cosine transform. We suppose that our signal has a period of $2N$ samples, and we extend it with a symmetry for $N < n \le 2N$: $d_n = d_{2N+1-n}$. By periodizing the signal with a symmetry, we avoid creating a discontinuity at the borders. The discrete wavelet coefficients are also $2N$ periodic. If the wavelet is antisymmetrical with respect to 0, as in Fig. 1(a), the wavelet coefficients are antisymmetrical at the borders. For the class of wavelets defined in Appendix A, one can also prove that when the scale is as large as the period ($2^J = 2N$), $S^d_{2^J} f$ is constant and equal to the mean value of the original signal D. We thus decompose any signal of N samples over $J = \log_2(N) + 1$ scales. Appendix B 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.

IV. 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 on the local Lipschitz regularity of the signal. This section explains what a Lipschitz exponent is and how this exponent is computed from the wavelet transform maxima.
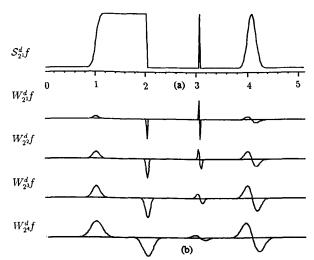

Fig. 3. (a) Four sharp variation points of this signal have a different Lipschitz regularity α_0 and smoothing variance σ^2. These values are given, respectively, by $(\alpha_0 = 0, \sigma = 3)$, $(\alpha_0 = 0, \sigma = 0)$, $(\alpha_0 = -1, \sigma = 0)$, and $(\alpha_0 = -1, \sigma = 4)$; (b) behavior of the modulus maxima across scales depends on the Lipschitz regularity α_0 and the smoothing factor σ.

A more detailed mathematical and numerical analysis of this topic can be found in [17]. 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 1: Let $0 \le \alpha \le 1$. A function $f(x)$ is uniformly Lipschitz α over an interval $]a, b[$ if and only if there exists a constant K such that for any $(x_0, x_1) \in]a, b[^2$

$$|f(x_0) - f(x_1)| \le K|x_0 - x_1|^\alpha. \tag{27}$$

We refer to the Lipschitz uniform regularity of $f(x)$ as the upper bound α_0 of all α such that $f(x)$ is uniformly Lipschitz α.

If $f(x)$ is differentiable at x_0, then it is Lipschitz $\alpha = 1$. If the uniform Lipschitz regularity α_0 is larger, the singularity at x_0 will be more "regular." If $f(x)$ is discontinuous but bounded in the neighborhood of x_0, its uniform Lipschitz regularity in the neighborhood of x_0 is 0. 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 that 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

$$|W_{2^j} f(x)| \le K(2^j)^\alpha. \tag{28}$$

The proof of this theorem can be found in [20]. From (28), we derive that

$$\log_2 |W_{2^j} f(x)| \le \log_2(K) + \alpha j. \tag{29}$$

If the uniform Lipschitz regularity is positive, (28) implies that the amplitude of the wavelet transform modulus maxima should decrease when the scale decreases. On the contrary, the singularity at the abscissa 3 of Fig. 3(b) produces wavelet

transform maxima that increase 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 then viewed as a tempered distribution. At the abscissa 3 of Fig. 3(b), this distribution is locally equal to a Dirac. The reader might want to consult Folland [7] for a quick presentation of the mathematical theory of distributions. The wavelet transform of tempered distributions is well defined if the wavelet $\psi(x)$ is smooth enough [17]. For example, if $\psi(x)$ is continuous, the wavelet transform of a Dirac $\delta(x)$ is given by

$$W_{2^j}\delta(x) = \delta * \psi_{2^j}(x) = \psi_{2^j}(x). \qquad (30)$$

To extend Lipschitz exponents to distributions, we say that a distribution has 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 that 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. Hence, 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 for any $\alpha < \alpha_0$ there exists K such that

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

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. 3(b).

In practice, we can only process discrete signals that approximate the original function at a finite resolution, which we normalize to 1. Strictly speaking, it is not meaningful to speak about singularities, discontinuities, or Diracs. In fact, we cannot compute the wavelet transform at scales finer than 1 and thus cannot verify (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 $(f_n)_{n \in Z}$, with $f_n = 0$ for $n < n_0$ and $f_n = 1$ for $n \geq n_0$, like at abscissa 2 of Fig. 3(a). At 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. 3(b), in the neighborhood of $x = 2$, 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 II, we only keep the local maxima of the wavelet transform modulus. It has been proved [17] that if a signal is singular at a point x_0, there exists a sequence of wavelet transform modulus maxima that converge to x_0 when the scale decreases. Hence, we detect all the singularities from the positions of the wavelet transform modulus maxima. Moreover, the decay of the wavelet transform is bounded by the decay of these modulus maxima, and we can thus measure the local 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. 3(a). It is generally important to estimate the smoothness of 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^2 is proportional to the time it would take to create a singularity at the point x_0 if we apply a backward 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

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

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 $\theta(x)$, (4) proves that the wavelet transform of $f(x)$ can be written

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

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

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

Equation (33) can thus be rewritten

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

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 nonsmoothed singularity $h(x)$ at the scale $s_0 = \sqrt{2^{2j} + \sigma^2}$. Equation (28) of Theorem 1 proves that the Lipschitz regularity is the upper bound of the set of α that satisfy

$$|W_{2^j}h(x)| \leq K(2^j)^\alpha. \qquad (36)$$

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 (35), we obtain

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

This equation is satisfied at all points x if and only if it is satisfied at the locations of all the local maxima of $|W_{2^j} f(x)|$. 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. Fig. 3 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 a smooth curve in the scale-space plane (s, x). These curves have been called "finger prints" by Witkin [25]. 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. 3, there is one sequence of maxima that belongs to the same maxima curve and converges to the position of the discontinuity at $x = 2$. For the Dirac at abscissa 3, there are two such sequences. Each one gives 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 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 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^I 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 (37) is as close as possible to an equality for each maximum a_j. These values

are obtained by minimizing

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

This is done with a steepest gradient descent algorithm. The value K gives the amplitude of the sharp variation. When computing the values of σ and α from the evolution of the maxima across scales in Fig. 3, 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 is only an approximation. In this case, $\theta(x)$ is the cubic spline shown in Fig. 1(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. 2(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 nonisolated singularities is studied in more detail in [17].

V. Signal Reconstruction from Multiscale Edges

Section IV 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 V-B describes an algorithm that reconstructs a close approximation of the original signal from the wavelet transform modulus maxima. Numerical results are presented in Section V-C.

A. Previous Results

The reconstruction of signals from multiscale edges has mainly been studied in the zero-crossing framework. We saw in Section II 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 the zero crossing is the Logan theorem [14]. However, as it is explained in [18], the hypotheses 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 [4], [24], [28]; Refer to a review by Hummel and Moniot [11] 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 $t = s^2$ [12]. With this approach, Hummel and Moniot [11], as well as Yuille and Poggio [27], have proved some completeness properties under restrictive conditions, like supposing that $f(x)$ is a polynomial

[29]. In general, there are known counterexamples that 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 [21] found a large family of such counterexamples.

To obtain a complete and stable zero-crossing representation, Mallat [16] conjectured that it is sufficient to record the zero-crossing positions of $W_{2^j}^b f(x)$ at all dyadic scales $2^j f$, as well as the integral values

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

between any pair of consecutive zero crossings $(z_n, z_n + 1)$. This conjecture was motivated by a reconstruction algorithm that is able to recover a close approximation of the original signal from these zero crossings and integral values [18]. We proved in Section II that the zero crossings of $W_{2^j}^b f(x)$ occur at the extrema points of the wavelet transform $W_{2^j}^a f(x)$, which are defined with respect to the wavelet $\psi^a(x) = \frac{d\theta(x)}{dx}$. From (4), (5), and (39), we derive that

$$e_n = W_{2^j}^a f(z_n + 1) - W_{2^j}^a f(z_n). \qquad (40)$$

To record the zero-crossing positions and integral values of $W_{2^j}^b f(x)$ is therefore equivalent to recording the positions where $W_{2^j}^a f(x)$ has local extrema and the value of $W_{2^j}^a f(x)$ at the corresponding locations. Meyer [21] proved that the completeness of this representation depends on the choice of the smoothing function $\theta(x)$ but that the conjecture is not valid in general. Let

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

For the wavelet shown in Fig. 1(a), Meyer [21] found a noncountable family of functions

$$f_\epsilon(x) = f_0(x) + \chi_\epsilon(x)$$

such that at all scales 2^j, $W_{2^j} f_\epsilon(x)$, and $W_{2^j} f_0(x)$ have the same extrema (positions and values). The functions $\chi_\epsilon(x)$ are small high-frequency perturbations, which are implicitly defined by constraint equations that guarantee 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 $\chi_\epsilon(x)$ must remain small, which would explain the quality of the signal reconstructions obtained in [16], 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 that contradict the completeness conjecture.

We explained in Section II 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. 1(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.

B. Reconstruction Algorithm

Let $f(x) \in L^2(\mathbf{R})$ and $(W_{2^j} f(x))_{j \in Z}$ be its dyadic wavelet transform. We describe an algorithm that reconstructs an approximation of $(W_{2^j} f(x))_{j \in Z}$, 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 $(x_n^j)_{n \in Z}$ 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.

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

Let us first analyze the condition 1. The value of $W_{2^j} f(x)$ at any x_0 can be written as an inner product in $L^2(\mathbf{R})$. Indeed

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

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

Condition 1 is, thus, equivalent to

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

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

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

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

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

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

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

This defines an affine space that we denote $f + O$. If $U = L^2(\mathbf{R})$, then $O = \{0\}$, which implies that $h(x)$ must be equal

to $f(x)$. In general, this is not the case; therefore, (43) does not characterize uniquely $f(x)$.

Condition 2 is more difficult to analyze because it is not convex. In order to solve this problem numerically, we approximate condition 2 with a convex constraint. Condition 2 defines the value of the wavelet transform at the points $(x_n^j)_{(j,n)\in Z^2}$. 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 $(x_n^j)_{(j,n)\in Z^2}$. The number of modulus maxima of $W_{2^j}f(x)$ depends on how much this function oscillates. To have as few modulus maxima as possible outside the abscissa $(x_n^j)_{(j,n)\in Z^2}$, 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

$$\|h\|^2 = \left| (W_{2^j}h(x))_{j\in Z} \right|^2$$

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

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 exist two constants $A_2 > 0$ and B_2 such that for all $\omega \in R$

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

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

$$A_2\|h\|^2 \le \|h\|^2 \le B_2\|h\|^2. \quad (49)$$

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

$$2^j\frac{dW_{2^j}h(x)}{dx} = f * \psi_{2^j}^1(x). \quad (50)$$

As for the norm equivalence equation (15), we then prove the implication by applying the Parseval theorem to each $L^2(R)$ norm component of the norm defined by (47). Equation (48) is valid for any dyadic wavelet $\psi(x)$ that satisfies (12) and is smooth enough. For example, there exist two such constant A_2 and B_2 for the wavelet shown in Fig. 1(a). By replacing condition 2 by the minimization of $\|h\|$, we define a problem that has a unique solution. Indeed, condition 1 imposes that $h(x)$ must belong to the closed affine space $f + O$, and the minimization of a norm over such a closed convex has a unique solution.

Although there exists a unique element of $f + O$ whose norm $\|\ \ \|$ is minimum, the computation of this function might not be stable. If the two constants A_2 and B_2 of (48) are equal, the norm $\|\ \ \|$ is proportional to the classical $L^2(R)$ norm. The solution of the minimization problem is therefore the orthogonal projection of $f(x)$ over U. The frame theory proves [6] that one can make a stable computation of the orthogonal projection of $f(x)$ onto U from the inner products $(\langle f(x), \psi_{2^j}(x_n^j - x)\rangle)_{(j,n)\in Z^2}$ if and only if the family of functions $\left(\sqrt{2^j}\psi_{2^j}(x_n^j - x) \right)_{(j,n)\in Z^2}$ is a frame of U. The

factor $\sqrt{2^j}$ normalizes the $L^2(R)$ norm of each function in the family. By definition, such a family is a frame of U [6] if and only if there exist two constants $A_3 > 0$ and B_3 such that for any function $g \in U$

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

When the two constants A_2 and B_2 of (48) are different, the norm $\|\ \ \|$ is not equal but is equivalent to the classical $L^2(R)$ norm. In this case, the stability also depends on whether the family of wavelets is a frame of U. The closer to 0 the value of $\frac{B_3-A_3}{B_3+A_3}$, the more stable the computations. Outside of a few particular cases, it is difficult to prove analytically whether a given family of wavelets $\left(\sqrt{2^j}\psi_{2^j}(x_n^j - x) \right)_{(j,n)\in Z^2}$ is or is not a frame of the space 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 algorithm based on alternate projections. Let K be the space of all sequences of functions $(g_j^{(x)})_{j\in Z}$ such that

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

The norm $|\ \ |$ defines a Hilbert structure over K. Let V be the space of all dyadic wavelet transforms of functions in $L^2(R)$. Equation (49) proves that V is included in K. Let Γ be the affine space of sequences of functions $(g_j(x))_{j\in Z} \in K$ 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 K. The dyadic wavelet transforms that satisfy condition 1 are the sequences of functions that belong to

$$\Lambda = V \cap \Gamma.$$

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

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

$$P_V = W \circ W^{-1}. \quad (53)$$

For any sequence $X = (g_j(x))_{j\in Z} \in K$, it is clear that $P_V X \in V$; therefore, P_V is a projector on V. We saw in (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 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. 1(a), the orthogonal projection on the space 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 B. Appendix E characterizes

the projection on the affine set Γ, which is orthogonal with respect to the norm $|\quad|$. We prove that this operator P_Γ is implemented by adding piecewise exponential curves to each function of the sequence that we project on Γ. Let $P = P_V \circ P_\Gamma$ be alternate projections on both spaces. Let $P^{(n)}$ be n iterations over the operator P. Since Γ is an affine space and V a Hilbert space, a classical result on alternate projections [26] proves that for any sequence of functions $X = (g_j(x))_{j \in Z} \in K$

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

Alternate projections on Γ and V converge strongly to the orthogonal projection on Λ. If X is the zero element of K, which means that $g_j(x) = 0$ for all $j \in Z$, the alternate projections converge to the element of Λ, which is the closest to zero, and thus whose norm $|\quad|$ is minimum. This is illustrated by Fig. 4. This iterative algorithm can be related to techniques based on frame operators for reconstructing signals from irregular samplings [9].

If the minimization problem is unstable, the convergence of the alternate projections is extremely slow. We saw that the numerical stability depends on whether $\left(\sqrt{2^j}\psi_{2^j}(x_n^j - x)\right)_{(n,j) \in Z^2}$ is a frame of U. Appendix F proves that if $\left(\sqrt{2^j}\psi_{2^j}(x_n^j - x)\right)_{(n,j) \in Z^2}$ is a frame of 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

$$|x_n^j - x_{n-1}^j| \geq D2^j$$

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

$$\left| P^{(n)} X - P_\Lambda X \right| \leq R(1 - \frac{DA_3}{2B_2})^{n/2} \tag{55}$$

where the A_3 is the frame bound defined in (51) and B_2 the norm equivalence bound defined in (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 $|\quad|$ 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 E 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.

C. Numerical Reconstruction of 1-D Signals from Local Maxima

There are several open issues behind the reconstruction algorithm that we described. From the results of Meyer's work [21], we know that, in general, we cannot 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 that yields a unique solution. We thus do not converge
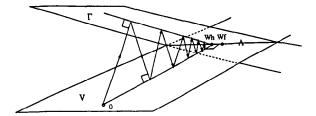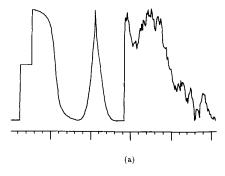

Fig. 4. Approximation of the wavelet transform of $f(x)$ is reconstructed by alternating orthogonal projections on an affine space Γ and on the space V of all dyadic wavelet transforms. The projections begin from the zero element and converge to its orthogonal projection on $\Gamma \cap V$.

toward the wavelet transform of the original signal but toward 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 is 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 III-B. Equation (53) proves that we can compute the projection P_V by implementing W^{-1} followed by W. With the fast algorithms described in Appendix B, this requires a total of $O(N \log_2(N))$ operations. Appendix E proves that the implementation of P_Γ 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 P involves $O(N \log_2(N))$ operations.

The rms signal-to-noise ratio (SNR) of the reconstruction is measured in decibels. At the scale 2^j for $1 \leq j \leq 6$, Fig. 6(a) gives the value of the SNR for the reconstruction of $W_{2^j}^d f$, after n iterations on the operator 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 P, the SNR increases when the scale increases. This proves that the remaining error is rather concentrated at fine scales, like in the counterexamples of Meyer [21]. After n iterations, we can reconstruct a signal by applying the inverse wavelet transform operator on the reconstructed wavelet transform. Fig. 6(b) shows the increase of the SNR, which is computed with respect to the original signal. This SNR is an aggregation of the wavelet transform SNR at all scales. The signal in Fig. 5(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 to which we converge. The other one is the distance between the wavelet transform to which we converge 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
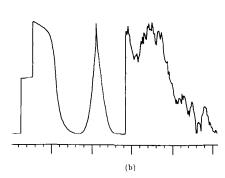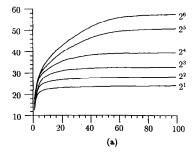
(a)

(b)

Fig. 5. (a) Original signal; (b) signal reconstructed with 20 iterations from the modulus maxima shown in Fig. 2(c).

(a)

(b)

Fig. 6. (a) SNR 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 $1 \leq j \leq 6$; (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.

defined by the maxima positions. In numerical computations, there is a finite number of maxima; therefore, the family of wavelets that generates U is finite. A finite family of vectors is always a frame, but the frame bound A_3 can be very small. The lower bound of the convergence rate given by (55) can thus also be very small. Fig. 7 is the SNR of the reconstructed signal computed with respect to the signal to which we converge. Instead of measuring the error with respect to the original signal as in Fig. 6(b), the error is measured with respect to the signal to which we converge. 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. 6(b) and Fig. 7, the increase of the SNR slows down after approximately 20 iterations. At this point, the distance between the reconstructed signal and the signal to which we converge is of the same order as the distance between the original signal and the signal to which we converge. Increasing the number of iterations slowly reduces the distance with respect to the point to which we converge but does not largely decrease the distance with respect to the original signal. This is why the SNR in Fig. 7 continues to increase slowly, whereas the SNR in Fig. 6(b) reaches a maximum on 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 (41). In all these examples, the SNR has the same type of behavior as in Figs. 6 and 7. In most cases, after 30 iterations, the relative increase of precision that is obtained
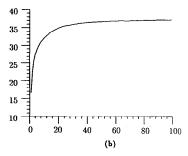
by increasing the number of iterations is negligible. Since each iteration requires $O(N \log(N))$ operations, these reconstructions do not require extensive computations and can be done in real time. The reconstructed functions are not equal to the original signal but are numerically close. They have no spurious oscillations and the same types 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. 5. We have no upper bound on the error due to the distance between the signal to which we converge 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.

VI. WAVELET TRANSFORM OF IMAGES

We explained in Section II 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 $\psi^2(x, y)$. The second part of this article extends our 1-D results for image processing applications.

A. General Properties

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

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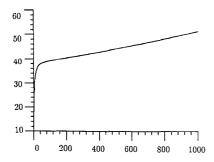(56)

Fig. 7.   SNR of the reconstructed signal computed with respect to the signal to which we converge as a function of the number of iterations on the operator P.

We refer to the *2-D dyadic wavelet transform* of $f(x,y)$ as the set of functions

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

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 $W_{2^j}^1 f(x,y)$ and $W_{2^j}^2 f(x,y)$ are, respectively, given by

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

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

To ensure that a dyadic wavelet transform is a complete and stable representation of $f(x,y)$, we impose that the 2-D 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 exist two strictly positive constants A_4 and B_4 such that

$$\forall(\omega_x,\omega_y) \in \boldsymbol{R}^2,$$

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

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

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

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

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

$$f(x,y) = \sum_{j=-\infty}^{+\infty} \left(W_{2^j}^1 f * \chi_{2^j}^1(x,y) + W_{2^j}^2 f * \chi_{2^j}^2(x,y)\right). \qquad (62)$$

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

$$\boldsymbol{W}^{-1}\left(g_j^1(x,y), g_j^2(x,y)\right)_{j \in Z}$$

$$= \sum_{j=-\infty}^{+\infty} \left(g_j^1 * \chi_{2^j}^1(x,y) + g_j^2 * \chi_{2^j}^2(x,y)\right). \qquad (63)$$

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

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

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

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

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

$$\begin{pmatrix} W_{2^j}^1 f(x,y) \\ W_{2^j}^2 f(x,y) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial x}(f * \theta_{2^j})(x,y) \\ \frac{\partial}{\partial y}(f * \theta_{2^j})(x,y) \end{pmatrix}$$

$$= 2^j \vec{\nabla}(f * \theta_{2^j})(x,y). \qquad (66)$$

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. 8, which shows the 2-D wavelet transform of the image of a circle. At each scale 2^j, the modulus of the gradient vector is proportional to

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

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

$$A_{2^j} f(x,y) = \text{argument}\,(W_{2^j}^1 f(x,y) + iW_{2^j}^2 f(x,y)). \qquad (68)$$

Like in the Canny algorithm [2], 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. 8 has 128 by 128 pixels. The first two columns of Fig. 8 give the discrete wavelet transform $W_{2^j}^{1,d} f$ and $W_{2^j}^{2,d} f$ 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 that is not interested by numerical implementations can skip Section VI-B. The discrete modulus images $M_{2^j}^d f$ and angle images $A_{2^j}^d f$ 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 is 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 first column of Fig. 9 displays its discrete modulus images $M_{2^j}^d f$, and the second 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
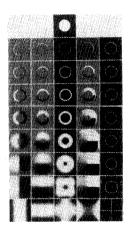
Fig. 8.   Original image at the top has 128 by 128 pixels. The first two columns from the left show, respectively, $\left(W_{2^j}^{1,d}f\right)_{1\leq j\leq 8}$ and $\left(W_{2^j}^{2,d}f\right)_{1\leq j\leq 8}$. The scale increases from top to bottom. Black, grey, and white pixels indicate, respectively, negative, zero, and positive pixel values. The third column displays the modulus images $\left(M_{2^j}^d f\right)_{1\leq j\leq 8}$. Black pixels indicate zero values, whereas white ones correspond to highest values. The fourth column shows the angle images $\left(A_{2^j}^d f\right)_{1\leq j\leq 8}$. The angle value turns from 0 (white) to 2π (black) along the circle contour. The images of the last column display, in white, the points where $M_{2^j}^d f$ has local maxima in the direction indicated $A_{2^j}^d f$.

has a small amplitude. The third column displays the maxima whose modulus is 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 $\theta_{2^j}(x,y)$.

B. Discrete Wavelet Transform of Images

Images are measured at a finite resolution; therefore, 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 is an aggregation of the wavelet components dilated by scales larger than 1

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

We shall suppose that $\phi(x,y)$ is real. From the admissibility condition (61), one can derive 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

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

The wavelet transform between the scales 1 and 2^J $\left(W_{2^j}^1 f(x,y), W_{2^j}^2 f(x,y)\right)_{1\leq j\leq J}$ provides the details that are available in $S_1 f(x,y)$ but that have disappeared in $S_{2^J}f(x,y)$.
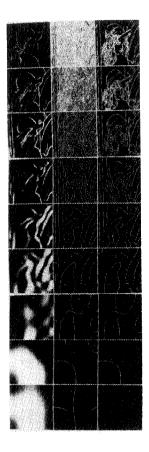

Fig. 9.   Original Lena image is at the top left of Fig. 11 and has 256 by 256 pixels. The first column gives its modulus images $\left(M_{2^j}^d f\right)_{1\leq j\leq 9}$. The second column displays the position of the local maxima of $M_{2^j}^d f$ for $1\leq j\leq 9$. The third column gives the positions 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.

At the output of a camera digitizer, an image is a finite energy 2-D discrete signal $D = (d_{n,m})_{(n,m)\in Z^2}$. Similar to the 1-D case, one can prove that there exists a function $f(x) \in L^2(R)$ (not unique) such that

$$\forall(n,m) \in Z^2, \quad S_1 f(n,m) = d_{n,m}. \tag{71}$$

The discrete image can thus be rewritten

$$D = (S_1 f(n,m))_{(n,m)\in Z^2}.$$

For the class of wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ given in Appendix C, from the sample values $(S_1 f(n,m))_{(n,m)\in Z^2}$, we can compute a uniform sampling of the wavelet transform of $f(x,y)$ at scales larger than 1. For any scale 2^j, we denote

$$W_{2^j}^{1,d}f = \left(W_{2^j}^1 f(n+w,m+w)\right)_{(n,m)\in Z^2},$$
$$W_{2^j}^{2,d}f = \left(W_{2^j}^2 f(n+w,m+w)\right)_{(n,m)\in Z^2} \text{ and}$$
$$S_{2^j}^d f = (S_{2^j}f(n+w,m+w))_{(n,m)\in Z^2}.$$

The sampling shift w depends on the choice of wavelets. Given an image D, Appendix D describes a fast algorithm

that computes the *discrete dyadic wavelet transform*

$$\left\{ S_{2^J}^d f, \left(W_{2^j}^{1,d} f\right)_{1 \le j \le J}, \left(W_{2^j}^{2,d} f\right)_{1 \le j \le J} \right\}. \quad (72)$$

Images are finite 2-D discrete signals $D = (d_{n,m})_{(n,m) \in N^2}$ of N by N pixels. We solve border problems such as those in a 2-D cosine transform. We suppose that the image is symmetrical with respect to each of its border and has a period of $2N$ by $2N$ pixels. 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. Figs. 8 and 9 give two examples of wavelet transforms computed over $\log_2(N) + 1$ scales. The numerical complexity of the fast discrete wavelet transform is $O(N^2 \log(N))$. The reconstruction of 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 (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 positions are in the direction indicated by the angle value of $A_{2^j}^d f$ at the corresponding location.

VII. CHARACTERIZATION OF IMAGE EDGES

Sharp variations of 2-D 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. As 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 a 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. As in one dimension, we discriminate different types of singularities by measuring their local Lipschitz regularity.

Definition 2: Let $0 < \alpha < 1$. A function $f(x, y)$ is uniformly Lipschitz α over an open set Ω of $\mathbf{R}^2$ if and only if there exists a constant K such that for all (x_0, y_0) and (x_1, y_1) in Ω

$$|f(x_0, y_0) - f(x_1, y_1)| \le K |(x_0 - x_1)^2 + (y_0 - y_1)^2|^{\alpha/2}. \quad (73)$$

The Lipschitz regularity of $f(x, y)$ over Ω is the superior bound of all α such that $f(x, y)$ is uniformly Lipschitz α.

In two dimensions, the Lipschitz regularity is characterized by the decay across scales of both $|W_{2^j}^1 f(x, y)|$ and $|W_{2^j}^2 f(x, y)|$. 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

$$M_{2^j} f(x, y) \le K(2^j)^\alpha. \quad (74)$$

This theorem is the 2-D extension of Theorem 1, and its proof is essentially the same [20]. The logarithm of (74) yields

$$\log_2(M_{2^j} f(x, y)) \le \log_2(K) + \alpha j. \quad (75)$$

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 between different types 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 2-D rotationally symmetric Gaussian of variance σ^2

$$f(x, y) = h * g_\sigma(x, y) \text{ with } g_\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x^2 + y^2}{2\sigma^2}). \quad (76)$$

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 symmetric Gaussian, then we can estimate the variance σ^2. The wavelet transform modulus of $f(x, y)$ is defined at any scale s by

$$M_s f(x, y) = \sqrt{|W_s^1 f(x, y)|^2 + |W_s^2 f(x, y)|^2}. \quad (77)$$

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

$$M_{2^j} f(x, y) \approx \frac{2^j}{s_0} M_{s_0} h(x, y) \text{ with } s_0 = \sqrt{2^{2j} + \sigma^2}. \quad (78)$$

Equation (74) 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

$$M_{s_0} h(x, y) \le K s_0^\alpha. \quad (79)$$

We thus derive from (78) that

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

Along a maxima chain, the singularity type varies smoothly; therefore, 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 the maxima chain that propagates between the scales 2^1 and 2^I. 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 (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 the chain. This is justified since we suppose that the singularity type does not vary greatly along this portion of chain. Let a_j be the average value of $M_{2^j} f(x, y)$. As in one dimension, we estimate the smoothing factor σ and the Lipschitz regularity α_0 by computing the values that minimize

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

This algorithm associates, with 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 at whether the image intensity is discontinuous or is smoothly varying. For the circle image of Fig. 8, 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 nonedge points but as a procedure that characterizes precisely the different types of image sharp variations.

VIII. RECONSTRUCTION OF IMAGES FROM MULTISCALE EDGES

A. Reconstruction Algorithm

The algorithm that reconstructs images from the local maxima of their wavelet transform modulus is an extension of the 1-D algorithm described in Section V-B. Let $f(x, y) \in L^2(R^2)$ and $\left(W_{2^j}^1 f(x, y), W_{2^j}^2 f(x, y) \right)_{j \in Z}$ 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 $\left((x_v^j, y_v^j) \right)_{v \in R}$ as well as $\left(M_{2^j} f(x_v^j, y_v^j), A_{2^j} f(x_v^j, y_v^j) \right)_{v \in R}$. In two dimensions, the number of modulus maxima is no longer countable. From $M_{2^j} f(x_v^j, y_v^j)$ and $A_{2^j} f(x_v^j, y_v^j)$, we can compute $W_{2^j}^1 f(x_v^j, y_v^j)$, $W_{2^j}^2 f(x_v^j, y_v^j)$, and vice versa. The inverse problem consists of finding the set of functions $h(x, y)$ that satisfy the following two constraints:

1. At each scale 2^j and for each modulus maxima location (x_v^j, y_v^j), we have $W_{2^j}^1 h(x_v^j, y_v^j) = W_{2^j}^1 f(x_v^j, y_v^j)$ and $W_{2^j}^2 h(x_v^j, y_v^j) = W_{2^j}^2 f(x_v^j, y_v^j)$.

2. At each scale 2^j, the modulus maxima obtained from $W_{2^j}^1 h(x, y)$ and $W_{2^j}^2 h(x, y)$ are located at the abscissa $\left((x_v^j, y_v^j) \right)_{v \in R}$.

Let us analyze property 1. At any point (x_0, y_0), the wavelet transform can be rewritten as inner products

$$W_{2^j}^1 h(x_0, y_0) = \langle f(x, y), \psi_{2^j}^1(x_0 - x, y_0 - y) \rangle,$$
$$W_{2^j}^2 h(x_0, y_0) = \langle f(x, y), \psi_{2^j}^2(x_0 - x, y_0 - y) \rangle. \quad (82)$$

Let U be the closure of the set of functions that are linear combinations of any function of the family

$$\left\{ 2^j \psi_{2^j}^1(x_v^j - x, y_v^j - y), 2^j \psi_{2^j}^2(x_v^j - x, y_v^j - y) \right\}_{(j, v) \in Z \times R}.$$

The factor 2^j normalizes the $L^2(R^2)$ norm of each function. One can prove that the set of functions $h(x, y)$ whose wavelet transform satisfies the condition 1 are the functions whose orthogonal projection on U is equal to the orthogonal projection of $f(x, y)$ on U. Let O be the orthogonal complement of U in $L^2(R^2)$. This set is therefore the affine space $f + O$ of functions that can be written

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

We replace condition 2 with a convex constraint that has a similar effect in order to solve the problem numerically. We do not impose that the points $\left((x_v^j, y_v^j) \right)_{(j, v) \in Z \times R}$ are the only modulus maxima of the wavelet transform but that they minimize a Sobolev norm defined by

$$\|h\|^2 = \left| \left(W_{2^j}^1 h(x, y), W_{2^j}^2 h(x, y) \right)_{j \in Z} \right|^2$$
$$= \sum_{j=-\infty}^{+\infty} \left(\|W_{2^j}^1 h\|^2 + \|W_{2^j}^2 h\|^2 \right.$$
$$\left. + 2^{2j} \left(\|\frac{\partial W_{2^j}^1 h}{\partial x}\|^2 + \|\frac{\partial W_{2^j}^2 h}{\partial y}\|^2 \right) \right). \quad (85)$$

The minimization of this norm creates a wavelet transform whose horizontal and vertical components have an $L^2(R^2)$ norm that is as small as possible. In conjunction with condition 1, 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 $\psi^3(x, y) = \frac{\partial \psi^1(x, y)}{\partial x}$ and $\psi^4(x, y) = \frac{\partial \psi^2(x, y)}{\partial y}$. If there exist two constants $A_5 > 0$ and $B_5 > 0$ such that for all $(\omega_x, \omega_y) \in R^2$

$$A_5 \leq \sum_{j=-\infty}^{+\infty} \left(|\hat{\psi}^1(2^j \omega_x, 2^j \omega_y)|^2 + |\hat{\psi}^2(2^j \omega_x, 2^j \omega_y)|^2 \right) +$$

$$\sum_{j=-\infty}^{+\infty} \left(|\hat{\psi}^3(2^j\omega_x, 2^j\omega_y)|^2 + |\hat{\psi}^4(2^j\omega_x, 2^j\omega_y)|^2 \right) \leq B_5$$

$$(86)$$

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

$$A_5\|h\|^2 \leq \|h\|^2 \leq B_5\|h\|^2. \qquad (87)$$

Similar to (49), we prove this implication by applying the Parseval theorem on each $L^2(R^2)$ norm component of the norm defined in (85). We saw that the set of functions $h(x,y)$ whose wavelet transform satisfies condition 1 is the closed affine space $f + O$. The minimization of the norm $\| \quad \|$ over this closed convex has a unique solution whose computation might, however, not be stable. As in one dimension, we can prove that the computation of this minimum is stable if and only if the family of functions

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

$$(88)$$

is a frame of the space U that they generate. The factor 2^j normalizes the $L^2(R^2)$ norm of the functions in this family. The frame condition expresses the equivalence of the $L^2(R^2)$ norm of any function in 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 just as in one dimension. Let K the space of all sequences of the function $(g^1_j(x,y), g^2_j(x,y))$ such that

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

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

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

$$(89)$$

The set Γ is an affine space that is closed in K. Let V be the space of dyadic wavelet transforms of all functions in $L^2(R^2)$. Equation (87) proves that $V \subset K$. The sequences of functions that satisfy condition 1 are the elements of K that belong to

$$\Lambda = V \cap \Gamma.$$

To reconstruct the element of $\Gamma \cap V$ that minimizes the norm $| \quad |$, we alternate projections on Γ and V that are orthogonal with respect to the norm $| \quad |$. As in one dimension, one can prove that the orthogonal projection on V is the operator $P_V = WoW^{-1}$ that was defined by (64). The orthogonal projections P_Γ on Γ are defined in Appendix E. For a discrete image of N^2 pixels, the implementations of both P_V and P_Γ require $O(N^2 \log_2(N))$ operations. Let $P = P_VoP_\Gamma$ be the alternate projection on both sets. Since Γ is an affine space and V a vector space, for any initial sequence $X = \left(g^1_j(x,y), g^2_j(x,y) \right)_{j\in Z}$, $\lim_{n\to+\infty} P^{(n)}X$ converges strongly to
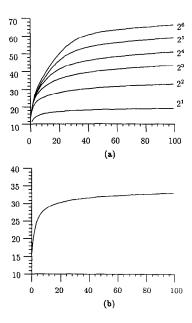

Fig. 10. (a) SNR when reconstructing the wavelet transform components $W^{1,d}_{2^j}f$ and $W^{2,d}_{2^j}f$ from the modulus maxima of the Lena image shown in Fig. 9. The abscissa gives the number of iterations on the operator P. Each curve is labeled by the scale 2^j for $1 \leq j \leq 6$; (b) SNR of the reconstructed Lena images computed with respect to the original image as a function of the number of iterations on the operator P.

the orthogonal projection of X onto $\Lambda = \Gamma \cap V$. Hence, if we begin the iteration from the zero element of K, the algorithm converges strongly to the element of Λ whose norm $| \quad |$ is minimum.

B. 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. 10(a) gives the evolution of the SNR when reconstructing the wavelet transform of the Lena image from the modulus maxima shown in the Fig. 9. After n iterations, we reconstruct an image by applying the inverse wavelet transform operator on the reconstructed wavelet transform. Fig. 10(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. 10 are very similar to the graphs of Fig. 6 that show the reconstruction SNR for a 1-D signal. The increase is fast during the first 20 iterations and then slows down. After a given number of iterations, Fig. 10(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 to which we converge, and the other one is the distance between the point to which we converge and the wavelet transform of the original image. As 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 to which we

Fig. 11. Top left: original image. Top right: image reconstructed from the maxima representation shown in the second column of Fig. 9. This reconstruction is performed with 10 alternate projections, and the SNR is 28 db. Lower left: image reconstructed from the thresholded modulus maxima shown in the third column of Fig. 9 with 10 iterations on the operator P.

converge is of the same order as the distance between the original image and the image to which we converge. Increasing the number of iterations thus does not greatly increase the SNR. The top right image in Fig. 11 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. 11, 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 singularities are not blured, 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 2-D functions such as Diracs, sinusoidal waves, step edges, Brownian noises, etc. For all these experiments, the SNR behaves similarly to Fig. 10. The visual quality of reconstructed images, with 10 iterations, is as good as in Fig. 11. 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 iteration 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 on 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. 11 is reconstructed from the modulus maxima shown in the third column of Fig. 9. 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 that is much smoother, whereas the boundaries of the shoulder are kept sharp. This thresholding can be viewed as a nonlinear noise removal technique. Hwang and Mallat [17] 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.

IX. 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 the parts of the image components that are not important for 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 [3] and Kunt et al. [13] 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. Selection of 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. 9, 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 bandwidth and, thus, cover 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 $S_{2^3}^d f$ defined in Section VI-B. The edge selection is first performed at the scale 2^2 because at the finer scale 2^1, edges are too greatly 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 efficiently code 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 $S_{2^3}^d f$. The geometry of the edge curves is coded only at the scale 2^2 because we neglect the differences between the maxima positions at the scales 2^1, 2^3, and 2^2. 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 [3] showed that this requires, on average, 10 b for the first point and 1.3 b 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 $S_{2^3}^d f$ has an energy that is 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 b.

We give, in Fig. 12, two 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 $S_{2^3}^d f$. Each original image has 256 by 256 pixels. The total amount of data to code the reconstructed images is 0.30 b per pixel for Fig. 12(a) and 0.24 b per pixel for Fig. 12(b). 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 many 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 Mallat [8] 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 distortions of edges, and a separate coding of these two types of features can be adapted to the specificity of the visual perception.

X. CONCLUSION

We showed that multiscale edges can be detected and characterized from the local maxima of a wavelet transform.
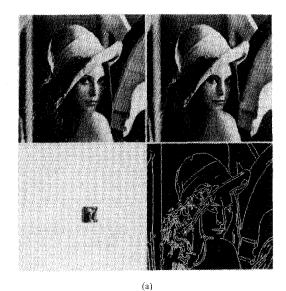

(a)

(b)

Fig. 12. Top left: original image of 256 by 256 pixels. Top right: reconstructed image from the coded multiscale edge representation. Image (a) requires 0.30 b per pixel and image and (b) requires 0.24 b per pixel. Bottom left: image $S_{2^3}^d f$ 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.

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 1-D and 2-D 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 that can be implemented in real time on a pipeline hardware architecture. 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 Mallat [8].

APPENDIX A
A PARTICULAR CLASS OF 1-D WAVELETS

This appendix defines a class of wavelets that can be used for a fast implementation of discrete algorithms. We first define the smoothing function $\phi(x)$ introduced in Section III-B, and then, we build the wavelet $\psi(x)$ and the reconstructing wavelet $\chi(x)$, which are associated with $\phi(x)$.

We impose that the Fourier transform of the smoothing function $\phi(x)$ defined by (20) 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 differentiable function such that

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

One can prove [15] that the conditions (91) are sufficient so that (90) defines a smoothing function $\phi(x)$, which is in $L^2(\mathbf{R})$. The parameter w is the sampling shift that was introduced in Section III-B. It is adjusted in order so that $\phi(x)$ 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. Equation (22) for $J = 1$ proves that $\hat{\psi}(\omega)$, $\hat{\chi}(\omega)$, and $\hat{\phi}(\omega)$ must satisfy

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

Let us impose 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. If we insert (92) and (95) into (94), we obtain

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

One can prove that (96) is sufficient to define $K(\omega)$ such that $\hat{\chi}(\omega)$ is the Fourier transform of a reconstructing wavelet that satisfies (13).

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 $|\hat{\phi}(0)| = 1$, (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 has a small compact support. A

TABLE I
FINITE IMPULSE RESPONSE OF THE FILTERS H, G, K, AND L THAT CORRESPOND TO THE QUADRATIC SPLINE WAVELET OF FIG. 1(a)

n	H	G	K	L
-3			0.0078125	0.0078125
-2			0.054685	0.046875
-1	0.125		0.171875	0.1171875
0	0.375	-2.0	-0.171875	0.65625
1	0.375	2.0	-0.054685	0.1171875
2	0.125		-0.0078125	0.046875
3				0.0078125

family of 2π periodic functions $H(\omega)$, $G(\omega)$, and $K(\omega)$ that satisfy these constraints is given by

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

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

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

From (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. 1, 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 (92) is equal to $1/2$. Equations (101) prove that $\psi(x)$ is a quadratic spline with compact support, whereas $\theta(x)$ 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 in Table I. These filters are used in fast wavelet transform computations.

APPENDIX B
FAST WAVELET ALGORITHMS FOR 1-D 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 A. We denote H_p, G_p, and K_p the discrete filters obtained by putting $2^p - 1$ zeros between each of the 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 $(j < J)$
$\qquad W_{2^{j+1}}^d f = \frac{1}{\lambda_j} \cdot S_{2^j}^d f * G_j$
$\qquad S_{2^{j+1}}^d f = S_{2^j}^d f * H_j$
$\qquad j = j + 1$
end while.

The proof of this algorithm is based on the properties of the wavelet $\psi(x)$ described in Appendix A. At each scale 2^j, we divide the values of the samples of $S_{2^j}^d f * G_j$ 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 compensate for this discrete effect. The values of λ_j that correspond to the filters of Table I are given in Table II. The border problems are treated by making symmetry and a periodization of $(S_1 f(n))_{1 \leq n \leq N}$, as explained in Section III-B. 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 nonzero coefficients in the impulse response of the filters H and G.

The inverse wavelet transform algorithm reconstructs $S_1^d f$ from the discrete dyadic wavelet transform. At each scale 2^j, it reconstructs $S_{2^{j-1}}^d f$ from $S_{2^j}^d f$ and $W_{2^j}^d f$. The complexity of this reconstruction algorithm is also $O(N \log(N))$.

$j = J$
while $(j > 0)$
$\qquad S_{2^{j-1}}^d f = \lambda_j \cdot W_{2^j}^d f * K_{j-1} + S_{2^j}^d f * \tilde{H}_{j-1}$
$\qquad j = j - 1$
end while.

APPENDIX C
A PARTICULAR CLASS OF 2-D DYADIC WAVELETS

In this appendix, we characterize the 2-D wavelets used for numerical computations. 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. Let $\psi(x)$ be a wavelet that belongs to the class described in Appendix A and whose Fourier transform is defined by

$$\hat{\psi}(2\omega) = e^{-iw\omega} G(\omega) \hat{\phi}_0(\omega) \text{ with } \hat{\phi}_0(\omega)$$
$$= e^{-iw\omega} \prod_{p=1}^{+\infty} H(2^{-p}\omega). \qquad (103)$$

We define

$$\psi^1(x, y) = \psi(x) 2\phi_0(2y) \text{ and } \psi^2(x, y) = 2\phi_0(2x)\psi(y). \qquad (104)$$

Since $\psi(x) = \frac{d\theta(x)}{dx}$, these two wavelets can be rewritten

$$\psi^1(x, y) = \frac{\partial \theta^1(x, y)}{\partial x} \text{ and } \psi^2(x, y) = \frac{\partial \theta^2(x, y)}{\partial y}, \text{ with} \qquad (105)$$

TABLE II
NORMALIZATION COEFFICIENTS λ_j FOR THE QUADRATIC
WAVELET OF FIG. 1(a). FOR $j > 5$, $\lambda_j = 1$.

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

$$\theta^1(x, y) = \theta(x) 2\phi_0(2y) \text{ and } \theta^2(x, y) = 2\phi_0(2x)\theta(y).$$

The fast discrete wavelet algorithm does not allow us to have $\theta^1(x, y) = \theta^2(x, y)$. However, the functions $\theta^1(x, y)$ and $\theta^2(x, y)$ are numerically close enough so that they can be considered to be equal to a single function $\theta(x, y)$ in a first approximation. 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

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

One can define the smoothing function $\phi(x, y)$ of Section VI-B as

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

Let us now define $\hat{\chi}^1(\omega_x, \omega_y)$ and $\hat{\chi}^2(\omega_x, \omega_y)$ such that

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

where the functions $K(\omega)$ and $L(\omega)$ are 2π periodic and satisfy

$$G(\omega)K(\omega) + |H(\omega)|^2 = 1, \qquad (107)$$
$$L(\omega) = \frac{1 + |H(\omega)|^2}{2}. \qquad (108)$$

One can prove that $\hat{\chi}^1(\omega_x, \omega_y)$ and $\hat{\chi}^2(\omega_x, \omega_y)$ are the Fourier transform of reconstructing wavelets that satisfy (61). As in Appendix A, we choose $G(\omega) = 4ie^{i\omega/2} \sin(\omega/2)$ to approximate a derivative. The 2-D wavelets used in the computations of this article are derived from the 1-D quadratic spline wavelet shown in Fig. 1. The values of the discrete filters H, G, K, and L are given in Table I.

APPENDIX D
FAST WAVELET ALGORITHMS FOR 2-D 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 C. We use the same notations as in Appendix B, 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 of the rows and columns, respectively, of the image A with the 1-D filters H and L.

The following algorithm computes the 2-D 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 $(j < J)$
$\quad W_{2^{j+1}}^{1,d} f = \frac{1}{\lambda_j} . S_{2^j}^d f * (G_j, D)$
$\quad W_{2^{j+1}}^{2,d} f = \frac{1}{\lambda_j} . S_{2^j}^d f * (D, G_j)$
$\quad S_{2^{j+1}}^d f = S_{2^j}^d f * (H_j, H_j)$
$\quad j = j + 1$
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 C. If the original image $(S_1 f(n, m))_{(n,m) \in Z^2}$ has N^2 nonzero pixels, the complexity of the algorithm is $O(N^2 \log(N))$. As explained in Section VI-B, 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.

As in the 1-D case, the reconstruction algorithm computes $S_1^d f$ by reconstructing at each scale 2^j the signal $S_{2^{j-1}}^d f$ from $S_{2^j}^d f$, $W_{2^j}^1, df$, and $W_{2^j}^2, df$. The complexity of this reconstruction algorithm is also $O(N^2 \log(N))$.

$j = J$
while $(j > 0)$
$\quad S_{2^{j-1}}^d f = \lambda_j \cdot W_{2^j}^{1,d} f * (K_{j-1}, L_{j-1}) + \lambda_j \cdot W_{2^j}^{2,d} f *$
$(L_{j-1}, K_{j-1}) + S_{2^j}^d f * (\tilde{H}_{j-1}, \tilde{H}_{j-1})$
$\quad j = j - 1$
end of while.

APPENDIX E
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 1-D reconstructions. We first study the 1-D case. The operator P_Γ transforms any sequence $(g_j(x))_{j \in Z} \in K$ into the closest sequence $(h_j(x))_{j \in Z} \in \Gamma$ with respect to the norm $| \ |$. Let $\epsilon_j(x) = h_j(x) - g_j(x)$. Each function $h_j(x)$ is chosen so that

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

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

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

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

$$\begin{cases} \epsilon_j(x_0) = W_{2^j} f(x_0) - g_j(x_0) \\ \epsilon_j(x1) = W_{2^j} f(x1) - g_j(x1). \end{cases} \qquad (110)$$

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

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

The Euler equation associated with this minimization is

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

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

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

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

In numerical computations, $W_{2^j}^d f$ 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 P_Γ modifies a discrete signal $g_j^d = (g_j(n))_{1 \leq n \leq N}$ by adding a discrete signal $\epsilon_j^d = (\epsilon_j(n))_{1 \leq n \leq N}$ that is computed from (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 P_Γ is $O(N \log_2(N))$.

We know that the modulus maxima of the original wavelet transform are only located at the positions x_n^j. 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 Y. Let $sign(x)$ be the sign of the real number x. Let Y be the set of sequences $(g_j(x))_{j \in Z} \in K$ such that for any pair of consecutive maxima positions (x_n^j, x_{n+1}^j) and $x \in [x_n^j, x_{n+1}^j]$

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

The set Y is a closed convex and $(W_{2^j} f)_{j \in Z} \in Y$. Instead of minimizing $| \ |$ over $\Gamma \cap V$ as explained in Section V-B, we can minimize it over $Y \cap \Gamma \cap V$. 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 P_Y on Y, which is not orthogonal with respect to the norm $| \ |$. Let $(g_j(x))_{j \in Z} \in K$ and $P_Y(g_j(x))_{j \in Z} = (h_j(x))_{j \in Z}$. For each index j, $h_j(x)$ is obtained by clipping the oscillations of $g_j(x)$. If the original signal has N samples, at each scale 2^j, the discrete implementation of the clipping procedure requires $O(N)$ computations. There are, at most, $\log_2(N) + 1$ scales; therefore, the total number of computations to implement P_Y is $O(N \log_2(N))$. Since this projector P_Y is not orthogonal, the iteration on the alternate projections operator $P = P_V \circ P_\Gamma \circ P_Y$ is not guaranteed to converge. Numerical experiments show that in most cases, after a few iterations, we stay inside Y, even after projections on Γ and V. Hence, the operator P_Y acts as the identity operator, and P can be rewritten $P = P_V \circ P_\Gamma$. The analysis of Section V-B proves that we are then guaranteed to converge strongly to an element in $Y \cap \Gamma \cap V$.

In two dimensions, the operator P_Γ transforms a sequence $(g_j^1(x, y), g_j^2(x, y))_{j \in Z} \in K$ into the closest sequence $(h_j^1(x, y), h_j^2(x, y))_{j \in Z} \in \Gamma$. Let $(\epsilon_j^1(x, y), \epsilon_j^2(x, y))_{j \in Z}$

be such that for any $j \in Z$, $\epsilon_j^1(x,y) = g_j^1(x,y) - h_j^1(x,y)$, and $\epsilon_j^2(x,y) = g_j^2(x,y) - h_j^2(x,y)$. The sequence $\left(h_j^1(x,y), h_j^2(x,y)\right)_{j \in Z}$ is chosen so that

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

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

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

and

$$\|\epsilon_j^2\|^2 + 2^{2j} \|\frac{\partial \epsilon_j^2}{\partial y}\|^2 \quad (116)$$

for all integers $j \in Z$. Let us concentrate on the minimization of (115). Let (x_0, y) and (x_1, y) be two consecutive modulus maxima position at a fixed y. The function $\epsilon_j^1(x,y)$ must satisfy

$$\begin{cases} \epsilon_j^1(x_0, y) = W_{2^j}^1 f(x_0, y) - g_j^1(x_0, y) \\ \epsilon_j^1(x_1, y) = W_{2^j}^1 f(x_1, y) - g_j^1(x_1, y) \end{cases} \quad (117)$$

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

$$\int_{x_0}^{x_1} \left(|\epsilon_j^1(x,y)|^2 + 2^{2j} \left| \frac{\partial \epsilon_j^1(x,y)}{\partial x} \right|^2 \right) dx. \quad (118)$$

For y fixed, we obtain a 1-D minimization problem, which is identical to the minimization of expression (111). The solution is a sum of two exponentials as in (113). This analysis shows that the solution of the 2-D minimization problem is obtained by fixing the parameter y for $\epsilon_j^1(x,y)$ and computing the 1-D solution along the x variable between two consecutive modulus maxima. The same analysis can be performed on the other component $\epsilon_j^2(x,y)$. The discrete implementation is, thus, a straightforward extension of the 1-D 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 P_Γ requires $O(N^2 \log(N))$ computations. In two dimensions, we do not introduce any sign constraint as it is done in 1-D reconstructions.

APPENDIX F
CONVERGENCE RATE OF THE ALTERNATE PROJECTIONS

We prove that if $\left(\sqrt{2^j}\psi_{2^j}(x_n^j - x) \right)_{(n,j) \in Z^2}$ is a frame of U, then the alternate projection converges exponentially, and we give a lower bound of the convergence rate. Let $X = (W_{2^j}g(x))_{j \in Z} \in V$ and $\epsilon_j(x_n^j) = W_{2^j}f(x_n^j) - W_{2^j}g(x_n^j)$ be the error at each modulus maxima location. We first prove that there exists a constant $C_1 > 0$ such that

$$\left| X - P_\Gamma X \right|^2 \geq C_1 \sum_{(n,j) \in Z^2} 2^j |\epsilon_j(x_n^j)|^2. \quad (119)$$

Let $(h_j(x))_{j \in Z} = P_\Gamma(W_{2^j}g(x))_{j \in Z} = P_\Gamma X$ and $\epsilon_j(x) = h_j(x) - W_{2^j}g(x)$. By definition

$$\left| X - P_\Gamma X \right|^2 = \left| (\epsilon_j(x))_{j \in Z} \right|^2$$
$$= \sum_{(n,j) \in Z^2} \left(\int_{x_n^j}^{x_{n+1}^j} |\epsilon_j(x)|^2 dx + 2^{2j} \int_{x_n^j}^{x_{n+1}^j} |\frac{d\epsilon_j(x)}{dx}|^2 dx \right). \quad (120)$$

We saw in Appendix E that $\epsilon_j(x)$ satisfies the differential equation (112); therefore, by integrating by parts, we obtain

$$\left| X - P_\Gamma X \right|^2$$
$$= \sum_{(n,j) \in Z^2} \left(2^{2j}\epsilon_j(x_{n+1}^j)\frac{d\epsilon_j(x_{n+1}^j)}{dx} - 2^{2j}\epsilon_j(x_n^j)\frac{d\epsilon_j(x_n^j)}{dx} \right). \quad (121)$$

We know that the function $\epsilon_j(x)$ is the sum of two exponentials given by (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 with their values specified by $\epsilon_j(x_n^j)$ and $\epsilon_j(x_{n+1}^j)$, with a few algebraic manipulations, we derive that

$$2^{2j}\epsilon_j(x_{n+1}^j)\frac{d\epsilon_j(x_{n+1}^j)}{dx} - 2^{2j}\epsilon_j(x_n^j)\frac{d\epsilon_j(x_n^j)}{dx} \geq$$
$$\frac{2^j}{4}(|\epsilon_j(x_{n+1}^j)|^2 + |\epsilon_j(x_n^j)|^2)\text{Min}(2^{-j}(x_{n+1}^j - x_n^j), 1.). \quad (122)$$

The derivative at x_{n+1}^j is the left derivative, whereas the derivative at x_n is the right derivative. Since we suppose that there exists a constant $D > 0$ such that $2^{-j}|x_n^j - x_{n-1}^j| \geq D \leq 1$, we obtain

$$\left| X - P_\Gamma X \right|^2 \geq \frac{D}{2} \sum_{(n,j) \in Z^2} 2^j |\epsilon_j(x_n^j)|^2 \quad (123)$$

which proves (119) for $C_1 = \frac{D}{2}$.

Let us now prove that there exists a constant C_2 such that for any $X = (W_{2^j}g(x))_{j \in Z} \in V$

$$\left| X - P_\Lambda X \right|^2 \leq C_2 \sum_{(n,j) \in Z^2} 2^j |\epsilon_j(x_n^j)|^2. \quad (124)$$

Let U and O be the spaces defined in Section V-B. The function $g(x)$ can be decomposed into

$$g(x) = g_1(x) + g_2(x) \quad (125)$$

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

$$f(x) = f_1(x) + f_2(x) \quad (126)$$

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

$$h(x) = f_1(x) + g_2(x). \quad (127)$$

Since $h(x) = f(x) + u(x)$, with $u(x) \in O$, we know from (46) that $h(x)$ satisfies constraint 1 and, thus, that $W_{2^j}h \in \Gamma$. We also have $h(x) - g(x) = f_1(x) - g_1(x) \in U$. Since we suppose

that $\left(\sqrt{2^j} \psi_{2^j} (x_n^j - x) \right)_{(n,j) \subset Z^2}$ is a frame of U, (53) implies that

$$\|h(x) - g(x)\|^2 \leq \frac{1}{A_3} \sum_{(n,j) \in Z^2} 2^j |\langle h(x) - g(x), \psi_{2^j}(x_n^j - x) \rangle|^2.$$

(128)

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

$$\langle h(x) - g(x), \psi_{2^j}(x_n^j - x) \rangle = W_{2^j} f(x_n^j) - W_{2^j} g(x_n^j) = \epsilon_j(x_n^j).$$

(129)

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

$$\left| (W_{2^j} h(x))_{j \in Z} - (W_{2^j} g(x))_{j \in Z} \right|^2 \leq B_2 \|h(x) - g(x)\|^2.$$

(130)

Since the projector P_Λ is orthogonal and $(W_{2^j} h(x))_{j \in Z} \in \Lambda$

$$\left| (W_{2^j} g(x))_{j \in jZ} - P_\Lambda (W_{2^j} g(x))_{j \in Z} \right|^2$$
$$\leq \left| (W_{2^j} g(x))_{j \in Z} - (W_{2^j} h(x))_{j \in Z} \right|^2.$$

(131)

Equations (128)–(131) imply that

$$\left| (W_{2^j} g(x))_{j \in Z} - P_\Lambda (W_{2^j} g(x))_{j \in Z} \right|^2$$
$$\leq \frac{B_2}{A_3} \sum_{(n,j) \in Z^2} 2^j |\epsilon_j(x_n^j)|^2.$$

(132)

This proves (120) for $C_2 = \frac{B_2}{A_3}$. From (123) and (132), we then derive that

$$|X - P_\Gamma X|^2 \geq \frac{D A_3}{2 B_2} |X - P_\Lambda X|^2.$$

(133)

This inequality gives a lower bound for the "angle" between the affine space Γ and the space V. Let $P = P_V \circ P_\Gamma$ be the alternate projection on both spaces. A classical result on alternate projections [26] enables us to derive that for any element $X \in K$, there exists a constant R such that

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

(134)

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

REFERENCES

[1] Z. Berman, "The uniqueness question of discrete wavelet maxima representation," Tech. Rep. TR 91-48, Syst. Res. Cent., Univ. of Maryland, College Park, Apr. 1991.
[2] J. Canny, "A computational approach to edge detection," IEEE Trans. Patt. Anal. Machine Intell., vol. PAMI-8, pp. 679–698, 1986.
[3] S. Carlsson, "Sketch based coding of grey level images," Signal Processing, vol. 15, pp. 57–83, July 1988.
[4] S. Curtis, S. Shitz, and V. Oppenheim, "Reconstruction of nonperiodic two-dimensional signals from zero-crossings," IEEE Trans. Acoustic Speech Signal Processing, vol. 35, pp. 890–893, 1987.
[5] I. Daubechies, "Ten lectures on wavelets," CBMS-NSF Series Appl. Math., SIAM, 1991.
[6] R. Duffin and A. Schaeffer, "A class of nonharmonic Fourier series," Trans. Amer. Math. Soc., vol. 72, pp. 341–366, 1952.
[7] G. Folland, Introduction to Partial Differential Equations. Princeton, NJ: Princeton University Press, 1976.
[8] J. Froment and S. Mallat, "Second generation compact image coding with wavelets," in Wavelets—A Tutorial in Theory and Applications(C. Chui, Ed.). New York: Academic, Jan. 1992, pp. 655–678.
[9] K. Grochenig, "Sharp results on random sampling of band-limited functions," in Proc. NATO ASI Workshop Stochastic Processes. New York: Kluwer, Aug. 1991.
[10] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," SIAM J. Math., vol. 15, pp. 723–736, 1984.
[11] R. Hummel and R. Moniot, "Reconstruction from zero-crossings in scale-space," IEEE Trans. Acoustic Speech Signal Processing, vol. 37, no. 12, Dec. 1989.
[12] J. Koenderink, "The structure of images," Biol. Cybern., 1984.
[13] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second generation image coding technics," Proc. IEEE, vol. 74, pp. 549–575, Apr. 1985.
[14] B. Logan, "Information in the zero-crossings of band pass signals," Bell Syst. Tech. J., vol. 56, pp. 510, 1977.
[15] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," IEEE Trans. Acoustic Speech Signal Processing, vol. 37, no. 12, pp. 2091–2110, Dec. 1989.
[16] ——, "Zero-crossings of a wavelet transform," IEEE Trans. Inform. Theory, vol. 37, pp. 1019–1033, July 1991.
[17] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," IEEE Trans. Inform. Theory, vol. 38, no. 2, Mar. 1992.
[18] D. Marr and E. Hildreth, "Theory of edge detection," Proc. Royal Soc. London, vol. 207, pp. 187–217, 1980.
[19] D. Marr, Vision. San Francisco: W.H. Freeman, 1982.
[20] Y. Meyer, Ondelettes et Operateurs. New York: Hermann, 1990.
[21] ——, "Un contre-exemple a la conjecture de Marr et a celle de S. Mallat," Preprint, 1991.
[22] O. Rioul and M. Vetterli, "Wavelets and signal processing," IEEE Signal Processing Mag., Oct. 1991.
[23] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," IEEE Trans. Comput., vol. C-20, pp. 562–569, 1971.
[24] J. Sanz and T. Huang, "Theorem and experiments on image reconstruction from zero-crossings," Res. Rep. RJ5460, IBM, Jan. 1987.
[25] A. Witkin, "Scale space filtering," in Proc. Int. Joint Conf. Artificial Intell., 1983.
[26] D. Youla and H. Webb, "Image restoration by the method of convex projections," IEEE Trans. Med. Imaging, vol. MI-1, pp. 81–101, Oct. 1982.
[27] A. Yuille and T. Poggio, "Scaling theorems for zero crossings," IEEE Trans. Patt. Anal. Machine Intell., vol. PAMI-8, Jan 1986.
[28] Y. Zeevi and D. Rotem, "Image reconstruction from zero-crossings," IEEE Acoustic Speech Signal Processing, vol. 34, pp. 1269–1277, 1986.
[29] S. Zhong, "Edges representation from wavelet transform maxima," Ph.D. thesis, New York Univ., Sept. 1990.

Stephane Mallat was born in Paris, France. He graduated from Ecole Polytechnique in 1984 and from Ecole Superieure des Telecommunications, Paris, in 1985. He received the Ph.D. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 1988.

Since September 1988, he has been Assistant Professor in the Computer Science Department of the Courant Institute of Mathematical Sciences, New York University, New York, NY. His research interests include computer vision, signal processing, and applied mathematics.

Dr. Mallat received the 1990 IEEE Signal Processing Society's paper award.

Sifen Zhong received the B.A. degree in 1985 in mathematics and computer science from Queen's College, New York, NY, and the M.S. and Ph.D. degrees in computer science from the Courant Institute of Mathematical Sciences, New York University, New York, NY, in 1987 and 1991, respectively.

He worked for two years as a research assistant at the New York University Robotics Research Laboratory. Presently, he is a visiting assistant researcher at the Department of Mathematics of the University of California, Los Angeles. He will join Cognitech Inc., an image processing company in Santa Monica, CA, in 1992.