# 4.2

# Multiscale Image Decompositions and Wavelets

Pierre Moulin
*University of Illinois at Urbana-Champaign*



## 1 Overview

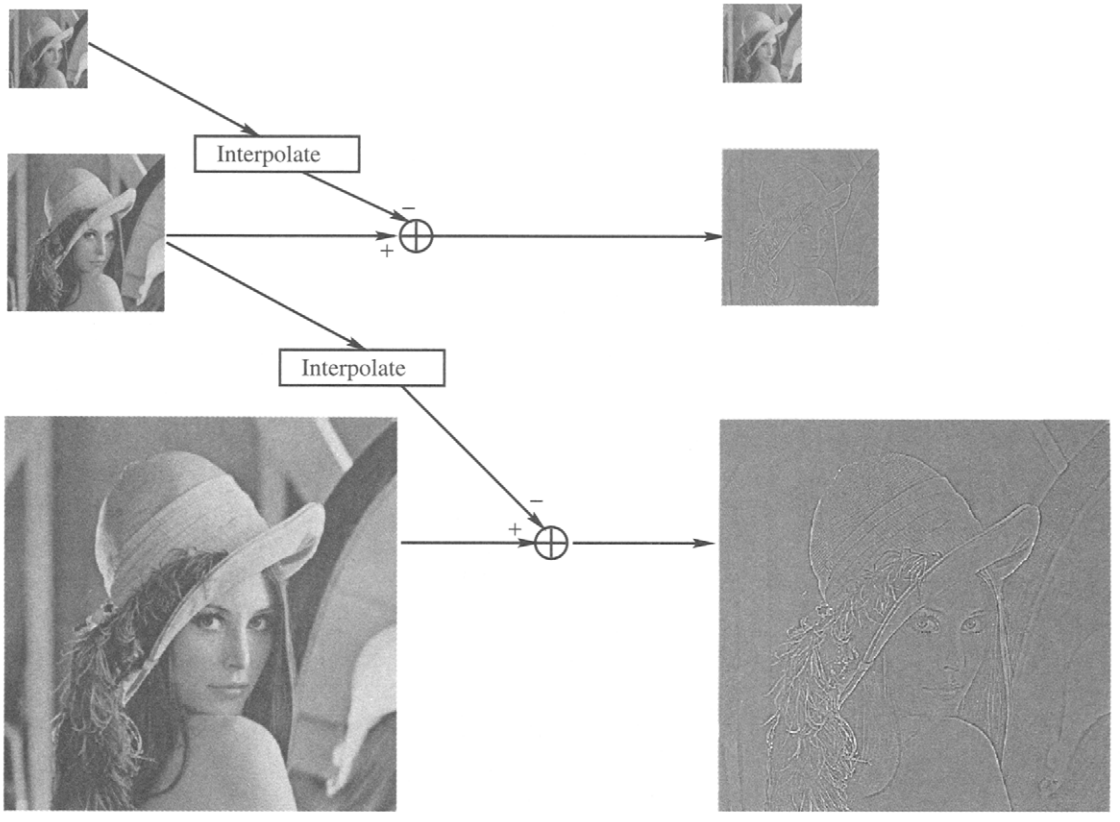
The concept of scale, or resolution of an image, is very intuitive. A person observing a scene perceives the objects in that scene at a certain level of resolution that depends on the distance to these objects. For instance, walking towards a distant building, she would first perceive a rough outline of the building. The main entrance becomes visible only in relative proximity to the buiding. Finally, the door bell is visible only in the entrance area. As this example illustrates, the notions of resolution and scale loosely correspond to the size of the details that can be perceived by the observer. It is of course possible to formalize these intuitive concepts, and indeed signal processing theory gives them a more precise meaning.

These concepts are particularly useful in image and video processing and in computer vision. A variety of digital image processing algorithms decompose the image being analyzed into several components, each of which captures information present at a given scale. While our main purpose is to introduce the reader to the basic concepts of multiresolution image decompositions and wavelets, applications will also be briefly discussed throughout this chapter. The reader is referred to other chapters of this book for more details.

Throughout, we assume that the images to be analyzed are rectangular with $N \times M$ pixels. While there exists several types of multiscale image decompositions, we consider three main methods [1–6]:

1. In a **Gaussian pyramid** representation of an image (Fig. 1a), the original image appears at the bottom of a pyramidal stack of images. This image is then lowpass filtered and subsampled by a factor of two in each coordinate. The resulting $N/2 \times M/2$ image appears at the second level of the pyramid. This procedure can be iterated several times. Here resolution can be measured by the size of the image at any given level of the pyramid. The pyramid in Fig. 1a has three resolution levels, or scales. In the original application of this method to computer vision, the lowpass filter used was often a Gaussian filter,[1] hence the terminology *Gaussian pyramid.* We shall use this terminology even

[1]This design was motivated by analogies to the human visual system; see Section 3.6.




(a) Gaussian pyramid

(b) Laplacian pyramid

(c)

**FIGURE 1** Three multiscale image representations applied to *Lena*: (a) Gaussian pyramid; (b) Laplacian pyramid; (c) wavelet representation.

when a lowpass filter is not a Gaussian filter. Another possible terminology in that case is simply *lowpass pyramid.* Note that the total number of pixels in a pyramid representation is $NM + NM/4 + NM/16 + \cdots \approx \frac{4}{3}NM$. This is said to be an *overcomplete* representation of the original image, due to the increase in the number of pixels.

2. The **Laplacian pyramid** representation of the image is closely related to the Gaussian pyramid, but here the difference between approximations at two successive scales is computed and displayed for different scales, see Fig. 1b. The precise meaning of the *interpolate* operation in the figure will be given in Section 2.1. The displayed images represent details of the image that are significant at each scale. An equivalent way to obtain the image at a given scale is to apply the difference between two Gaussian filters to the original image. This is analogous to filtering the image using a Laplacian filter, a technique commonly employed for edge detection (see Chapter 4.20). Laplacian filters are bandpass, hence the name Laplacian pyramid, also termed *bandpass pyramid.*
3. In a **wavelet decomposition,** the image is decomposed into a set of subimages (or subbands) which also represent details at different scales (Fig. 1c). Unlike pyramid representations, the subimages also represent details with different spatial orientations (such as edges with horizontal, vertical, and diagonal orientations). The number of pixels in a wavelet decomposition is only $NM$. As we shall soon see, the signal processing operations involved here are more sophisticated than those for pyramid image representations.

The pyramid and wavelet decompositions are presented in more detail in Sections 2 and 3, respectively. The basic concepts underlying these techniques are applicable to other multiscale decomposition methods, some of which are listed in Section 4.

Hierarchical image representations such as those in Fig. 1 are useful in many applications. In particular, they lend themselves to effective designs of reduced-complexity algorithms for texture analysis and segmentation, edge detection, image analysis, motion analysis, and image understanding in computer vision. Moreover, the Laplacian pyramid and wavelet image representations are *sparse* in the sense that most detail images contain few significant pixels (little significant detail). This sparsity property is very useful in image compression, as bits are allocated only to the few significant pixels; in image recognition, because the search for significant image features is facilitated; and in the restoration of images corrupted by noise, as images and noise possess rather distinct properties in the wavelet domain. The recent JPEG2000 international standard for image compression is based on wavelets [7], unlike its predecessor JPEG which was based on the discrete cosine transform [8].

# 2 Pyramid Representations

In this section, we shall explain how the Gaussian and Laplacian pyramid representations in Fig. 1 can be obtained from a few basic signal processing operations. To this end, we first describe these operations in Section 2.1 for the case of one-dimensional (1D) signals. The extension to two-dimensional (2D) signals is presented in Sections 2.2 and 2.3 for Gaussian and Laplacian pyramids, respectively.

## 2.1 Decimation and Interpolation

Consider the problem of decimating a 1D signal by a factor of two, namely, reducing the sample rate by a factor of two. This operation generally entails some loss of information, so it is desired that the decimated signal retain as much fidelity as possible to the original. The basic operations involved in decimation are lowpass filtering (using a digital anti-aliasing filter) and subsampling, as shown in Fig. 2. The impulse response of the lowpass filter is denoted by $h(n)$, and its discrete-time fourier transform [9] by $H(e^{j\omega})$. The relationship between input $x(n)$ and output $y(n)$ of the filter is the convolution equation

$$y(n) = x(n) * h(n) = \sum_k h(k)x(n-k).$$

The downsampler discards every other sample of its input $y(n)$. Its output is given by

$$z(n) = y(2n).$$

Combining these two operations, we obtain

$$z(n) = \sum_k h(k)x(2n-k). \tag{1}$$

Downsampling usually implies a loss of information, as the original signal $x(n)$ cannot be exactly reconstructed from its decimated version $z(n)$. The traditional solution for reducing this information loss consists in using an "ideal" digital anti-aliasing filter $h(n)$ with cutoff frequency

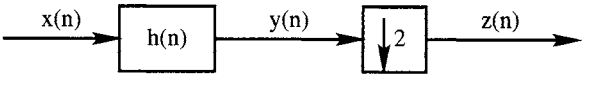



**FIGURE 2** Decimation of a signal by a factor of two, obtained by cascade of a lowpass filter $h(n)$ and a subsampler $\downarrow 2$.

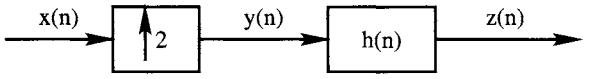



**FIGURE 3** Interpolation of a signal by a factor of two, obtained by cascade of an upsampler ↑ 2 and a lowpass filter $h(n)$.

$\omega_c = \pi/2$ [9][2]. However such "ideal" filters have infinite length. In image processing, short finite impulse response (FIR) filters are preferred for obvious computational reasons. Furthermore, approximations to the "ideal" filters above have an oscillating impulse response, which unfortunately results in visually annoying ringing artifacts in the vicinity of edges. The FIR filters typically used in image processing are symmetric, with length between three and twenty taps. Two common examples are the 3-tap FIR filter $h(n) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, and the length-$(2L+1)$ truncated Gaussian, $h(n) = Ce^{-n^2/(2\sigma^2)}$, $|n| \leq L$, where $C = 1/\sum_{|n|\leq L} e^{-n^2/(2\sigma^2)}$. The coefficients of both filters add up to one: $\sum_n h(n) = 1$, which implies that the DC response of these filters is unity.

Another common image processing operation is *interpolation*, which increases the sample rate of a signal. Signal processing theory tells us that interpolation may be performed by cascading two basic signal processing operations: upsampling and lowpass filtering, see Fig. 3. The upsampler inserts a zero between every other sample of the signal $x(n)$:

$$y(n) = \begin{cases} x(n/2) & : n \text{ even} \\ 0 & : n \text{ odd} \end{cases}$$

The upsampled signal is then filtered using a lowpass filter $h(n)$. The interpolated signal is given by $z(n) = h(n) * y(n)$ or, in terms of the original signal $x(n)$,

$$z(n) = \sum_k h(k)x(n-2k). \tag{2}$$

The so-called ideal interpolation filters have infinite length. Again, in practice, short FIR filters are used.

## 2.2 Gaussian Pyramid

The construction of a Gaussian pyramid involves 2D lowpass filtering and subsampling operations. The 2D filters used in image processing practice are *separable*, which means that they can be implemented as the cascade of 1D filters operating along image rows and columns. This is a convenient choice in many respects, and the 2D decimation scheme is then separable as well. Specifically, 2D decimation is implemented by applying 1D decimation to each row of the image (using Eq. 1) followed by 1D decimation to each column of the resulting image (using Eq. 1 again). The same result would be obtained by first processing columns and then rows. Likewise, 2D interpolation is obtained by first applying Eq. 2 to each row of the image, and then again to each column of the resulting image, or vice-versa.

This technique was used at each stage of the Gaussian pyramid decomposition in Fig. 1a. The lowpass filter used for both horizontal and vertical filtering was the 3-tap filter $h(n) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

Gaussian pyramids have found applications to certain types of image storage problems. Suppose for instance that remote users access a common image database (say an Internet site) but have different requirements with respect to image resolution. The representation of image data in the form of an image pyramid would allow each user to directly retrieve the image data at the desired resolution. While this storage technique entails a certain amount of redundancy, the desired image data are available directly and are in a form that does not require further processing. Another application of Gaussian pyramids is in motion estimation for video [1, 2]: in a first step, coarse motion estimates are computed based on low-resolution image data, and in subsequent steps, these initial estimates are refined based on higher-resolution image data. The advantages of this multiresolution, coarse-to-fine, approach to motion estimation are a significant reduction in algorithmic complexity (as the crucial steps are performed on reduced-size images) and the generally good quality of motion estimates, as the initial estimates are presumed to be relatively close to the ideal solution. Another closely related application that benefits from a multiscale approach is pattern matching [1].

## 2.3 Laplacian Pyramid

We define a *detail image* as the difference between an image and its approximation at the next coarser scale. The Gaussian pyramid generates images at multiple scales, but these images have different sizes. In order to compute the difference between a $N \times M$ image and its approximation at resolution $N/2 \times M/2$, one should interpolate the smaller image to the $N \times M$ resolution level before performing the subtraction. This operation was used to generate the Laplacian pyramid in Fig. 1b. The interpolation filter used was the 3-tap filter $h(n) = (\frac{1}{2}, 1, \frac{1}{2})$.

As illustrated in Fig. 1b, the Laplacian representation is *sparse* in the sense that most pixel values are zero or near zero. The significant pixels in the detail images correspond to edges and textured areas such as *Lena*'s hair. Just like the Gaussian pyramid representation, the Laplacian representation is also *overcomplete*, as the number of pixels is greater (by a factor $\approx 33\%$) than in the original image representation.

Laplacian pyramid representations have found numerous applications in image processing, and in particular in texture analysis and segmentation [1]. Indeed, different textures often present very different spectral characteristics which can be

[2]The paper [10] derives the filter that actually minimizes this information loss in the mean-square sense, under some assumptions on the input signal.

analyzed at appropriate levels of the Laplacian pyramid. For instance, a nearly uniform region such as the surface of a lake contributes mostly to the coarse-level image, while a textured region like grass often contributes significantly to other resolution levels. Some of the earlier applications of Laplacian representations include image compression [11, 12], but the emergence of wavelet compression techniques has made this approach somewhat less attractive. However, a Laplacian-type compression technique was adopted in the hierarchical mode of the lossy JPEG image compression standard [8], also see Chapter 5.5.

# 3 Wavelet Representations

While the sparsity of the Laplacian representation is useful in many applications, overcompleteness is a serious disadvantage in applications such as compression. The wavelet transform offers both the advantages of a sparse image representation and a complete representation. The development of this transform and its theory has had a profound impact on a variety of applications. In this section, we first describe the basic tools needed to construct the wavelet representation of an image. We begin with filter banks, which are elementary building blocks in the construction of wavelets. We then show how filter banks can be cascaded to compute a wavelet decomposition. We then introduce wavelet bases, a concept that provides additional insight into the choice of filter banks. We conclude with a discussion of the relation of wavelet representations to the human visual system, and a brief overview of some applications.

## 3.1 Filter Banks

Figure 4a depicts an *analysis filter bank,* with one input $x(n)$ and two outputs $x_0(n)$ and $x_1(n)$. The input signal $x(n)$ is processed through two paths. In the upper path, $x(n)$ is passed through a lowpass filter $H_0(e^{j\omega})$ and decimated by a factor of 2. In the lower path, $x(n)$ is passed through a highpass filter $H_1(e^{j\omega})$ and also decimated by a factor of 2. For convenience, we make the following assumptions. First, the number $N$ of available samples of $x(n)$ is even. Second, the filters perform a circular convolution (see Chapter 2.3), which is equivalent to assuming that $x(n)$ is a periodic signal. Under these assumptions, the output of each path is periodic with period equal to $N/2$ samples. Hence the analysis filter bank can be thought of as a *transform* that maps the original set $\{x(n)\}$ of $N$ samples into a new set $\{x_0(n), x_1(n)\}$ of $N$ samples.

Figure 4b shows a *synthesis filter bank.* Here there are two inputs $y_0(n)$ and $y_1(n)$, and one single output $y(n)$. The input signal $y_0(n)$ (resp. $y_1(n)$) is upsampled by a factor of two and filtered using a lowpass filter $G_0(e^{j\omega})$ (resp. highpass filter $G_1(e^{j\omega})$.) The output $y(n)$ is obtained by summing the two filtered signals. We assume that the input signals $y_0(n)$ and $y_1(n)$ are periodic with period $N/2$. This implies that the output $y(n)$ is periodic with period equal to $N$. So the synthesis filter bank can also be thought of as a transform that maps the original set of $N$ samples $\{y_0(n), y_1(n)\}$ into a new set of $N$ samples $\{y(n)\}$.

What happens when the output $x_0(n), x_1(n)$ of an analysis filter bank is applied to the input of a synthesis filter bank? As it turns out, under some specific conditions on the four filters $H_0(e^{j\omega})$, $H_1(e^{j\omega})$, $G_0(e^{j\omega})$ and $G_1(e^{j\omega})$, the output $y(n)$ of the resulting *analysis/synthesis* system is *identical* (possibly up to a constant delay) to its input $x(n)$. This condition is known as *Perfect Reconstruction.* It holds, for instance, for the following trivial set of 1-tap filters: $h_0(n)$ and $g_1(n)$ are unit impulses, and $h_1(n)$ and $g_0(n)$ are unit delays. In this case, the reader can verify that $y(n) = x(n-1)$. In this simple example, all four filters are allpass. It is however not obvious to design more useful sets of FIR filters that also satisfy the *perfect reconstruction* condition. A general methodology for doing so was discovered in the mid eighties. We refer the reader to [4, 5] for more details.

Under some additional conditions on the filters, the transforms associated with both the analysis and the synthesis filter banks are orthonormal. Orthonormality implies that the energy of the samples is preserved under the transformation. If these conditions are met, the filters possess the following remarkable properties: the synthesis filters are a time-reversed version of the analysis filters, and the highpass filters are modulated versions of the lowpass filters, namely, $g_0(n) = (-1)^n h_1(n)$, $g_1(n) = (-1)^{n+1} h_0(n)$, and $h_1(n) = (-1)^{-n} h_0 \times (K-n)$, where $K$ is an integer delay. Such filters are often known as quadrature mirror filters (QMF), or conjugate

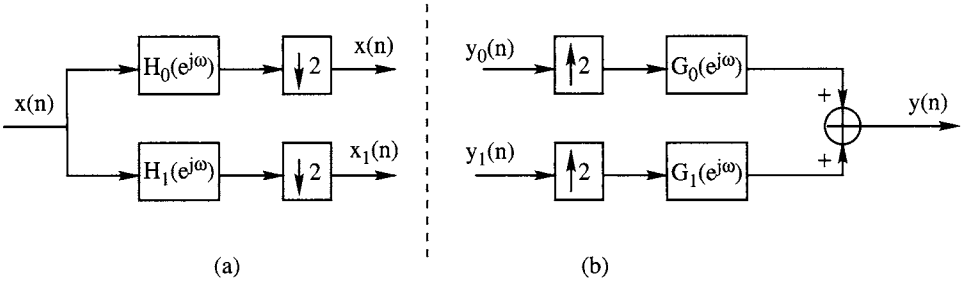



**FIGURE 4** (a) Analysis filter bank, with lowpass filter $H_0(e^{j\omega})$ and highpass filter $H_1(e^{j\omega})$. (b) Synthesis filter bank, with lowpass filter $G_0(e^{j\omega})$ and highpass filter $G_1(e^{j\omega})$.

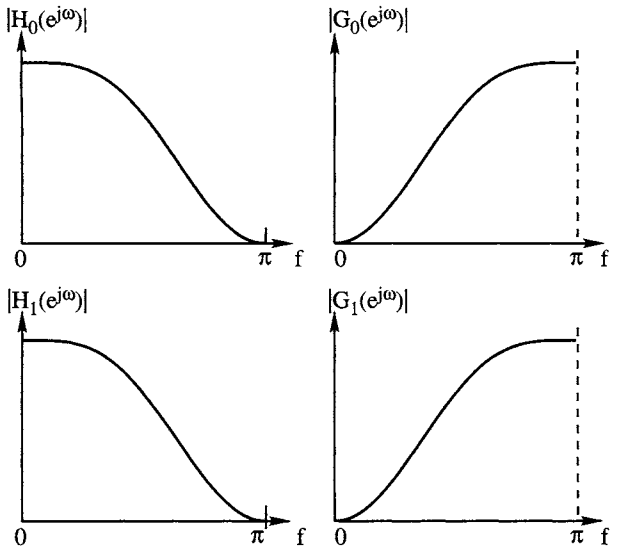



**FIGURE 5** Magnitude frequency response of the four subband filters for a QMF filter bank generated from the prototype Daubechies' 4-tap lowpass filter.

quadrature filters (CQF), or power-complementary filters [5], because both lowpass (resp. highpass) filters have the same frequency response, and the frequency responses of the lowpass and highpass filters are related by the power-complementary property $|H_0(e^{j\omega})|^2 + |H_1(e^{j\omega})|^2 = 2$, valid at all frequencies. The filter $h_0(n)$ is viewed as a prototype filter, because it automatically determines the other three filters.

Finally, if the prototype lowpass filter $H_0(e^{j\omega})$ has a zero at frequency $\omega = \pi$, the filters are said to be *regular filters*, or *wavelet filters*. The meaning of this terminology will become apparent in Section 3.4. Figure 5 shows the frequency responses of the four filters generated from a famous 4-tap filter designed by Daubechies [4, p. 195]:

$$h_0(n) = \frac{1}{4\sqrt{2}}(1+\sqrt{3},\ 3+\sqrt{3},\ 3-\sqrt{3},\ 1-\sqrt{3}).$$

This filter is the first member of a family of FIR wavelet filters that have been constructed by Daubechies and possess nice properties (such as shortest support size for a given number of vanishing moments, see Section 3.4).

There also exist *biorthogonal wavelet filters*, a design that sets aside degrees of freedom for choosing the synthesis lowpass filter $h_1(n)$ given the analysis lowpass filter $h_0(n)$. Such filters are subject to regularity conditions [4]. The transforms are no longer orthonormal, but the filters can have linear phase (unlike nontrivial QMF filters).

## 3.2 Wavelet Decomposition

An analysis filter bank decomposes 1D signals into lowpass and highpass components. One can perform a similar decomposition on images by first applying 1D filtering along rows of the image and then along columns, or vice-versa [13]. This operation is illustrated in Fig. 6a. The same filters $H_0(e^{j\omega})$ and $H_1(e^{j\omega})$ are used for horizontal and vertical filtering. The output of the analysis system is a set of four $N/2 \times M/2$ subimages: the so-called LL (low low), LH (low high), HL (high low) and HH (high high) *subbands*, which correspond to different spatial frequency bands in the image. The decomposition of *Lena* into four such subbands is shown in Fig. 6b. Observe that the LL subband is a coarse (low resolution) version of the original image, and that the HL, LH and HH subbands respectively contain details with vertical, horizontal and diagonal orientations. The total number of pixels in the four subbands is equal to the original number of pixels, $NM$.

In order to perform the wavelet decomposition of an image, one recursively applies the scheme of Fig. 6a to the LL subband. Each stage of this recursion produces a coarser version of the image as well as three new detail images at that particular scale. Figure 6 shows the cascaded filter banks that implement this wavelet decomposition, and Fig. 1c shows a 3-stage wavelet decomposition of *Lena*. There are seven subbands, each corresponding to a different set of scales and orientations (different spatial frequency bands).

Both the Laplacian decomposition in Fig. 1b and the wavelet decomposition in Fig. 1c provide a coarse version of the image as well as details at different scales, but the wavelet representation is complete and provides information about image components at different spatial orientations.

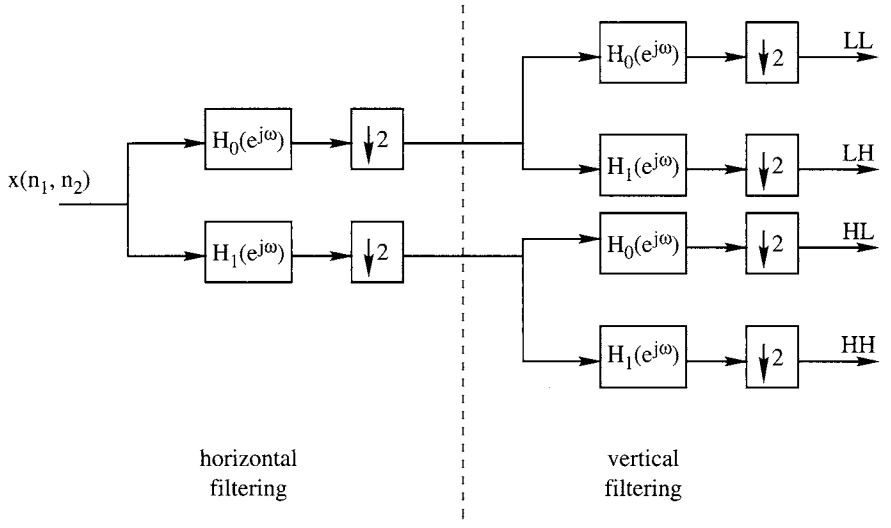
## 3.3 Discrete Wavelet Bases

So far we have described the mechanics of the wavelet decomposition in Fig. 7, but we are yet to explain what wavelets are, and how they relate to the decomposition in Fig. 7. In order to do so, we first introduce discrete wavelet bases. Consider the following representation of a signal $x(t)$ defined over some (discrete or continuous) domain $\mathcal{T}$:

$$x(t) = \sum_k a_k \varphi_k(t), \quad t \in \mathcal{T}. \tag{3}$$

Here $\varphi_k(t)$ are termed *basis functions*, and $a_k$ are the coefficients of the signal $x(t)$ in the *basis* $\mathcal{B} = \{\varphi_k(t)\}$. A familiar example of such signal representations is the Fourier series expansion for periodic real-valued signals with period $T$, in which case the domain $\mathcal{T}$ is the interval $[0, T)$, $\varphi_k(t)$ are sines and cosines, and $k$ represents frequency. It is known from Fourier series theory that a very broad class of signals $x(t)$ can be represented in this fashion.

For discrete $N \times M$ images, we let the variable $t$ in (3) be the pair of integers $(n_1, n_2)$, and the domain of $x$ be $\mathcal{T} = \{0, 1, \ldots, N-1\} \times \{0, 1, \ldots, M-1\}$. The basis $\mathcal{B}$ is then said to be discrete. Note that the wavelet decomposition of an image, as described in Section 3.2, can be viewed as a




(a)

(b)

**FIGURE 6** Decomposition of $N \times M$ image into four $N/2 \times M/2$ subbands: (a) basic scheme, (b) application to *Lena*, using Daubechies' 4-tap wavelet filters.

linear transformation of the original $NM$ pixel values $x(t)$ into a set of $NM$ wavelet coefficients $a_k$. Likewise, the synthesis of the image $x(t)$ from its wavelet coefficients is also a linear transformation, and hence $x(t)$ is the sum of contributions of individual coefficients. The contribution of a particular coefficient $a_k$ is obtained by setting all inputs to the synthesis filter bank to zero, except for one single sample with amplitude $a_k$, at a location determined by $k$. The output is $a_k$ times the response of the synthesis filter bank to a unit impulse at location $k$. We now see that the signal $x(t)$ takes the form (3), where $\varphi_k(t)$ are the spatial impulse responses above.

The index $k$ corresponds to a given location of the wavelet coefficient within a given subband. The discrete basis functions $\varphi_k(t)$ are translates of each other for all $k$ within a given subband. However, the shape of $\varphi_k(t)$ depends on the scale and orientation of the subband. Figures 8a–d shows discrete basis functions in the four coarsest subbands. The

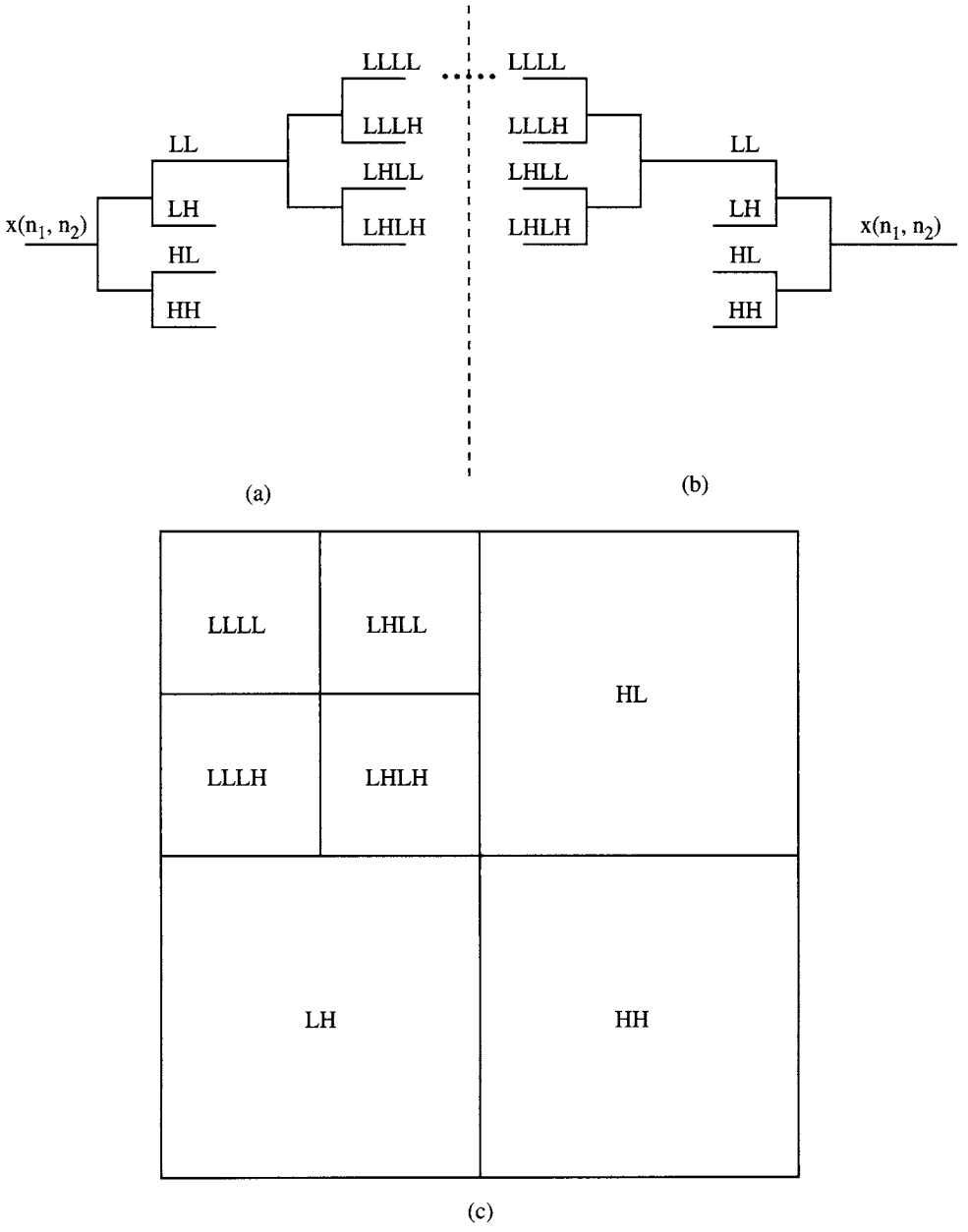



**FIGURE 7** Implementation of wavelet image decomposition using cascaded filter banks: (a) wavelet decomposition of input image $x(n_1, n_2)$; (b) reconstruction of $x(n_1, n_2)$ from its wavelet coefficients; (c) nomenclature of subbands for a 3-level decomposition.

basis function in the LL subband (Fig. 8a) is characterized by a strong central bump, while the basis functions in the other three subbands (detail images) have zero mean. Notice that the basis functions in the HL and LH subbands are related through a simple 90-degree rotation. The orientation of these basis functions make them suitable to represent patterns with the same orientation. For reasons that will become apparent in the next section, the basis functions in the low subband are called *discrete scaling functions,* while those in the other subbands are called *discrete wavelets.* The size of the support set of the basis functions is determined by the length of the wavelet filter, and essentially quadruples from one scale to the next.

## 3.4 Continuous Wavelet Bases

Basis functions corresponding to different subbands with the *same orientation* have a similar shape. This is illustrated in Fig. 9 which shows basis functions corresponding to two subbands with vertical orientation (Figs. 9a,b,c). The shape of the basis functions converges to a limit (Fig. 9d) as the scale becomes coarser. This phenomenon is due to the regularity

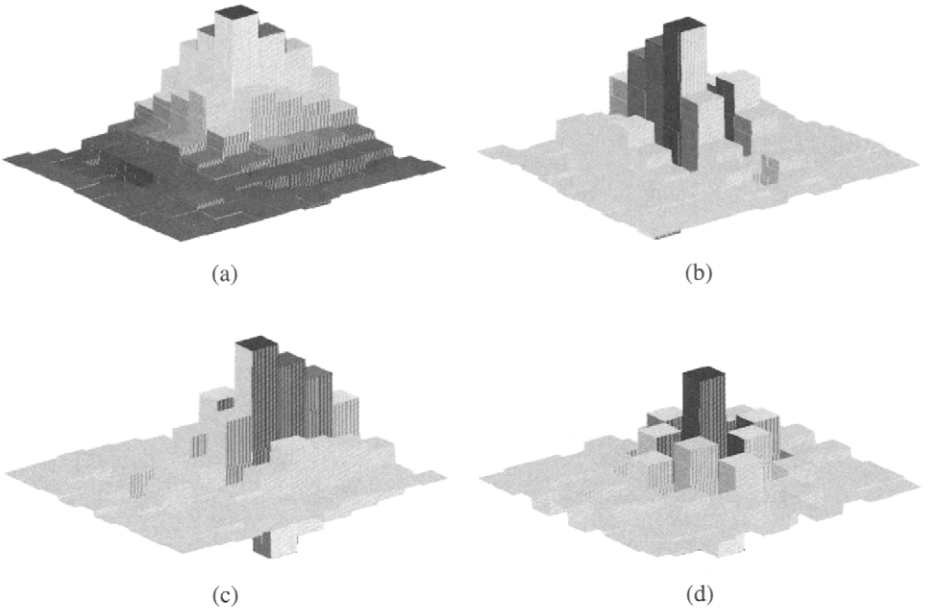



**FIGURE 8** Discrete basis functions for image representation: (a) discrete scaling function from LLLL subband; (b)–(d) discrete wavelets from LHLL, LLLH, and LHLH subbands. These basis functions are generated from Daubechies' 4-tap filter.

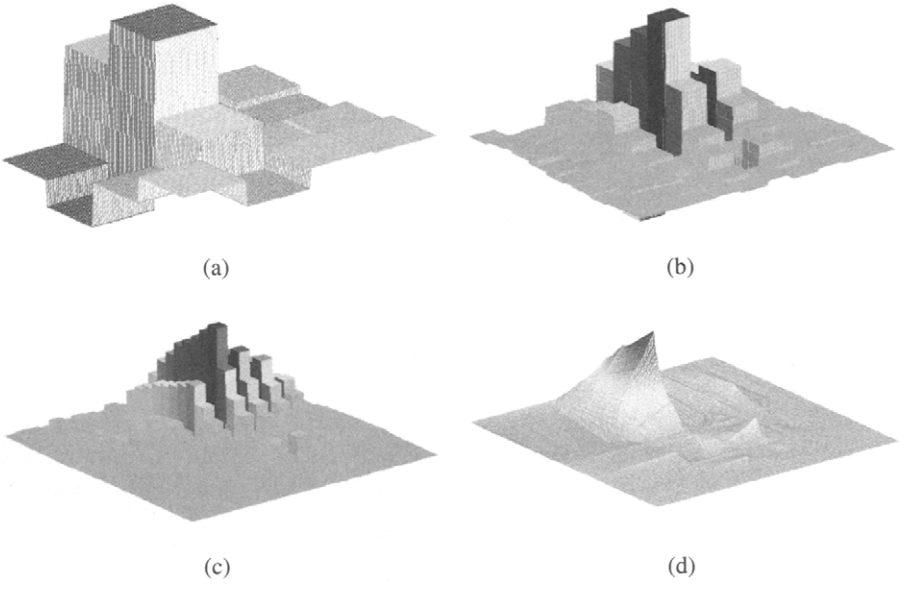



**FIGURE 9** Discrete wavelets with vertical orientation at three consecutive scales: (a) in HL band; (b) in LHLL band; (c) in LLHLLL band. (d) Continuous wavelet is obtained as a limit of (normalized) discrete wavelets as scale becomes coarser.

of the wavelet filters used (Section 3.1). One of the remarkable results of Daubechies' wavelet theory [4] is that under regularity conditions, the shape of the impulse responses corresponding to subbands with the same orientation does converge to a limit shape at coarse scales. Essentially the basis functions come in four shapes, which are displayed in Figs. 10a–d. The limit shapes corresponding to the vertical, horizontal and diagonal orientations are called *wavelets.* The limit shape corresponding to the coarse scale is called *scaling function.* The three wavelets and the scaling function depend on the wavelet filter $h_0(n)$ used (in Fig. 10, Daubechies' 4-tap filter). The four functions in Fig. 10a–d are separable and are respectively of the form $\phi(x)\phi(y)$, $\phi(x)\psi(y)$, $\psi(x)\phi(y)$, and $\psi(x)\psi(y)$. Here $(x, y)$ are horizontal and vertical coordinates,

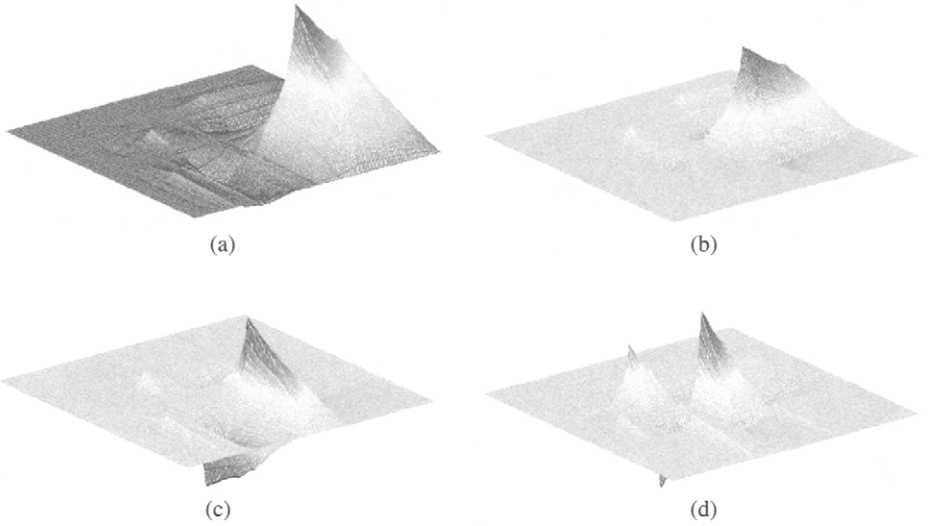



**FIGURE 10** Basis functions for image representation: (a) scaling function; (b)–(d) wavelets with horizontal, vertical and diagonal orientations. These four functions are tensor products of the 1D scaling function and wavelet in Fig. 11. The horizontal wavelet has been rotated by 180 degrees so that its negative part is visible on the display.

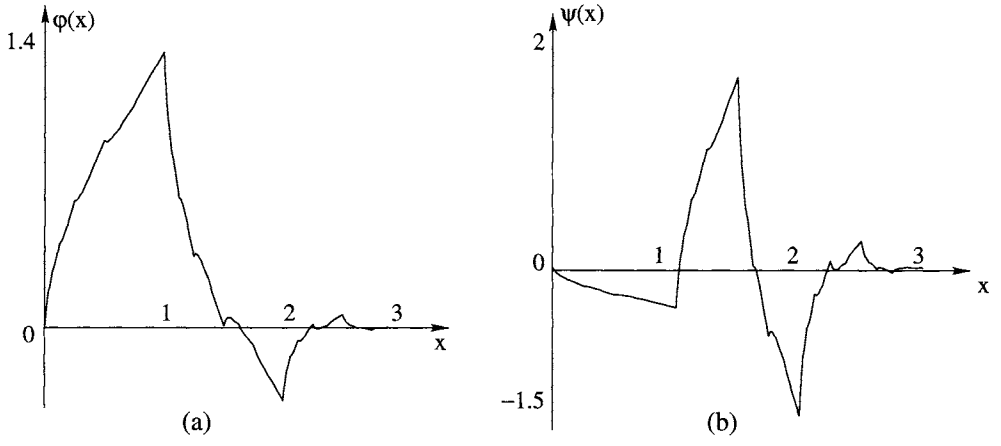



**FIGURE 11** (a) 1D scaling function and (b) 1D wavelet generated from Daubechies' D4 filter.

and $\phi(x)$ and $\psi(x)$ are, respectively, the 1D scaling function and the 1D wavelet generated by the filter $h_0(n)$. These two functions are shown in Fig. 11, respectively. While the aspect of these functions is somewhat rough, Daubechies' theory shows that the smoothness of the wavelet increases with the number $K$ of zeroes of $H_0(e^{j\omega})$ at $\omega = \pi$. In this case, the first $K$ moments of the wavelet $\psi(x)$ are zero:

$$\int x^k \psi(x) dx = 0, \quad 0 \leq k < K.$$

The wavelet is then said to possess $K$ vanishing moments.

## 3.5 More on Wavelet Image Representations

The connection between wavelet decompositions and bases for image representation shows that images are sparse linear combinations of elementary images (discrete wavelets and scaling functions) and provides valuable insights for selecting the wavelet filter. Some wavelets are better able to compactly represent certain types of images than others. For instance, images with sharp edges would benefit from the use of short wavelet filters, due to the spatial localization of such edges. Conversely, images with mostly smooth areas would benefit from the use of longer wavelet filters with several vanishing moments, as such filters generate smooth wavelets. See [14] for a performance comparison of wavelet filters in image compression.

## 3.6 Relation to Human Visual System

Experimental studies of the human visual system (HVS) have shown that the eye's sensitivity to a visual stimulus strongly depends upon the spatial frequency contents of this stimulus.

Similar observations have been made about other mammals. Simplified linear models have been developed in the psychophysics community to explain these experimental findings. For instance, the modulation transfer function describes the sensitivity of the HVS to spatial frequency, see Chapter 1.2. Additionally, several experimental studies have shown that images sensed by the eye are decomposed into bandpass channels as they move towards and through the visual cortex of the brain [15]. The bandpass components correspond to different scales and spatial orientations. Figure 5 in [16] shows the spatial impulse response and spatial frequency response corresponding to a channel at a particular scale and orientation. While the Laplacian representation provides a decomposition based on scale (rather than orientation), the wavelet transform has a limited ability to distinguish between patterns at different orientations, as each scale is comprised of three channels which are respectively associated with the horizontal, vertical, and diagonal orientations. This may not be not sufficient to capture the complexity of early stages of visual information processing, but the approximation is useful. Note there exist linear multiscale representations that more closely approximate the response of the HVS. One of them is the *Gabor transform*, for which the basis functions are Gaussian functions modulated by sine waves [17]. Another one is the *cortical transform* developed by Watson [18]. However, as discussed by Mallat [19], the goal of multiscale image processing and computer vision is not to design a transform that mimics the HVS. Rather, the analogy to the HVS motivates the use of multiscale image decompositions as a front end to complex image processing algorithms, as Nature already contains successful examples of such a design.

## 3.7 Applications

We have already mentioned several applications in which a wavelet decomposition is useful. This is particularly true of applications where the completeness of the wavelet representation is desirable. One such application is image and video compression, see Chapters 5.4 and 6.2. Another one is image denoising, as several powerful methods rely on the formulation of statistical models in an orthonormal transform domain [20], also see Chapter 3.4. There exist other applications in which wavelets present a plausible (but not necessarily superior) alternative to other multiscale decomposition techniques. Examples include texture analysis and segmentation [3, 21, 22] which is also discussed in Chapter 4.7, recognition of handwritten characters [23], inverse image halftoning [24], and biomedical image reconstruction [25].

# 4 Other Multiscale Decompositions

For completeness, we also mention two useful extensions of the methods covered in this chapter.

## 4.1 Undecimated Wavelet Transform

The wavelet transform is not invariant to shifts of the input image, in the sense that an image and its translate will in general produce different wavelet coefficients. This is a disadvantage in applications such as edge detection, pattern matching, and image recognition in general. The lack of translation invariance can be avoided if the outputs of the filter banks are not decimated. The undecimated wavelet transform then produces a set of bandpass images which have the same size as the original dataset ($N \times M$).

## 4.2 Wavelet Packets

Although the wavelet transform often provides a sparse representation of images, the spatial frequency characteristics of some images may not be best suited for a wavelet representation. Such is the case of fingerprint images, as ridge patterns constitute relatively narrowband bandpass components of the image. An even sparser representation of such images can be obtained by recursively splitting the appropriate subbands (instead of systematically splitting the low–frequency band as in a wavelet decomposition). This scheme is simply termed *subband decomposition*. This approach was already developed in signal processing during the 1970s [5]. In the early 1990s, Coifman and Wickerhauser developed an ingenious algorithm for finding the subband decomposition that gives the sparsest representation of the input signal (or image) in a certain sense [26]. The idea has been extended to find the best subband decomposition for compression of a given image [27].

## 4.3 Geometric Wavelets

One of the main strengths of one-dimensional wavelets is their ability to represent abrupt transitions in a signal. This property does not extend straightforwardly to higher dimensions. In particular, the extension of wavelets to two dimensions, using tensor-product constructions, has two shortcomings: (1) limited ability to represent patterns at arbitrary orientations, and (2) limited ability to represent image edges. For instance, the tensor-product construction is suitable for capturing the discontinuity across an edge, but is ineffective for exploiting the smoothness along the edge direction. To represent a simple, straight edge, one needs many wavelets.

To remedy this problem, several researchers have recently developed improved two-dimensional multiresolution representations. The idea was pioneered by Candès and Donoho [28]. They introduced the *ridgelet transform*, which decomposes images as a superposition of ridgelets such as the one shown in Fig. 12. A ridgelet is parameterized by three parameters: resolution, angle, and location. Ridgelets are also known as geometric wavelets, a growing family which includes exotically named functions such as curvelets,

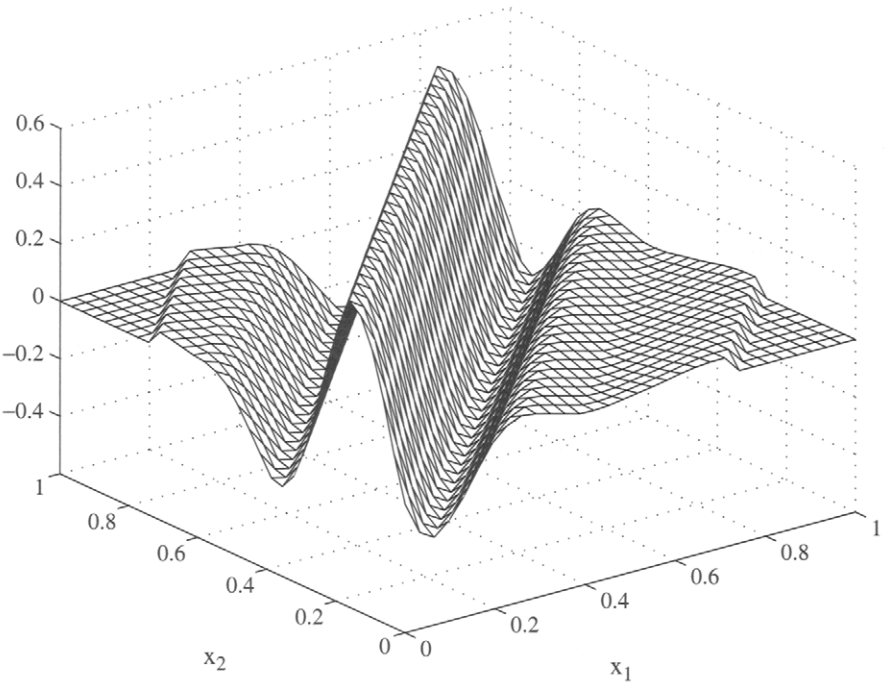



**FIGURE 12** Ridgelet. (*Courtesy of M. Do.*)

bandelets, and contourlets. Signal processing algorithms for discrete images and applications to denoising and compression have been developed by Starck et al. [29], Do and Vetterli [30, 31], and Le Pennec and Mallat [32]. Remarkable results have been obtained by exploiting the sparse representation of object contours offered by geometric wavelets.

## 5 Conclusion

We have introduced basic concepts of multiscale image decompositions and wavelets. We have focused on three main techniques: Gaussian pyramids, Laplacian pyramids, and wavelets. The Gaussian pyramid provides a representation of the same image at multiple scales, using simple lowpass filtering and decimation techniques. The Laplacian pyramid provides a coarse representation of the image as well as a set of detail images (bandpass components) at different scales. Both the Gaussian and the Laplacian representation are overcomplete, in the sense that the total number of pixels is approximately 33% higher than in the original image.

Wavelet decompositions are a more recent addition to the arsenal of multiscale signal processing techniques. Unlike the Gaussian and Laplacian pyramids, they provide a complete image representation and perform a decomposition according to both scale and orientation. They are implemented using cascaded filter banks in which the lowpass and highpass filters satisfy certain specific constraints. While classical signal processing concepts provide an operational understanding of such systems, there exist remarkable connections with work in applied mathematics (by Daubechies, Mallat, Meyer, and others) and in psychophysics, which provide a deeper understanding of wavelet decompositions and their role in vision. From a mathematic standpoint, wavelet decompositions are equivalent to signal expansions in a wavelet basis. The regularity and vanishing-moment properties of the lowpass filter impact the shape of the basis functions and hence their ability to efficiently represent typical images. From a psychophysical perspective, early stages of human visual information processing apparently involve a decomposition of retinal images into a set of bandpass components corresponding to different scales and orientations. This suggests that multiscale/multiorientation decompositions are indeed natural and efficient for visual information processing.

## Acknowledgment

I would like to thank Juan Liu for generating the figures and plots in this chapter.

## References

[1] A. Rosenfeld, *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed., Springer-Verlag, 1984.

[2] P. Burt, "Multiresolution Techniques for Image Representation, Analysis, and 'Smart' Transmission," *SPIE*, 1199, 1989.

[3] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Transform," *IEEE Trans. Patt. Anal. and Mach. Intell.*, 11, 7, 674–693, 1989.

[4] I. Daubechies, *Ten Lectures on Wavelets*, CBMS–NSF Regional Conference Series in Applied Mathematics, 61, SIAM, Philadelphia, 1992.

[5] M. Vetterli and J. Kovačevič, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[6] S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

[7] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Kluwer, Norwell, MA, 2001.

[8] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Data Compression Standard*, Van Nostrand Reinhold, 1993.

[9] J. Proakis and Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd Ed., Prentice-Hall, 1996.

[10] M. K. Tsatsanis and G. B. Giannakis, "Principal Component Filter Banks for Optimal Multiresolution Analysis," *IEEE Trans. Sig. Proc.*, 43, 8, 1766–1777, Aug. 1995.

[11] P. Burt and A. H. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Trans. Comm.*, 31, 532–540, 1983.

[12] M. Vetterli and K. M. Uz, "Multiresolution coding techniques for digital video: a review," *Multidimensional Sys. and Signal Proc.*, Special Issue on Multidimensional Proc. of Video Signals, 3, 161–187, 1992.

[13] M. Vetterli, "Multi-dimensional Sub–band Coding: Some Theory and Algorithms," *Signal Processing*, 6, 2, 97–112, 1984.

[14] J. D. Villasenor, B. Belzer and J. Liao, "Wavelet Filter Comparison for Image Compression," *IEEE Trans. Im. Proc.*, 4, 8, 1053–1060, 1995.

[15] F. W. Campbell and J. G. Robson, "Application of Fourier Analysis to Cortical Cells," *J. Physiology*, 197, 551–566, 1968.

[16] M. Webster and R. De Valois, "Relationship Between Spatial-Frequency and Orientation Tuning of Striate-Cortex Cells," *J. Opt. Soc. Amer.*, July 1985.

[17] J. G. Daugmann, "Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile," *Vision Research*, 20, 847–856, 1980.

[18] A. B. Watson, "The Cortex Transform: Rapid Computation of Simulated Neural Images," *Computer Graphics and Image Processing*, 39, 1987.

[19] S. G. Mallat, "Multifrequency Channel Decompositions of Images and Wavelet Models," *IEEE Trans. Acoustics, Speech and Sig. Proc.*, 37, 12, Dec. 1989.

[20] P. Moulin and J. Liu, "Analysis of Multiresolution Image Denoising Schemes Using Generalized-Gaussian and Complexity Priors," to appear in *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Analysis, Apr. 1999.

[21] M. Unser, "Texture Classification and Segmentation Using Wavelet Frames," *IEEE Trans. on Image Processing*, 4, 11, 1549–1560, 1995.

[22] R. Porter and N. Canagarajah, "A Robust Automatic Clustering Scheme for Image Segmentation Using Wavelets," *IEEE Trans. on Image Processing*, 5, 4, 662–665, 1996.

[23] Y. Qi and B. R. Hunt, "A Multiresolution Approach to Computer Verification of Handwritten Signatures," *IEEE Trans. on Image Processing*, 4, 6, 870–874, 1995.

[24] J. Luo, R. de Queiroz and Z. Fan, "A Robust Technique for Image Descreening Based on the Wavelet Transform," *IEEE Trans. on Signal Processing*, 46, 4, 1179–1184, Apr. 1998.

[25] A. H. Delaney and Y. Bresler, "Multiresolution Tomographic Reconstruction Using Wavelets," *IEEE Trans. on Image Processing*, 4, 6, 799–813, 1995.

[26] R. R. Coifman and M. V. Wickerhauser, "Entropy-Based Algorithms for Best Basis Selection," *IEEE Trans. on Info. Theory*, Special Issue on Wavelet Tranforms and Multiresolution Signal Analysis, 38, 2, 713–718, 1992.

[27] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Im. Proc.*, 2, 160–175, Apr. 1993.

[28] E. J. Candès, "Ridgelets: Theory and Applications," *Ph.D. Thesis*, Department of Statistics, Stanford University, 1998.

[29] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The Curvelet Transform for Image Denoising," *IEEE Trans. on Image Processing*, 11, 6, 670–684, June 2002.

[30] M. N. Do and M. Vetterli, "The Finite Ridgelet Transform for Image Representation," *IEEE Trans. on Image Processing*, 12, 1, 16–28, Jan. 2003.

[31] M. N. Do and M. Vetterli, "Contourlets" in G. V. Welland, editor, *Beyond Wavelets*, Academic Press, New York, 2003.

[32] E. Le Pennec and S. G. Mallat, "Sparse Geometrical Image Approximation with Bandelets," to appear in *IEEE Trans. on Image Processing*, 2005.