Multifrequency Channel Decompositions of Images and Wavelet Models

STEPHANE G. MALLAT

Abstract—**In this paper we review recent multichannel models developed in psychophysiology, computer vision, and image processing. In psychophysiology, multichannel models have been particularly successful in explaining some low-level processing in the visual cortex. The expansion of a function into several frequency channels provides a representation which is intermediate between a spatial and a Fourier representation. We describe the mathematical properties of such decompositions and introduce the wavelet transform. We review the classical multiresolution pyramidal transforms developed in computer vision and show how they relate to the decomposition of an image into a wavelet orthonormal basis. In the last section we discuss the properties of the zero crossings of multifrequency channels. Zero-crossings representations are particularly well adapted for pattern recognition in computer vision.**

I. Introduction

WITHIN the last 10 years, multifrequency channel decompositions have found many applications in image processing. In the psychophysiology of human vision, multichannel models have also been particularly successful in explaining some low-level biological processes. The expansion of a function into several frequency channels provides a representation which is intermediate between a spatial and a Fourier representation. In harmonic analysis, this kind of transform appeared in the work of Littlewood and Payley in the 1930's. More research has recently been focused on this domain with the modeling of a new decomposition called the wavelet transform. In this paper we review the recent multichannel models developed in psychophysiology, computer vision, and image processing. We describe the motivations of the models within each of these disciplines and show how they relate to the wavelet transform.

In psychophysics and the physiology of human vision, evidence has been gathered showing that the retinal image is decomposed into several spatially oriented frequency channels. In the first section of this paper, we describe the experimental motivations for this model. Biological studies of human vision have always been a source of ideas for computer vision and image processing research. Indeed, the human visual system is generally considered to be an optimal image processor. The goal is not to imitate the processings implemented in the human brain, but rather to understand the motivations of such processings and analyze their application to computer vision problems. From this point of view, the recent experimental findings in psychophysics and physiology open challenging questions. In order to get a better understanding of multichannel decompositions, we review the main mathematical results in this domain. The best-known decomposition which is intermediate between a spatial and a frequency representation is the window Fourier transform. The window Fourier transform is used in signal processing for coding and pattern detection [47]. We describe its properties but also show why it is not a convenient decomposition for image analysis. The wavelet transform was introduced by Morlet to overcome the shortcomings of the window Fourier transform. It is computed by expanding the signal into a family of functions which are the dilations and translations of a unique function $\psi(x)$. Grossmann and Morlet [20] have shown that any function in $L^2(\boldsymbol{R})$ can be characterized from its decomposition on the wavelet family $(\sqrt{s}\,\psi(s(x-u)))_{(s,u)\in\boldsymbol{R}^2}$. A wavelet transform can be interpreted as a decomposition into a set of frequency channels having the same bandwidth on a logarithmic scale. We review the most important properties of a wavelet transform and describe its discretization as studied by Daubechies [11]. A very important particular case of discrete wavelet transform was found by Meyer [45] and Stromberg [55]. They proved that there exist some wavelets $\psi(x)$ such that $(\sqrt{2^j}\psi(2^j(x-2^{-j}n)))_{(j,n)\in\boldsymbol{Z}^2}$ is an orthonormal basis of $L^2(\boldsymbol{R})$. Wavelet orthonormal bases provide an important new tool in functional analysis. Indeed, it was believed that we could not build simple orthonormal bases of $L^2(\boldsymbol{R})$ whose elements have a good localization both in the spatial and Fourier domains. These bases have already found many applications in pure and applied mathematics [27], [33], [57], in quantum mechanics [15], [48], and in signal processing [30].

In computer vision, multifrequency channel decompositions are interpreted through the concept of multiresolution. Generally, the structures that we want to recognize have very different sizes. Hence, it is not possible to define *a priori* an optimal resolution for analyzing images. Several researchers [22], [42], [52] have developed pattern matching algorithms which process the image at different resolutions. Some pyramidal implementations have been developed for computing these decompositions [4], [10], [50]. A multiresolution transform also decomposes the signal into a set of frequency channels of constant

Manuscript received March 17, 1989. This work was supported by NSF Grant IRI-8903331.

The author is with the Computer Science Department, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012.

IEEE Log Number 8931327.

bandwidth on a logarithmic scale. It can be interpreted as a discrete wavelet transform. We review the wavelet multiresolution model [38] which provides a mathematical interpretation of the concept of resolution. We see in particular that a large class of wavelet orthonormal bases can be computed from quadrature mirror filters [39].

Multifrequency channel decompositions are well adapted for data compression in image coding. We show that this efficiency is due to the intrinsic statistical properties of images and to the ability of such representations to match the sensitivity of human vision. For pattern recognition applications, it is also necessary to build a signal representation which translates when the signal translates. Indeed, the representation of a pattern should not depend upon its position. When a pattern is translated, its representation should be translated without being modified. The pyramidal multiresolution representations as well as discrete wavelet transforms do not have this translation property. In the last section, we study the properties of representations based on zero crossings of multifrequency channels. These representations do translate, and for a particular class of band-pass filters, the zero crossings provide the location of the signal edges. It remains to show that a zero-crossing representation can provide a complete and stable signal decomposition. We review previous results on zero-crossings properties and explain how the problem can be expressed through the wavelet model.

II. Multichannel Models in Psychophysics and Physiology of Vision

In this section, we summarize some experimental results showing that a multifrequency channel decomposition seems to be taking place in the human visual cortex. For further details, we refer to tutorials by Georgeson [18] and Levine [34]. Over the past 20 years, a large effort has been devoted in psychophysics and physiology to analyze the response of the human visual system to stimuli having particular orientation and frequency tunings. Linear models have been partly successful in explaining some experimental data. The simplest, which was first developed in psychophysiology, approximates the human visual system with a linear filter. Fig. 1 illustrates the anatomical pathway in the human visual system. Photoreceptors in the eyes measure the light input intensity. This information is processed by bipolar and ganglion cells in the retina and is transmitted through the optic nerve. The optic nerve ends in a relay station (the lateral geniculate nucleus) whose axons extend to the visual cortex.

Replacing these different stages by a global linear filter is clearly an extremely simplified model, but it gives some insights about the visual system sensitivity. Given this hypothesis, Campbell and Green [6] tried to measure the global transfer function of the visual system. In their experiments, the visual stimuli shown to the observer were vertical sinusoidal gratings of different spatial frequencies (see Fig. 2).

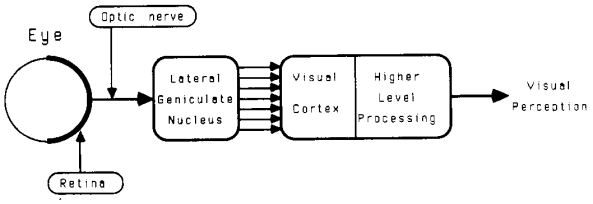


Fig. 1. Illustration of the anatomical visual pathway. The higher level processes are the least understood and are difficult to evaluate in psychophysical experiments.

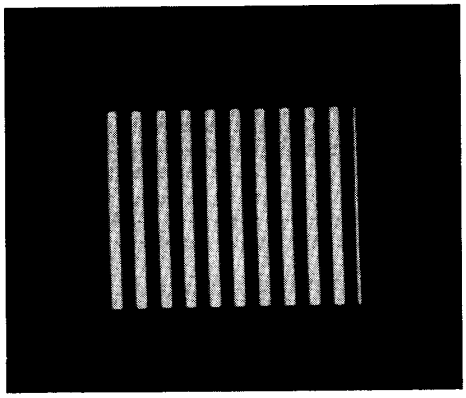


Fig. 2. This image is a typical visual stimulus used in psychological experiments for computing the transfer function of the visual system. It consists of a sinusoidal grating whose frequency varies during the experiment. In order to evaluate the sensitivity to orientation, these gratings are rotated.

In psychophysics, frequencies are measured in cycles per degree of visual angle subtended on the eye. The transfer function $H(\omega)$ of the visual system is defined as the ratio of the contrast perceived by the observer to the real contrast of the stimulus for sinusoidal gratings of frequency ω. The contrast is given by

$$C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}},$$

where $L_{\max}$ and $L_{\min}$ are the maximum and minimum luminance of the stimuli. In order to estimate this transfer function, a solution which is widely adopted is to measure the *Contrast Sensitivity Function*. At each frequency ω, we measure the minimum contrast $C_t(\omega)$ necessary to distinguish the sinusoidal gratings from a uniform background. This contrast is called the contrast threshold. The contrast sensitivity function is then defined by

$$CSF(\omega) = \frac{1}{C_t(\omega)}, \quad \text{and} \quad H(\omega) = CSF(\omega).$$

Many experiments [5], [6], [31] have been performed to measure the function $CSF(\omega)$ and they agree approximately with the function shown in Fig. 3. Although this linear model is clearly oversimplified, it shows qualitatively the sensitivity of the human system to stimuli of different frequencies.

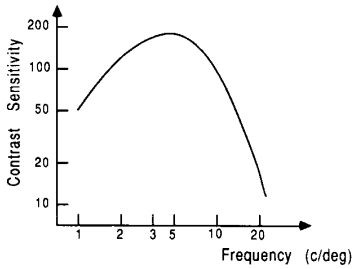


Fig. 3. Contrast Sensitivity Function (redrawn from Kulikowski and King-Smith [31]). The visual system has the maximum sensitivity to contrast when the frequency of the stimulus is around 5 cycles/deg.

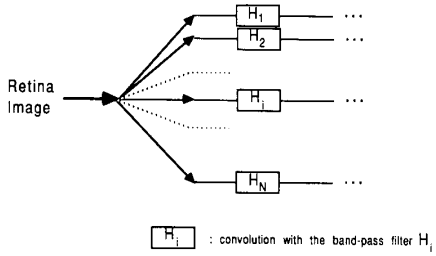


Fig. 4. Multichannel model. The retinal image is supposed to be filtered by independent band-pass filters. These filters have approximately the same bandwidth on a logarithmic scale and have a spatial orientation selectivity.

With further experiments, Campbell and Robson [8] have shown that the retinal image is likely to be processed in separate frequency channels. These experiments were based on adaptation techniques. If a stimulus is shown to an observer for a long time, the visual sensitivity for the same kind of stimuli decreases. This behavior is called an adaptation process. Campbell and Robson [8] have shown that if the visual system adapts to a sinusoidal grating of a given frequency ω_0, the sensitivity decreases for any stimuli whose frequency is in a frequency band around ω_0. However, outside this frequency band, the sensitivity is not affected. These experiments indicate that at some stage, the visual information in different frequency bands is processed separately. Researchers in psychophysics have tried to measure the width of these bands. In order to simplify the analysis of the problem, Campbell and Robson supposed that the retinal image is decomposed through independent band-pass linear filters as shown in Fig. 4. Their first estimate of the frequency bandwidth of these filters was very narrow. However, other experiments by Georgeson [17] and Nachmias [46] have since contradicted their results. They showed that the frequency bandwidth of these filters is more likely to be around one octave. In other words, the retina image seems to be decomposed in several frequency bands having approximately the same width on a logarithmic scale.

Other psychophysical experiments have shown that the visual sensitivity to a sinusoidal grating also depends upon its spatial orientation. The results of Campbell and Kulikowski [7] show that the human visual system has a maximum sensitivity when the signal has an orientation of 0° or 90°. In between, the sensitivity decreases monotonically reaching a minimum at 45°. The filters of the model shown in Fig. 4 must therefore have a spatial orientation selectivity.

This filter bank model only provides a qualitative description of some low-level processing of the visual system. In particular, it does not take into account the nonlinearities of the biological processes. However, recent physiological experiments support such approaches. Cell recordings are generally performed on cats and monkeys which have a visual cortex similar to the human one. In the cat's visual cortex, Hubel and Wiesel [23] discovered a class of cells whose response depends upon the frequency and orientation of the visual stimuli. These cells are called simple cells. Maffei and Fiorentini [35] have shown that their response is reasonably linear and that they can be modeled with linear filters. Several groups of researchers have recorded the impulse responses of simple cells [2], [36], [59]. These studies showed that the bandwidths of simple cells range from 0.6 to 2.0 octaves with an average value of 1.3 octaves. The response of simple cells also depends upon the spatial orientation of the stimuli. Fig. 5 shows the two-dimensional impulse response of simple cells measured by Webster and De Valois [61]. These impulse responses have been modeled by Daugmann [12], [13] with Gaussians modulated by sinusoidal waves. As explained in the next section, these functions generate a particular window Fourier transform called the Gabor transform. Fig. 5 shows the comparison between the impulse response of a simple cell and the corresponding Gabor function model. These graphs clearly show that a simple cell behaves like a band-pass filter with a spatial orientation tuning. The support of the impulse response of a cell is called the receptive field. It corresponds to the domain of the retina where the input light influences the cell firings. Simple cells have a receptive field of varying size depending on their frequency tuning [49].

Much evidence has now been gathered about this multifrequency channel modeling of the low-level visual cortex processing. However, we do not know what type of information is extracted from this decomposition and how it relates to further processing by complex and hypercomplex cells [49]. Since the human visual cortex is an excellent image processor, this low-level biological model raises important questions from an image processing point of view. What is the advantage of decomposing a signal into several frequency channels? Is it related to the intrinsic statistical properties of images? Does it lead to a better reorganization of the image information? If we do accept that such a decomposition offers a useful representation of images, it remains to find out how to process these different frequency channels. What type of information do we want to extract? Should we process each channel independently or compare the values of the signal from band to band? In the following sections, we show that some

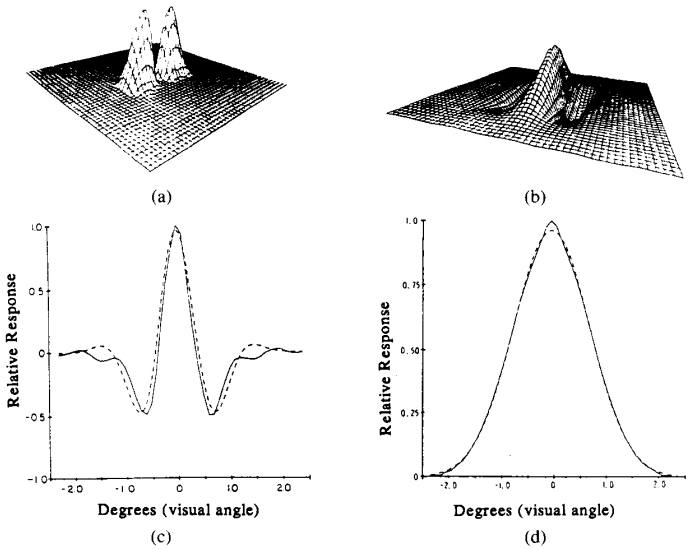


Fig. 5. (Reprint from Webster and De Valois [61].) (a) This surface is the two-dimensional transfer function of a simple cell. It is a band-pass oriented filter. Its bandwidth is 0.94 octaves. (b) Impulse response computed by taking the inverse Fourier transform of (a). (c) and (d) Cross sections of the impulse response respectively along the x and y axes. The dashed lines give the best fitting Gabor functions.

results in mathematics, computer vision, and image coding give elements of answers to these questions. Our primary goal is not to build a model of the human visual cortex but rather to justify the use of such decompositions in image processing.

III. Mathematical Analysis of Multichannel Models

In this section we review the mathematical properties of multifrequency channel decompositions. We do believe that a good mathematical understanding of these decompositions is necessary in order to evaluate their range of applications in image processing. We summarize the most relevant mathematical results in this domain. No proof is written, but references to original works are given. Most results are first introduced for one-dimensional functions and then generalized to two dimensions if needed. We review the properties of the window Fourier transform which is the most well-known intermediate decomposition between spatial and Fourier representations. This decomposition has already found many applications in signal coding and pattern detection [47]. We describe the drawbacks of the window Fourier transform for analyzing signals like images. The wavelet transform is then introduced and compared to the window Fourier transform. More details can be found in a complete article by Daubechies [11] and an advanced functional analysis book by Meyer [44].

Notation: $\boldsymbol{Z}$, $\boldsymbol{R}$, and $\boldsymbol{R}^+$ denote, respectively, the sets of integers, real numbers, and positive real numbers. $\boldsymbol{L}^2(\boldsymbol{R})$ denotes the Hilbert space of measurable, square-integrable one-dimensional functions $f(x)$. We suppose that our signals are finite energy functions $f(x) \in \boldsymbol{L}^2(\boldsymbol{R})$. For a pair of functions $f(x) \in \boldsymbol{L}^2(\boldsymbol{R})$, $g(x) \in \boldsymbol{L}^2(\boldsymbol{R})$, the inner product of $f(x)$ with $g(x)$ is written

$$\langle g(x), f(x) \rangle = \int_{-\infty}^{+\infty} g(x)\overline{f(x)}\, dx, \tag{1}$$

where $\overline{f(x)}$ is the complete conjugate of $f(x)$. The norm of $f(x)$ in $\boldsymbol{L}^2(\boldsymbol{R})$ is given by

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

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

$$f * g(u) = \int_{-\infty}^{+\infty} f(x)g(u - x)\, dx. \tag{3}$$

The dilation of a function $f(x) \in \boldsymbol{L}^2(\boldsymbol{R})$ by a scaling factor s is written

$$f_s(x) = \sqrt{s}\, f(sx). \tag{4}$$

The reflection of $f(x)$ about 0 is written

$$\tilde{f}(x) = f(-x). \tag{5}$$

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

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)e^{-i\omega x}\, dx. \tag{6}$$

A. Definition of a Window Fourier Transform

From the Fourier transform of a function $f(x)$, we get a measure of the irregularities (high frequencies) but this information is not spatially localized. Indeed, the Fourier transform $\hat{f}(\omega)$ is defined through an integral which covers the whole spatial domain. It is therefore difficult to find the position of the irregularities. In order to localize the information provided by the Fourier transform, Gabor [16] defined a new decomposition using a spatial window $g(x)$ in the Fourier integral. This window is translated along the spatial axis in order to cover the whole signal. At a position u and for a frequency ω, the window Fourier transform of a function $f(x) \in L^2(R)$ is defined by

$$Gf(\omega, u) = \int_{-\infty}^{+\infty} e^{-i\omega x}\, g(x-u)\, f(x)\, dx. \qquad (7)$$

It measures locally, around the point u, the amplitude of the sinusoidal wave component of frequency ω. In the original Gabor transform, the window function $g(x)$ is a Gaussian. It has since been generalized for any type of window function and is called a window Fourier transform [28]. The window function is generally a real even function and the energy of its Fourier transform is concentrated in the low frequencies (see Fig. 6). It can be viewed as the impulse response of a low-pass filter. For normalization purposes, we suppose that the energy of $g(x)$ is equal to 1:

$$\|g\|^2 = \int_{-\infty}^{+\infty} |g(x)|^2\, dx = 1.$$

Let us denote

$$g_{\omega_0, u_0}(x) = e^{i\omega_0 x} g(x - u_0).$$

A window Fourier transform can also be interpreted as the inner products of the function $f(x)$ with the family of functions $(g_{\omega,u}(x))_{(\omega,u)\in R^2}$:

$$Gf(\omega, u) = \langle f(x), g_{\omega,u}(x) \rangle. \qquad (8)$$

In quantum physics, such a family of functions is called a family of coherent states. The Fourier transform $g_{\omega,u_0}(x)$ is given by

$$\hat{g}_{\omega_0,u_0}(\omega) = e^{-iu_0\omega} \hat{g}(\omega - \omega_0), \qquad (9)$$

where $\hat{g}(\omega)$ is the Fourier transform of $g(x)$. A family of coherent states thus corresponds to a translation in the spatial domain (parameter u) and in the frequency domain (parameter ω) of the function $g(x)$ (see Fig. 6). This double translation is represented in a phase-space where one axis corresponds to the spatial parameter u and the other to the frequency parameter ω (see Fig. 7). Families of coherent states have found many applications in quantum physics because they make it possible to analyze simultaneously a physical phenomena in both the spatial and frequency domains.

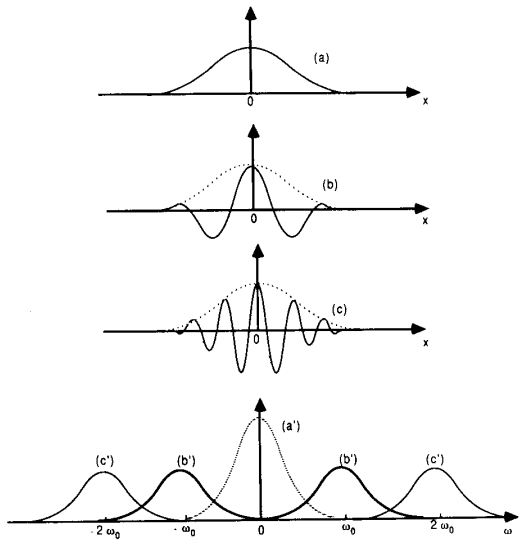


Fig. 6. (a) Window function $g(x)$. (b) Graph of $g(x)\cos(\omega_0 x)$. (c) Graph of $g(x)\cos(2\omega_0 x)$. All these curves have the same support but the number of cycles varies with the frequency of the sinusoidal modulation. The curves (a'), (b'), (c') are, respectively, the Fourier transform of $g(x)$, $g(x)\cos(\omega_0 x)$, and $g(x)\cos(2\omega_0 x)$. They have the same bandwidth but different positions on the frequency axis.

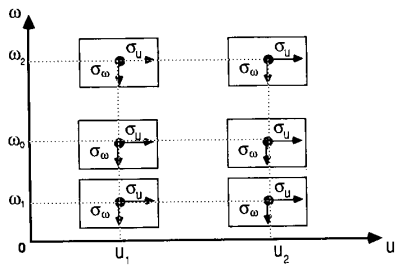


Fig. 7. Phase-space representation. The vertical axis gives the frequency ω whereas the horizontal axis gives the spatial position u. A window Fourier coefficient $Gf(\omega_0, u_0)$ provides a description of $f(x)$ within the resolution cell of $[u_0 - \sigma_u, u_0 + \sigma_u] \times [\omega_0 - \sigma_\omega, \omega_0 + \sigma_\omega]$.

Let us now describe how a window Fourier transform relates to a spatial or a frequency representation. Let σ_u be the standard deviation of $g(x)$

$$\sigma_u^2 = \int_{-\infty}^{+\infty} x^2 |g(x)|^2\, dx. \qquad (10)$$

Let σ_ω be the standard deviation of the Fourier transform of $g(x)$

$$\sigma_\omega^2 = \int_{-\infty}^{+\infty} \omega^2 |\hat{g}(\omega)|^2\, d\omega. \qquad (11)$$

The function $g_{\omega_0,u_0}(x)$ is centered in u_0 and has a standard deviation σ_u in the spatial domain. Its Fourier transform given by (9) is centered in ω_0 and has a standard deviation

σ_ω. By applying the Parseval theorem on (8), we get

$$Gf(\omega_0, u_0) = \int_{-\infty}^{+\infty} f(x)\overline{g_{\omega_0, u_0}(x)}\, dx$$

$$= \int_{-\infty}^{+\infty} \hat{f}(\omega)\overline{g_{\omega_0, u_0}(\omega)}\, d\omega. \qquad (12)$$

The first integral shows that in the spatial domain, $Gf(\omega_0, u_0)$ essentially depends upon the values of $f(x)$ for $x \in [u_0 - \sigma_u, u_0 + \sigma_u]$. The second integral proves that in the frequency domain, $Gf(\omega_0, u_0)$ depends upon the values of $\hat{f}(\omega)$ for $\omega \in [\omega_0 - \sigma_\omega, \omega_0 + \sigma_\omega]$. The spatio-frequency domain which is covered by $Gf(\omega_0, u_0)$ can thus be represented in the phase-space by the resolution cell $[u_0 - \sigma_u, u_0 + \sigma_u] \times [\omega_0 - \sigma_\omega, \omega_0 + \sigma_\omega]$ as shown in Fig. 7. The surface and shape of the resolution cell is independent from u_0 and ω_0. The uncertainty principle applied to the function $g(x)$ implies that

$$\sigma_u^2 \sigma_\omega^2 \geq \frac{\pi}{2}. \qquad (13)$$

The resolution cell can therefore not be smaller than $2\sqrt{2\pi}$. The uncertainty inequality reaches its upper limit if and only if $g(x)$ is a Gaussian. Hence, the resolution in the phase-space is maximized when the window function is a Gaussian as in the Gabor transform.

B. Properties of a Window Fourier Transform

A window Fourier transform is an isometry (to a proportionality coefficient) from $L^2(R)$ into $L^2(R^2)$

$$\int_{-\infty}^{+\infty} \left| f(x) \right|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| Gf(\omega, u) \right|^2 d\omega\, du. \qquad (14)$$

The function $f(x)$ is reconstructed from $Gf(\omega, u)$ with the formula

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Gf(\omega, u) g(u - x) e^{i\omega x}\, d\omega\, du. \qquad (15)$$

Equations (14) and (15) are proved by applying the Parseval theorem and using the definition of $Gf(\omega, x)$ given in (7).

A window Fourier transform is a redundant representation. If instead of computing $Gf(\omega, u)$ for all values $(\omega, u) \in R^2$ we sample uniformly both ω and u, the representation can still be complete and stable. Let u_0 and ω_0 be the sampling intervals in both domains. A discrete Fourier transform is defined by

$$\forall n \in Z, \quad \forall m \in Z \quad G_d f(m, n) = Gf(m\omega_0, nu_0)$$

$$= \int_{-\infty}^{+\infty} e^{-im\omega_0 x} g(x - nu_0) f(x)\, dx. \qquad (16)$$

This discretization corresponds to a uniform sampling of the phase-space as shown in Fig. 8. A discrete window Fourier transform is equivalent to a division of the frequency axis into intervals separated by ω_0 (see Fig. 6). In each of these intervals, the signal is sampled at a rate $1/u_0$. Daubechies [11] made a thorough study of the completeness and stability of a discrete window Fourier transform. Intuitively, the sampling intervals u_0 and ω_0 must be chosen in order to cover the whole phase-space with the resolution cells shown in Fig. 7. Formally, to reconstruct any function $f(x) \in L^2(R)$ from the set of sample $(G_d f(n, m))_{(n,m)\in Z^2}$, the operator

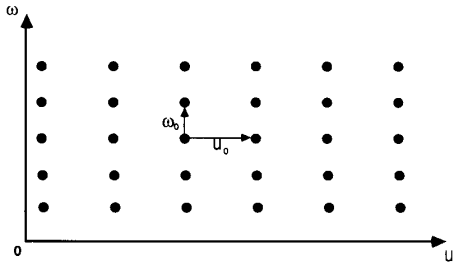


Fig. 8. Sampling pattern of a discrete window Fourier transform in the phase-space. Since the resolution cells are identical everywhere in the phase-space, the sampling is uniform.

$$L^2(R) \xrightarrow{G_d} l^2(Z^2)$$

must be invertible on its range and have a bounded inverse. Each sample $G_d f(n, m)$ can also be expressed as an inner product in $L^2(R)$

$$G_d f(m, n) = \langle f(x), e^{im\omega_0 x} g(x - nu_0) \rangle$$

$$= \langle f(x), g_{m\omega_0, nu_0}(x) \rangle. \qquad (17)$$

The properties of a discrete Fourier transform thus depend upon the family of functions $(g_{m\omega_0, nu_0}(x))_{(n,m)\in Z^2}$. In order to invert G_d, Daubechies [11] has shown that ω_0 and u_0 must verify

$$\omega_0 u_0 < 2\pi.$$

When $\omega_0 u_0 = 2\pi$, we reach the Nyquist frequency limit and G_d does not have a bounded inverse. When $\omega_0 u_0 < 2\pi$, the range of Gd has a complicated structure.

Although several researchers have tried to model the impulse response of simple cells with Gabor functions, it is unlikely that the human visual cortex implements some type of window Fourier transform. Indeed, we saw that a window Fourier transform decomposes a function into a set of frequency intervals having the same size. On the other hand, experimental data indicate that the retinal image is decomposed into a set of frequency channels having approximately a constant bandwidth on a logarithmic scale (octave). The measured impulse responses of simple cells do not have an increasing number of cycles for a constant envelope as in a window Fourier transform (see Fig. 6). Rather, they have a support (receptive field) of varying size.

Although some researchers [58] have been using the Gabor transform in computer vision, this decomposition has several drawbacks when applied to image analysis.

We saw that the spatial and frequency resolution of a window Fourier transform is constant. In the spatial domain, the information provided by this decomposition is therefore unlocalized within intervals of size σ_u. The standard deviation σ_u of $g(x)$ defines a resolution of reference. If the signal has a discontinuity such as an edge, with a window Fourier transform, it is difficult to locate this edge with a precision better than σ_u (see Fig. 9). This localization limit is generally not acceptable. If the signal has important features of very different sizes, we cannot define an optimal resolution for analyzing the signal. This is typically the case with images. For example, in the image of a house, the pattern we want to analyze might range from the overall structure of the house to the details on one of the window curtains. With a given window size, it is difficult to analyze both the fine and the large structures. This fixed resolution also introduces misleading high frequencies when decomposing local features. Let $e(x)$ be an edge as shown in Fig. 9, and suppose that

$$e(x) = \begin{cases} 0 & \text{if } x \leq x_0 - \dfrac{\Delta x}{2} \\ \dfrac{1}{2} + \dfrac{1}{2}\sin\left(\dfrac{\pi}{\Delta x}(x - x_0)\right) & \text{if } x_0 - \dfrac{\Delta x}{2} < x < x_0 + \dfrac{\Delta x}{2} \\ 1 & \text{if } x \geq x_0 + \dfrac{\Delta x}{2}. \end{cases}$$

Let us denote $\omega_0 = \pi/\Delta x$. One would expect that at the point x_0, the decomposition coefficients $Ge(\omega, x_0) = \langle e(x), e^{i\omega x} g(x - x_0) \rangle$ decrease very quickly when ω gets larger than ω_0. Indeed, in the neighborhood of x_0, the edge $e(x)$ is a sinusoidal wave of frequency ω_0. In reality, this property does not hold because the edge is very localized and has only half of the sinusoidal wave period. As a consequence, when the frequency ω is large with respect to ω_0, the modulus of the coefficients $Ge(x_0, \omega)$ decreases slowly. Although the signal $e(x)$ is locally a pure sinusoidal wave of frequency ω_0, at a frequency $2\omega_0$, the window Fourier coefficient $|Ge(x_0, 2\omega_0)|$ is still about half the value of $|Ge(x_0, \omega_0)|$. This numerical property makes it hard to interpret the window Fourier coefficients when the features are very localized with respect to the size of the support of $g(x)$. More details about this property can be found in the article of Daubechies [11]. A window Fourier transform is better suited for analyzing signals where all the patterns appear approximately at the same scale.

In order to avoid the inconvenience of a transform having a fixed resolution in the spatial and frequency domains, Morlet defined a decomposition based on dilations. In the next section, we describe the properties of this decomposition which is called the wavelet transform.

C. Definition of a Wavelet Transform

Morlet [20] defined the wavelet transform by decomposing the signal into a family of functions which are the translation and dilation of a unique function $\psi(x)$. The

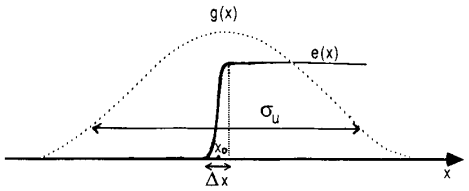


Fig. 9. With a window Fourier transform, a local feature such as an edge $e(x)$ cannot be located with a precision better than the variance σ_u of the window function $g(x)$. Since the variation step Δx of the edge $e(x)$ is small with respect to σ_u, in the neighborhood of x_0, the window Fourier transform of $e(x)$ decreases slowly when the frequency ω gets larger than $\omega_0 = \pi/\Delta x$.

function $\psi(x)$ is called a wavelet and the corresponding wavelet family is given by $(\sqrt{s}\,\psi(s(x - u)))_{(s,u)\in R^2}$. The wavelet transform of a function $f(x) \in L^2(R)$ is defined by

$$Wf(s, u) = \int_{-\infty}^{+\infty} f(x)\sqrt{s}\,\psi(s(x - u))\,dx. \quad (19)$$

The idea behind the wavelet decomposition is not new. It is very much related to some other types of spatial-frequency decompositions such as the Wigner–Ville transform. Some versions of the wavelet transform have been studied independently under other names such as the scale-space decomposition of Witkin [62], and in mathematics its origin can be traced back to be beginning of the century. However, the formalization effort of Morlet and Grossmann [20] opened a broader field of applications and has led to important new mathematical results. Let us denote the dilation of $\psi(x)$ with a factor s by

$$\psi_s(x) = \sqrt{s}\,\psi(sx). \quad (20)$$

A wavelet transform can be rewritten as inner products in $L^2(R)$

$$Wf(s, u) = \langle f(x), \psi_s(x - u) \rangle.$$

It thus corresponds to a decomposition of $f(x)$ on the family of functions $(\psi_s(x - u))_{(s,u)\in R^2}$. As shown in Fig. 10, the functions $\psi_s(x)$ have the same type as $\psi(x)$, but have a support s times smaller. In the following, we suppose that the wavelet $\psi(x)$ and the signal $f(x)$ have real values. As explained later, in order to reconstruct $f(x)$ from its wavelet transform, the Fourier transform $\hat{\psi}(\omega)$ of $\psi(x)$ must satisfy

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\,d\omega < +\infty. \quad (21)$$

This condition implies that $\hat{\psi}(0) = 0$, and that $\hat{\psi}(\omega)$ is small enough in the neighborhood of $\omega = 0$. The function $\psi(x)$ can be interpreted as the impulse response of a bandpass filter. For normalization purposes, we suppose that the energy of $\psi(x)$ is equal to 1. Let us denote $\tilde{\psi}_s(x) = \psi_s(-x)$. We can rewrite the wavelet transform at a point u and a scale s as a convolution product with $\tilde{\psi}_s(x)$

$$Wf(s, u) = f * \tilde{\psi}_s(u). \quad (22)$$

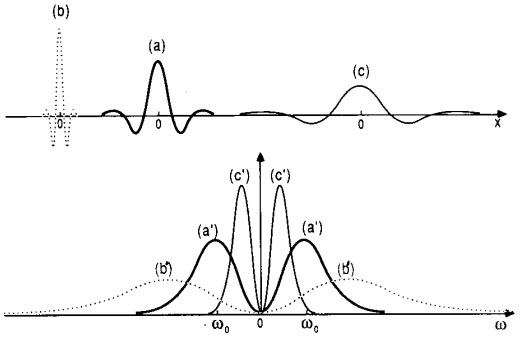


Fig. 10. (a) Graph of a wavelet $\psi(x)$. (b) Graph of $\psi_{s_1}(x)$ for $s_1 > 1$. (c) Graph of $\psi_{s_2}(x)$ for $s_2 < 1$. The curves (a'), (b'), and (c') are, respectively, the Fourier transform of the function shown in (a), (b), and (c). They have the same rms bandwidth on a logarithmic scale.

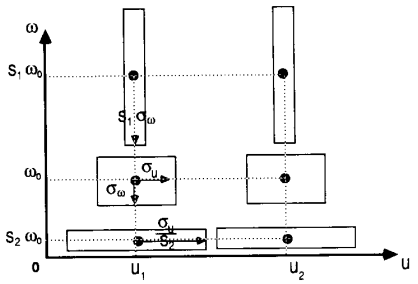


Fig. 11. In the phase-space, the shape of a wavelet resolution cell depends upon the scale. When the scale increases, the resolution increases in the spatial domain and decreases in the frequency domain. The surface of all the resolution cells is the same.

A wavelet transform can therefore be viewed as a filtering of $f(x)$ with a band-pass filter whose impulse response is $\tilde{\psi}_s(x)$. From (20), we derive that the Fourier transform of $\psi_s(x)$ is given by

$$\hat{\psi}_s(\omega) = \frac{1}{\sqrt{s}}\hat{\psi}\left(\frac{\omega}{s}\right).$$

In opposition to a window Fourier transform which has a fixed resolution in the spatial and frequency domain, the resolution of a wavelet transform varies with the scale parameter s. Since $\psi(x)$ is real, $|\hat{\psi}(\omega)| = |\hat{\psi}(-\omega)|$. Let ω_0 be the center of the passing band of $\hat{\psi}(\omega)$

$$\int_0^{+\infty} (\omega - \omega_0)\left|\hat{\psi}(\omega)\right|^2 d\omega = 0.$$

Let σ_ω be the rms bandwidth around ω_0

$$\sigma_\omega^2 = \int_0^{+\infty} (\omega - \omega_0)^2\left|\hat{\psi}(\omega)\right|^2 d\omega.$$

It is clear that the center of the passing band of $\hat{\psi}_s(\omega)$ is $s\omega_0$ and that its rms bandwidth is $s\sigma_\omega$. On a logarithmic scale, the rms bandwidth of $\hat{\psi}_s(\omega)$ is the same for all $s \in \boldsymbol{R}^+$. Hence, a wavelet transform decomposes the signal into a set of frequency bands having a constant size on a logarithmic scale (see Fig. 10).

Let σ_u be the standard deviation of $|\psi(x)|^2$ around zero. One can also show easily that the wavelet $\psi_s(x - u_0)$ has an energy concentrated around u_0 within a standard deviation σ_u/s. In the frequency domain, we saw that its energy is concentrated around $s\omega_0$ within a standard deviation $s\sigma_\omega$. In the phase-space, the resolution cell of this wavelet is therefore equal to $[u_0 - (\sigma_u/s), u_0 + (\sigma_u/s)] \times [s\omega_0 - s\sigma_\omega, s\omega_0 + s\sigma_\omega]$. As opposed to a window Fourier transform, the shape of the resolution cell varies with the scale s. This is illustrated in Fig. 11. When the scale s is small, the resolution is coarse in the spatial domain and fine in the frequency domain. If the scale s increases, the resolution increases in the spatial domain and decreases in the frequency domain (see Fig. 11). In the next section, we show that this variation of resolution enables the wavelet transform to zoom into the irregularities of the signal and characterize them locally.

For some applications, it can be useful to use a complex wavelet $\psi(s)$ in order to separate a phase and modulus component from the wavelet transform. For this purpose, Morlet and Grossmann are using wavelets whose Fourier transform $\hat{\psi}(\omega)$ is equal to zero for $\omega < 0$ [20]. Such functions are called Hardy functions. The wavelet transform $Wf(s, u)$ is then a complex number. When the scale s is fixed and u varies, the function $Wf(s, u)$ is also a Hardy function. The phase and the modulus of the wavelet transform can easily be separated for any given scale s and position u. Separating the phase and energy component of the wavelet transform signal has found some applications in speech processing [30].

Remark: There is a common misunderstanding in the psychophysiological and computer vision literature around Gabor and wavelet transforms. A Gabor function is a Gaussian modulated by a sinusoidal wave. A Gabor function satisfies the condition (21) and is therefore an admissible wavelet. If we build a transform based on a dilation of this function, it will be a wavelet transform and not a Gabor transform (window Fourier transform). Indeed, in order to define a Gabor transform, we must modify the frequency of the sinusoidal modulation without changing the size of the window function. This is much more than a terminology problem since the properties of a wavelet transform and a Gabor transform are very different.

D. Properties of a Wavelet Transform

Morlet and Grossmann [20] have shown that the wavelet transform is an isometry (to a proportionality coefficient) from $\boldsymbol{L}^2(\boldsymbol{R})$ into $\boldsymbol{L}^2(\boldsymbol{R}^+ \times \boldsymbol{R})$

$$\int_{-\infty}^{+\infty}\int_0^{+\infty} \left|Wf(s, u)\right|^2 ds\, du = C_\psi \int_{-\infty}^{+\infty} \left|f(x)\right|^2 dx. \tag{23}$$

The constant C_ψ is defined by

$$C_\psi = \int_0^{+\infty} \frac{\left|\hat{\psi}(s\omega)\right|^2}{s} ds = \int_0^{+\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{\omega} d\omega < +\infty.$$

Equation (23) is proved by applying the Parseval theorem and using the definition of $Wf(s, u)$ given in (19). Similarly, we can derive that the reconstruction of $f(x)$ from $Wf(s, u)$ is given by

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} Wf(s, u)\psi_s(x - u)\, ds\, du. \quad (24)$$

Like a window Fourier transform, a wavelet transform is redundant. In other words, the value of $Wf(s', u')$ depends upon the values of $Wf(s, u)$ for $s \neq s'$ and $u \neq u'$. By inserting (24) in the definition (19) of a wavelet transform, one can show that the function $Wf(s, u)$ satisfies the following reproducing kernel equation [21]:

$$\forall (s', u') \in \boldsymbol{R}^+ \times \boldsymbol{R}, \quad Wf(s', u') = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} Wf(s, u)K(s, s', u, u')\, ds\, du, \quad (25)$$

where

$$K(s, s', u, u') = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \psi_s(x - u)\psi_{s'}(x - u')\, dx.$$

$K(s, s', u, u')$ is called a reproducing kernel. It expresses the redundancy between $Wf(s, u)$ and $Wf(s', u')$ for any two pairs of points (s, u) and (s', u'). Equation (25) shows that, *a priori*, any function $F(s, u) \in \boldsymbol{L}^2(\boldsymbol{R}^+ \times \boldsymbol{R})$ is not the wavelet transform of some function $f(x) \in \boldsymbol{L}^2(\boldsymbol{R})$. One can easily prove that there exists a function $f(x) \in \boldsymbol{L}^2(\boldsymbol{R})$ such that $F(s, u) = Wf(s, u)$, if and only if

$$\forall (s', u') \in \boldsymbol{R}^+ \times \boldsymbol{R}, \quad F(s', u') = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} F(s, u)K(s, s', u, u')\, ds\, du. \quad (26)$$

The function $f(x)$ is then given by

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} F(s, u)\psi_s(x - u)\, dx\, du. \quad (27)$$

The reproducing kernel equation is an important characterization of a wavelet transform that we use later.

The wavelet transform can be discretized by sampling both the scale parameter s and the translation parameter u. In order to build a complete representation, we must cover the phase-space with the resolution cells shown in Fig. 11. This can be done with an exponential sampling of the scale parameter. We first select a sequence of scales $(\alpha^j)_{j \in \boldsymbol{Z}}$, where α is the elementary dilation step. We saw in (22) that the wavelet transform $Wf(\alpha^j, u)$ can be rewritten

$$Wf(\alpha^j, u) = f * \tilde{\psi}_{\alpha^j}(u). \quad (28)$$

For each scale α^j, $\tilde{\psi}_{\alpha^j}(x)$ has a Fourier transform centered in $\alpha^j\omega_0$ with an rms bandwidth of $\alpha^j\sigma_\omega$. Equation (28) can therefore be interpreted as a decomposition of $f(x)$ in a set of frequency channels centered in $\alpha^j\omega_0$ and whose rms bandwidth is $\alpha^j\sigma_\omega$. In order to characterize the decomposed signal in each channel, we must sample it uniformly at a rate proportional to α^j. Let α^j/β be the sampling rate at the scale α^j. The discrete wavelet transformed is defined by

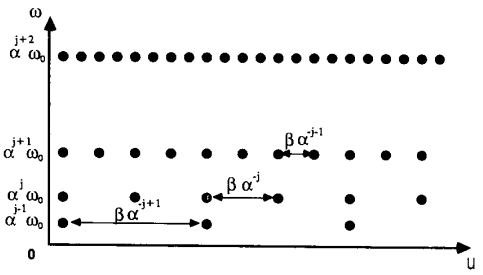


Fig. 12. Sampling of the phase-space corresponding to a discrete wavelet transform (adapted from Daubechies [11]). Each sample corresponds to an inner product with a particular wavelet. This sampling pattern is adapted to the shape of the wavelet resolution cells at the different scales (see Fig. 11).

$$W_d f(j, n) = Wf\left(\alpha^j, \frac{n\beta}{\alpha^j}\right) = \int_{-\infty}^{+\infty} f(x)\psi_{\alpha^j}\left(x - \frac{n\beta}{\alpha^j}\right) dx = f * \tilde{\psi}_{\alpha^j}\left(\frac{n\beta}{\alpha^j}\right). \quad (29)$$

Fig. 12 illustrates this sampling pattern in the phase-space. When the scale increases, the density of samples increases.

It is not possible to understand the properties of this transform by using the Nyquist theorem since the Fourier transform of $\psi(x)$ does not have a compact support (it is not strictly band-limited). With an approach similar to her study of the discrete window Fourier transform, Daubechies [11] analyzed the main properties of a discrete wavelet transform. She made a clear comparison of these two types of multichannel decompositions from a mathematical point of view. In order to reconstruct a function $f(x)$ from the discrete wavelet transform $(W_d f(j, n))_{(n,j) \in \boldsymbol{Z}^2}$, the operator

$$\boldsymbol{L}^2(\boldsymbol{R}) \xrightarrow{W_d} \boldsymbol{l}^2(\boldsymbol{Z}^2) \quad (30)$$

must be invertible on its range and have a bounded inverse. Since

$$W_d(j, n) = \left\langle f(x), \psi_{\alpha^j}\left(x - \frac{n\beta}{\alpha^j}\right)\right\rangle, \quad (31)$$

the properties of the operator W_d depend upon the family of functions $(\psi_{\alpha^j}(x - (n\beta/\alpha^j)))_{(n,j) \in \boldsymbol{Z}^2}$. Daubechies [11] studied the properties of this family of functions and gave some necessary and sufficient conditions on α, β, and $\psi(x)$ so that the operator W_d admits a bounded inverse.

A very important class of discrete wavelet transform was found independently by Meyer [45] and Stromberg [55]. They showed that there exist some wavelets $\psi(x) \in \boldsymbol{L}^2(\boldsymbol{R})$ such that $(\psi_{2^j}(x - (n/2^j)))_{(n,j) \in \boldsymbol{Z}^2}$ is an orthonormal basis of $\boldsymbol{L}^2(\boldsymbol{R})$. These particular wavelets are called *orthogonal wavelets*. A wavelet orthonormal basis corresponds to a discrete wavelet transform for $\alpha = 2$ and

$\beta = 1$. Wavelet orthonormal bases can be built for sequences of scales other than $(2^j)_{j\in Z}$, but we will concentrate on dyadic scales which lead to simpler decomposition algorithms. These new orthonormal bases had a striking impact in functional analysis. It was indeed believed that one could not find simple orthonormal bases whose elements have a good localization both in the spatial and frequency domains. Any function can be reconstructed from its decomposition into a wavelet orthonormal basis with the classical expansion formula of a vector into an orthonormal basis

$$f(x) = \sum_{j\in Z}\sum_{n\in Z} \langle f(u), \psi_{2^j}(u - n2^{-j})\rangle \psi_{2^j}(x - n2^{-j}). \tag{32}$$

The Haar basis is a well-known particular case of wavelet orthonormal basis. The orthogonal wavelet corresponding to the Haar basis is given by

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \le x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

The Haar wavelet is not continuous, which is a major inconvenience for many applications. Meyer [45] showed that we can find some orthogonal wavelets $\psi(x)$ which are infinitely continuously differentiable and whose decay at infinity are faster then any power x^{-n}, $n > 0$. In Section IV-A, we show that the Fourier transform of a large class of orthogonal wavelets can be expressed from the transfer function of a quadrature mirror filter [38]. The decomposition of a function in such a wavelet orthonormal basis can be computed with a quadrature mirror filter bank. Fig. 13 gives the graph of a particular orthogonal wavelet and its Fourier transform. This wavelet is a cubic spline studied independently by Lemarie [32] and Battle [3].

An important property of a wavelet transform is to easily characterize the local regularity of a function. This can have a particularly interesting application for discriminating image textures. In mathematics, it leads to a simple characterization of the classical functional spaces such as the $L^p(R)$ spaces, the Sobolev spaces, the Holder spaces, etc. Let us give an example. One way to measure the local regularity of a function is to measure the lipschitz exponent. A function $f(x)$ is lipschitz α in the neighborhood of a point x_0, if and only if, for any point x in a neighborhood of x_0,

$$|f(x) - f(x_0)| = O(|x - x_0|^{\alpha}). \tag{34}$$

A function which is differentiable in x_0 is lipschitz 1. The larger the lipschitz coefficient α, the smoother the function is in the neighborhood of x_0. Let us now suppose that the wavelet $\psi(x)$ is continuously differentiable. We also assume that our signal $f(x)$ is continuous and that there exist $\epsilon > 0$ such that $f(x)$ is lipschitz ϵ everywhere. Jaffard [26] proved that for any $\alpha > 0$, one can find whether

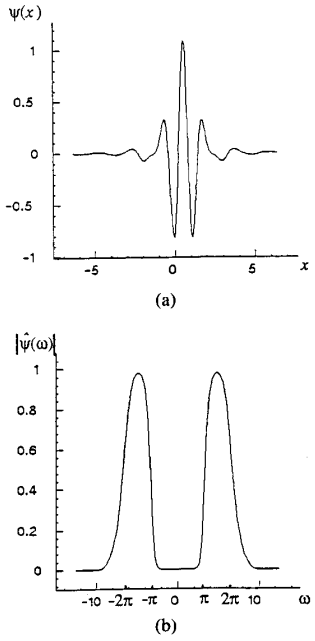


Fig. 13. (a) Example of orthogonal wavelet $\psi(x)$. (b) Modulus of its Fourier transform. The wavelet $\psi(x)$ can be interpreted as the impulse response of a band-pass filter. This particular wavelet is a cubic spline.

$f(x)$ is lipschitz α at x_0 by measuring the decay of wavelet coefficients in the neighborhood of x_0. More precisely, $f(x)$ is lipschitz α at x_0 if and only if

$$\exists C > 0, \quad \forall n \in Z, \quad \left|\langle f(x), \psi_{2^j}(x - n2^{-j})\rangle\right| \le C2^{-j(1/2+\alpha)}(1 + |2^j x_0 - n|). \tag{35}$$

The regularity of a function at a point x_0 thus depends upon the decay rate of the wavelet coefficients in the neighborhood of x_0, when the scale increases. Other kinds of regularity, such as the derivability at any order (in the sense of Sobolev), can be derived similarly [33]. These results show that it is necessary to combine the information at different scales in order to analyze the local properties of a function. In the next section, we describe the extension of the wavelet model to two-dimensional signals. We come back to orthonormal wavelets in Section IV-A to explain their relation to the concept of multiresolution in computer vision.

E. Wavelet Transform in Two Dimensions

The wavelet transform can be generalized in R^n, but we only consider the two-dimensional case for image processing applications. The model can first be extended without distinguishing any spatial orientation. Let $\Psi(x, y) \in L^2(R^2)$ be a function whose Fourier transform $\hat{\Psi}(\omega_x, \omega_y)$ satisfies

$$\forall(\omega_x, \omega_y) \in R^2 \int_0^{+\infty} \frac{\left|\hat{\Psi}(s\omega_x, s\omega_y)\right|^2}{s} ds = C_\Psi < +\infty. \tag{36}$$

The value of the integral (36) must be finite and constant for all $(\omega_x, \omega_y) \in \boldsymbol{R}^2$. For example, this property is satisfied for a wavelet $\Psi(x, y)$ which is isotropic ($\Psi(x, y) = \rho(\sqrt{x^2 + y^2})$) and whose Fourier transform is null at the origin ($\hat{\Psi}(0, 0) = 0$). For normalization purposes, we suppose that $\| \Psi \| = 1$. The function $\Psi(x, y)$ can be interpreted as the impulse response of a band-pass filter having no preferential spatial orientation. The wavelet transform of a function $f(x, y) \in \boldsymbol{L}^2(\boldsymbol{R}^2)$ at the scale s and a point (u, v) is defined by

$$Wf(s, (u, v)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) s\Psi(s(x - u), s(y - v))\, dx\, dy. \tag{37}$$

Let $\Psi_s(x, y) = s\Psi(sx, sy)$ and $\tilde{\Psi}_s(x, y) = \Psi_s(-x, -y)$. The wavelet transform of $f(x, y)$ at the scale s and a point (u, v) can be rewritten as a convolution product

$$Wf(s, (u, v)) = f * \tilde{\Psi}_s(u, v). \tag{38}$$

It can be interpreted as a two-dimensional band-pass filtering with no orientation selectivity. The wavelet transform in two dimensions has the same properties as a one-dimensional wavelet transform. There is an energy conservation equation

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \left| Wf(s, (u, v)) \right|^2 s\, ds\, du\, dv = C_\Psi \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| f(x, y) \right|^2 dx\, dy. \tag{39}$$

As in the one-dimensional case, this equation is proved with the Parseval theorem. We can also reconstruct a function $f(x, y)$ from its wavelet transform with a simple two-dimensional extension of (24):

$$f(x, y) = \frac{1}{C_\Psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} Wf(s, (u, v)) \cdot \Psi_s(x - u, y - v) s\, ds\, du\, dv. \tag{40}$$

In two dimensions, a wavelet transform also satisfies a reproducing kernel equation similar to (25).

For image recognition applications, it is often necessary to have a decomposition which differentiates the local orientation of the image features. Let us define N wavelet functions $\Psi^i(x, y)$ $(1 \le i \le N)$ whose Fourier transform $\hat{\Psi}^i(\omega_x, \omega_y)$ satisfies

$$\sum_{i=1}^{N} \left| \hat{\Psi}^i(\omega_s, \omega_y) \right|^2 = \left| \hat{\Psi}(\omega_s, \omega_y) \right|^2. \tag{41}$$

Fig. 14 shows an example of decomposition of $\hat{\Psi}(\omega_x, \omega_y)$ into the different functions $\hat{\Psi}^i(\omega_x, \omega_y)$. In the example shown in Fig. 14, the decomposition is symmetrical, but this is not a constraint of the model. Each function $\Psi^i(x, y)$ can be viewed as the impulse response of a band-pass filter having a particular orientation tuning. The wavelet

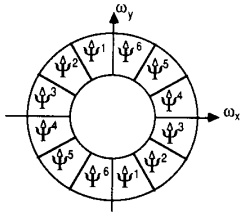


Fig. 14. Decomposition in the Fourier domain of the support of $\hat{\Psi}(\omega_x, \omega_y)$ into 6 wavelets $\hat{\Psi}^i(\omega_x, \omega_y)$ $(1 \le i \le 6)$ having different orientation selectivities. In this example, the supports of the functions $\hat{\Psi}^i(\omega_x, \omega_y)$ are symmetrical about 0 and are rotated from one another.

transform within the orientation i is defined by

$$W^i f(s, (u, v)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) s\Psi^i(s(x - u), s(y - v))\, dx\, dy. \tag{42}$$

Let $\Psi_s^i(x, y) = s\Psi^i(sx, sy)$ and $\tilde{\Psi}_s^i(x, y) = \Psi_s^i(-x, -y)$. The wavelet transform of $f(x, y)$ at the scale s and a point (u, v), within the orientation i, can be rewritten

$$W^i f(s, (u, v)) = f * \tilde{\Psi}_s^i(u, v). \tag{43}$$

It can thus be interpreted as a filtering of $f(x, y)$ with a band-pass filter having an orientation selectivity. Similar to (37), the wavelet decomposition in several orientations defines an isometry from $\boldsymbol{L}^2(\boldsymbol{R}^2)$ into $\boldsymbol{L}^2(\boldsymbol{R}^+ \times \boldsymbol{R}^2)$

$$\sum_{i=1}^{N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \left| W^i f(s, (u, v)) \right|^2 s\, ds\, du\, dv = C_\Psi \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| f(x, y) \right|^2 dx\, dy. \tag{44}$$

We can also reconstruct a function $f(x, y)$ from its wavelet transform decomposed into several directions

$$f(x, y) = \frac{1}{C_\Psi} \sum_{i=1}^{N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} Wf^i(s, (u, v)) \cdot \Psi_s^i(x - u, y - v) s\, ds\, du\, dv.$$

The discretization of a wavelet transform in two dimensions is similar to the discretization in one dimension. We choose a sequence of scales $(\alpha^j)_{j \in Z}$ where α is the elementary dilation step. For each scale α^j, the translation vector (u, v) is uniformly sampled on a two-dimensional grid at a rate proportional to α^j. In the next section, we study the two-dimensional extension of the orthonormal wavelet decomposition and its implementation.

IV. Computer Vision and Multiresolution Decomposition

Let us now analyze the multiresolution approach to image interpretation. A multiresolution decomposition is also an image decomposition in frequency channels of constant bandwidth on a logarithmic scale. It provides a different perspective on this kind of transform. We describe the classical pyramidal implementation of multi-

resolution transforms and show how it relates to a discrete wavelet decomposition.

Multiresolution transforms have been thoroughly studied in computer vision since the work of Rosenfeld and Thurston [51] on multiscale edge detection, and the Marr theory of low-level vision [40]. At different resolutions, the details of an image generally characterize different types of physical structures. For example, a coarse resolution satellite image of a coast gives a description of only the overall shape of the coast. When the resolution of the image is increased, we are able to successively distinguish the local relief of the region, and if the resolution gets even finer, we can recognize the different types of local vegetation. In order to process these different structures separately, researchers in computer vision have tried to extract the difference of information between the approximation of an image at two different resolutions. Given a sequence of increasing resolutions $(r_j)_{j\in Z}$, the details of $f(x)$ at the resolution r_j are defined as the difference of information between the approximation of $f(x)$ at the resolution r_{j+1} and the approximation at the resolution r_j.

A multiresolution representation also provides a simple hierarchical framework for interpreting the image information [29]. In some sense, the details of the image at a coarse resolution provide the "context" of the image, whereas the finer details correspond to the particular "modalities." For example, it is difficult to recognize that a small rectangle inside an image is the window of a house if we did not previously recognize the house "context." It is therefore natural to first analyze the image details at a coarse resolution and then increase the resolution. This is called a coarse-to-fine processing strategy. At a coarse resolution, the image details are characterized by very few samples. Hence, the coarse information processing can be performed quickly. The finer details are characterized by more samples, but the prior information, derived from the context, constrains and thus speeds up the computations. With a coarse-to-fine strategy, we process the minimum amount of details which are necessary to perform a recognition task. Indeed, if we can recognize an object from a coarse description, we do not need to analyze the finer details. For example, in order to distinguish a car from a house, the coarse details of the image should be enough. Such a strategy is efficient for pattern recognition algorithms. It has already been widely studied for low-level image processing tasks such as stereo matching and template matching [19], [22].

A. Pyramidal Multiresolution Decompositions

The approximation of a signal $f(x)$ at a resolution r is defined as an estimate of $f(x)$ derived from r measurements per unit length. These measurements are computed by uniformly sampling at a rate r the function $f(x)$ smoothed by a low-pass filter whose bandwidth is proportional to r. In order to be consistent when the resolution varies, these low-pass filters are derived from a unique function $\theta(x)$ which is dilated by the resolution factor $r: \theta_r = \sqrt{r}\theta(rx)$. The set of measurements $A_r f = (f * \theta_r(n/r))_{n\in Z}$ is called a discrete approximation of $f(x)$ at the resolution r. In the following, we study the approximation of a function on a dyadic sequence of resolutions $(2^j)_{j\in Z}$. The discrete approximation of a function $f(x)$ at the resolution 2^j is thus given by

$$A_{2^j} f = \left(f * \theta_{2^j}\left(\frac{n}{2^j}\right)\right)_{n\in Z}. \tag{45}$$

Tanimoto and Pavlidis [56], Burt [4], and Crowley [10] have developed efficient algorithms to compute the approximation of a function at different resolutions. We first describe these decompositions and then explain the Burt and Crowley algorithms for computing the details at different resolutions. The details are regrouped in a pyramid data structure called a Laplacian pyramid. This simple and elegant algorithm does not define the details from the difference of information between $A_{2^{j+1}} f$ and $A_{2^j} f$. At different resolutions, the details computed with this algorithm are correlated. It is thus difficult to know whether a similarity between the image details at different resolutions is due to a property of the image itself or to the intrinsic redundancy of the representation. We review the multiresolution wavelet model which shows that the difference of information between two successive resolutions can be computed by decomposing the signal in a wavelet orthonormal basis.

In pyramidal multiresolution algorithms, the low-pass filter function $\theta(x)$ is chosen such that its Fourier transform can be written

$$\hat{\theta}(\omega) = \prod_{p=1}^{+\infty} U(e^{-i2^{-p}\omega}), \tag{46}$$

where $U(e^{-i\omega})$ is the transfer function of a low-pass discrete filter $U = (u_n)_{n\in Z}$. Daubechies [11] studied the regularity and decay at infinity of the function $\theta(x)$ depending upon the properties of the filter $U(e^{-i\omega})$. In general, we want to have a function $\theta(x)$ which is as smooth as possible and which is well concentrated around 0 in the spatial domain.

Let us suppose that we have already computed the discrete approximation of a function $f(x) \in L^2(R)$ at the resolution $2^{j+1}: A_{2^{j+1}} f = (f * \theta_{2^{j+1}}(n/2^{j+1}))_{n\in Z}$. One can show [4], [11], [38] that the discrete approximation of $f(x)$ at a resolution 2^j is calculated by filtering $A_{2^{j+1}} f$ with the discrete low-pass filter $U = (u_n)_{n\in Z}$ and keeping every other sample of the convolution product. Let $\Lambda = (\lambda_n)_{n\in Z}$ be such that

$$\Lambda = A_{2^{j+1}} f * U, \tag{47}$$

then

$$A_{2^j} f = (\lambda_{2n})_{n\in Z}. \tag{48}$$

A measuring device provides the approximation of an input signal at a finite resolution. Let us suppose for normalization purposes that this resolution is equal to one. The approximation of this signal at any resolution 2^{-J}, J

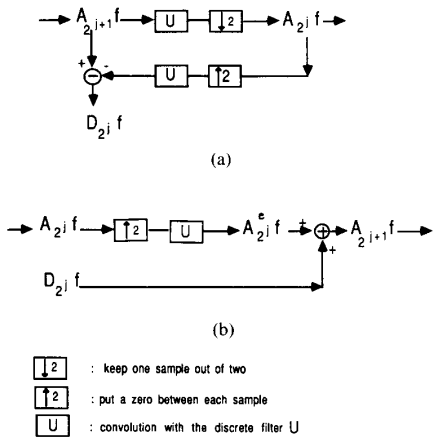


Fig. 15. (a) Decomposition of $A_{2^{j+1}}f$ into $A_{2^j}f$ and $D_{2^j}f$ when computing a Laplacian pyramid. (b) Reconstruction of $A_{2^{j+1}}f$ from $A_{2^j}f$ and $D_{2^j}f$ when reconstructing the original signal from a Laplacian pyramid.

> 0, can be computed by iterating on (47) and (48), and j varying between 0 and $J + 1$. This pyramidal algorithm is illustrated in Fig. 15(a). The set of discrete approximations $(A_{2^j}f)_{0 \geq j \geq -J}$ was called a Gaussian pyramid by Burt [4].

We now describe the algorithm of Burt [4] and Crowley [10] in order to extract the details of $f(x)$ which appear in $A_{2^{j+1}}f$ but not in $A_{2^j}f$. The discrete approximation $A_{2^{j+1}}f$ has twice as many samples as $A_{2^j}f$, so we first expand $A_{2^j}f$ by a factor of two. This is performed with a classical interpolation procedure [9]. We put a zero between each sample of $A_{2^j}f$ and filter the resulting signal with a low-pass filter. In this algorithm, the low-pass filter is the filter U defined previously. Let $A^e_{2^j}f$ be the expanded discrete signal. The details $D_{2^j}f$ at the resolution 2^j are then computed by subtracting $A^e_{2^j}f$ from $A_{2^{j+1}}f$

$$D_{2^j}f = A_{2^{j+1}}f - A^e_{2^j}f. \tag{49}$$

This algorithm decomposes a discrete approximation $A_1 f$ at a resolution of 1 into an approximation $A_{2^{-J}}f$ at a coarse resolution 2^{-J} and the successive detail signals $(D_{2^j}f)_{0<j\leq -J}$. If the signal $A_1 f$ has N nonzero samples, each detail signal $D_{2^j}f$ has $2^{j+1}N$ samples, whereas the coarse signal $A_{2^{-j}}f$ has $2^{-j}N$ samples. Hence, the total number of samples of this representation is approximately $2N$. The signals $\{A_{2^{-J}}f, (D_{2^j}f)_{0<j\leq -J}\}$ are regrouped in a data structure called a Laplacian pyramid [4].

The original signal can easily be reconstructed from such a decomposition. At each resolution, we compute $A_{2^{j+1}}f$ by expanding $A_{2^j}f$ by a factor two and adding the details $D_{2^j}f$. By repeating this algorithm when j is varying between $-J$ and 0, we reconstruct $A_1 f$. The reconstruction algorithm is illustrated by a block diagram in Fig. 15(b).

In two dimensions, the discrete approximation of a signal $f(x, y) \in L^2(\boldsymbol{R}^2)$ at the resolution 2^j is similarly defined by

$$A_{2^j}f = \left(f * \Theta_{2^j}(2^{-j}n, 2^{-j}m)\right)_{(n,m)\in \boldsymbol{Z}^2}, \tag{50}$$

where $\Theta(x, y)$ is a two-dimensional low-pass filter, and $\Theta_{2^j}(x, y) = 2^j\Theta(2^j x, 2^j y)$. For image processing, the pyramidal algorithm is extended with separable convolutions along the rows and columns of the image [4]. The low-pass filter $\Theta(x, y)$ is chosen such that its Fourier transform can be written

$$\hat{\Theta}(\omega_x, \omega_y) = \prod_{p=1}^{+\infty} U(e^{-i2^{-p}\omega_x})U(e^{-i2^{-p}\omega_y}).$$

Let us suppose that the video camera provides an image approximated at the resolution 1: $A_1 f = (f * \Theta(n, m))_{(n,m)\in \boldsymbol{Z}^2}$. With a separable extension of the algorithm described in (47) and (48), we can compute the approximation of an image at any resolutions 2^j, $(j < 0)$. Fig. 16 shows an image approximated at the resolution 2^j for $0 \geq j \geq -3$ (Gaussian pyramid). The detail signals $(D_{2^j}f)_{0<j\leq -3}$ can also be computed with a straightforward extension of the one-dimensional algorithm. Fig. 17 shows the Laplacian pyramid of the image given in Fig. 16. If the original image has N^2 pixels, each detail image $D_{2^j}f$ has $2^{j+1}N^2$ pixels and $A_{2^{-j}}f$ has $2^{-J}N^2$ pixels. Hence, the total number of pixels of this representation is approximately $\frac{4}{3}N^2$.

In a Laplacian pyramid, the signals $D_{2^j}f$ do not correspond to the difference of information between $A_{2^{j+1}}f$ and $A_{2^j}f$. If they did, the total number of pixels representing the signal would be the same as in the original signal. We saw that the number of samples representing the signal is increased by a factor of 2 in one dimension and by a factor of $\frac{4}{3}$ in two dimensions. This is due to the correlation between the detail signals $D_{2^j}f$ at different resolutions. The correlation can be understood and suppressed with the multiresolution wavelet model described in [39] and [38]. It is indeed possible to extract exactly the difference of information between $A_{2^{j+1}}f$ and $A_{2^j}f$ by decomposing the signal into a wavelet orthonormal basis.

Let us first explain the multiresolution wavelet model in one dimension. We saw in (45) that the discrete approximation of a function $f(x)$ at the resolution 2^j is defined by $A_{2^j}f = (f * \theta_{2^j}(2^{-j}n))_{n\in \boldsymbol{Z}}$. Let us denote $\tilde{\theta}_{2^j}(x) = \theta_{2^j}(-x)$. Each convolution product in a point can be rewritten as an inner product in $L^2(\boldsymbol{R})$

$$A_{2^j}f = \left(\langle f(x), \tilde{\theta}_{2^j}(x - 2^{-j}n)\rangle\right)_{n\in \boldsymbol{Z}}. \tag{51}$$

Let us call the *continuous approximation* of $f(x)$ at the resolution of 2^j the best estimate of $f(x)$ given the sequence of inner products $A_{2^j}f$. By "best" we mean as close as possible to $f(x)$ with respect to the $L^2(\boldsymbol{R})$ distance (mean square distance). One can easily derive from the projection theorem that this best estimate is equal to the orthogonal projection of $f(x)$ on the vector space V_{2^j} generated by the family of functions $(\tilde{\theta}_{2^j}(x - 2^{-j}n))_{n\in \boldsymbol{Z}}$. The vector space V_{2^j} can be viewed as the set of all possible approximations of functions at the resolution 2^j. The sequence of vector spaces $(V_{2^j})_{j\in \boldsymbol{Z}}$ is called a *multiresolution approximation* of $L^2(\boldsymbol{R})$. The proper-

Fig. 16. Gaussian pyramid. The image is approximated at the resolutions 1, $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{8}$. As the resolution decreases, higher resolution details are lost and the image is characterized by fewer pixels.

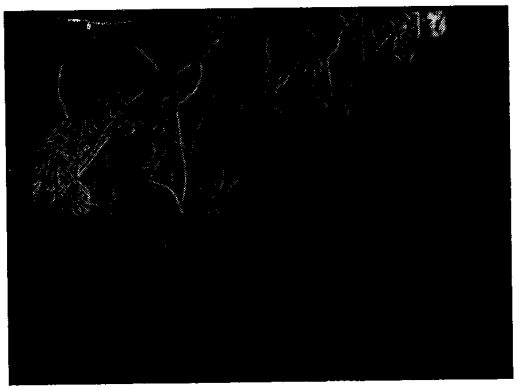


Fig. 17. Laplacian pyramid. This figure shows the detail images at the resolution $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and the coarse image approximated at the resolution $\frac{1}{8}$. At each resolution, the pixels of the detail image have a large amplitude when the original image is not "smooth" at the corresponding location.

ties of the vector space V_{2^j} are further studied in [38] and [39]. For any function $f(x) \in L^2(R)$, the continuous approximation of $f(x)$ at the resolution 2^j is thus given by the orthogonal projection of $f(x)$ on V_{2^j}. In order to compute this approximation, we need an orthonormal basis of V_{2^j}. One can show [39] that we can build such an orthonormal basis by dilating and translating a particular function $\phi(x)$ called a *scaling function*. For any resolution 2^j, let us denote $\phi_{2^j}(x) = \sqrt{2^j}\phi(2^j x)$. The family of functions $(\phi_{2^j}(x - 2^{-j}n))_{n\in Z}$ is then an orthonormal basis of V_{2^j}. The Fourier transform of $\phi(x)$ is characterized by

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} H(e^{-i2^{-p}\omega}), \tag{52}$$

where $H(e^{-i\omega})$ is the transfer function of a discrete filter [39]. One can show that $H(e^{-i\omega})$ satisfies the condition

$$\left|H(e^{-i\omega})\right|^2 + \left|H(-e^{-i\omega})\right|^2 = 1. \tag{53}$$

The discrete filters $H = (h_n)_{n\in Z}$ whose transfer function satisfy (53) are called *quadrature mirror* filters [14].

The orthogonal projection of a function $f(x) \in L^2(R)$ on V_{2^j} can now be computed by decomposing $f(x)$ into the orthonormal basis $(\phi_{2^j}(x - 2^{-j}n))_{n\in Z}$. Let $P_{V_{2^j}}$ be

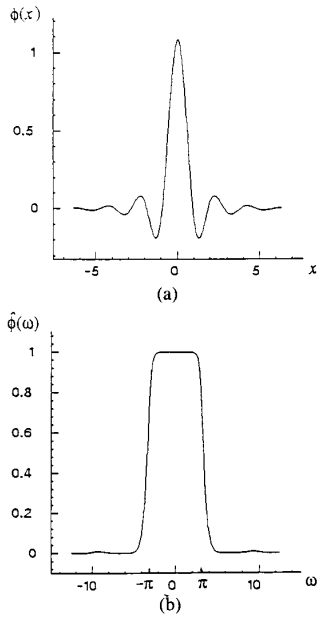


Fig. 18. (a) Example of scaling function $\phi(x)$. (b) Fourier transform $\hat{\phi}(\omega)$. A scaling function can be interpreted as the impulse response of a low-pass filter. The computation of this particular function is described in [38]. The corresponding orthogonal wavelet is shown in Fig. 13.

the orthogonal projection operator on V_{2^j}:

$$P_{V_{2^j}}(f)(x) = \sum_{n\in Z} \langle f(u), \phi_{2^j}(u - 2^{-j}n)\rangle \, \phi_{2^j}(x - 2^{-j}n). \tag{54}$$

Let us denote $\tilde{\phi}(x) = \phi(-x)$. Since $\phi(x)$ is a low-pass filter, we can redefine the discrete approximation $A_{2^j}f$ with the function $\tilde{\phi}(x)$ instead of $\theta(x)$

$$A_{2^j}f = \left(f * \tilde{\phi}_{2^j}(2^{-j}n)\right)_{n\in Z} = \left(\langle f(x), \phi_{2^j}(x - 2^{-j}n)\rangle\right)_{n\in Z}. \tag{55}$$

The best estimate of $f(x)$ can easily be derived from this discrete approximation by using (54). Let $\tilde{H}$ be the discrete filter whose impulse response is $(h_{-n})_{n\in Z}$. From (52) and (55), one can show [38] that the discrete approximations, $A_{2^j}f$, are computed with the same pyramidal algorithm described in (47) and (48), by using the discrete filter $\tilde{H}$ instead of U. Fig. 18 gives the graph of a scaling function $\phi(x)$.

Let us now explain how to extract exactly the difference of information between the approximations of a function at the resolutions 2^j and 2^{j+1}. The approximations of a function $f(x) \in L^2(R)$ at the resolutions 2^j and 2^{j+1} are given by the orthogonal projection of $f(x)$ on the vector spaces V_{2^j} and $V_{2^{j+1}}$, respectively. Intuitively, the approximation at the resolution 2^{j+1} must give a better estimate of $f(x)$ than the approximation at the resolution 2^j. Hence, the vector spaces V_{2^j} and $V_{2^{j+1}}$ should satisfy

$$V_{2^j} \subset V_{2^{j+1}}. \tag{56}$$

The difference of information between the approximations at the resolutions 2^j and 2^{j+1} is therefore equal to the

orthogonal projection of $f(x)$ on the orthogonal complement of V_{2^j} in $V_{2^{j+1}}$. Let O_{2^j} be this orthogonal complement. The vector space O_{2^j} is orthogonal to V_{2^j} and satisfies

$$O_{2^j} \oplus V_{2^j} = V_{2^{j+1}}.$$

To compute the orthogonal projection of a function $f(x)$ on O_{2^j}, we need to find an orthonormal basis of O_{2^j}. One can show [39] that such an orthonormal basis can be built by dilating and translating a particular wavelet $\psi(x)$. For any resolution 2^j, let us denote $\psi_{2^j}(x) = \sqrt{2^j}\psi(2^j x)$. The family of functions $(\psi_{2^j}(x - 2^{-j}n))_{n\in Z}$ is then an orthonormal basis of O_{2^j}. The Fourier transform of $\psi(x)$ is given by

$$\hat{\psi}(2\omega) = G(e^{-i\omega})\hat{\phi}(\omega)$$

$$\text{with } G(e^{-i\omega}) = e^{-i\omega}\overline{H(e^{-i\omega})}. \quad (57)$$

$G(e^{-i\omega})$ is the transfer function of a discrete filter $G = (g_n)_{n\in Z}$. The filters G and H make a pair of quadrature mirror filters [54].

When the resolution 2^j varies between 0 and $+\infty$, the family of functions $(\psi_{2^j}(x - 2^{-j}n))_{(n,j)\in Z^2}$ constitutes a wavelet orthonormal basis of $L^2(R)$ [39]. This shows that the multiresolution concept and quadrature mirror filters are directly related to wavelet orthonormal bases.

Let $P_{O_{2^j}} f(x)$ be the orthonormal projection of a function $f(x) \in L^2(R)$ on the vector space O_{2^j}. $P_{O_{2^j}} f(x)$ gives the difference of information between the approximations of $f(x)$ at the resolutions of 2^j and 2^{j+1}. It can be computed by expanding $f(x)$ in the orthonormal basis of O_{2^j}

$$P_{O_{2^j}} f(x) = \sum_{n\in Z} \langle f(u), \psi_{2^j}(u - 2^{-j}n)\rangle \psi_{2^j}(x - 2^{-j}n). \quad (58)$$

This difference of information is characterized by the set of inner products

$$D_{2^j} f = \left(\langle f(x), \psi_{2^j}(x - 2^{-j}n)\rangle\right)_{n\in Z}. \quad (59)$$

Let $\tilde{G}$ be the filter whose impulse response is given by $\tilde{G} = (g_{-n})_{n\in Z}$. From (55), (57), and (59), one can derive that $D_{2^j} f$ is computed by filtering $A_{2^j} f$ with $\tilde{G}$ and keeping every other sample of the convolution product [38]. This algorithm is illustrated by the block diagram shown in Fig. 19(a); it is essentially similar to a quadrature mirror filter bank decomposition [14].

Let us now describe a simple two-dimensional extension of the one-dimensional multiresolution wavelet model. We saw that a separable multiresolution representation is computed by filtering the signal with a low-pass filter $\Theta(x, y) = \theta(x)\theta(y)$ [(50)]. Let $\tilde{\Theta}(x, y) = \Theta(-x, -y)$. The discrete approximation of a function $f(x, y) \in L^2(R^2)$ at the resolution 2^j can also be rewritten

$$A_{2^j} f = \left(\langle f(x, y), \tilde{\Theta}_{2^j}(x - 2^{-j}n, y - 2^{-j}m)\rangle\right)_{(n,m)\in Z^2}. \quad (60)$$

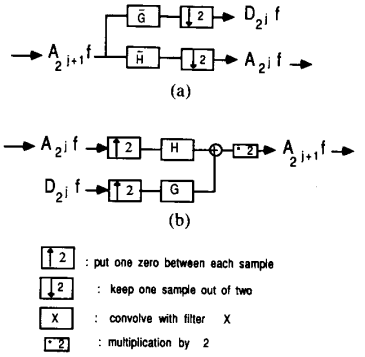


Fig. 19. (a) Decomposition of $A_{2^{j+1}} f$ into $A_{2^j} f$ and $D_{2^j} f$ when computing an orthogonal wavelet representation. The filters H and G make a pair of quadrature mirror filters. (b) Reconstruction of $A_{2^{j+1}} f$ from $A_{2^j} f$ and $D_{2^j} f$ when reconstructing the original signal from an orthogonal wavelet representation.

The extension of the one-dimensional model is straightforward. The best estimate of $f(x, y)$ given the inner products of $A_{2^j} f$ is equal to the orthogonal projection of $f(x, y)$ on the vector space V_{2^j} generated by the family of functions

$$\left(\tilde{\Theta}_{2^j}(x - 2^{-j}n, y - 2^{-j}m)\right)_{(n,m)\in Z^2}. \quad (61)$$

The sequence of vector spaces $(V_{2^j})_{j\in Z}$ is called a multiresolution approximation of $L^2(R^2)$. Similarly to the one-dimensional model, the difference of information between the approximation of a signal $f(x, y)$ at the resolutions 2^j and 2^{j+1} is equal to the orthogonal projection of $f(x, y)$ on the orthogonal complement O_{2^j} of V_{2^j} in $V_{2^{j+1}}$. We can build [45] an orthonormal basis of O_{2^j} by scaling and translating three wavelets: $\Psi^1(x, y)$, $\Psi^2(x, y)$, and $\Psi^3(x, y)$. Let us denote $\Psi^i_{2^j}(x, y) = 2^j\Psi^i(2^j x, 2^j y)$ for $1 \le i \le 3$. The family of functions

$$\begin{Bmatrix} 2^{-j}\Psi^1_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \\ 2^{-j}\Psi^2_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \\ 2^{-j}\Psi^3_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \end{Bmatrix}_{(n,m)\in Z^2} \quad (62)$$

is an orthonormal basis of O_{2^j}. When the resolution 2^j varies between 0 and $+\infty$, the family of functions

$$\begin{Bmatrix} 2^{-j}\Psi^1_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \\ 2^{-j}\Psi^2_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \\ 2^{-j}\Psi^3_{2^j}(x - 2^{-j}n, y - 2^{-j}m) \end{Bmatrix}_{(n,m,j)\in Z^3} \quad (63)$$

is a wavelet orthonormal basis of $L^2(R^2)$. Fig. 20 shows approximately the frequency support of the three wavelets $\Psi^1(x, y)$, $\Psi^2(x, y)$, $\Psi^3(x, y)$. Each wavelet $\Psi^i(x, y)$ can be interpreted as the impulse response of a band-pass filter having a specific orientation selectivity. This corresponds to a particular case of oriented two-dimensional discrete wavelet transform.

In two dimensions, the difference of information between the approximations $A_{2^{j+1}} f$ and $A_{2^j} f$ is therefore

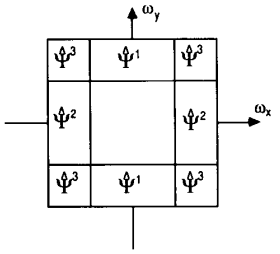


Fig. 20. Approximate repartition of the frequency support of $\hat{\Psi}^1(\omega_x, \omega_y)$, $\hat{\Psi}^2(\omega_x, \omega_y)$, and $\hat{\Psi}^3(\omega_x, \omega_y)$ in the frequency plane.

characterized by the sequences of inner products

$$D_{2^j}^1 f = \left(\langle f(x, y), \Psi_{2^j}^1(x - 2^{-j}n, y - 2^{-j}m)\rangle\right)_{(n,m)\in Z^2},$$

$$D_{2^j}^2 f = \left(\langle f(x, y), \Psi_{2^j}^2(x - 2^{-j}n, y - 2^{-j}m)\rangle\right)_{(n,m)\in Z^2},$$

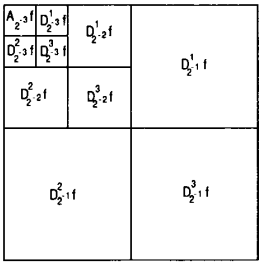
$$D_{2^j}^3 f = \left(\langle f(x, y), \Psi_{2^j}^3(x - 2^{-j}n, y - 2^{-j}m)\rangle\right)_{(n,m)\in Z^2}.$$

Each of these sequences of inner products can be considered as an image. $D_{2^j}^1 f$ gives the vertical higher frequencies (horizontal edges), $D_{2^j}^2 f$ gives the horizontal higher frequencies (vertical edges), and $D_{2^j}^3 f$ gives the higher frequencies in both directions (corners) (see Fig. 21). Let us suppose that initially we have an image $A_1 f$ measured at the resolution 1. For any $J > 0$, this discrete image can be decomposed between the resolutions 1 and 2^{-J}, and completely represented by the $3J + 1$ discrete images

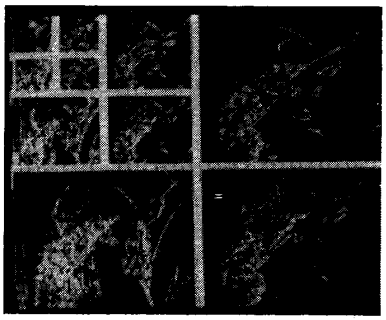
$$\left(A_{2^{-J}} f, (D_{2^j}^1 f)_{-J\le j\le -1}, (D_{2^j}^2 f)_{-J\le j\le -1}, (D_{2^j}^3 f)_{-J\le j\le -1}\right).$$

This set of images is called an *orthogonal wavelet representation* in two dimensions [38]. The image $A_{2^{-J}} f$ is a coarse approximation, and the images $D_{2^j}^i f$ give the image details for different orientations and resolutions. If the original image has N^2 pixels, each image $A_{2^j} f$, $D_{2^j}^1 f$, $D_{2^j}^2 f$, $D_{2^j}^3 f$ has $2^j \cdot N^2$ pixels ($j < 0$). The total number of pixels of an orthogonal wavelet representation is therefore equal to N^2. It does not increase the volume of data. This is due to the orthogonality of the representation.

A wavelet representation can be computed with a separable extension of the algorithm illustrated in Fig. 19(a) [38]. This extension corresponds to a separable quadrature mirror filter decomposition as described by Woods [63]. Fig. 21(b) gives the wavelet representation of the image in Fig. 16. From this representation, we can reconstruct the original image with a two-dimensional separable extension of the algorithm illustrated in Fig. 19(b) [38]. Fig. 21(c) is the reconstructed image from the wavelet representation shown in Fig. 21(b). The reconstruction is numerically stable. It enables us to use this type of representation for image coding. A more general nonseparable extension of the wavelet model was studied by Meyer [43]. Such extensions are, however, more difficult to implement and are computationally more expensive.



(a)



(b)

(c)

Fig. 21. (a) Labeling of the detail images shown in the wavelet representation. (b) Orthogonal wavelet representation of the lady image for $J = 3$. At a given resolution, each detail image corresponds to a particular spatial orientation tuning. (c) Reconstruction of the original image from the orthogonal wavelet representation. The reconstruction is numerically stable.

B. Applications of Multiresolution Transforms

The wavelet model gives a precise understanding of the concept of multiresolution by introducing the sequence of vector spaces $(V_{2^j})_{j\in Z}$. A noncorrelated multiresolution representation can be built by decomposing the signal into a wavelet orthonormal basis. A difficult problem when using a multiresolution representation for analyzing a scene is to relate the details appearing at different resolutions. Many ad hoc techniques have been developed for this purpose. We saw in Section III-D that the local regularity of a function is provided by the decay rate of the

wavelet coefficients when the resolution increases. These theorems give a first approach for comparing the value of the decomposition at different resolutions.

Multiband image decompositions are also well adapted for coding images because it is possible to match the human visual system sensitivity and take advantage of the intrinsic statistical properties of images. The contrast sensitivity function (Fig. 3) shows that the sensitivity of human vision depends upon the frequency of the stimulus. We want to quantize each frequency band with the minimum number of bits, and at the same time try to reconstruct the best possible image for the human visual perception. For this purpose, we adapt the quantization noise to the human sensitivity along each frequency band. The more sensitive the human system, the less quantization noise is introduced. This enables us to introduce a minimum amount of perceivable distortion in the reconstructed image. Watson has done some particularly detailed psychophysical experiments to test this type of approach for image coding [60].

The statistical properties of images give another reason for using multiband decompositions in image coding. It is well known that the intensity of images is locally correlated. Predictive codings have been particularly successful to compress the number of bits used in coding an image. The wavelet coefficients give a measure of the local contrast at different scales. Since the image intensity is locally correlated, these local contrasts generally have a small amplitude [38]. We can take advantage of this property for coding the wavelet coefficients on fewer bits without introducing any noticeable distortion. As explained in the previous section, a wavelet orthogonal representation can also be imterpreted as a decomposition into a quadrature mirror filter bank. Several studies in image processing have already shown the efficiency of these filter banks for data compression [1], [63].

In order to use a multiresolution representation for pattern recognition applications, we must be able to build models of patterns within the multiresolution representation. The patterns might be located anywhere in the image. Hence, the models must be independent from the pattern location. When a pattern is translated, its model should only be translated but not modified. Let us show that a multiresolution representation does not verify this translation property. To simplify the explanation, we consider the particular case of a one-dimensional orthogonal wavelet decomposition. At the resolution 2^j, the details of a signal $f(x) \in L^2(\boldsymbol{R})$ are defined by

$$D_{2^j} f = \left(\langle f(x), \Psi_{2^j}(x - 2^{-j}n) \rangle \right)_{n \in \boldsymbol{Z}}.$$

$D_{2^j} f$ can be expressed as a uniform sampling of the wavelet transform at the scale 2^j

$$D_{2^j} f = \left(Wf(2^j, 2^{-j}n) \right)_{n \in \boldsymbol{Z}}.$$

Let $g(x) = f(x - \tau)$ be a translation of $f(x)$ by τ. Since a wavelet transform can be written as a convolution product [(22)], it is shift invariant

$$Wg(2^j, u) = Wf(2^j, u - \tau).$$

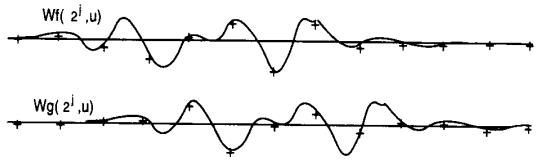


Fig. 22. This drawing shows that the sampling of a wavelet transform (given by the crosses) can be very different after translating the signal. The wavelet transform is translated but the sampling does not translate if the translation is not proportional to the sampling interval (adapted from [37]).

However, the sampling of $Wg(2^j, u)$ does not correspond to a translation of the sampling of $Wf(2^j, u)$ unless $\tau = k2^{-j}$, $k \in \boldsymbol{Z}$ (see Fig. 22).

This distortion through translation implies that the wavelet coefficients of a pattern at the resolution 2^j depend upon the position of the pattern modulo 2^{-j}. This property is inherent to the notion of resolution. Indeed, at the resolution 2^j, we cannot measure anything smaller than 2^{-j} so we cannot represent a displacement smaller than 2^{-j}. One can find the same problem in all the pyramidal multiresolution representations and any uniform sampling of a wavelet transform.

A first solution to this translation problem is to sample the wavelet transform $Wf(2^j, u)$ at a rate much larger than 2^j. The samples then translate approximately when the signal translates. However, this solution considerably increases the redundancy of the representation and the translation is still not perfect. This technique is often adopted for pattern recognition algorithms based on pyramid decompositions. A second solution consists of defining a representation based on an adaptive sampling of the functions $Wf(2^j, u)$ which translates when the signal translates.

V. Zero Crossings of Multifrequency Channels

In the previous sections we studied the properties of the decomposition of a function into multifrequency channels of constant size on a logarithmic scale. We saw that such a decomposition can be interpreted as a wavelet transform. We then described the properties and applications of a discrete wavelet transform built from a uniform sampling of the continuous wavelet transform. However, we showed that such a discretization is difficult to use for pattern recognition applications because it is not invariant through translation. Here, we review the characterization of a signal from the zero crossings of a wavelet transform. Such a characterization defines a discrete representation which translates when the signal translates.

If a function $f(x)$ is translated, for each scale s, the function $Wf(s, u)$ is translated along the parameter u. Hence, the zero crossings of $Wf(s, u)$ are translated as well. Let us suppose that $\psi(x)$ is equal to the second derivative of a smoothing function $\xi(x)$

$$\psi(x) = \xi''(x).$$

A smoothing function is a function which can be interpreted as the impulse response of a low-pass filter. Any

zero crossing of $Wf(s, u)$ corresponds to a point of abrupt change in the function $f(x)$ smoothed by $\xi_s(x) = \sqrt{s}\xi(sx)$. Indeed, if $\psi(x) = \xi''(x)$

$$Wf(s, u) = f * \tilde{\psi}_s(u) = s^2(f * \xi_s)''(u).$$

Hence, a zero crossing of $Wf(s, u)$ is an inflection point of the function $f(x)$ smoothed by $\xi_s(x)$. Fig. 23 illustrates this on a straight edge. This zero-crossing detection is a standard edge finding operation in computer vision [41].

Let us now study the completeness of stability of such a representation. Is it possible to reconstruct $f(x)$ from the zero crossings of $Wf(s, u)$? We know that a wavelet transform $Wf(s, u)$ defines a stable and complete representation of $f(x)$. It is therefore equivalent to study the reconstruction of $Wf(s, u)$ from its own zero crossings. If the function $Wf(s, u)$ was *a priori* any function of $L^2(R^+ \times R)$, it is clear that such a reconstruction would not be possible. Indeed, for a given set of zero crossings, there is an infinite number of functions in $L^2(R^+ \times R)$ whose zero crossings correspond to this set. However, we saw that a wavelet transform $Wf(s, u)$ is not any function of $L^2(R^+ \times R)$. It verifies the constraint of the reproducing kernel [(25)]. We must therefore study whether the constraint of the reproducing kernel plus the information on the zero-crossing positions is enough to have a stable characterization of $Wf(s, u)$.

An interesting particular case of wavelet transform consists of choosing a wavelet equal to the Laplacian of a Gaussian. Since a Gaussian is a smoothing function, the zero crossings of such a wavelet transform can also be interpreted as signal edges [41]. In this particular case, the intrinsic redundancy of the wavelet transform $Wf(s, u)$ can be expressed with the differential equation of heat diffusion [29]. By applying the maximum principles to the solutions of the heat differential equation, Hummel [24] proved that a function $f(x)$ is indeed characterized by the zero crossings of $Wf(s, u)$. However, Hummel also showed that this characterization is not stable. So a slight perturbation of the zero crossings may correspond a substantial perturbation of the high frequencies of the reconstructed function. Reconstruction algorithms have been developed on images by Sanz and Huang [53] as well as Zeevi and Rotem [64]. These reconstruction algorithms are iterative. They were not able to reconstruct the image perfectly in both cases. Hummel and Moniot [25] tried to stabilize the zero-crossings representation by also recording the value of the gradient of $Wf(s, u)$ along each zero crossing. By adding the gradient information, they have shown experimentally that one can then compute a stable reconstruction of $f(x)$ from the zero crossings of $Wf(s, u)$. In this algorithm, the position of the zero crossings and the value of the gradients are kept along a uniform sequence of scales: $(j\alpha)_{j\in Z}$ with $\alpha > 0$. Such a sequence is much more dense than the dyadic sequences $(2^j)_{j\in Z}$ used when we discretized the wavelet transform.

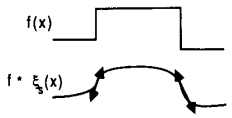


Fig. 23. The zero crossings of a wavelet transform provide the location of the inflection points (edges) of $f * \xi_s(x)$ (adapted from [37]).

Another way to stabilize a zero-crossing representation is to record the energy of $Wf(s, u)$ between two consecutive zero crossings appearing at the same scale [37]. This energy preserves an $L^2(R)$ structure to the zero-crossing representation. In particular, we can then define an $L^2(R)$ distance for pattern recognition applications. By keeping the position of the zero crossings of $Wf(s, u)$ and the local energies only along a dyadic sequence of scales $(2^j)_{j\in Z}$, we showed that the original signal can be reconstructed exactly in few iterations [37]. The reconstruction uses the reproducing kernel equation which is valid for any type of wavelet transform. We believe that the mathematical tools developed within the wavelet model give a simpler approach for analyzing the zero-crossing problem. From a practical point of view, the reconstruction algorithm developed from the reproducing kernel is simple to implement in both one and two dimensions and converges quickly (about 10 iterations).

Representations based on zero crossings of multifrequency channels are still not well understood. They are built with a nonlinear transform which is difficult to model. However, they have very good potential for pattern characterization. They characterize the position of the signal edges and are translation invariant.

VI. Conclusion

In this paper, we reviewed the application of multifrequency decompositions to image processing from several viewpoints. We covered some psychophysical and physiological data showing that such a decomposition seems to be implemented in the human visual cortex. We then described the mathematical properties of these decompositions. We first reviewed the properties of a window Fourier transform and explained why this decomposition is not convenient for analyzing signals such as images. We then introduced the wavelet transform and described its most important properties. Although the goal of this paper was not to build any psychophysiological model of the human visual system, it would be interesting to further investigate the relevance of the wavelet model to some low-level processes in the visual cortex.

In computer vision, multifrequency channel decompositions are interpreted through the concept of multiresolution. We described the classical pyramidal multiresolution algorithms and the wavelet approach to multiresolution decompositions. This model shows that the difference of information between the approximation of a function at two different resolutions is computed by decomposing the function into a wavelet orthonormal basis. We also explained the relationship between orthonormal

wavelets and quadrature mirror filters. We can compute the decomposition of a function into a wavelet orthonormal basis with a quadrature mirror filter bank. A third motivation for using multiband decomposition is due to the intrinsic statistical properties of images. Images have a relatively simple decomposition into frequency subbands. These bands can be coded on fewer bits with no visible distortions.

A uniform sampling of each multifrequency channel defines a representation which is not translation invariant. It is therefore difficult to build pattern recognition algorithms from such decompositions. We reviewed the properties of zero crossing in multiband decompositions. This adaptive sampling is translation invariant but is much more difficult to analyze. We described some previous results and gave the wavelet formalization of this problem through the reproducing kernel equation.

Acknowledgment

I would like to thank particularly R. Bajcsy for her advice on the writing of this paper, and N. Treil for helping me draw the figures.

References

[1] E. Adelson and E. Simoncelli, "Orthogonal pyramid transform for image coding," in *Proc. SPIE, Visual Commun. and Image Processing*, 1987.

[2] B. Andrew and D. Pollen, "Relationship between spatial frequency selectivity and receptive field profile of simple cells," *J. Physiol.*, vol. 287, pp. 163-176, 1979.

[3] G. Battle, "A block spin construction of ondelettes, Part 1: Lemarie functions," *Commun. Math. Phys.*, vol. 110, pp. 601-615, 1987.

[4] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532-540, Apr. 1983.

[5] F. Campbell, R. Carpenter, and J. Levinson, "Visibility of aperiodic patterns compared with cortical cells compared with sinusoidal gratings," *J. Physiol.*, vol. 204, pp. 283-298, 1969.

[6] F. Campbell and D. Green, "Optical and retina factors affecting visual resolution," *J. Physiol.*, vol. 181, pp. 576-593, 1965.

[7] F. Campbell and J. Kulikowski, "Orientation selectivity of the human visual system," *J. Physiol.*, vol. 197, pp. 437-441, 1966.

[8] F. Campbell and J. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol.*, vol. 197, pp. 551-566, 1968.

[9] R. C. Crochiere and L. R. Rabiner, "Interpolation and decimation in signal processing," *Proc. IEEE*, vol. 69, Mar. 1981.

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

[11] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909-996, Nov. 1988.

[12] J. G. Daugmann, "Two-dimensional spectral analysis of cortical receptive field profile," *Vis. Res.*, vol. 20, pp. 847-856, 1980.

[13] ——, "Six formal properties of two dimensional anisotropic visual filter. Structural principles and frequency/orientation selectivity," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, Sept. 1983.

[14] D. Esteban and C. Galand, "Applications of quadrature mirror filters to split band voice coding schemes," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, May 1977.

[15] P. Federbush, "Quantum field theory in ninety minutes," *Bull. Amer. Math. Soc.*, 1987.

[16] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.*, London, vol. 93, pp. 429-457, 1946.

[17] M. Georgeson, "Mechanisms of visual image processing: studies of pattern interaction and selective channels in human vision," Ph.D. dissertation, Univ. Sussex, Brighton, England, 1975.

[18] ——, "Spatial Fourier analysis and human vision," in *Tutorial Essays in Psychology, A Guide to Recent Advances.* N. Sutherland, Ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.

[19] W. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 17-34, Jan. 1985.

[20] 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.

[21] A. Grossman, J. Morlet, and T. Paul, "Transforms associated to square integrable group representations," *Int. J. Math. Phys.*, vol. 26, pp. 2473-2479, 1986.

[22] E. Hall, J. Rouge, and R. Wong, "Hierarchical search for image matching," in *Proc. Conf. Decision Contr.*, 1976, pp. 791-796.

[23] D. Hubel and T. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, 1962.

[24] R. Hummel, "Representations based on zero-crossings in scale-space," Tech. Rep. 225, Courant Inst. Dep. Comput. Sci., June 1986.

[25] R. Hummel and R. Moniot, "A network approach to reconstruction from zero-crossings," in *Proc. IEEE Workshop Comput. Vis.*, Dec. 1987.

[26] S. Jaffard, "Estimations Holderiennes ponctuelles des fonctions au moyen des coefficients d'ondelettes," *Notes au Compte-Rendu de l'Academie Des Sciences*, France, 1989.

[27] S. Jaffard and Y. Meyer, "Bases d'ondelettes dans des ouverts de Rn," *J. Mathematiques Pures et Appliquees*, 1987.

[28] J. Klauder and B. Skagerstam, in *Coherent States.* Singapore: World Scientific, 1985.

[29] J. Koenderink, "The structure of images," in *Biological Cybernetics.* New York: Springer-Verlag, 1984.

[30] R. Kronland-Martinet, J. Morlet, and A. Grossmann, "Analysis of sound patterns through wavelet transform," *Int. J. Pattern Recogn. Artificial Intell.*, 1988.

[31] J. Kulikowski and P. King-Smith, "Orientation selectivity of grating and line detectors in human vision," *Vis. Res.*, vol. 13, pp. 1455-1478, 1973.

[32] P. G. Lemarie, "Ondelettes a localisation exponentielles," *J. Math. Pures et Appliquees*, 1988.

[33] P. G. Lemarie and Y. Meyer, "Ondelettes et bases Hilbertiennes," *Revista Matematica Ibero Americana*, vol. 2, 1986.

[34] M. D. Levine, *Vision in Man and Machine.* New York: McGraw-Hill, 1985.

[35] L. Maffei and A. Fiorentini, "The unresponsive regions of visula cortical receptive fields," *Vis. Res.*, vol. 16, pp. 1131-1139, 1976.

[36] L. Maffei, C. Morrone, M. Pirchio, and G. Sandini, *J. Physiol.*, vol. 296, pp. 24-47, 1979.

[37] S. Mallat, "Dyadic wavelets energy zero-crossings," Tech. Rep. MS-CIS-88-30, U. Penn., 1988.

[38] ——, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674-693, July 1989.

[39] ——, "Multiresolution approximation and wavelet orthonormal bases of L2," *Trans. Amer. Math. Soc.*, vol. 3-15, pp. 69-87, Sept. 1989.

[40] D. Marr, in *Vision.* San Francisco, CA: Freeman, 1982.

[41] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London*, vol. 207, pp. 187-217, 1980.

[42] D. Marr and T. Poggio, "A theory of human stereo vision," *Proc. Roy. Soc. London*, vol. B 204, pp. 301-328, 1979.

[43] Y. Meyer, "Ondelettes et fonctions splines," presented at the Seminaire Equations aux Derivees Partielles, Ecole Polytechnique, Paris, France, Dec. 1986.

[44] ——, in *Ondelettes et Operateurs.* Paris, France: Hermann, 1988.

[45] ——, "Principe d'incertitude, bases hilbertiennes et algebres d'operateurs," presented at the Bourbaki seminar, 1985-1986, Paper 662.

[46] J. Nachmais and A. Weber, "Discrimination of simple and complex gratings," *Vis. Res.*, vol. 15, pp. 217-223, 1975.

[47] J. Oppenheim and J. Lim, in *Advanced Topics in Signal Processing*, Signal Processings Series. Englewood Cliffs, NJ: Prentice-Hall, 1988, pp. 289-336.

[48] T. Paul, "Affine coherent states and the radial Schrodinger equation. Radial harmonic oscillator and hydrogen atom," preprint.

[49] D. A. Pollen and S. F. Ronner, "Visual corical neurons as localized spatial frequency filter," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, Sept. 1983.

[50] A. Rosenfeld, *Multiresolution Image Processing and Analysis.* New York: Springer-Verlag, 1982.

[51] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. C-20, 1971.

[52] A. Rosenfeld and G. J. Vanderburg, "Coarse-fine template matching," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, pp. 104–107, 1977.
[53] J. Sanz and T. Huang, "Theorem and experiments on image reconstruction from zero-crossings," IBM, Res. Rep. RJ5460.
[54] M. J. Smith and T. P. Barnwell, "Exact reconstruction techniques for tree-structured subband coders," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, June 1986.
[55] J. Stromberg, "A modified Franklin system and higher-order systems of Rn as unconditional bases for Hardy spaces," in *Proc. Conf. Harmonic Anal. Honor of a. Zygmund*, vol. 2, Wadsworth Math Series, pp. 475–493.
[56] S. Tanimoto and T. Pavlidis, "A hierarchical data structure for image processing," *Comput. Graphics Image Processing*, vol. 4, pp. 104–119, 1975.
[57] P. Tchamitchian, "Biorthogonalite et theorie des operateurs," *Revista Matematica Ibero Americana*, vol. 2, 1986.
[58] M. Turner, "Texture discrimination by Gabor functions," *Biological Cybern.*, vol. 55, pp. 71–82, 1986.
[59] K. De Valois, R. De Valois, and E. Yund, *J. Physiol.*, vol. 291, pp. 483–505, 1979.
[60] A. Watson, "Efficiency of a model human image code," *J. Opt. Soc. Amer.*, vol. 4, pp. 2401–2417, Dec. 1987.
[61] M. Webster and R. De Valois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," *J. Opt. Soc. Amer.*, July 1985.
[62] A. Witkin, "Scale space filtering," in *Proc. Int. Joint Conf. Artificial Intell.*, 1983.
[63] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, Oct. 1986.
[64] Y. Zeevi and D. Rotem, "Image reconstruction from zero crossings," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1269–1277, Oct. 1986.

Stephane G. Mallat was born in Paris, France. He graduated from Ecole Polytechnique, Paris, in 1984, and from Ecole Nationale 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 an Assistant Professor in the Computer Science Department of the Courant Institute of Mathematical Sciences, New York University, New York. His research interests include computer vision, signal processing, and applied mathematics.