MIT Media Laboratory Vision and Modeling Technical Report #137.
Appears in: "Subband Coding", edited by John Woods, Kluwer Academic Press, 1990.

Chapter 4

Subband Transforms

Eero P. Simoncelli[†] and Edward H. Adelson[‡]
Vision Science Group, The Media Laboratory, and
[†]Department of Electrical Engineering and Computer Science
[‡]Department of Brain and Cognitive Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Linear transforms are the basis for many techniques used in image processing, image analysis, and image coding. Subband transforms are a subclass of linear transforms which offer useful properties for these applications. In this chapter, we discuss a variety of subband decompositions and illustrate their use in image coding. Traditionally, coders based on linear transforms are divided into two categories: transform coders and subband coders. This distinction is due in part to the nature of the computational methods used for the two types of representation.

Transform coding techniques are usually based on *orthogonal* linear transforms. The classic example of such a transform is the discrete Fourier transform (DFT), which decomposes a signal into sinusoidal frequency components. Two other examples are the discrete cosine transform (DCT) and the Karhunen-Loeve transform (KLT). Conceptually, these transforms are computed by taking the inner product of the finite-length signal with a set of basis functions. This produces a set of coefficients, which are then passed on to the

This work was supported by contracts with the IBM Corporation (agreement dated 1/1/84) and DARPA (Rome Airforce F30602-89-C-0022) and a grant from the NSF (IRI 871-939-4). The opinions expressed are those of the authors and do not necessarily represent those of the sponsors.

quantization stage of the coder. In practice, many of these transforms have efficient implementations as cascades of "butterfly" computations. Furthermore, these transforms are usually applied independently to non-overlapping sub-blocks of the signal.

Subband transforms are generally computed by convolving the input signal with a set of bandpass filters and decimating the results. Each decimated subband signal encodes a particular portion of the frequency spectrum, corresponding to information occurring at a particular spatial scale. To reconstruct the signal, the subband signals are upsampled, filtered, and then combined additively. For purposes of coding, subband transforms can be used to control the relative amounts of error in different parts of the frequency spectrum. Most filter designs for subband coders attempt to minimize the "aliasing" resulting from the subsampling process. In the spatial domain, this aliasing appears as evidence of the sampling structure in the output image. An ideal subband system incorporates "brick-wall" bandpass filters which avoid aliasing altogether. Such filters, however, produce ringing (Gibbs phenomenon) in the spatial domain which is perceptually undesirable.

Although coders are usually classified in one of these two categories, there is a significant amount of overlap between the two. In fact, the latter part of this chapter will focus on transforms which may be classified under either category. As an example, consider the block discrete cosine transform (DCT), in which the signal (image) is divided into non-overlapping blocks, and each block is decomposed into sinusoidal functions. Several of these sinusoidal functions are depicted in figure 4.1. The basis functions are orthogonal, since the DCT is orthogonal and the blocks are chosen so that they do not overlap. Coders employing the block DCT are typically classified as transform coders.

We may also view the block DCT as a subband transform. Computing a DCT on non-overlapping blocks is equivalent to convolving the image with each of the block DCT basis functions and then subsampling by a factor equal to the block spacing. The Fourier transform of the basis functions (also shown in figure 4.1) indicates that each of the DCT functions is selective for a particular frequency subband, although it is clear that the subband localization is rather poor. Thus, the DCT also qualifies as a subband transform.

4.1 Subband Transform Properties

Given the overlap between the categories of transform and subband coders, what criteria should be used in choosing a linear transformation for coding purposes? We will consider a set of properties which are relevant to the problem

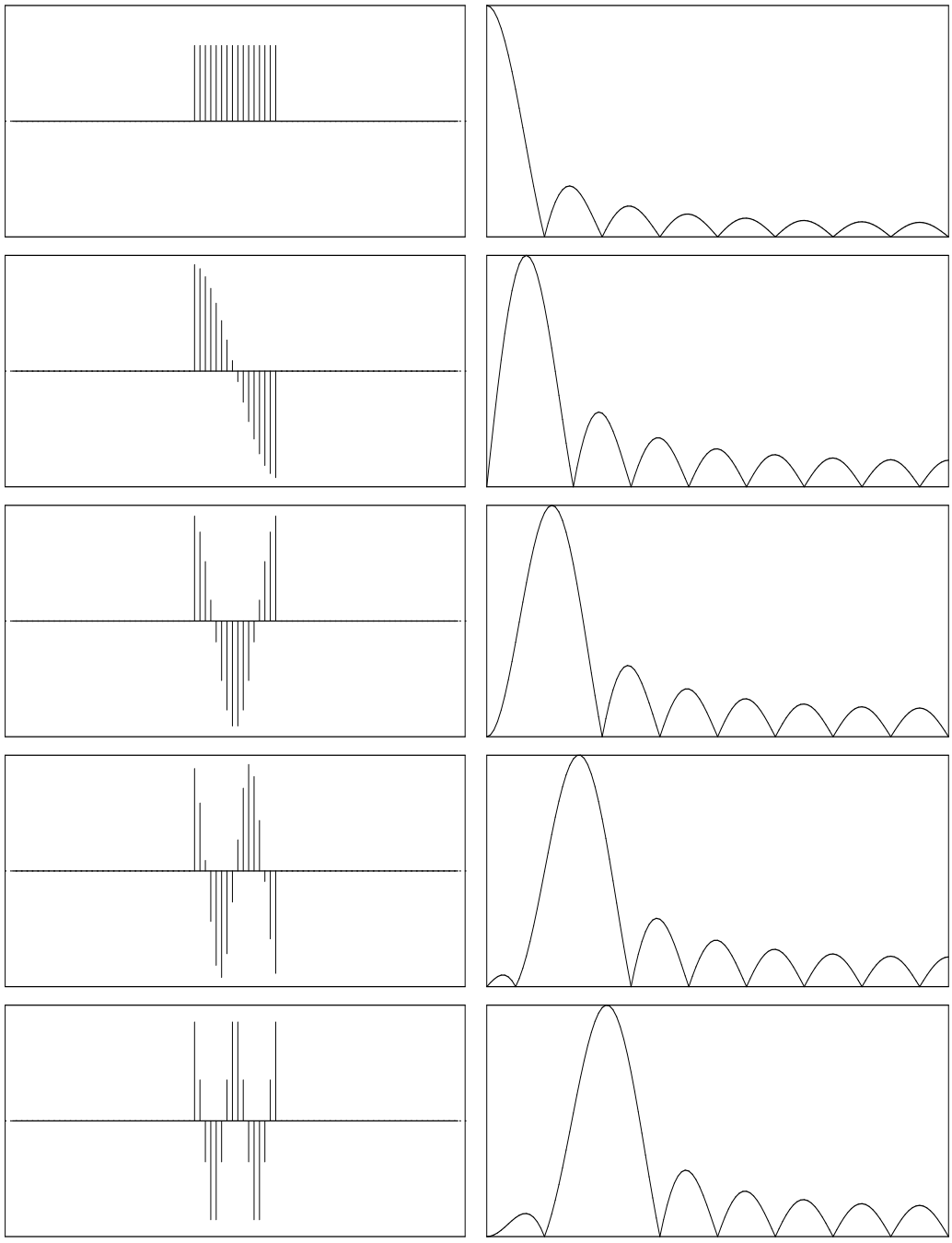


Figure 4.1: Several of the 16-point DCT basis functions (left) with their corresponding Fourier transforms (right). The Fourier transforms are plotted on a linear scale over the range from 0 to π.

of image coding.

Scale and Orientation

An explicit representation of scale is widely accepted as being important for effective image representation [1, 2, 3, 4, 5, 6]. Images contain objects and features of many different sizes which may be viewed over a large range of distances, and therefore, a transformation should analyze the image simultaneously (and independently) at different scales. Several authors have argued that the correct partition in terms of scale is one in which the scales are related by a fixed constant of proportionality. In the frequency domain, this corresponds to a decomposition into localized subbands with equal widths on a logarithmic scale.

For two-dimensional signals, a localized region in the frequency plane corresponds spatially to a particular scale and *orientation*. Orientation specificity allows the transform to extract higher order oriented structures typically found in images, such as edges and lines. Thus, it is useful to construct transformations which partition the input signal into localized patches in the frequency domain.

Spatial localization

In addition to localization in frequency, it is advantageous for the basis functions to be spatially localized; that is, the transform should encode positional information. The necessity of spatial localization is particularly apparent in machine vision systems, where information about the location of features in the image is critical. This localization should not, however, occur abruptly as in the block DCT example given earlier – abrupt transitions lead to poor localization in the frequency domain.

The concept of joint localization in the spatial and spatial-frequency domains may be contrasted with the two most common representations used for the analysis of linear systems: the sampled or *discrete* signal, and its Fourier transform. The first of these utilizes the standard basis set for discrete signals consisting of impulses located at each sample location. These basis functions are maximally localized in space, but convey no information about scale. On the other hand, the Fourier basis set is composed of even and odd phase sinusoidal sequences, whose usefulness is primarily due to the fact that they are the eigenfunctions of the class of linear shift-invariant systems. Although they are maximally localized in the frequency domain, each one covers the entire spatial extent of the signal.

It is clear that representation in the space or frequency domains is extremely useful for purposes of system analysis, but this does *not* imply that impulses or sinusoids are the best way to encode signal information. In a number of recent papers [7, 8, 9], the importance of this issue is addressed and related to a 1946 paper by Dennis Gabor [10], who showed that the class of linear transformations may be considered to span a range of joint localization with the impulse basis set and the Fourier basis set at the two extremes. He demonstrated that one-dimensional signals can be represented in terms of basis functions which are localized both in space *and* frequency. We will return to Gabor's basis set in section 4.3.

Orthogonality

A final property to be considered is orthogonality. The justification usually given for the orthogonality constraint is in terms of decorrelation. Given a signal with prescribed second order statistics (i.e. a covariance matrix), there is an orthogonal transform (the Karhunen-Loeve transform) which will decorrelate the signal (i.e. diagonalize the covariance matrix). In other words, the second order correlations of the transform coefficients will be zero. Orthogonality is usually not discussed in the context of subband transforms, although many such transformas are orthogonal. The examples in the next section will demonstrate that although orthogonality is not strictly necessary, a transform that is strongly non-orthogonal may be undesirable for coding.

4.2 Linear Transformations on Finite Images

The results presented in this chapter are based on analysis in both the spatial and the frequency domains, and thus rely on two separate notational frameworks: the standard matrix notation used in linear algebra, and the Fourier domain representations commonly used in digital signal processing. In this section, we describe the two types of notation and make explicit the connection between them. For simplicity, we will restrict the discussion to analysis of one dimensional systems, although the notation may be easily extended to multiple dimensions.

4.2.1 Analysis/Synthesis Filter Bank Formulation

We will be interested in linear transformations on images of a *finite* size which may be expressed in terms of convolutions with finite impulse response (FIR)

filters. The schematic diagram in figure 4.2 depicts a convolution-based system known as an analysis/synthesis (A/S) filter bank [11]. The notation in the diagram is standard for digital signal processing [12], except that for the purposes of this paper, the boxes $\boxed{H_i(\omega)}$ indicate *circular* convolution of a finite input image of size N with a filter with impulse response $h_i(n)$ and Fourier transform

$$H_i(\omega) = \sum_n h_i(n) e^{-j\omega n}$$

We do not place a causality constraint on the filter impulse responses, since they are meant for application to images. We do, however, assume that the region of support of the filter is smaller than the image size. The boxes $\boxed{k_i \downarrow}$ indicate that the sequence is subsampled by a factor of k_i where k_i is an integer for all i. The boxes $\boxed{k_i \uparrow}$ indicate that the sequence should be upsampled by inserting $k_i - 1$ zeros between each sample. We will assume that the integers k_i are divisors of N.

The analysis section of the A/S system takes an input sequence $x(n)$ of length N and performs a linear transformation to decompose it into M sequences $y_i(n)$ of length N/k_i. The synthesis section performs the inverse operation of the analysis transformation. Here the M sequences $y_i(n)$ are upsampled and, after filtering with filters $g_i(n)$, are combined additively to give an approximation $\hat{x}(n)$ to the original sequence, $x(n)$. Note that although one-dimensional signals are indicated in the diagram, the system is equally valid for multi-dimensional signals if we replace occurences of the scalars n, ω, k_i with vectors $\boldsymbol{n}$, $\boldsymbol{\omega}$, and a matrix $\mathbf{K}_i$, respectively.

The use of the A/S formulation emphasizes the computation of the transform coefficients through convolution. This is intuitively desirable since different regions of the image should be processed in the same manner. Furthermore, the expression of the problem in the frequency domain allows us to easily separate the error $e(n) = \hat{x}(n) - x(n)$ into two parts: an aliasing component and a shift-invariant component. To see this, we write the contents of the intermediate signals $y_i(n)$ in the frequency domain as

$$Y_i(\omega) = \frac{1}{k} \sum_{j=0}^{k-1} H_i\left(\tfrac{\omega}{k} + \tfrac{2\pi j}{k}\right) X\left(\tfrac{\omega}{k} + \tfrac{2\pi j}{k}\right) \tag{4.1}$$

and the A/S system output as

$$\hat{X}(\omega) = \sum_{i=0}^{M-1} Y_i(k\omega) G_i(\omega)$$

where we have used well-known facts about the effects of upsampling and downsampling in the frequency domain [12]. Combining the two gives

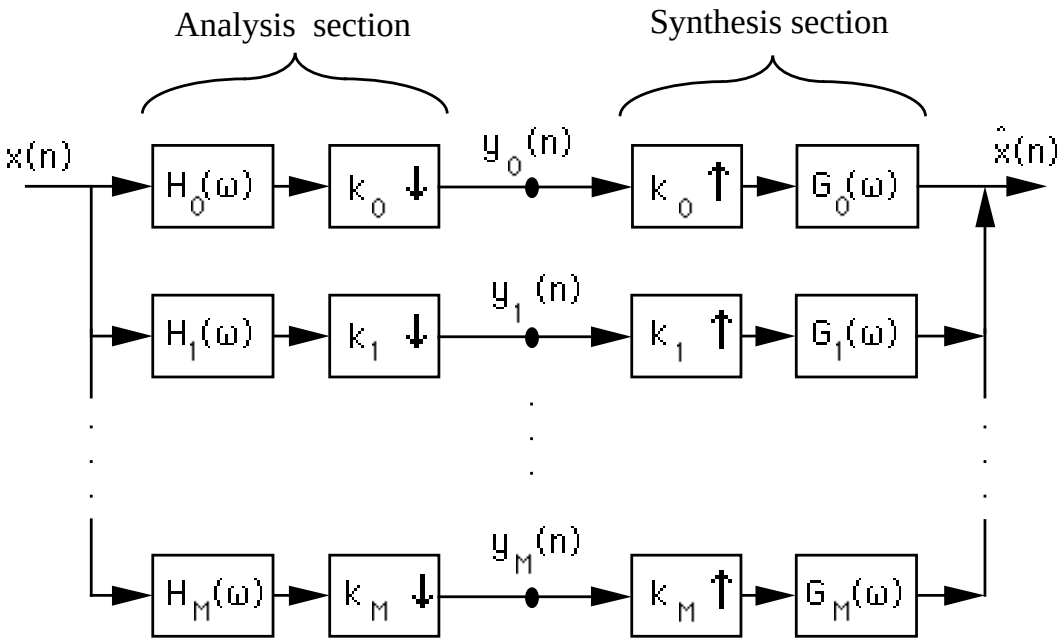


Figure 4.2: An analysis/synthesis filter bank.

$$\begin{aligned}\hat{X}(\omega) &= \frac{1}{k}\sum_{i=0}^{M-1}\left[\sum_{j=0}^{k-1} H_i\left(\omega+\tfrac{2\pi j}{k}\right) X\left(\omega+\tfrac{2\pi j}{k}\right)\right] G_i(\omega) \\ &= \frac{1}{k}\sum_{i=0}^{M-1} H_i(\omega)G_i(\omega)X(\omega) \\ &\quad + \frac{1}{k}\sum_{j=1}^{k-1} X\left(\omega+\tfrac{2\pi j}{k}\right)\sum_{i=0}^{M-1} H_i\left(\omega+\tfrac{2\pi j}{k}\right) G_i(\omega) \qquad (4.2)\end{aligned}$$

The first sum corresponds to a linear shift-invariant system response, and the second contains the system aliasing.

4.2.2 Cascaded Systems

A further advantage of the A/S system is that it allows explicit depiction and analysis of hierarchically constructed transformations. If we assume that we are dealing with A/S systems with perfect response (that is, $\hat{x}(n) = x(n)$), then any intermediate signal $y_i(n)$ of an A/S system may be further decomposed by application of any other A/S system. To make this notion more precise, an example is given in the diagram of figure 4.3 in which an A/S system has been re-applied to its own intermediate signal $y_0(n)$. If the original A/S system (as shown in figure 4.2) had a perfect response then it is clear that the

two-stage system shown in figure 4.3 will also have a perfect response. If the cascading is applied to each of the M intermediate signals $y_i(n)$, we will call the system a *uniform* cascade system. Otherwise, it will be termed a *non-uniform* or *pyramid* cascade. A system which we will discuss in section 4.3 is based on pyramid cascades of two-band A/S systems. Such a cascade produces an octave-width subband decomposition, as illustrated in the idealized frequency diagram in figure 4.4.

4.2.3 Matrix Formulation

An alternative to the frequency domain notation associated with the A/S filter bank is the matrix notation of linear algebra. An image of finite extent which has been sampled on a discrete lattice may be written as a finite length column vector $\boldsymbol{x}$ which corresponds to a point in $\mathbf{R}^N$, the set of all real N-tuples. The value of each component of $\boldsymbol{x}$ is simply the corresponding sample value in the image. Multi-dimensional images are converted to this vector format by imposing an arbitrary but fixed order on the lattice positions. If we let N be the length of the vector $\boldsymbol{x}$, a linear transformation on the image corresponds to multiplication of $\boldsymbol{x}$ by some matrix $\mathbf{M}$ with N columns.

Since the analysis and synthesis stages of the system in figure 4.2 each correspond to linear transformations, we may represent the same transformations using matrix notation. Using the definition of convolution, and assuming (for simplicity) a one-dimensional system, we may write

$$y_i(m) = \sum_{l=0}^{N-1} x(l) h_i(k_i m - l)$$

and

$$\hat{x}(n) = \sum_{i=0}^{M-1} \sum_{m=0}^{\frac{N}{k_i}-1} y_i(m) g_i(n - k_i m)$$

where the filter and image sample locations $(k_i m - l)$ and $(n - k_i m)$ are computed modulo N. These expressions may be formulated as matrix-vector products

$$\boldsymbol{y} = \mathbf{H}^t \boldsymbol{x}$$

and

$$\hat{\boldsymbol{x}} = \mathbf{G}\boldsymbol{y}$$

or combining these two equations

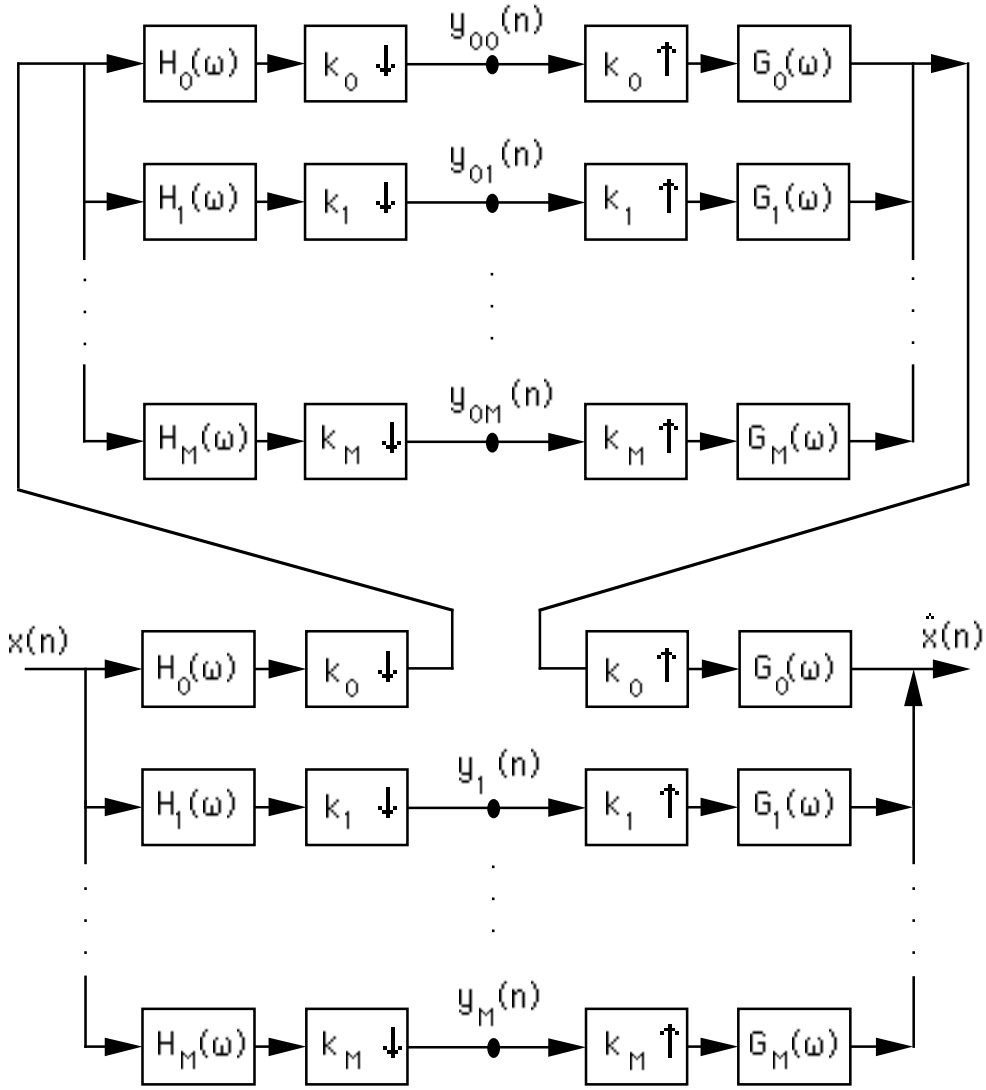


Figure 4.3: A non-uniformly cascaded analysis/synthesis filter bank.

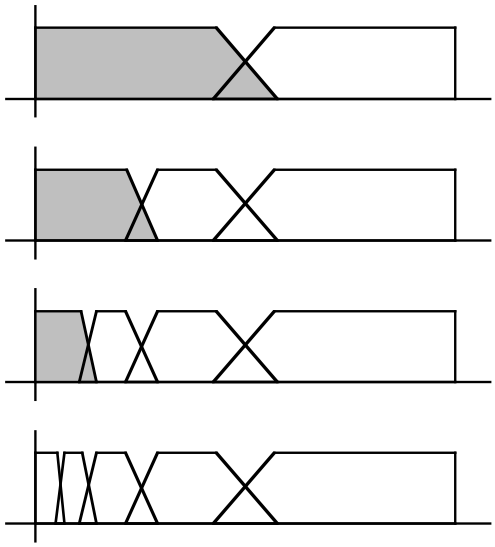


Figure 4.4: Octave band splitting produced by a four-level pyramid cascade of a two-band A/S system. The top picture represents the splitting of the two-band A/S system. Each successive picture shows the effect of re-applying the system to the lowpass subband (indicated in grey) of the previous picture. The bottom picture gives the final four-level partition of the frequency domain. All frequency axes cover the range from 0 to π.

$$\hat{\boldsymbol{x}} \;=\; \mathbf{G}\mathbf{H}^t\boldsymbol{x} \tag{4.3}$$

where $\boldsymbol{y}$ and $\hat{\boldsymbol{x}}$ are N-vectors, the superscript t indicates matrix transposition, and

$$\mathbf{H} \;=\; \begin{bmatrix} h_0(0) & h_0(k_0) & & h_1(0) & h_1(k_1) & \\ h_0(-1) & h_0(k_0-1) & & h_1(-1) & h_1(k_1-1) & \\ h_0(-2) & h_0(k_0-2) & & h_1(-2) & h_1(k_1-2) & \\ \vdots & h_0(k_0-3) & \cdots & \vdots & h_1(k_1-3) & \cdots \\ & h_0(k_0-4) & & & h_1(k_1-4) & \\ h_0(2) & \vdots & & h_1(2) & \vdots & \\ h_0(1) & & & h_1(1) & & \end{bmatrix} \tag{4.4}$$

and

$$\mathbf{G} \;=\; \begin{bmatrix} g_0(0) & g_0(k_0) & & g_1(0) & g_1(k_1) & \\ g_0(1) & g_0(k_0+1) & & g_1(1) & g_1(k_1+1) & \\ g_0(2) & g_0(k_0+2) & & g_1(2) & g_1(k_1+2) & \\ \vdots & g_0(k_0+3) & \cdots & \vdots & g_1(k_1+3) & \cdots \\ & g_0(k_0+4) & & & g_1(k_1+4) & \\ g_0(-2) & \vdots & & g_1(-2) & \vdots & \\ g_0(-1) & & & g_1(-1) & & \end{bmatrix} \tag{4.5}$$

The columns of $\mathbf{G}$, composed of copies of the filter kernels shifted by increments of k_i and imbedded in vectors of length N, are known as the *basis* functions of the transformation, and the columns of $\mathbf{H}$, composed of copies of the *time-inverted* filters $h_i(-n)$ shifted by increments of k_i, are the *sampling* functions of the transformation.

From the discussion above, it is clear that we can express any linear A/S system in matrix form. The converse of this result is also true: there is an A/S system corresponding to the linear transformation and inverse transformation defined by any invertible matrix $\mathbf{M}$. Given a transformation matrix $\mathbf{M}$ with l rows, we trivially create an analysis filter bank with $k_i = N$ for each i, containing l different filters, each defined by a row of the matrix $\mathbf{M}$.

4.2.4 Inverse Transforms

A primary advantage of the matrix notation is the ease with which it can express the conditions for transform invertibility. From equation (4.3), we see that in order for the A/S system to perfectly reconstruct the original signal $x(n)$, the corresponding matrices must obey

$$\mathbf{G}\mathbf{H}^t = \mathbf{I} \tag{4.6}$$

where $\mathbf{I}$ is the identity matrix. If $\mathbf{H}$ has rank N and is square, we may choose a synthesis matrix

$$\mathbf{G} = (\mathbf{H}^{-1})^t \tag{4.7}$$

which will also be square with rank N. Thus, transform inversion in the spatial domain is a conceptually simple procedure and we will find it useful in the analysis of A/S systems. Furthermore, it should be clear that $\mathbf{H}$ and $\mathbf{G}$ may be interchanged, thus using the basis functions as sampling functions and vice versa.

If the matrix $\mathbf{H}$ is of rank N but is *not* square (that is, the representation is *overcomplete*), we may always build a perfect reconstruction system by choosing $\mathbf{G}$ to be the generalized inverse or pseudo-inverse [13] of $\mathbf{H}$:

$$\mathbf{G} = (\mathbf{H}\mathbf{H}^t)^{-1}\mathbf{H} \tag{4.8}$$

If $\mathbf{H}$ is square, equation (4.8) reduces to the solution given in equation (4.7). Similarly, if we start with a (possibly non-square) matrix $\mathbf{G}$ of rank N, we may choose $\mathbf{H} = (\mathbf{G}\mathbf{G}^t)^{-1}\mathbf{G}$.

4.2.5 Orthogonal Transforms

As mentioned in the introduction, the issue of orthogonality is usually not considered when discussing subband filters. It is, however, a property which is relevant to image coding, as we will discuss in the next section. A matrix $\mathbf{M}$ corresponding to an orthogonal transformation is a square matrix with the property that

$$\mathbf{M}\mathbf{M}^t = \mathbf{M}^t\mathbf{M} = \mathbf{I} \tag{4.9}$$

In terms of the columns or basis functions of M, this means that the inner product of any two distinct columns must be zero, and the inner product of a column with itself must be unity.

The orthogonality condition places a number of restrictions on the corresponding A/S system. Since the transformation matrix must be square, the number of samples in the transformed signal must be equal to N, the number of samples in the original image. For the A/S system, this means that

$$\sum_{i=0}^{M-1} \frac{1}{k_i} = 1$$

where we have assumed that N is divisible by all of the k_i. Such a system has been termed a *maximally decimated* or *critically sampled* filter bank [11].

A second, more important constraint is placed on the A/S system by orthogonality. Combining the perfect reconstruction requirement in (4.6) with the orthogonality constraint in (4.9) gives

$$\mathbf{G} = \mathbf{H}$$

If we consider the relationships between the A/S filters h and g and the matrices H and G described by equations (4.4) and (4.5), this means that the filters must obey

$$g_i(n) = h_i(-n), \qquad \text{for all } i \tag{4.10}$$

In other words, the synthesis filters of an orthogonal transform are *time-inverted* versions of the analysis filters.

4.3 Some Example Transforms

In this section, we will briefly discuss three one-dimensional transforms to illustrate some of the points made in the previous sections. Each transform will demonstrate both advantageous and disadvantageous properties for coding.

The Gabor Transform

In the introduction to this chapter, we argued that the basis functions of a useful decomposition should be localized in both the spatial and the spatial-frequency domains. One solution to the problem of spatially localized subband decomposition is that proposed by Dennis Gabor [10]. Gabor introduced a one-dimensional transform in which the basis functions are sinusoids weighted by Gaussian windows. The Gabor transform can be considered to perform a localized frequency decomposition in a set of overlapping windows. The resulting basis functions are localized in both space and spatial frequency; in fact, Gabor showed that this joint localization was optimal with respect to a measure that he chose (although Lerner [14] later noted that altering the measure of joint localization produces different optimal functions). The first five basis functions of a Gabor transform are shown in figure 4.5, along with their frequency spectra. Both the basis functions and their transforms are smooth and compact. In two dimensions, the Gabor basis functions are directional sinusoids weighted by gaussian windows. Daugman [15, 16] has used two-dimensional Gabor transforms for image compression.

The primary difficulty with the Gabor transform is that it is strongly non-orthogonal (i.e. the sampling functions are drastically different from the basis functions). The sampling functions corresponding to the Gabor transform (computed by inverting the transformation matrix) are depicted in figure 4.6. These functions are extremely poorly behaved, both in the spatial and spatial-frequency domains. In a coding application, errors introduced by quantization of the coefficients will be distributed throughout the spatial and frequency domains, even though the coefficient values are computed based on information in localized spatial and frequency regions.

It is interesting to note that the localization of the inverse Gabor functions can be substantially improved if one uses an overcomplete Gabor basis set. This can be accomplished by spacing the Gaussian windows more closely than is required, or by dividing each window into more frequency bands. This results in an increase in the number of coefficients, however, which may be disadvantageous for coding systems. The use of overcomplete Gabor sets for coding remains a topic for further research.

Several authors have discussed related overcomplete oriented transforms for use in image coding. Kunt [17] advocated the use of directional (i.e. orientation) subdivision for image coding, and used an oriented decomposition for this purpose. Watson [18] developed the Cortex transform, an overcomplete transform which decomposes the image into oriented octave-bandwidth subbands, and used it to compress image data.

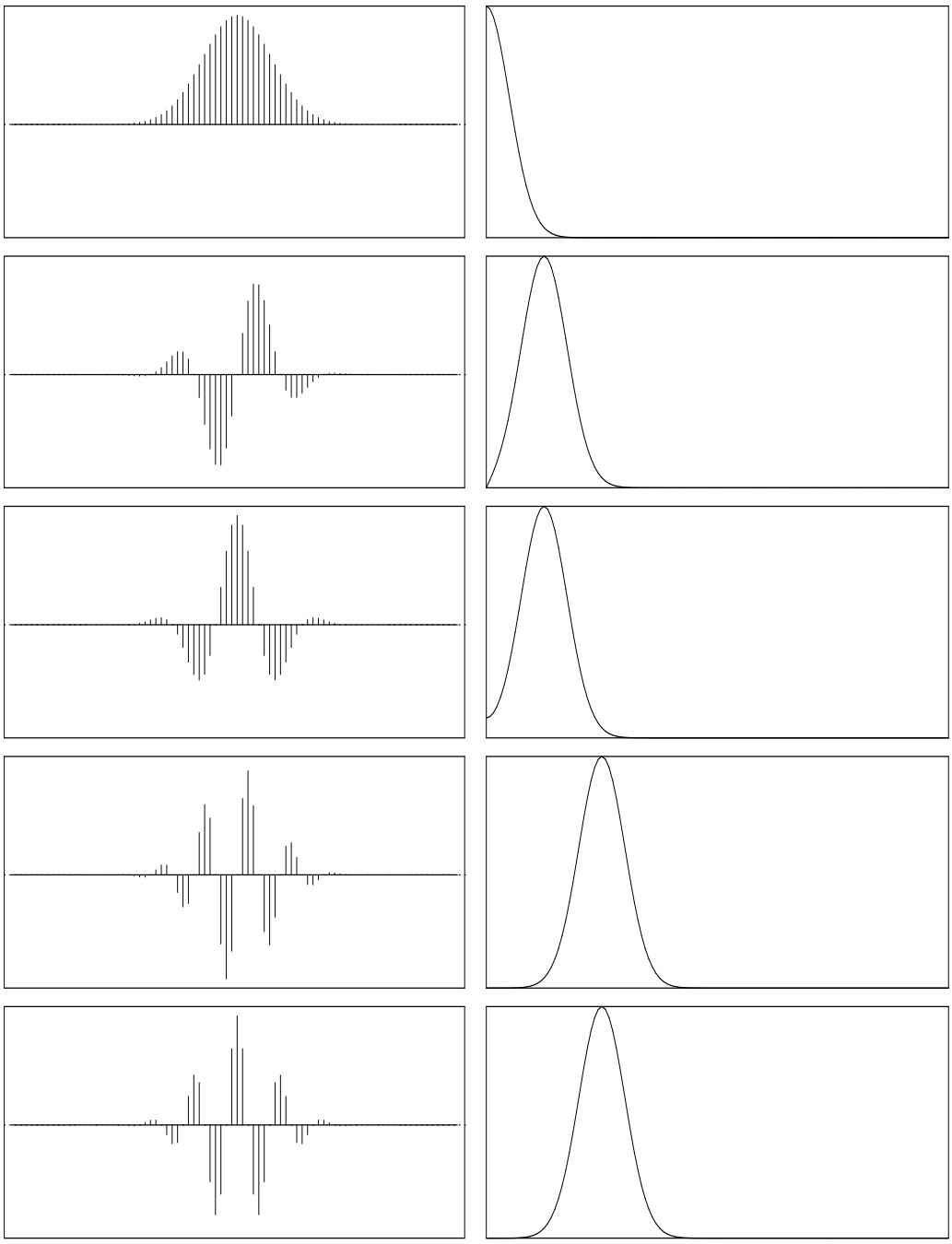


Figure 4.5: Five of the sixteen basis functions of a Gabor filter set, with their corresponding Fourier transforms. The transforms are plotted on a linear scale over the range from 0 to π.

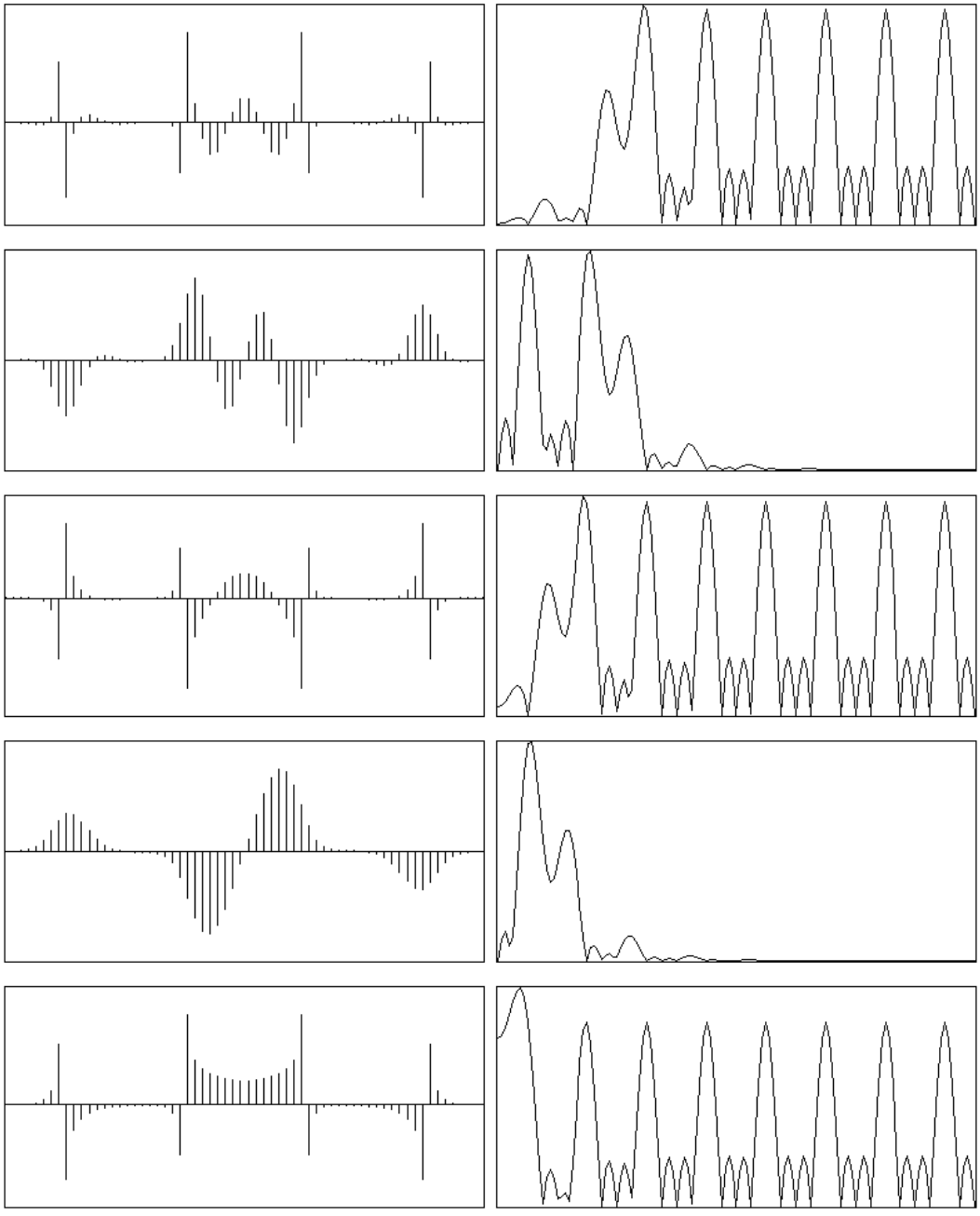


Figure 4.6: The inverse (sampling) functions of the Gabor filter set given in figure 4.5

The DCT and LOT Transforms

The use of the discrete cosine transform (DCT) in image coding systems is often justified with the statement that it approximates the optimal transform for a signal with first-order Gauss-Markov statistics [19]. In practice, the transform is usually not computed globally, but is applied independently to non-overlapping sub-blocks of the image. As illustrated in figure 4.1, the resulting block DCT basis functions constitute a subband transform, but the subbands are not very well localized. Considered in the framework of the A/S system, the subsampled subband images will contain severe amounts of aliasing. Since the transform is invertible (in fact, orthogonal), it should be clear that this subband aliasing is cancelled in the synthesis stage. However, if the transform coefficients are quantized or discarded (e.g. in a coding system), the aliasing no longer cancels, and the errors appear as block edge artifacts in the reconstructed image.

Recent work by Cassereau et. al. [20] describes an elegant technique for reducing the aliasing of the block DCT. They perform an orthogonal transformation on the block DCT coefficients which combines coefficients computed from adjacent blocks. In the resulting transform, which they have called a *Lapped Orthogonal Transform* (LOT), the basis functions from adjacent blocks overlap each other, and their impulse responses are tapered at the edges. Malvar [21] has implemented an efficient version of this transform, known as the fast LOT, in which the additional orthogonal transformation is computed using a butterfly network of simple rotation transformations. Several of the even-symmetric basis functions of the fast LOT are shown in figure 4.7. One limitation which applies to both the DCT and the LOT is that the transforms are limited to equal-sized subbands. As discussed previously, it may be advantageous to subdivide the spectrum into equal log-width subbands.

The Laplacian Pyramid

One of the first techniques for octave subband decomposition was developed by Burt [22] and applied to image coding by Burt and Adelson [23]. They used a pyramid cascade of small Gaussian-like filters to create an overcomplete subband representation which they called a *Laplacian pyramid.* A system for constructing one level of this pyramid (in one dimension) is illustrated in figure 4.8. The signal is blurred with a lowpass filter, $B(\omega)$, and then subsampled to produce a lowpass subband $W_0(\omega)$. A highpass subband, $W_1(\omega)$, is formed by upsampling $W_0(\omega)$, convolving with an interpolation filter $A(\omega)$, and subtracting from the original signal. The signal is reconstructed by upsampling and filtering $W_0(\omega)$ with $A(\omega)$ and adding it to $W_1(\omega)$. This reconstruction is

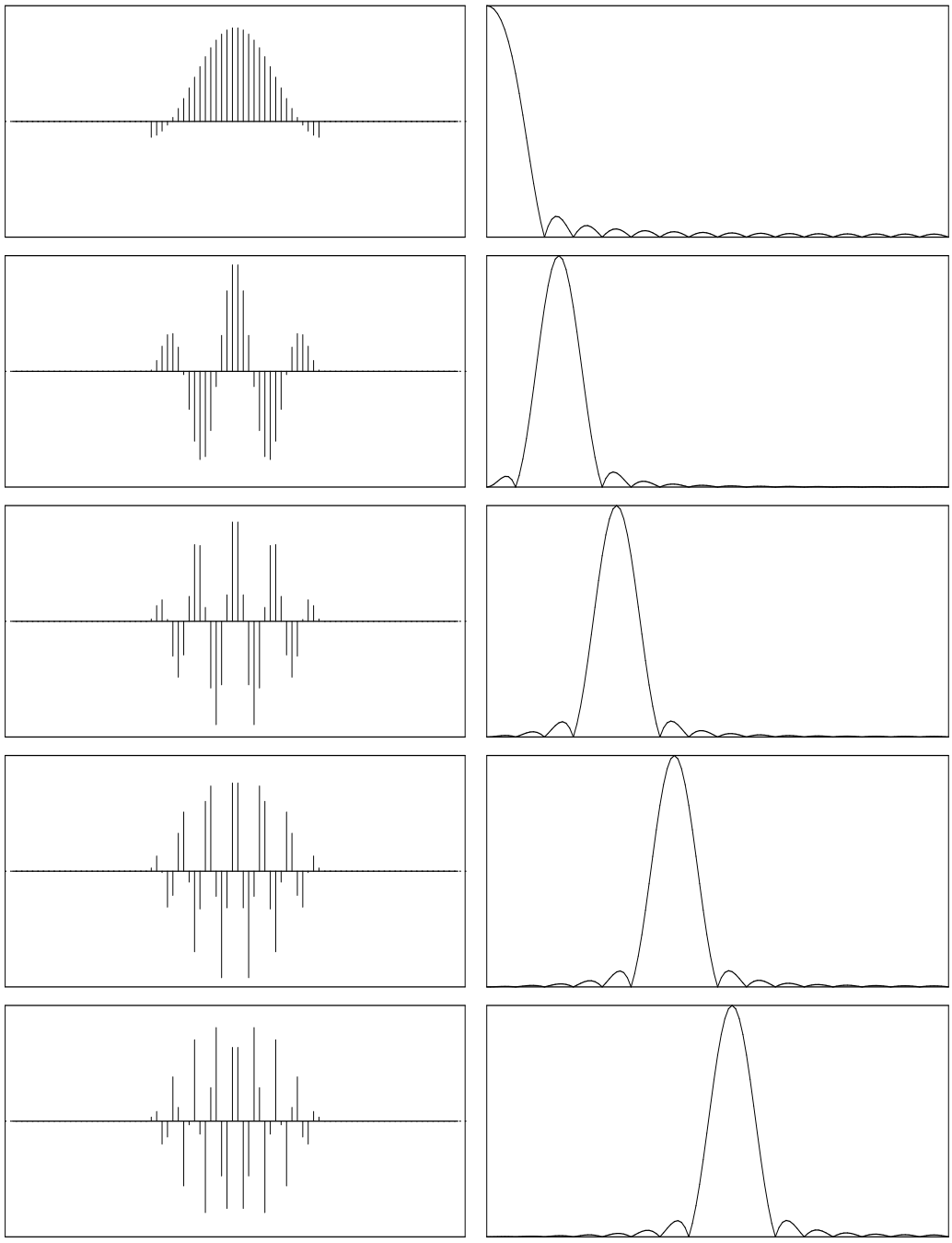


Figure 4.7: Five of the eight even-symmetric basis functions of a LOT. The basis functions are illustrated on the left, and their Fourier transforms on the right. The transforms are plotted on linear axes and cover the range from 0 to π.

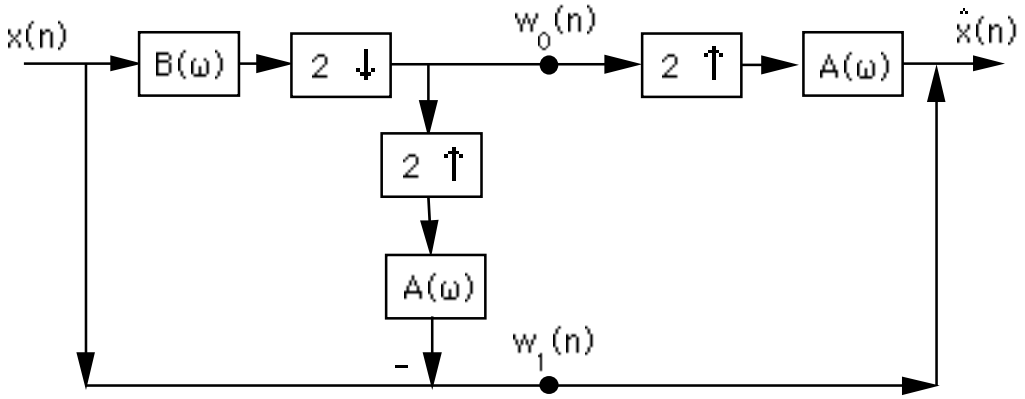


Figure 4.8: Signal processing diagram depicting the standard construction technique for one level of the Laplacian pyramid. A full pyramid is built by non-uniformly cascading this system. This transformation may also be described as an A/S filter bank (see text).

exact, regardless of the choice of the filters $B(\omega)$ and $A(\omega)$. The full pyramid is constructed recursively, by re-applying the system to the lowpass subband. Typically, the filters $A(\omega)$ and $B(\omega)$ are set to some common, compact lowpass filter, although better coding results are obtained by choosing the two filters independently. Some example basis and sampling functions (with $A(\omega) = B(\omega)$) are plotted in figures 4.9 and 4.10, respectively.

In addition to its suitability for data compression, the multi-scale nature of the pyramid makes it particularly useful for the task of progressive transmission. Progressive transmission is a process by which an image is sent through a low-capacity channel so that a low resolution or blurred version of the image becomes available quickly, and higher resolution information is added in a gradual manner. In the case of a pyramid, this is easily accomplished by sending the transform coefficients in order from lowest to highest resolution.

For comparison to other subband transforms, we have re-formulated the Laplacian pyramid scheme as a three-band A/S system (see diagram in figure 4.2) by separating $W_1(\omega)$ into two subsignals: $Y_1(\omega)$ contains the even-numbered samples, and $Y_2(\omega)$ contains the odd-numbered samples. The subsampling factors are $k = 2$ for all three A/S branches, thus producing a representation that is overcomplete by a factor of 3/2. The appropriate filters for the A/S system are defined in terms of the original filters $A(\omega)$ and $B(\omega)$ as follows:

$$\begin{aligned}
H_0(\omega) &= B(\omega), & G_0(\omega) &= A(\omega) \\
H_1(\omega) &= \tfrac{1}{2}\Big[1 - B(\omega)A(\omega) - B(\omega)A(\omega+\pi)\Big], & G_1(\omega) &= 1 \\
H_2(\omega) &= \tfrac{e^{j\omega}}{2}\Big[1 - B(\omega)A(\omega) + B(\omega)A(\omega+\pi)\Big], & G_2(\omega) &= e^{-j\omega}
\end{aligned}$$

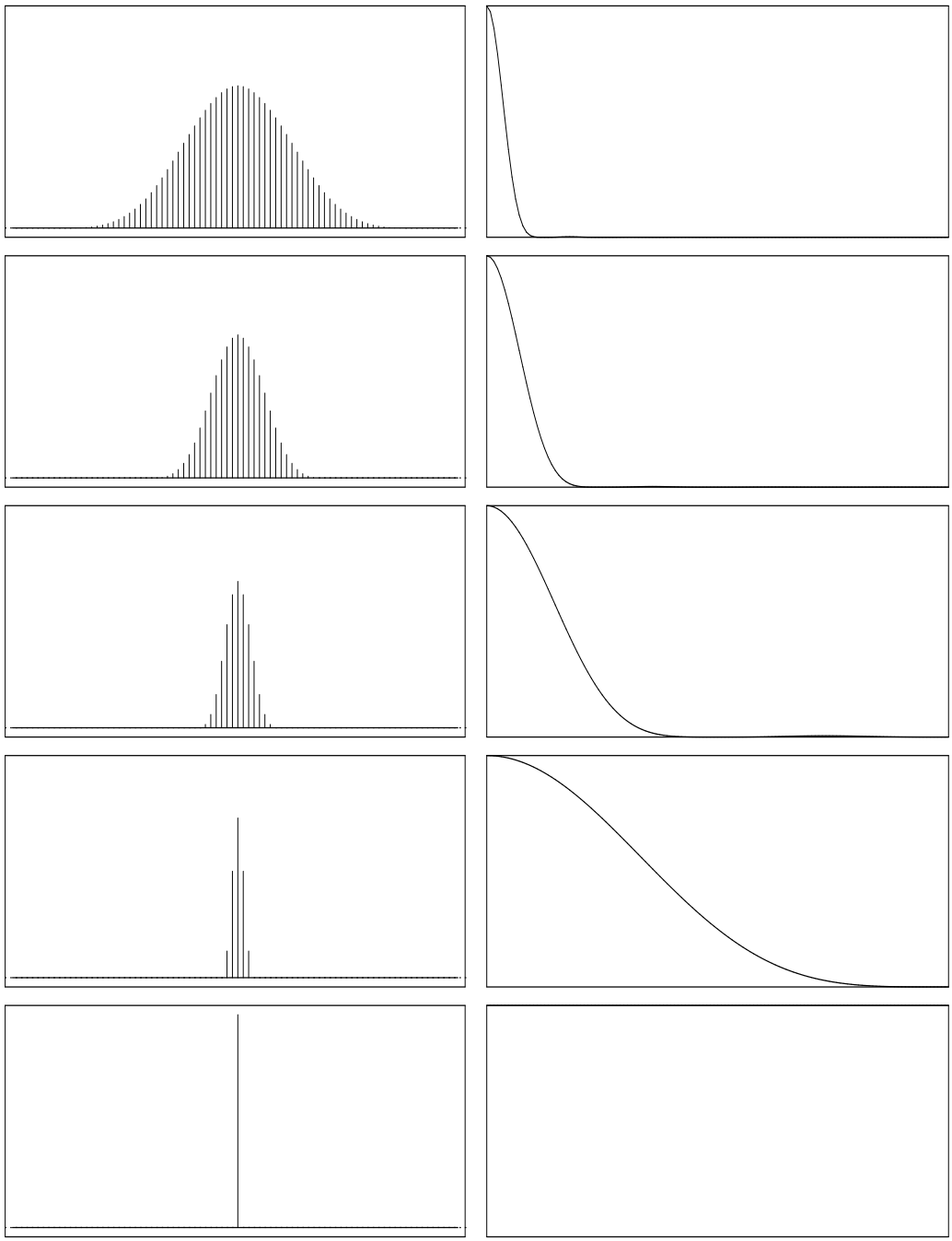


Figure 4.9: Five example basis functions of a four level Laplacian pyramid.

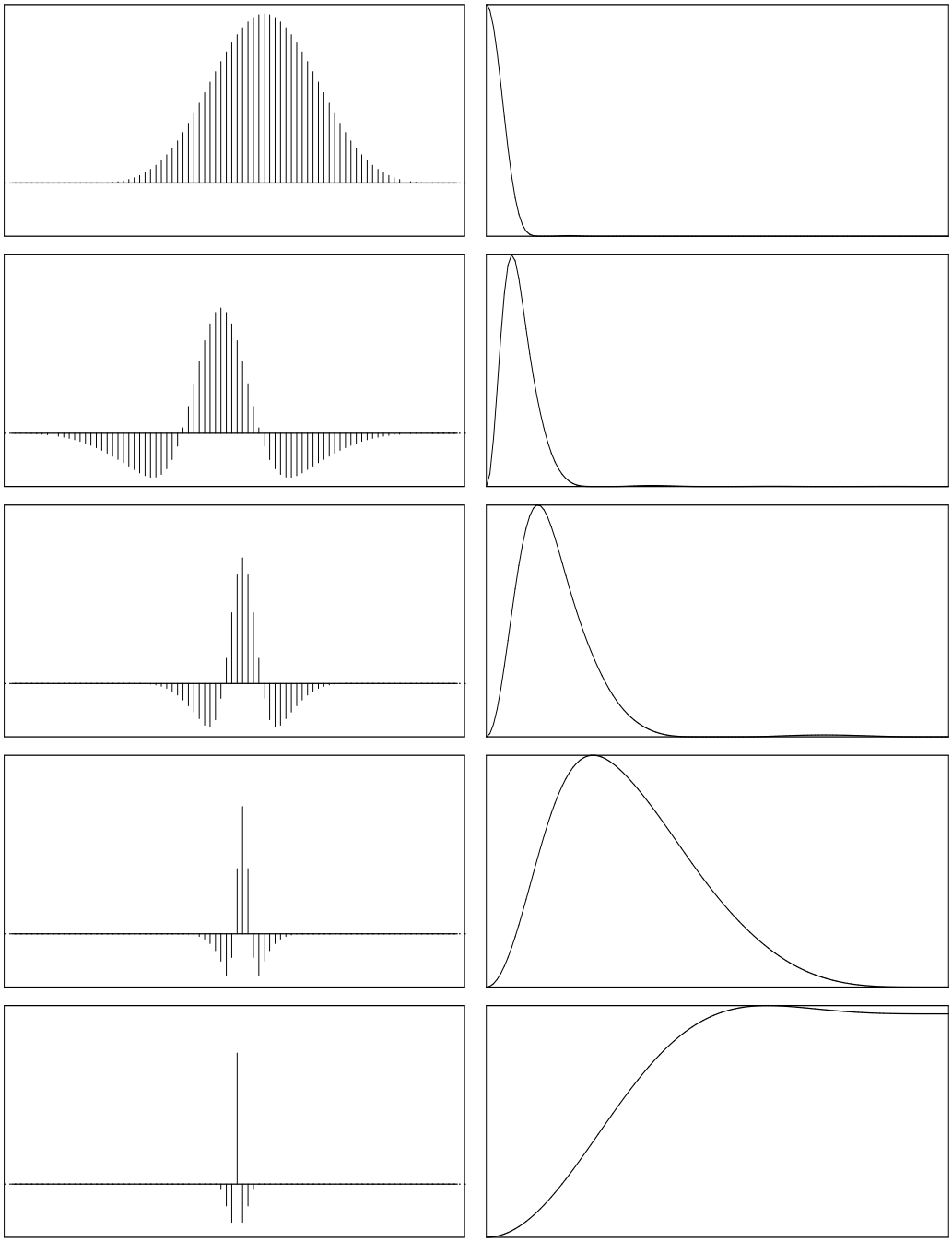


Figure 4.10: Five example inverse (sampling) functions of the Laplacian pyramid.

Notice that when $A(\omega)$ and $B(\omega)$ are lowpass filters, the sampling functions are bandpass and the basis functions are broadband. Since the resulting A/S system violates the constraint in equation 4.10, the transform is clearly not orthogonal. In two dimensions, Burt and Adelson constructed Laplacian pyramids using the same algorithm, but with a separable two-dimensional blurring filter. The two dimensional Laplacian pyramid may be re-formulated as a 5-band A/S system with each band subsampled by a factor of two both horizontally and vertically.

The Laplacian pyramid has certain disadvantages for image coding. The most serious of these is the fact that quantization errors from highpass subbands do not remain in these subbands. Instead, they appear in the reconstructed image as broadband noise. As with the Gabor transform, the non-orthogonality of the transform is the source of the difficulty. Furthermore, the basis set is overcomplete, requiring an increase (in two dimensions) by a factor of $\frac{4}{3}$ in the number of sample points over the original image. Finally, the two-dimensional basis functions are not oriented, and thus will not extract the oriented structural redundancy typically found in natural images. Despite these disadvantages for still-image coding, the Laplacian pyramid has been effectively used for motion-compensated video coding, where its overcompleteness makes it robust in to motion-compensation errors [24].

4.4 Quadrature Mirror Filters

In the previous section, we described three example transforms, each demonstrating useful properties for coding. Now we consider a transform which captures the advantages of the previous examples, while avoiding the disadvantages.

As was illustrated with the Laplacian pyramid, an octave subband transform may be constructed by cascading a two-band A/S system in a non-uniform manner. A useful two-band subband transform which was developed for speech coding is based on banks of quadrature mirror filters (QMF), developed by Croiser et. al. [25, 26]. They discovered a class of non-ideal FIR bandpass filters that could be used in an A/S system while still avoid aliasing in the overall system output. Although they did not describe them as such, these filters form an *orthogonal* subband transform, as was discussed by Adelson et al. [27] and Mallat [3, 28]. Mallat related QMFs to the mathematical theory of wavelets. Vetterli [29] was the first to suggest the the application of QMFs to two-dimensional images. In this section, we give a brief review of QMFs in one dimension. A more thorough review may be found in [30] or [31].

The original QMF problem was formulated as a two-band critically sampled analysis/synthesis filter bank problem. The overall system response of the filter bank is given by equation (4.2), with subsampling factor on each branch set to $k = 2$:

$$\hat{X}(\omega) = \frac{1}{2}\Big[H_0(\omega)G_0(\omega) + H_1(\omega)G_1(\omega)\Big]X(\omega) + \frac{1}{2}\Big[H_0(\omega+\pi)G_0(\omega) + H_1(\omega+\pi)G_1(\omega)\Big]X(\omega+\pi). \quad (4.11)$$

The first term is a linear shift-invariant (LSI) system response, and the second is the system aliasing.

The term QMF refers to a particular choice of filters that are related by spatial shifting and frequency modulation. We define

$$\begin{aligned} H_0(\omega) &= G_0(-\omega) = F(\omega) \\ H_1(\omega) &= G_1(-\omega) = e^{j\omega}F(-\omega+\pi) \end{aligned} \quad (4.12)$$

for $F(\omega)$ an arbitrary function of ω. This is a more general definition than that originally provided by Croisier et. al., and makes explicit the orthogonality of the transform (see section 4.2.5). In particular, the analysis and synthesis filters satisfy the relationship in equation (4.10), and the relationship between the filters in the two branches (i.e. H_0 and H_1) ensures that the corresponding basis functions are orthogonal.

With the choice of filters given in (4.12), equation (4.11) becomes

$$\hat{X}(\omega) = \frac{1}{2}\Big[H(\omega)H(-\omega) + H(-\omega+\pi)H(\omega+\pi)\Big]X(\omega) + \frac{1}{2}\Big[H(\omega+\pi)H(-\omega) + e^{j\pi}H(-\omega)H(\omega+\pi)\Big]X(\omega+\pi).$$

The second (aliasing) term cancels, and the remaining LSI system response is

$$\hat{X}(\omega) = \frac{1}{2}\Big[H(\omega)H(-\omega) + H(-\omega+\pi)H(\omega+\pi)\Big]X(\omega). \quad (4.13)$$

Note that the aliasing cancellation is exact, independent of the choice of the function $F(\omega)$. We should emphasize, however, that it is the overall system aliasing that cancels — the individual subbands *do* contain aliasing.

4.4.1 QMF Design

The design problem is now reduced to finding a lowpass filter with Fourier transform $H(\omega)$ that satisfies the constraint

$$\frac{1}{2}\Big[H(\omega)H(-\omega) + H(-\omega+\pi)H(\omega+\pi)\Big] = 1$$

or

$$\left|H(\omega)\right|^2 + \left|H(\omega+\pi)\right|^2 = 2. \tag{4.14}$$

Several authors have studied the design and implementation of these filters [32, 33, 34, 35, 36]. Johnston [32] designed a set of widely used even-length filters by minimizing an error function containing a shift-invariant error term and a weighted stopband ripple term for a fixed number of filter taps. Jain and Crochiere [33, 34] used a similar error criterion in the time domain, and formulated an iterative design scheme which in which each iteration required the constrained minimization of a quadratic function.

A technique for design of perfect reconstruction filter sets is given by Smith and Barnwell in [37]. They first design a lowpass *product filter* $F(\omega)$ which is factorable as

$$F(\omega) = \left|H(\omega)\right|^2$$

and which satisfies

$$f(n) \cdot \frac{\left(1+(-1)^n\right)}{2} = \delta(n)$$

The resulting $F(\omega)$ is factored to get $h(n)$, the lowpass filter. Wackersreuther [35] independently arrived at a similar design method in the time domain. The problem with these design methods is the somewhat arbitrary choice of the product filter.

Simoncelli [36] proposed an exploratory design method utilizing an iterative matrix averaging technique, and designed a set of odd-length filters using a frequency-sampling method with error criteria similar to Johnston. The design constraints for QMFs do not necessitate sharp transitions and thus frequency-sampling designs perform quite well. Furthermore, it was found that odd-length filters could be made smaller for a given transition band width. The basis functions for a four-level QMF pyramid based on a 9-tap kernel are shown in figure 4.11. A set of example QMF kernels and a more detailed description of this design technique are given in the appendix to this chapter.

QMFs are typically applied to images in a separable manner. In order to compute a multi-scale pyramid, the transform is applied recursively to the lowpass subimage. Such a cascaded transformation partitions the frequency domain into octave-spaced oriented subbands, as illustrated in the idealized frequency diagram of figure 4.12. Thus, the QMF pyramid satisfies the properties described in the introduction to this chapter: it is multi-scale and oriented, it is spatially localized, and it is an orthogonal transformation, and so constrains quantization errors to remain within subbands. One unfortunate aspect of the transform is that the orientation decomposition is incomplete.

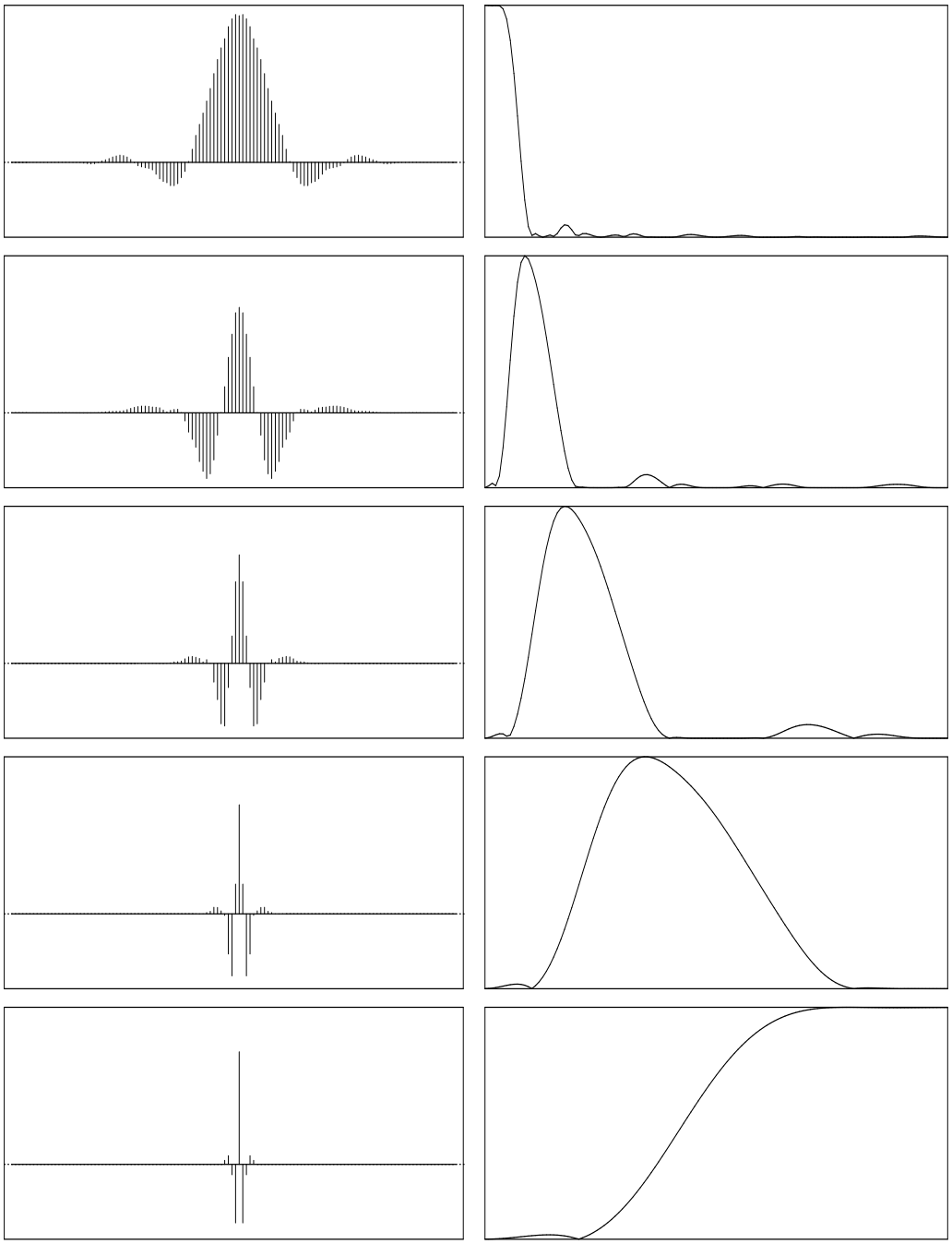


Figure 4.11: Five of basis functions of a 9-tap QMF pyramid transform.

The two diagonal orientations are lumped together in a single subband. We will address this problem in section 4.5.

4.4.2 An Asymmetrical System

Thus far, we have ignored the issue of computational efficiency. For many applications, this is relatively unimportant due to the steady increase in the speed of signal processing hardware. There are, however, situations where it is desirable to quickly decode or encode a coded image using conventional or general-purpose hardware. For example, an image data base that will be accessed by millions of users with personal computers should be quickly decodable on standard hardware; the cost of encoding these images is of relatively minor importance as long as the decoding is simple. At the other extreme, a remotely piloted vehicle demands a very simple encoding scheme in order to minimize weight and power requirements.

For these situations, it is advantageous to develop asymmetric coding techniques in which simplicity is emphasized at one end at the expense of complexity at the other end. For a QMF transform, the computational complexity is directly proportional to the size of the filters employed. Thus, we wish to relax the orthogonality constraint which forces the synthesis filters to be time-reversed copies of the analysis filters. Consider the situation in which we require efficient decoding. The increase in efficiency can be accomplished by using a very compact filter pair in the synthesis stage of an A/S system [27, 38]. In particular, one can choose the 3-tap lowpass filter $g_0(n) = [1, 2, 1]$, with a highpass counterpart $g_1(n) = [-1, 2, -1]$. Convolutions with these filters may be performed using only arithmetic shifting and addition operations.

The relationship $G_1(\omega) = e^{j\omega}G_0(-\omega + \pi)$, as in equation (4.12) ensures that the linear subspaces spanned by the basis functions corresponding to each filter will be orthogonal. Conceptually, a set of inverse filters $h_i(n)$ is found by forming a square matrix of the $g_i(n)$ as in equation (4.5) and inverting it. The size of the matrix determines the size of the resulting inverse filters. In practice, a better design technique is to minimize an error function for a given kernel size. We have designed a set of inverse filters (given in the appendex) by minimizing the maximal reconstruction error for a step edge input signal. These kernels are given in the appendix.

Another highly efficient A/S system was proposed by LeGall. He derived the following set of simple filters for use in an A/S filter bank:

$$\begin{aligned} H_0(\omega) &= A(\omega), & G_0(\omega) &= B(\omega) \\ H_1(\omega) &= e^{j\omega}B(\omega + \pi), & G_1(\omega) &= e^{-j\omega}A(\omega + \pi) \end{aligned}$$

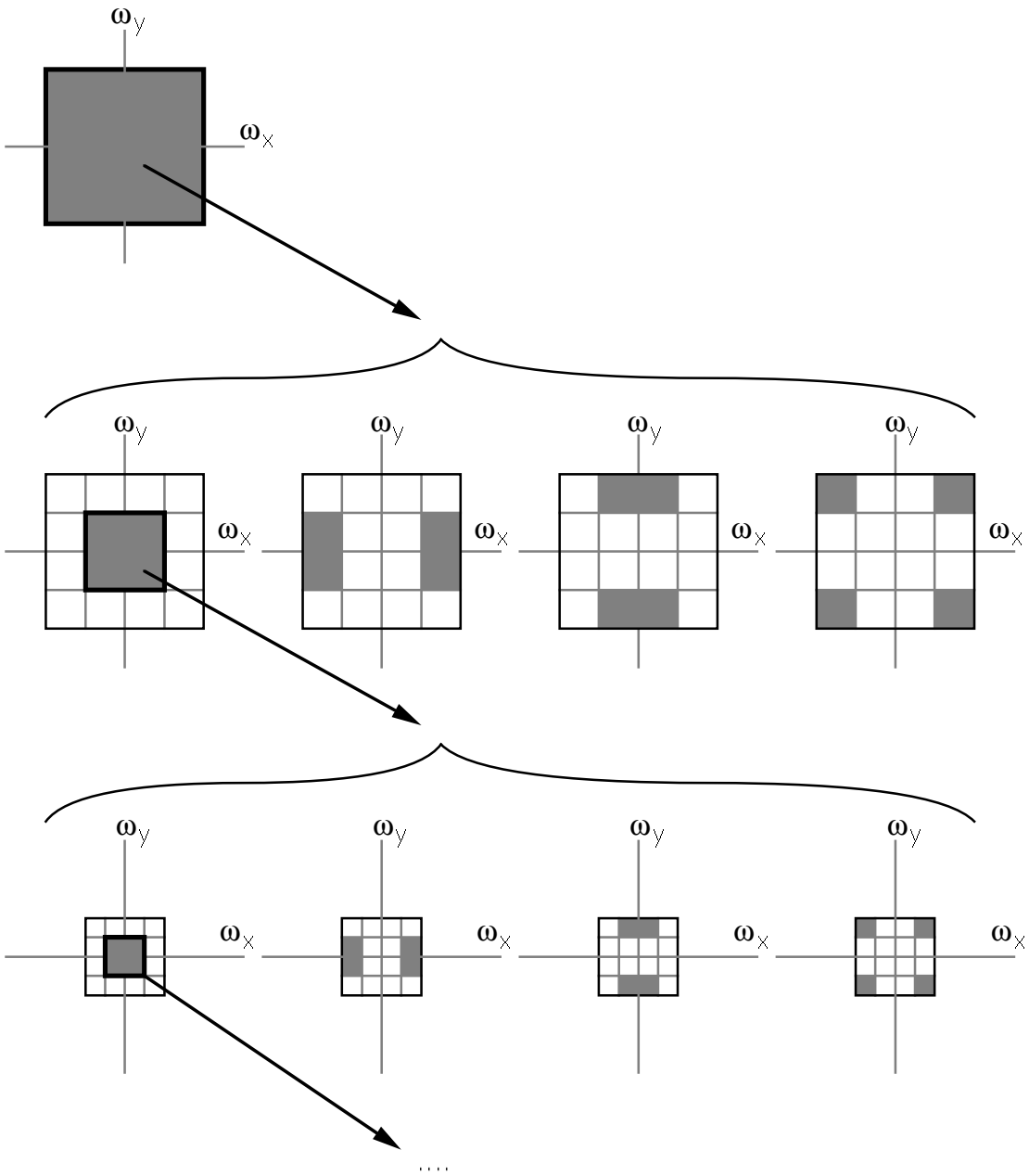


Figure 4.12: Idealized diagram of the partition of the frequency plane resulting from a 4-level pyramid cascade of separable 2-band filters. The top plot represents the frequency spectrum of the original image, with axes ranging from $-\pi$ to π. This is divided into four subbands at the next level. On each subsequent level, the lowpass subband (outlined in bold) is subdivided further.

where the filter kernels (impulse responses) corresponding to $A(\omega)$ and $B(\omega)$ are

$$\begin{aligned} a(n) &= [1, 2, 1] \\ b(n) &= [-1, 2, 6, 2, -1] \end{aligned}$$

Note that the definitions of $a(n)$ and $b(n)$ may be interchanged. These filters allow efficient encoding *and* decoding, and provide exact reconstruction (with perfect aliasing cancellation).

4.5 Non-separable QMF Transforms

In the previous section, we described the separable QMF pyramid transform. Most two-dimensional work with QMFs has employed separable filters or non-oriented non-separable filters [29]. As discussed in the previous section, separable application of one-dimensional QMFs produces a representation in which one of the subbands contains a mixture of two orientations. This problem is inherent in the rectangular sampling scheme. Rectangular sampling of a signal in the spatial domain corresponds to summing aliased or modulated copies of the spectrum in the frequency domain. Thus, the frequency response of any rectangularly sampled function has the same value at the points (π, π), $(-\pi, \pi)$, $(\pi, -\pi)$, and $(-\pi, -\pi)$ (i.e. this point corresponds to two opposing orientations). Splitting the frequencies in the neighborhood of this point into different orientation bands requires the use of very large filters. In general, the high-frequency diagonal regions of the spectra of natural images are relatively insignificant. But if the filter bank is cascaded to form a pyramid, then the lower frequency diagonals (where there *is* significant power) will also be mixed.

4.5.1 Hexagonal Systems

In this section, we will discuss the use of hexagonal sampling systems and filters. We will show that the mixed orientation problem discussed above can be avoided by using hexagonally symmetric filters. This non-separable extension of the QMF concept was first described by Adelson et al. [27] and improved and generalized in later work [36, 39]. Other authors have also explored the use of hexagonal sampling systems for image representation. Crettez and Simon [40] and Watson [41] describe decompositions on hexagonal lattices using non-overlapping basis functions. The blocked nature of these functions suggests that they are unlikely to offer efficient image compression.

Figure 4.13 shows a hexagonal sampling lattice and its Fourier transform. The sampling lattice is defined by a pair of *sampling vectors* in the plane:

$$\boldsymbol{v}_0 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}, \qquad \boldsymbol{v}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The locations of the lattice points consist of all linear combinations of these vectors with integer coefficients. In the frequency domain, the effect of this sampling is to convolve the original frequency spectrum of the image with a *modulation* or *reciprocal* lattice which is the Fourier transform of the sampling lattice. The modulation lattice is defined by a pair of *modulation vectors* in the frequency plane:

$$\tilde{\boldsymbol{v}}_0 = \begin{pmatrix} 4\pi/\sqrt{3} \\ 0 \end{pmatrix}, \qquad \tilde{\boldsymbol{v}}_1 = \begin{pmatrix} -2\pi/\sqrt{3} \\ 2\pi \end{pmatrix}.$$

Thus if $F(\omega)$ is the Fourier transform of a hexagonally sampled signal (image) then it is invariant to translations by multiples of the vectors $\tilde{\boldsymbol{v}}_i$:

$$F(\omega) = F(\omega + n_0\tilde{\boldsymbol{v}}_0 + n_1\tilde{\boldsymbol{v}}_1) \tag{4.15}$$

for n_0 and n_1 any two integers.

In general, the relationship between the sampling vectors and modulation vectors is easily described in terms of matrices [42, 43]. If we consider the *sampling matrix* $\mathbf{V}$ with columns containing the vectors $\boldsymbol{v}_i$ and the *modulation matrix* $\tilde{\mathbf{V}}$ with columns containing the vectors $\tilde{\boldsymbol{v}}_i$, then the two matrices are related by the equation

$$\tilde{\mathbf{V}} = 2\pi(\mathbf{V}^{-1})^t. \tag{4.16}$$

Note that we know $\mathbf{V}$ is invertible since we assume that the sampling vectors span the space (i.e. they are linearly independent).

As stated in section 4.2.1, the A/S system depicted in figure 4.2 is valid for two-dimensional signals, but the filtering and subsampling is done in two dimensions: $\boldsymbol{\omega}$ is now a two-dimensional vector, and the subsampling is parameterized by a non-singular two-by-two *subsampling matrix*, $\mathbf{K}$, with integer entries. Figure 4.14 illustrates two-dimensional subsampling in both the spatial and frequency domains.

In order to write a general expression for the output of a multi-dimensional analysis/synthesis system, we need a frequency-domain equation analogous to that given in (4.1) relating the subsampled signal to the sampled signal. We also need an equation relating an upsampled signal to the original sampled signal. For rectangular sampling lattices in d dimensions, the relationship is simple. The sampling matrix $\mathbf{K}$ generates a sublattice defined by

$$\{\boldsymbol{n} : \boldsymbol{n} = \mathbf{K}\boldsymbol{m}, \boldsymbol{m} \in \mathbf{Z}^d\},$$

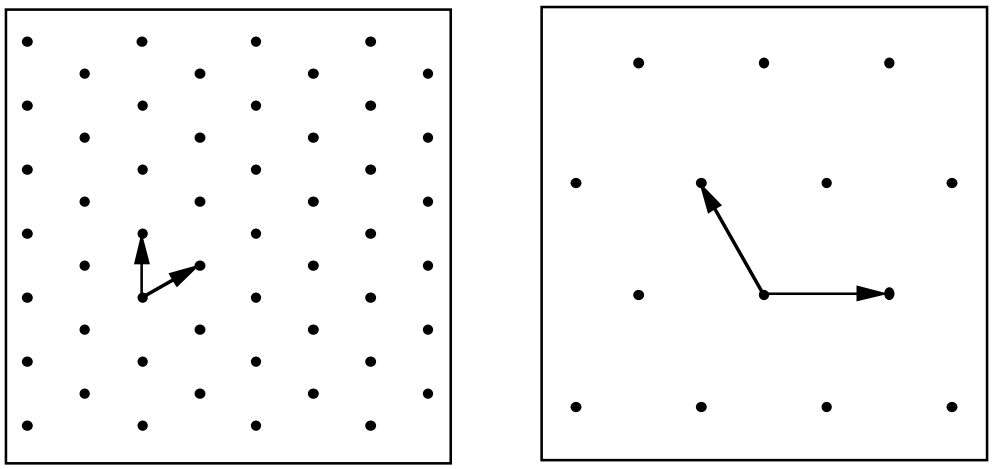


Figure 4.13: Relationship between hex sampling lattices in the spatial and spatial-frequency domains. On the left is the lattice defined by the sampling vectors. On the right is the Fourier transform of this lattice, defined by the modulation vectors.

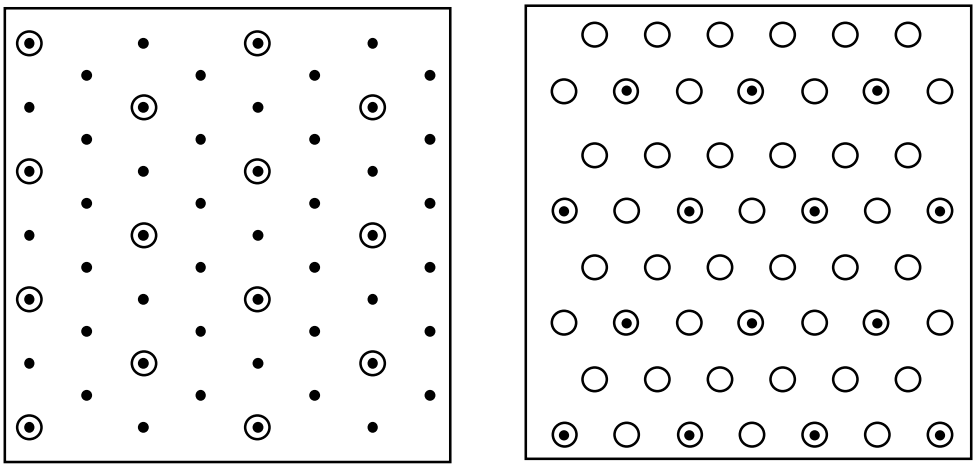


Figure 4.14: Illustration of subsampling on a hexagonal lattice. The points in the diagram on the left represent the original sampling lattice and the circles represent the subsampled lattice points. The picture on the right shows the Fourier transform of the lattice (points) and the Fourier transform of the subsampled lattice (circles).

where $\mathbf{Z}^d$ is the set of all d-dimensional vectors with integer components. The sublattice has $|\mathbf{K}|$ distinct cosets, each coset being a copy of the sublattice translated by an integer vector, and the union of the cosets is the original sampling lattice [44]. Consider two signals related by subsampling: $s(\boldsymbol{n}) = r(\mathbf{K}\boldsymbol{n})$. Then their Fourier transforms are related by the expression

$$S(\boldsymbol{\omega}) = \frac{1}{|\mathbf{K}|} \sum_{i=0}^{|\mathbf{K}|-1} R\left((\mathbf{K}^{-1})^t(\boldsymbol{\omega} - 2\pi \boldsymbol{k}_i)\right)$$

where $S(\boldsymbol{\omega})$ and $R(\boldsymbol{\omega})$ are the Fourier transforms of $s(\boldsymbol{n})$ and $r(\boldsymbol{n})$ respectively, and the $\boldsymbol{k}_i$ are a set of polyphase shift vectors corresponding to each of the $|\mathbf{K}|$ sublattice cosets [44]. A simple example of a set of shift vectors is the following:

$$\left\{\boldsymbol{k} : (\mathbf{K}^{-1})^t \boldsymbol{k} \in [0,1)^d, \boldsymbol{k} \in \mathbf{Z}^d\right\},$$

where $[0,1)^d$ is the half-open unit interval in d dimensions.

The corresponding expression for non-rectangular sampling lattices is obtained by mapping from the rectangular case. The result of subsampling in the analysis/synthesis system may then be written as a convolution of the sampled spectrum with a set of *subsampling modulation vectors* $\tilde{\boldsymbol{\kappa}}_j$:

$$Y_i(\boldsymbol{\omega}) = \frac{1}{|\mathbf{K}|} \sum_{j=0}^{|\mathbf{K}|-1} H_i\left((\mathbf{K}^{-1})^t \boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j\right) X\left((\mathbf{K}^{-1})^t \boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j\right), \tag{4.17}$$

where the $\tilde{\boldsymbol{\kappa}}_j$ are defined as

$$\begin{aligned} &\left\{\tilde{\boldsymbol{\kappa}}_j : \; j = 0, 1, \ldots, |\mathbf{K}| - 1\right\} \\ &\qquad = \; \left\{\tilde{\mathbf{V}}(\mathbf{K}^{-1})^t \boldsymbol{n} : \; (\mathbf{K}^{-1})^t \boldsymbol{n} \in [0,1)^d, \boldsymbol{n} \in \mathbf{Z}^d\right\}. \end{aligned} \tag{4.18}$$

The effect of upsampling in the frequency domain is the same as for the rectangular case [43]. Combining equation (4.17) with the frequency domain upsampling relationship gives an expression for the overall filter bank response:

$$\begin{aligned} \hat{X}(\boldsymbol{\omega}) &= \frac{1}{|\mathbf{K}|} \sum_{i=0}^{|\mathbf{K}|-1} G_i(\boldsymbol{\omega}) Y_i(\mathbf{K}^t \boldsymbol{\omega}) \\ &= \frac{1}{|\mathbf{K}|} \sum_{i=0}^{|\mathbf{K}|-1} G_i(\boldsymbol{\omega}) \sum_{j=0}^{|\mathbf{K}|-1} H_i(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j) X(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j) \\ &= \frac{1}{|\mathbf{K}|} \sum_{j=0}^{|\mathbf{K}|-1} X(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j) \left[\sum_{i=0}^{|\mathbf{K}|-1} G_i(\boldsymbol{\omega}) H_i(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j)\right]. \end{aligned} \tag{4.19}$$

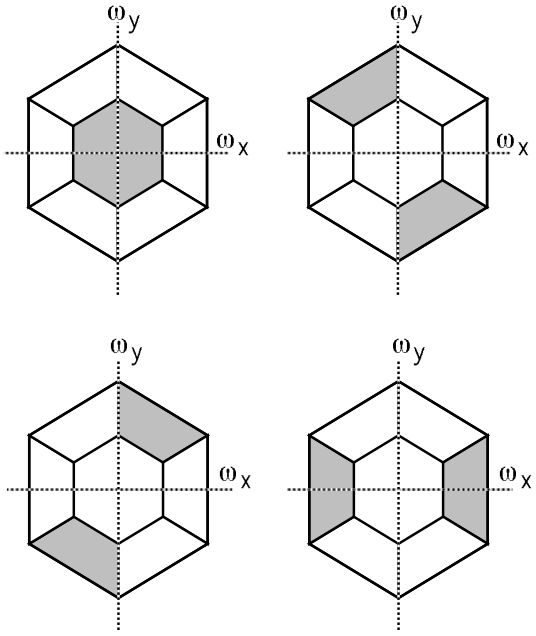


Figure 4.15: Illustration of the modulating effect of subsampling in the frequency domain. Assume that the sampled image has a spectrum bandlimited to the gray region in the upper left frequency diagram. Subsampling will modulate the spectrum to the gray regions in the other three diagrams. The resulting spectrum will be the sum of the four spectra.

As in equation (4.11), the first term of the sum ($j = 0$) corresponds to the LSI system response, and the remaining terms are the system aliasing.

Returning now to the specific case of the hexagonal sampling lattice, we describe a system obtained by using a specific sampling matrix $\mathbf{K}$. Since we want to be able to apply the transform recursively, we choose a subsampling scheme which preserves the geometry of the original sampling lattice:

$$\mathbf{K} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

On the hexagonal sampling lattice with this subsampling scheme, the definition given in (4.18) produces the following modulation vectors:

$$\tilde{\boldsymbol{\kappa}}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_1 = \begin{pmatrix} 2\pi/\sqrt{3} \\ 0 \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_2 = \begin{pmatrix} \pi/\sqrt{3} \\ \pi \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_3 = \begin{pmatrix} -\pi/\sqrt{3} \\ \pi \end{pmatrix}.$$

Figure 4.15 offers an idealized picture of this modulation.

Analogous to the one-dimensional case, we can choose the filters to eliminate the aliasing terms in equation (4.19):

$$H_0(\boldsymbol{\omega}) \;=\; G_0(-\boldsymbol{\omega}) = F(\boldsymbol{\omega}) = F(-\boldsymbol{\omega})$$

$$\begin{aligned} H_1(\boldsymbol{\omega}) &= G_1(-\boldsymbol{\omega}) = e^{j\boldsymbol{\omega}\cdot\boldsymbol{s}_1} F(\boldsymbol{\omega}+\tilde{\boldsymbol{\kappa}}_1) \\ H_2(\boldsymbol{\omega}) &= G_2(-\boldsymbol{\omega}) = e^{j\boldsymbol{\omega}\cdot\boldsymbol{s}_2} F(\boldsymbol{\omega}+\tilde{\boldsymbol{\kappa}}_2) \\ H_3(\boldsymbol{\omega}) &= G_3(-\boldsymbol{\omega}) = e^{j\boldsymbol{\omega}\cdot\boldsymbol{s}_3} F(\boldsymbol{\omega}+\tilde{\boldsymbol{\kappa}}_3) \end{aligned} \tag{4.20}$$

where H is a function that is invariant under negation of its argument, the $\boldsymbol{s}_i$ are a set of spatial shift vectors (defined in the next paragraph), and the expressions $\boldsymbol{\omega}\cdot\boldsymbol{s}_i$ indicates an inner product of the two vectors. As in equation (4.12), the filters are related by spatial shifting and frequency modulation. For the subsampling matrix we are using here, there are four sublattice cosets and therefore four distinct shifting vectors (including the zero vector). Two assignments of the $\boldsymbol{s}_i$ lead to system aliasing cancellation, and these two assignments are related by reflection through the origin. So without loss of generality, we choose the shifting vectors to be

$$\boldsymbol{s}_1 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}, \qquad \boldsymbol{s}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{s}_3 = \begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix}.$$

After cancellation of the aliasing terms in equation (4.19), the remaining LSI system response is

$$\begin{aligned} \hat{X}(\boldsymbol{\omega}) &= \frac{1}{4} X(\boldsymbol{\omega}) \sum_{i=0}^{3} G_i(\boldsymbol{\omega}) H_i(\boldsymbol{\omega}) \\ &= \frac{1}{4} X(\boldsymbol{\omega}) \sum_{i=0}^{3} F(-\boldsymbol{\omega}+\tilde{\boldsymbol{\kappa}}_i) F(\boldsymbol{\omega}+\tilde{\boldsymbol{\kappa}}_i) \\ &= \frac{1}{4} X(\boldsymbol{\omega}) \sum_{i=0}^{3} \left| F(\boldsymbol{\omega}+\tilde{\boldsymbol{\kappa}}_i) \right|^2 . \end{aligned} \tag{4.21}$$

As in one dimension, the aliasing cancellation is exact, independent of the choice of $F(\boldsymbol{\omega})$, and the design problem is reduced to finding a filter with Fourier transform $F(\boldsymbol{\omega})$ satisfying the constraint

$$\sum_{i=0}^{3} \left| F(\boldsymbol{\omega}+\tilde{\boldsymbol{\kappa}}_i) \right|^2 = 4. \tag{4.22}$$

This is analogous to the one-dimensional equation (4.14). Again, a lowpass solution will produce a band-splitting system which may be cascaded hierarchically to produce an octave-bandwidth decomposition in two dimensions. An idealized illustration of this is given in figure 4.16. Finer frequency and orientation subdivisions may be achieved by recursively applying the filter bank to some of the high frequency subbands, as illustrated in figure 4.17.

Filters may be designed using the methods described in section 4.4.1 [39]. Several example filter sets are given in the appendix to this chapter. The

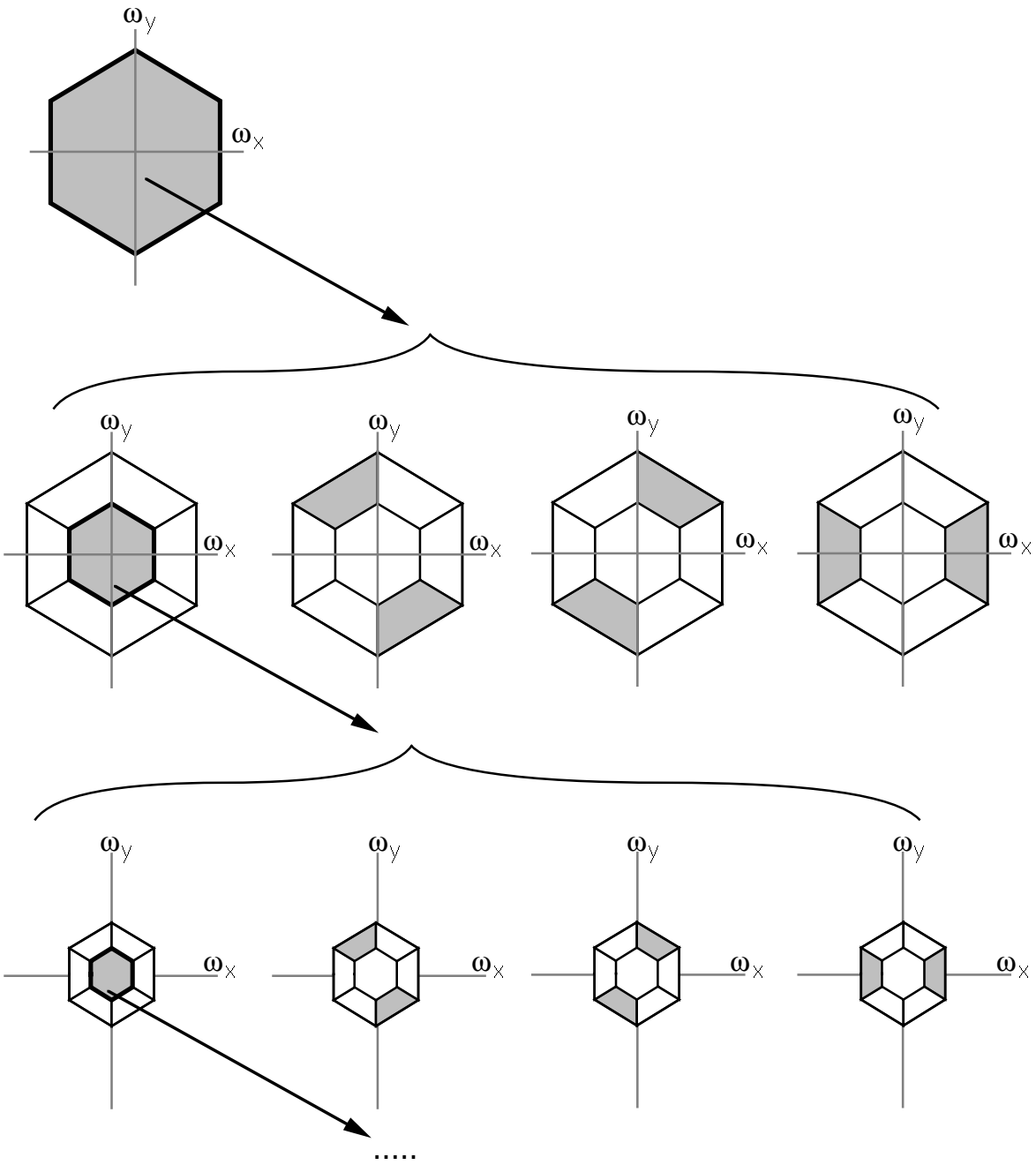


Figure 4.16: Idealized diagram of the partition of the frequency plane resulting from a four-level pyramid cascade of hexagonal filters. The top plot represents the frequency spectrum of the original image. This is divided into four subbands at the next level. On each subsequent level, the lowpass subband (outlined in bold) is sub-divided further.

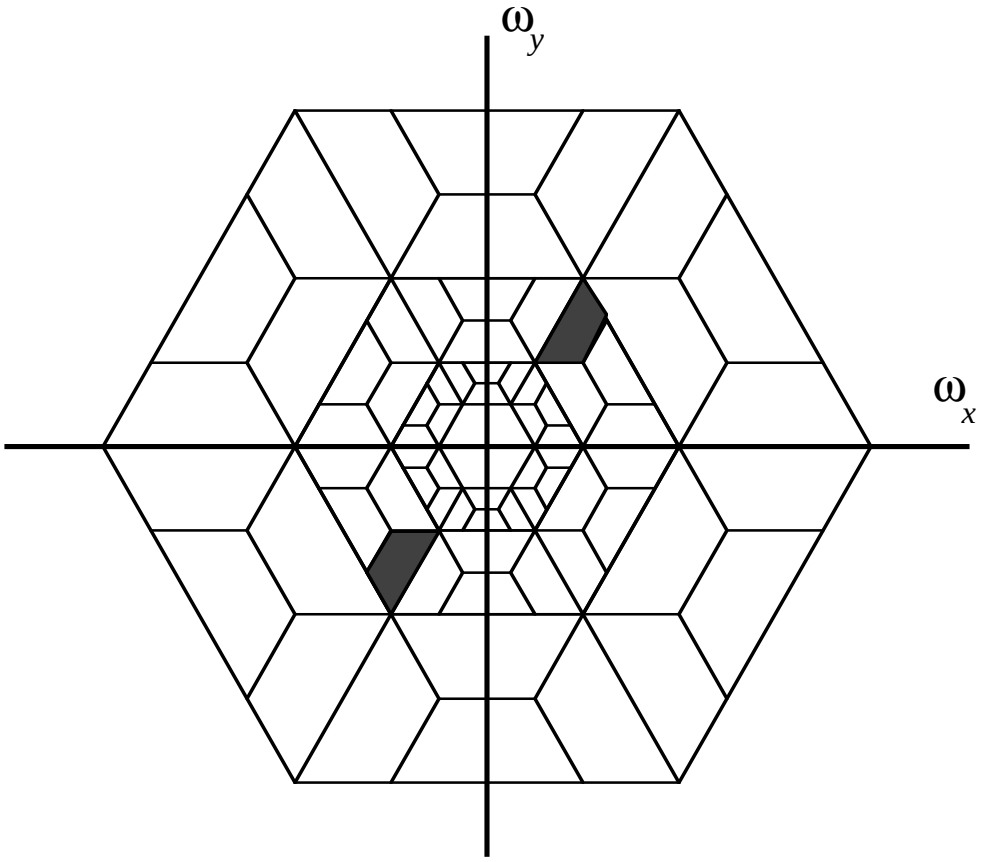


Figure 4.17: An example of a frequency-domain partition which could be computed using a non-pyramid cascade of the hexagonal filter bank transform described in the text. The shaded region indicates the frequency region associated with one of the transform subbands.

power spectra of an example set of filters (the "4-ring" filters) are plotted in figure 4.18. These filters are extremely compact, requiring only nine multiplications per convolution point (assuming one takes advantage of the twelve-fold hexagonal symmetry). Figure 4.19 shows the results of applying this bank of filters recursively to an image of a disk. Examples of images coded using these filters will be given in section 4.6.

4.5.2 Rhombic Dodecahedral Systems

The extension of the concepts developed in the previous section to three-dimensional signal processing is fairly straightforward. Such systems are useful for applications such as compression of medical images or video motion sequences. Analogous to the two-dimensional hexagonal case, one can choose a periodic sampling lattice which corresponds to the densest packing of spheres

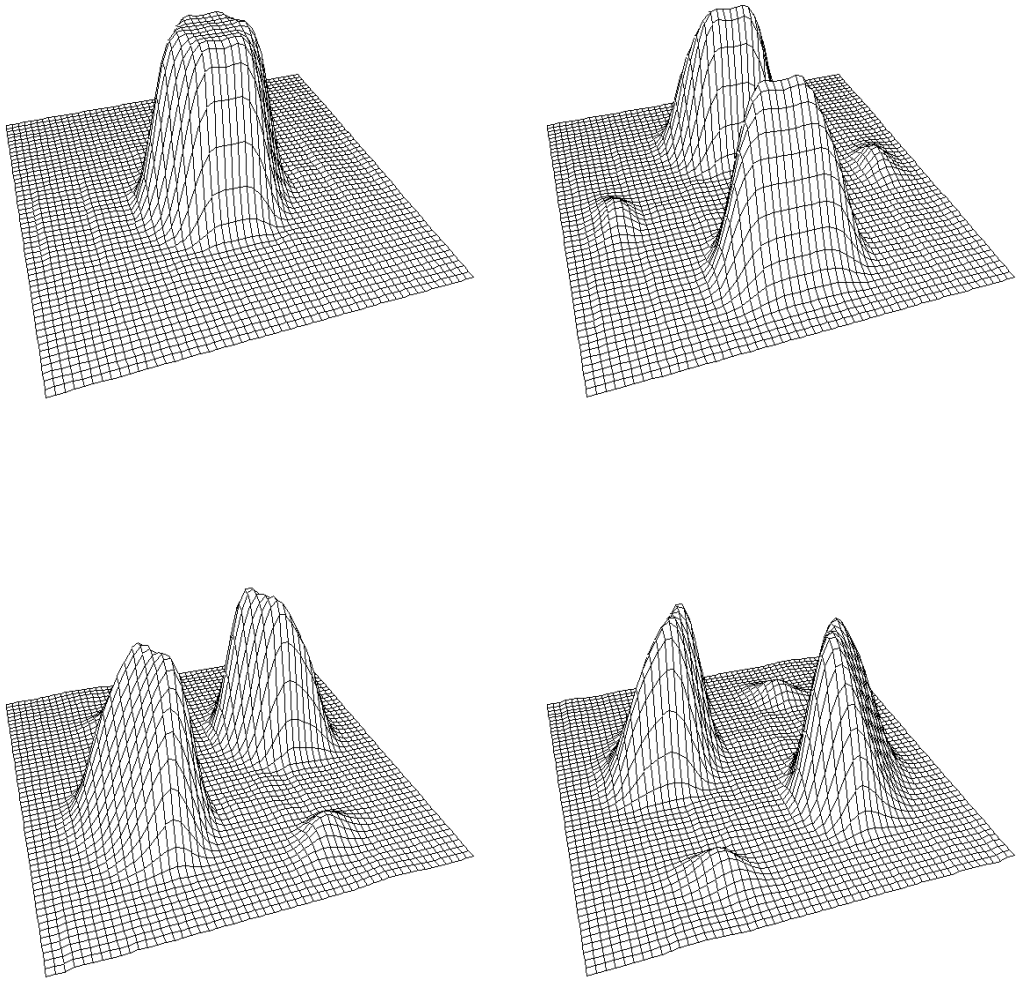
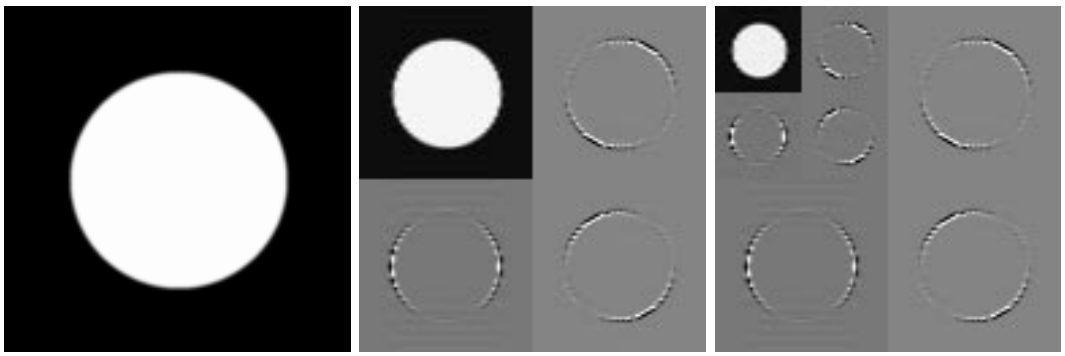


Figure 4.18: The power spectra for the "4-ring" set of hexagonal QMF filters. The filter kernels are given in the appendix.



(a) (b) (c)

Figure 4.19: Results of applying a hexagonal QMF bank to an image of a disk. (a) The original image. (b) The result after one application of the analysis section of the filter bank. The image has been decomposed into a lowpass and three oriented high-pass images at 1/4 density. (c) The result of applying the filter bank recursively to the lowpass image to produce a two-level pyramid decomposition.

in three dimensions. This packing corresponds to the crystal structure of garnet. We choose as a band limiting region the *Voronoi* region of this lattice (a rhombic dodecahedron) which is illustrated in figure 4.20. The sampling matrix for the lattice is

$$\mathbf{V} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}.$$

Using equation (4.16), the modulation matrix is then

$$\tilde{\mathbf{V}} = \begin{bmatrix} \pi & 0 & 0 \\ -\pi & 2\pi & 0 \\ -\sqrt{2}\pi & 0 & 2\sqrt{2}\pi \end{bmatrix}.$$

To preserve the geometry of the original sampling lattice, we choose an eight-band A/S system with subsampling matrix

$$\mathbf{K} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

This produces the following subsampling modulation points, as determined by equation (4.18):

$$\tilde{\boldsymbol{\kappa}}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_1 = \begin{pmatrix} 0 \\ \pi \\ \sqrt{2}\pi \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_2 = \begin{pmatrix} 0 \\ \pi \\ 0 \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_3 = \begin{pmatrix} \pi/2 \\ \pi/2 \\ -\pi/\sqrt{2} \end{pmatrix},$$

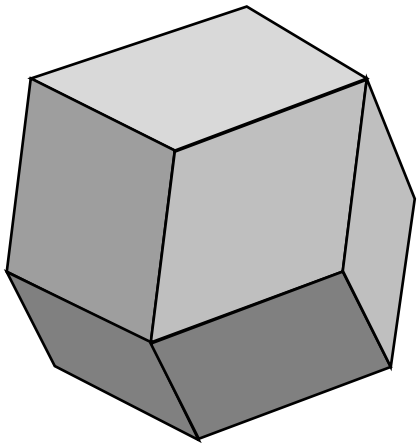


Figure 4.20: A rhombic dodecahedron. This is the shape of the bandlimiting frequency region for the "garnet" filter.

$$\tilde{\boldsymbol{\kappa}}_4 = \begin{pmatrix} \pi/2 \\ -\pi/2 \\ \pi/\sqrt{2} \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_5 = \begin{pmatrix} \pi/2 \\ -\pi/2 \\ -\pi/\sqrt{2} \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_6 = \begin{pmatrix} \pi/2 \\ \pi/2 \\ \pi/\sqrt{2} \end{pmatrix}, \quad \tilde{\boldsymbol{\kappa}}_7 = \begin{pmatrix} 0 \\ 0 \\ \sqrt{2}\pi \end{pmatrix}.$$

When applied to video motion sequences, these modulation vectors correspond to a decomposition into the following subbands: lowpass, stationary vertical, stationary horizontal, motion up/right, motion up/left, motion down/right, motion down/left, and combined stationary diagonals and full-field flicker. Unfortunately, there seems to be no way to avoid the last filter which contains mixed orientations. The overall system response of the filter bank is

$$\hat{X}(\boldsymbol{\omega}) = \frac{1}{8} \sum_{j=0}^{7} X(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j) \left[\sum_{i=0}^{7} G_i(\boldsymbol{\omega}) H_i(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_j) \right]. \tag{4.23}$$

where the first term is the LSI system response, and the remaining terms are aliasing terms.

Once again, we can choose filters related by shifts and modulations that will cancel the system aliasing terms:

$$\begin{aligned} H_0(\boldsymbol{\omega}) &= G_0(-\boldsymbol{\omega}) = F(\boldsymbol{\omega}) = F(-\boldsymbol{\omega}) \\ H_i(\boldsymbol{\omega}) &= G_i(-\boldsymbol{\omega}) = e^{j\boldsymbol{\omega}\cdot\boldsymbol{s}_i} F(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_i), \quad i \in \{1, 2, ...7\} \end{aligned}$$

where the shift vectors $\boldsymbol{s}_i$ are defined as

$$\boldsymbol{s}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{s}_1 = \begin{pmatrix} 1 \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, \quad \boldsymbol{s}_2 = \begin{pmatrix} 1 \\ 1 \\ \sqrt{2} \end{pmatrix}, \quad \boldsymbol{s}_3 = \begin{pmatrix} 1 \\ 2 \\ 1/\sqrt{2} \end{pmatrix},$$

$$\boldsymbol{s}_4 = \begin{pmatrix} 2 \\ 1 \\ 1/\sqrt{2} \end{pmatrix}, \quad \boldsymbol{s}_5 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{s}_6 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{s}_7 = \begin{pmatrix} 0 \\ 1 \\ 1\sqrt{2} \end{pmatrix}.$$

Note that as in the hexagonal case, the choice of shift vectors is not unique.

With the choice of filters given above, the aliasing terms in equation (4.23) cancel and the remaining LSI system response is

$$\hat{X}(\boldsymbol{\omega}) = \frac{1}{8} X(\boldsymbol{\omega}) \sum_{i=0}^{7} |F(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_i)|^2 ,$$

independent of the choice of the function $F(\boldsymbol{\omega})$. The design constraint equation is now

$$\sum_{i=0}^{7} |F(\boldsymbol{\omega} + \tilde{\boldsymbol{\kappa}}_i)|^2 = 8.$$

To illustrate the use of the garnet filter, we apply it to an image sequence of a sinusoidal pinwheel rotating in a counterclockwise direction. One frame of the sequence is shown in figure 4.21(a). The squared responses of the four different motion-selective filters (filters $H_3(\boldsymbol{\omega})$ through $H_6(\boldsymbol{\omega})$) are shown in figure 4.21(b-e).

4.6 Image Coding Examples

Several authors have used QMFs for purposes of image coding. Woods and O'Neill [45] were the first to implement an image coding system using QMFs. They constructed a separable sixteen-band decomposition using a uniform cascade of 32- and 80-tap filters designed by Johnston [32], and then coded the bands using adaptive DPCM. Gharavi and Tabatabai [46] used a pyramid of separable filters and in [47], applied it to color images. Tran et. al. [48] used an extension of Chen and Pratt's [49] combined Huffman and run-length coding scheme to code QMF pyramids. Adelson et al. have used both separable and hexagonal QMF pyramids for image coding [27, 36, 39]. Mallat [3] used filters derived from wavelet theory to code images. Westerink et. al. [50] have used vector quantization for subband coding of images.

In figures 4.22 and 4.23, we give examples of data compression of the 256×256 "Lena" image using a separable 9-tap QMF bank, a 3-tap asymmetric filter bank (described in section 4.4.2), and a hexagonal "4-ring" QMF bank. In all cases, a four-level pyramid transform was computed by recursive application of the analysis portion of the A/S system to the lowpass image. The total bit

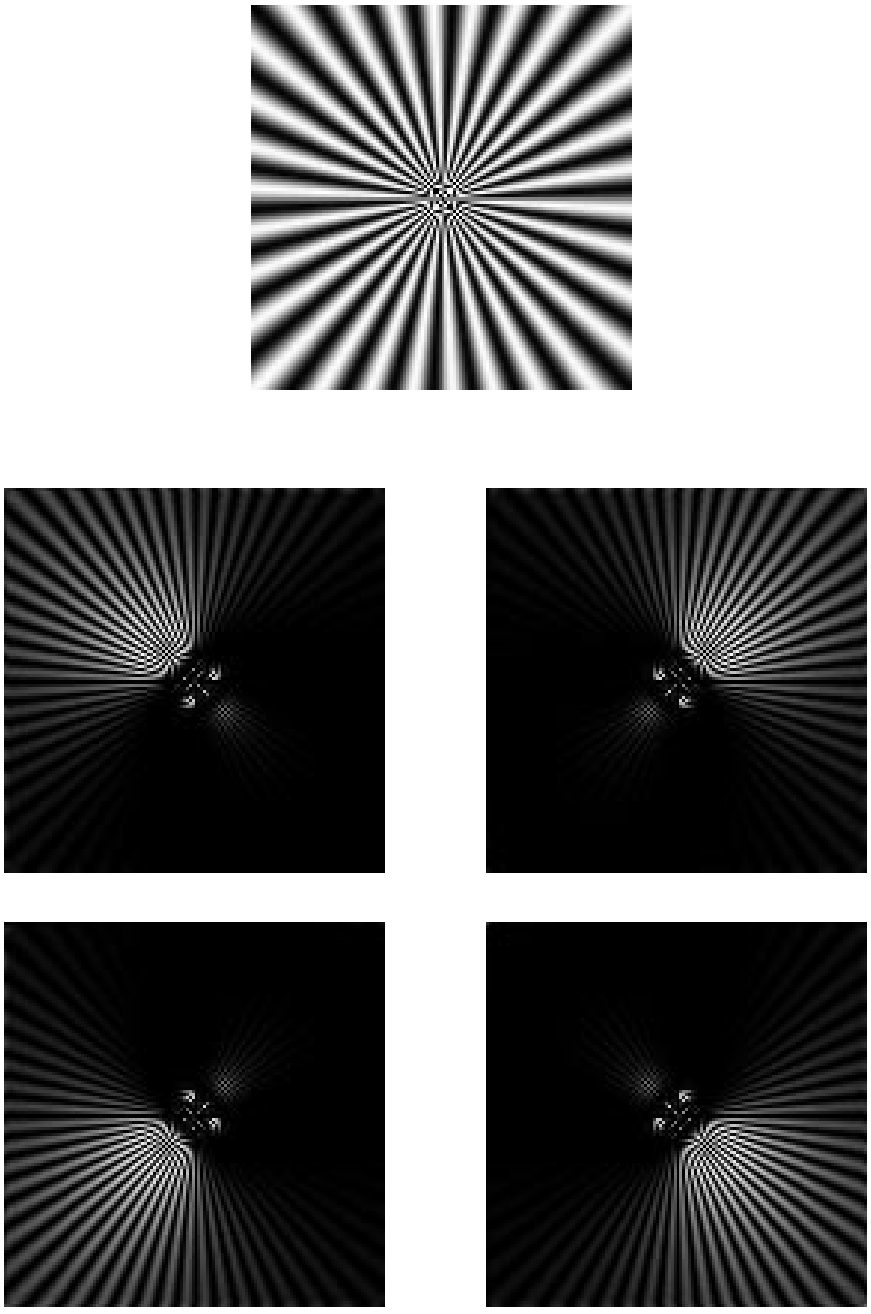


Figure 4.21: At the top is one image from a rotating pinwheel image sequence. The four lower images are the squared result of convolving the sequence with four of the "garnet" filters described in the text. Each filter responds preferentially to one direction of motion.

rate R was fixed and the bit rates assigned to the coefficients of the transform were determined using the standard optimal allocation formula [19]:

$$R_k = R + \frac{1}{2}\log_2 \frac{\sigma_k^2}{\left[\prod_{j=0}^{N-1} \sigma_j^2\right]^{1/N}} \tag{4.24}$$

where, as before, σ_k^2 is the variance of the kth coefficient in the transform. Negative values of R_k were set to zero and the other bit rates raised to maintain the correct overall bit rate R.

Note that if we assume stationary image statistics, the σ_k^2 are the same for all coefficients belonging to the same sub-image of the transform. It has been shown [19] that the optimal quantizer for entropy coding is nearly uniform for bit rates which are high enough that the image probability distribution is approximately constant over each bin. Even though the examples shown were compressed to relatively low bit rates, uniform quantization was used due to its simplicity. Each sub-image was quantized with the bin size chosen to give a first order entropy equal to the optimal bit rate R_k for that subimage.

For the hexagonal pyramid, additional pre- and post-processing was necessary to resample the image on a hexagonal grid. Before building the pyramid, we resampled the original image vertically by a factor of 7/4 using sinc interpolation. We then multiplied by the function $f(\boldsymbol{n}) = 1 + (-1)^{(n_x+n_y)}$. This method, which is similar to one suggested in [42], gives a reasonable geometric approximation to a hexagonal sampling lattice. After re-synthesizing the image, we interpolated the zero-valued pixels and vertically resampled by a factor of 4/7.

The hexagonal QMF system generally offers coding performance perceptually superior to that of the separable system, perhaps because the aliasing errors are not as visually disturbing as those of separable QMFs. Of course, the hexagonal system has the disadvantage of being more inconvenient to use in conjunction with standard hardware.

4.7 Conclusion

We have discussed the properties of linear transforms that are relevant to the task of image compression. In particular, we have suggested that the basis and sampling functions of the transform should be localized in both the spatial and the spatial-frequency domains. We have also suggested that it is desirable for the transform to be orthogonal.

(a) (b)

(c) (d)

Figure 4.22: Data compression example using four-level pyramids. The pyramid data was compressed to a total of 65536 bits (i.e. total first-order entropy was 1.0 bit/pixel). (a) Original "Lena" image at 256×256 pixels. (b) Compressed using 9-tap separable QMF bank. (c) Compressed using 3-tap asymmetrical filter bank. (d) Compressed using "4-ring" hexagonal QMF bank.

Figure 4.23: Data compression example using four-level pyramids. The pyramid data was compressed to a total of 16384 bits (i.e. total first-order entropy was 0.25 bit/pixel). (a) Original "Lena" image at 256×256 pixels. (b) Compressed using 9-tap separable QMF bank. (c) Compressed using 3-tap asymmetrical filter bank. (d) Compressed using "4-ring" hexagonal QMF bank.

Several examples serve to illustrate these properties. The Gabor basis functions are well-localized, but the severe non-orthogonality of the transform leads to sampling functions which are very poorly localized. The block DCT is an equal-width subband transform with poor frequency localization. The LOT enhancement provides improved frequency localization. The Laplacian pyramid is an example of an octave-width subband transform that is non-orthogonal, non-oriented and overcomplete; its properties are non-optimal for coding still images, but may be advantageous for coding moving images.

Subband transforms based on banks of QMFs are well-localized, orthogonal, and can be applied recursively to form octave-width subbands. Separable application of these transforms offers orientation specificity in some but not all of the subbands. Non-separable orthogonal subband transforms based on hexagonal sampling offer orientation specificity in all of the subbands, although they are more difficult to implement. These orthogonal subband transforms are highly effective in image coding applications, and may also be appropriate for applications in image enhancement and machine vision tasks.

Appendix: Filters

In this appendix, we discuss the design of QMFs, present a set of example filter kernels, and compare their theoretical energy compaction properties to those of the DCT and LOT transforms.

Filter Design

A "good" QMF is one that satisfies the constraint given in equation (4.14). In addition, one would like the sub-band images to have a minimal amount of aliasing. The objective then is to design filters with small regions of support that satisfy both of these constraints. Assuming symmetric (linear phase) filter designs, a filter of size N is determined by a set of $\lceil N/2 \rceil$ free parameters, where $\lceil \cdot \rceil$ indicates the ceiling function. Therefore, filters may be designed by minimizing an error function defined on the space of these free parameters.

For a fixed filter size, we define a frequency-domain filter bank error function as the maximal deviation of the overall filter bank response given in equation (4.13) from its ideal value:

$$E_1 = \max_{\omega} \left\{ f_1(\omega) \left| |F(\omega)|^2 + |F(\omega + \pi)|^2 - 2 \right| \right\}$$

where ω ranges over the samples in the frequency spectrum. The function $f_1(\omega)$ is a frequency weighting function roughly matched to the sensitivity of

n	QMF-5	QMF-9	QMF-13
0	0.8593118	0.7973934	0.7737113
1	0.3535534	0.41472545	0.42995453
2	-0.0761025	-0.073386624	-0.057827797
3		-0.060944743	-0.09800052
4		0.02807382	0.039045125
5			0.021651438
6			-0.014556438

Table 4.1: Odd-length QMF kernels. Half of the impulse response sample values are shown for each of the normalized lowpass QMF filters (All filters are symmetric about $n = 0$). The appropriate highpass filters are obtained by delaying by one sample and multiplying with the sequence $(-1)^n$.

the human visual system and the statistics of images:

$$f_1(\omega) = 1/|\omega|.$$

We also define an intra-band aliasing error function:

$$E_2 = \max_{\omega'} \{f_2(\omega')\,|F(-\omega')F(\omega'+\pi)|\}$$

where the function $f_2(\omega')$ is defined as

$$f_2(\omega') = 1/|\omega'|^2.$$

The frequency vector ω' ranges over all of the samples in the frequency spectrum, except for the point at $\pi/2$. Aliasing within subbands cannot be eliminated at this point because the overall filter bank response at this point would then be forced to zero, violating the constraint in equation (4.14).

Finally, we combine the two error functions as a weighted sum:

$$E = \alpha E_1 + (1-\alpha)E_2, \quad \alpha \in [0,1].$$

Given a set of values for the free parameters, we can construct a kernel and compute the value of the error function E. To design filters, we used a downhill simplex method to search the space of free parameters for minima in E. The weighting factor α was adjusted to give a filter bank response error E_1 less than a fixed threshold. A set of example odd-length filter kernels are given in table 4.1.

The same design technique was used for multi-dimensional non-separable filters. For the hexagonally symmetric filters, the free parameters comprise a wedge-shaped region covering approximately one twelfth of the kernel. The

two error functions are defined in the same manner as for the one-dimensional filters. The frequency vector $\boldsymbol{\omega}'$ now ranges over all of the samples in the two-dimensional frequency spectrum, except for those in a hexagonal boundary containing the point $(0, \frac{\pi}{2})$. A set of kernel values is given in table 4.3

In table 4.4 we give several inverse filters for the 3-tap asymmetrical system described in section 4.4.2. These kernels were designed by minimizing the maximal absolute-value reconstruction error for a step edge input signal.

Filter Compaction Properties

An optimal transform for data compression should minimize the bit rate for a given allowable error in the reconstructed image. If the basis functions of the transform are orthonormal, and if expected mean square difference is used as an error measure, this is equivalent to maximizing the following expression for the gain in coding over PCM [19, 45]:

$$G = \frac{\frac{1}{N}\sum_{j=0}^{N-1}\sigma_j^2}{\left[\prod_{j=0}^{N-1}\sigma_j^2\right]^{1/N}}$$

where σ_j^2 is the variance of the jth transform coefficient.

This measure was computed for some of the one-dimensional QMF filters given in this Appendix, as displayed in table 4.5. Values were computed assuming Markov second order signal statistics, where the autocorrelation matrix $\mathbf{R}_{xx}$ is a symmetric Toeplitz matrix of the form

$$\mathbf{R}_{xx} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^N \\ \rho & 1 & \rho & & \rho^{(N-1)} \\ \rho^2 & \rho & 1 & & \rho^{(N-2)} \\ \vdots & & & \ddots & \\ \rho^N & \rho^{(N-1)} & \rho^{(N-2)} & & 1 \end{bmatrix}$$

and where ρ is the inter-sample correlation coefficient. A value of $\rho = 0.95$ was used to compute the numbers given in table 4.5. The compaction values are given for a one-dimensional image of size $N = 256$ with the QMF filter kernels reflected at the edges in a manner that preserves the orthogonality of the basis set [36]. Comparable values for the 16-point LOT (with kernel sizes L=32) and a 16-point block DCT and a 32-point block DCT are also given.

				j		k		l		l		k		j						
			k		g		h		i		h		g		k					
		l		h		e		f		f		e		h		l				
	l		i		f		c		d		c		f		i		l			
k		h		f		d		b		b		d		f		h		k		
j	g		e		c		b		a		b		c		e		g		j	
k		h		f		d		b		b		d		f		h		k		
	l		i		f		c		d		c		f		i		l			
		l		h		e		f		f		e		h		l				
			k		g		h		i		h		g		k					
				j		k		l		l		k		j						

Table 4.2: A hexagonal filter. The letters refer to the free parameters (see text). Only the low-pass filter is shown. The three highpass filters are formed by modulating and shifting the low-pass.

Parameter	3-ring	4-ring	5-ring
a	0.59290695	0.6066799	0.60879886
b	0.32242984	0.3162482	0.31689283
c	-0.016686682	-0.028019974	-0.027267352
d	-0.061579883	-0.0016289932	-0.012790751
e	-0.0020203826	-0.02741341	-0.03874194
f	-0.0038235565	-0.038143888	-0.02383056
g		-0.005958891	0.0008673751
h		0.019682134	0.015554102
i		0.016045252	0.0080001475
j			-0.0009099232
k			0.0022140248
l			-0.0010486352

Table 4.3: Some example hexagonal filter coefficient values. The parameter letters correspond to the diagram shown in table 4.2

n	15	17	21
0	0.8648855700	0.8662753700	0.8660005000
1	0.3589060300	0.3588442800	0.3586960400
2	-0.1476441600	-0.1488108800	-0.1486006000
3	-0.0618851260	-0.0616580880	-0.0615359620
4	0.0244434030	0.0257062400	0.0255328510
5	0.0106931890	0.0102884290	0.0105768030
6	-0.0030558493	-0.0044906090	-0.0043832410
7	-0.0015278960	-0.0012884160	-0.0017810371
8		0.0006442405	0.0007449251
9			0.0002303323
10			-0.0001151661

Table 4.4: Filter impulse response values for 15, 17, and 21-tap inverses for the 3-tap system described in section 4.4.2. Half of the impulse response sample values are shown for each of the normalized lowpass filters (All filters are symmetric about $n = 0$). The appropriate highpass filters are obtained by multiplying with the sequence $(-1)^n$ and shifting by one pixel.

The 9-tap subband filter gives slightly better value than the 16-point DCT, and the 13-tap subband filter is substantially better. These comparisons do not necessarily correspond to measurements of subjective quality, however, since they are based on a crude Markov statistical model of images, and since they assume an MSE error measure. We have found that images compressed with a 9-tap subband transform are perceptually superior to the 32-point DCT, primarily because of the absence of the block artifacts. We also find that the 9-tap QMF is preferable to the 13-tap QMF: the 9-tap filter produces more aliasing, but the Gibbs ringing is more noticeable with the 13-tap filter. We have not performed any coding experiments using the LOT, and so cannot comment on its performance.

filter	G_{PCM}
QMF-5	8.07
QMF-9	9.05
QMF-13	9.28
fast-LOT-16	9.32
DCT-16	8.82
DCT-32	9.49

Table 4.5: Theoretical coding gains over PCM for four-level QMF pyramids, the fast LOT (with $N = 16$ and $L = 32$), and the block DCT. Values were computed assuming first-order Gauss-Markov signal statistics with $\rho = 0.95$, on a a one-dimensional image of size 256.

Bibliography

[1] David Marr, Tomaso Poggio, and Shimon Ullman. Bandpass channels, zero-crossings, and early visual information processing. *J. Opt. Soc. Am.*, 69:914–916, 1977.

[2] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6):33–41, November/December 1984.

[3] S. G. Mallat. *A theory for multiresolution signal decomposition: the wavelet representation.* GRASP Lab Technical Memo MS-CIS-87-22, University of Pennsylvania, Department of Computer and Information Science, 1987.

[4] Andrew P. Witkin. Scale-space filtering. In *Proc. Int. Joint Conf. Artificial Intelligence*, pages 1019–1021, 1985.

[5] Jan J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[6] Alex Pentland. Fractal based description of natural scenes. *IEEE Trans. PAMI*, 6(6):661–674, 1984.

[7] S. Marcelja. Mathematical description of the response of simple cortical cells. *J. Opt. Soc. Am.*, 70:1297–1300, 1980.

[8] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, July 1985.

[9] John G. Daugman and Daniel M. Kammen. Pure orientation filtering: a scale-invariant image-processing tool for perception research and data compression. *Behavior Research Methods, Instruments, & Computers*, 18(6):559–564, 1986.

[10] Dennis Gabor. Theory of communication. *J. IEE*, 93:492–457, 1946.

[11] R. E. Crochiere and L. R. Rabiner. *Multirate Digital Signal Processing. Signal Processing Series*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[12] Alan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing.* Prentice-Hall, Inc., Englewood Cliffs, 1975.

[13] Gilbert Strang. *Linear Algebra and Its Applications.* Academic Press, Orlando, 1980.

[14] Robert M. Lerner. *Lectures on Communication System Theory*, chapter 10. McGraw-Hill, New York, 1961.

[15] John G. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. ASSP*, ASSP-36:1169–1179, 1988.

[16] John G. Daugman. Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Trans. Biomedical Engineering*, 36(1):107–114, 1989.

[17] M. Kunt, A. Ikonomopoulos, and M. Kocher. Second generation image-coding techniques. In *Proceedings IEEE*, pages 549–574, 1985.

[18] A. B. Watson. Efficiency of a model human image code. *J. Opt. Soc. Am. A*, 12:2401–2417, 1987.

[19] N.S. Jayant and Peter Noll. *Digital Coding of Waveforms. Signal Processing Series*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[20] P. M. Cassereau, D. H. Staelin, and G. de Jager. Encoding of images based on a lapped orthogonal transform. *IEEE Trans. Communications*, 37(2):189–193, 1988.

[21] Henrique S. Malvar and David H. Staelin. Reduction in blocking effects in image coding with a lapped orthogonal transform. In *Proc. ICASSP*, pages 781–784, New York, 1988.

[22] Peter J. Burt. Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16:20–51, 1981.

[23] P. J. Burt and Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Trans. Communications*, COM-31(4):532–540, April 1983.

[24] R. Schäfer, P. Kauff, and U. Gölz. On the application of spatio-temporal contrast sensitivity functions to HDTV. In *Conference on Applied Vision*, pages 118–121, Optical Society of America, San Fransico, July 1989.

[25] A. Croisier, D. Esteban, and C. Galand. Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques. In *International Conference on Information Sciences and Systems*, pages 443–446, Patras, August 1976.

[26] D. Esteban and C. Galand. Application of quadrature mirror filters to split band voice coding schemes. In *Proceedings ICASSP*, pages 191–195, 1977.

[27] Edward H. Adelson, Eero Simoncelli, and Rajesh Hingorani. Orthogonal pyramid transforms for image coding. In *Proceedings of SPIE*, pages 50–58, Cambridge, MA, October 1987.

[28] Stephane G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. PAMI*, 11:674–693, July 1989.

[29] Martin Vetterli. Multi-dimensional sub-band coding: some theory and algorithms. *Signal Processing*, 6(2):97–112, February 1984.

[30] P. P. Vaidyanathan. Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques. *IEEE ASSP Magazine*, 4–20, July 1987.

[31] Martin Vetterli. A theory of multirate filter banks. *IEEE Trans. ASSP*, ASSP-35(3):356–372, March 1987.

[32] J. D. Johnston. A filter family designed for use in quadrature mirror filter banks. In *Proceedings ICASSP*, pages 291–294, 1980.

[33] Vijay K. Jain and Ronald E. Crochiere. A novel approach to the design of analysis/synthesis filter banks. In *Proceedings ICASSP*, pages 5.10–5.10, 1983.

[34] Vijay K. Jain and Ronald E. Crochiere. Quadrature mirror filter design in the time domain. *IEEE Trans. ASSP*, ASSP-32(2):353–360, April 1984.

[35] Gunter Wackersreuther. On the design of filters for ideal QMF and polyphase filter banks. *Arch. Elekt. Ubertrag*, 39:123–130, 1985.

[36] Eero P. Simoncelli. *Orthogonal sub-band image transforms.* Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, May 1988.

[37] M. J. T. Smith and T. P. Barnwell, III. A procedure for designing exact reconstruction filter banks for tree-structured subband coders. In *Proceedings ICASSP*, pages 27.1.1–27.1.4, 1984.

[38] Edward H. Adelson and Eero P. Simoncelli. Subband image coding with three-tap pyramids. In *Picture Coding Symposium*, Cambridge, MA, March 1990.

[39] Eero P. Simoncelli and Edward H. Adelson. Non-separable extensions of quadrature mirror filters to multiple dimensions. In *Proceedings of the IEEE: Special Issue on Multi-dimensional Signal Processing*, April 1990.

[40] J. P. Crettez and J. C. Simon. A model for cell receptive fields in the visual striate cortex. *Computer Graphics and Image Processing*, 20:299–318, 1982.

[41] Andrew B. Watson and Albert J. Ahumada. A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex. *IEEE Trans. Biomedical Engineering*, 36(1):97–106, January 1989.

[42] Russell M. Mersereau and Theresa C. Speake. The processing of periodically sampled multidimensional signals. *IEEE Trans. ASSP*, ASSP-31(1):188–194, February 1983.

[43] Dan E. Dudgeon and Russell M. Mersereau. *Multidimensional Digital Signal Processing. Signal Processing Series*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[44] Eric Viscito and Jan Allebach. Design of perfect reconstruction multi-dimensional filter banks using cascaded Smith form matrices. In *Proceedings ISCAS*, pages 831–834, 1988.

[45] John W. Woods and Sean D. O'Neil. Subband coding of images. *IEEE Trans. ASSP*, ASSP-34(5):1278–1288, October 1986.

[46] H. Gharavi and Ali Tabatabai. Sub-band coding of digital images using two-dimensional quadrature mirror filtering. In *Proceedings of SPIE*, pages 51–61, 1986.

[47] H. Gharavi and A. Tabatabai. Application of quadrature mirror filters to the coding of monochrome and color images. In *Proceedings ICASSP*, pages 32.8.1–32.8.4, 1987.

[48] Anh Tran, Kwun-Min Liu, Kou-Hu Tzou, and Eileen Vogel. An efficient pyramid image coding system. In *Proceedings ICASSP*, pages 18.6.1–18.6.4, 1987.

[49] W. Chen and W.K. Pratt. Scene adaptive coder. *IEEE Trans. Communications*, COM-32(3):225–232, March 1984.

[50] Peter H. Westerink, Dick E. Boekee, Jan Biemond, and John W. Woods. Subband coding of images using vector quantization. *IEEE Trans. COM*, COM-36(6):713–719, June 1988.