Pyramids and multiscale representations

Edward H. Adelson†, Eero P. Simoncelli‡ and William T. Freeman

MIT Media Laboratory,
†Brain and Cognitive Sciences Department,
‡Electrical Engineering and Computer Science Department
Cambridge, Massachusetts 02139

Introduction

Images are composed of features of many sizes, and there is no particular scale or spatial frequency that has a special status in natural scenes. Therefore a visual system, whether natural or artificial, should offer a certain uniformity in the representation and processing of visual information over multiple scales.

Primate visual systems achieve a multiscale character in two ways. First, in the retina, there is a continuous variation in the sizes of the receptive fields of ganglion cells, with the size increasing roughly in proportion to distance from the fovea (and spatial resolution descreasing accordingly); a similar scaling is reflected in cortex. And second, for a given patch of the visual field there are numerous cells in striate cortex which are tuned for different bands of spatial frequency. The decomposition of each part of the image into a set of spatial frequency tuned responses seems to be critical to vision systems in nature, and has been found to be very useful in many artificial settings as well.

One approach to understanding the issues that a natural vision system must face is to build artificial systems and discover the power and limits of different representational schemes. This paper will present a brief overview of some of the multiscale representations that we have explored in our laboratory, and will describe some of the lessons we have learned about representing and using multiscale image information. In this limited space it is impossible to present much detail, so readers may wish to consult the original sources for further information.

Pyramids

A pyramid is a multiscale representation that is constructed with a recursive method that leads naturally to self-similarity. The first basic idea is shown in figure 1, which shows a "Gaussian" pyramid (Burt, 1981; Burt and Adelson, 1983). The original test image is convolved with a low-pass filter and subsampled by a factor of two; the filter-subsample operation is repeated recursively to produce the sequence of images shown. Such a pyramid can be useful for operations that require access to information about low frequencies. The pyramid is also highly efficient to compute.

This research was supported by grants from NSF (IRI 871-939-4), and DARPA (Rome Airforce F30602-89-C-0022).

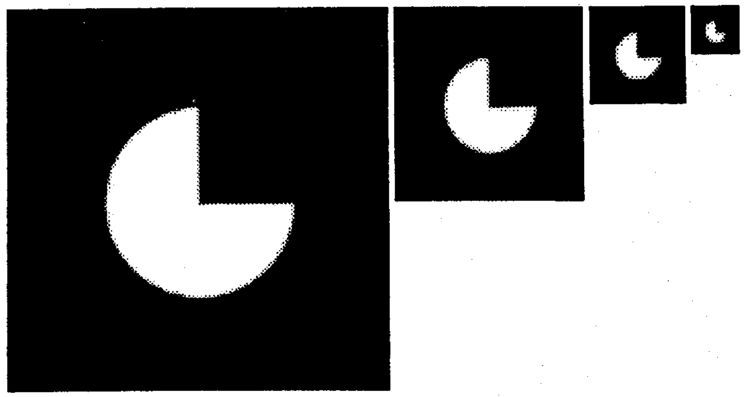


Figure 1: A 4-level gaussian pyramid constructed on a test image.

Figure 2 shows a "Laplacian" pyramid of the same image, in which a bandpass filter is used rather than a low-pass filter. A Laplacian pyramid is a complete representation of an image, in the sense that one can perfectly reconstruct the original image given the coefficients in the pyramid. The reconstruction process is straightforward: one simply interpolates ("expands") each image up to the full size of the original image using the correct interpolation filter, and then sums all of the interpolated images.

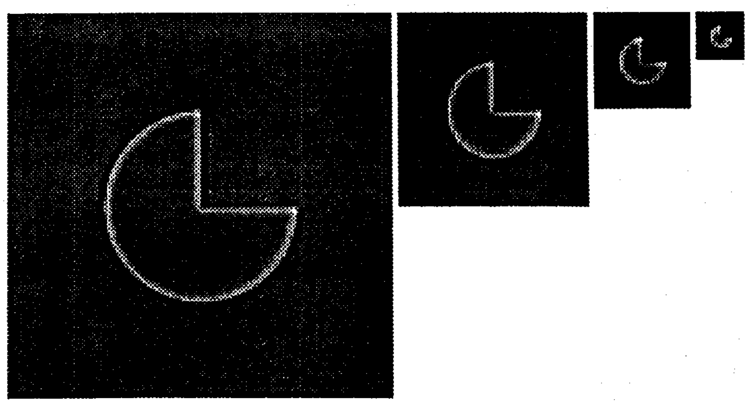
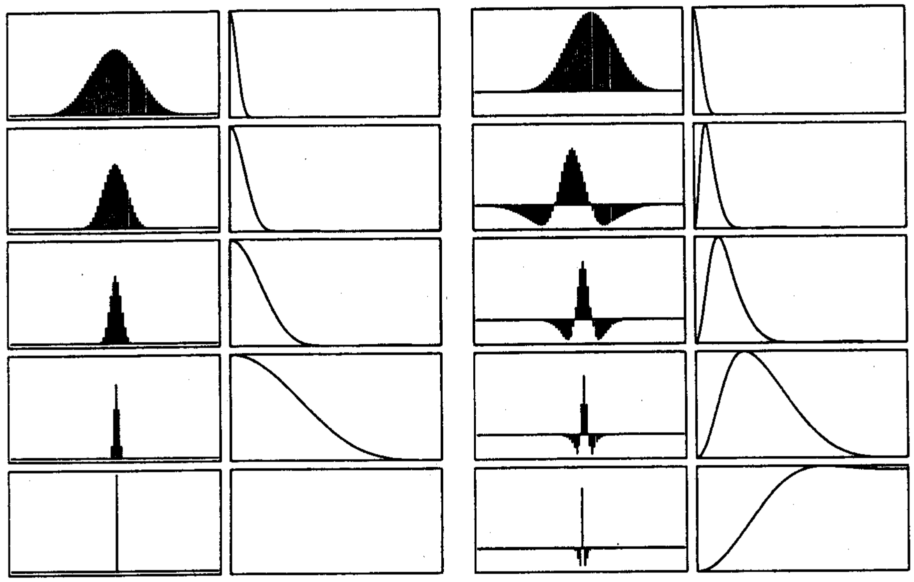


Figure 2: A 4-level laplacian pyramid constructed on a test image.

The hierarchical filtering procedures lead to equivalent filters which are illustrated in figure 3. The equivalent filters used in building the Gaussian pyramid are shown in figure 3(a), while the equivalent filters used in building the Laplacian pyramid are shown in figure 3(b).

Completeness is a valuable property for a representation in early vision, not because a visual system needs to literally reconstruct the image from its representation, but rather because completeness guarantees that no information has been lost, i.e. that if two images are different



(a) (b)

Figure 3: (a) Five example basis functions of a four level Laplacian pyramid, along with their Fourier transforms. (b) The corresponding inverse (sampling) functions of the pyramid, with their Fourier transforms. The transforms are plotted on a linear scale over the range from 0 to π.

then their representations are different.

A complete representation can also be used as a method of storing image information, and the Laplacian pyramid offers an efficient means of storage (Burt and Adelson, 1983). The image encoding procedure is as follows: first the Laplacian pyramid is constructed; then the coefficients in each level are quantized into a fairly small number of bins; then entropy coding techniques such as Huffman coding are applied to the quantized values. Because the Laplacian pyramid values tend to cluster around zero, and because the higher frequencies coefficients tend to have low variance, this technique allows a 256x256 8-bit greyscale image to be stored with about 1.5 bits per pixel with little degradation in image quality.

What can the research on image data compression tell us about multiscale representations in biological system? A biological system is not trying to store an image for reconstruction and display, and cannot use digital techniques such as Huffman coding to gain efficiency. Nonetheless there are important lessons to be learned. The first task of most image coding schemes is to find a representation that is robust and is well-matched to scene statistics. The quantization step noted above leads to random perturbations in the pyramid coefficients, and these perturbations translate into local contrast errors in the bandpassed images. In spite of the random perturbations, it is possible to reconstruct the image with little degradation due to the robust nature of the pyramid. Moreover we find that, from the standpoint of scene statistics, one can allow the perturbations to be quite large in the high frequency bands, while one needs more accuracy in the medium and low frequency bands. For a biological representation this means that it is possible to get away with noisy neurons without losing very much image information, and that the representation of the high frequencies can be particularly tolerant of error.

The efficiency of pyramid representation has relevance in other domains as well. For example, many computational operations such as coarse-to-fine motion processing or stereo matching can be accomplished very efficiently in a pyramid structure. Computational advantages will also be found in such applications as texture analysis, orientation analysis, and pattern matching. A pyramid typically uses as few coefficients as are possible at a given scale, and this reduces both the storage requirements and the number of operations that must be performed in a given task.

Gabor Functions and Orientation Tuning

Orientation tuning is one of the most salient aspects of the cells found in striate cortex, and so it would be useful to understand how to build and use oriented multiscale image representations. Two-dimensional Gabor functions have been the most popular idealized receptive field models (e.g. Granlund, 1978; Marcelja, 1980; Daugman, 1985). One difficulty with the Gabor transform, at least in its original formulation, is that it is highly non-orthogonal. To understand what this means, we have to discuss some general properties of linear transforms.

A linear transform expresses a given (discrete) signal, $f(n)$, as a sum of a set of basis functions, $b_i(n)$:

$$f(n) = \sum_i c_i b_i(n)$$

In the familiar case of the Fourier transform, the $b_i(n)$'s are sinusoids. The c_i's are the coefficients indicating the amount of each basis function that must be added in order to synthesize the original signal.

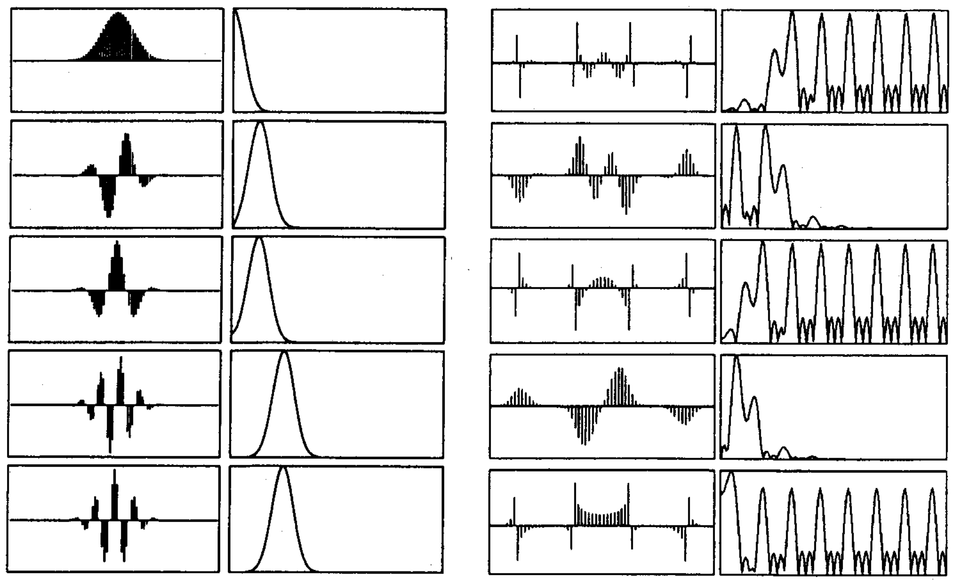
The value of each coefficient c_i can be determined by taking a weighted sum of the pixels in the input signal, i.e. by taking a dot product of the input and a sampling function which represents a "receptive field." That is, for the ith coefficient there is a sampling function $s_i(n)$ such that

$$c_i = \sum_n f(n) s_i(n)$$

In the case of an orthogonal transform, such as the Fourier transform, the sampling functions $s_i(n)$ and the basis functions $b_i(n)$ are identical, so that one detemines the coefficient of a given sinusoid by computing the dot product of the image with that same sinusoid. But in the case of non-orthogonal transforms, the sampling and basis functions can be quite different.

The Gabor transform is invertible because its basis functions are linearly independent; howevever it is not orthogonal and the sampling functions are quite different from the basis functions. Figure 4 shows the Gabor basis functions, along with their Fourier transforms, and also the inverse functions. The inverse set is quite poorly behaved and not at all like one expects to find in a biological system. If one wants to use the Gabor functions as a basis set with which to build images, then one must derive the coefficients by applying the inverse functions, i.e. one would have to use a visual system with these bizarre receptive fields. Or, if one builds an image representation by applying the receptive fields comprising the Gabor set, then the resulting coefficients implicitly represent the image as a sum of the unpleasant inverse functions.

The original Gabor transform has some additional difficulties, one being that it is not self-similar since all the Gabor functions are windowed by a Gaussian of the same width. Many of the investigators who have used Gabor functions in their work have devised self-similar, pyramid-like, approaches.



(a) (b)

Figure 4: (a) Five of the sixteen basis functions of a Gabor filter set, with their corresponding Fourier transforms. (b) The inverse (sampling) functions of the Gabor filter set. The transforms are plotted on a linear scale over the range from 0 to π.

Quadrature Mirror Filter Pyramids and Wavelets

Is it possible to construct a representation that has many of the aspects of a self-similar Gabor transform, and yet which is orthogonal? The answer is yes, as we will now discuss.

Quadrature mirror filters (QMF's) are a class of band-pass filters that were described in the speech domain by Croisier, Esteban and Galand (Croisier *et. al.*, 1976; Esteban and Galand, 1977), and have more recently been applied to the decomposition of images (Vetterli, 1984;

Woods and O'Neil, 1986; Adelson *et al.*, 1987). Although the filters were originally developed using signal processing concepts, they can be easily understood in terms of orthogonal linear transforms (Simoncelli, 1988; Simoncelli and Adelson, 1990b). There has been considerable theoretical and applied work on QMF's in recent years, much of which is reviewed in a book edited by Woods (Woods, 1990). In addition, it has been shown that QMF pyramids are a discrete orthogonal form of wavelets (Mallat, 1989), and for image representation the terms "wavelets" and "QMF's" are sometimes used interchangeably.

Figure 5 shows a self-similar set of QMF's derived from a basic one that has 9 coefficients. These filters can be used as a self-similar basis set for an orthogonal pyramid, where the sampling density of each level is one-half that of the previous level. The result is a pyramid which is "critically sampled," i.e. the number of coefficients is equal to the number of pixels in the original image. The filters are not perfect, in that the reconstructed image will differ from the original very slightly.

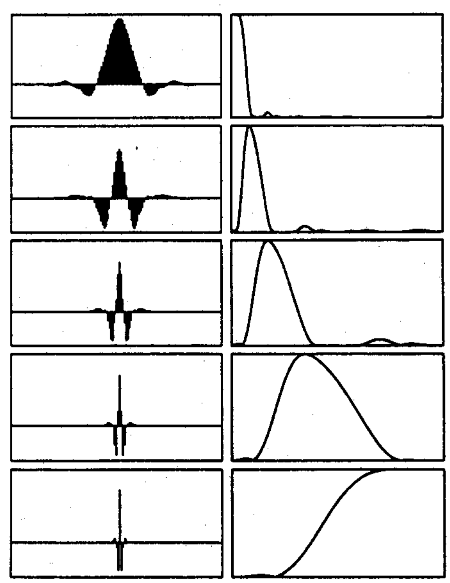


Figure 5: Five of basis functions of a 9-tap QMF/wavelet pyramid transform, along with their Discrete Time Fourier Transforms.

In addition to offering a self-similar orthogonal basis set, the QMF's shown in figure 5 are compact in both space and spatial frequency. Since they are (approximately) orthogonal, the sampling functions are identical to the basis functions. Filters like these form a promising set for application to many problems in image processing. Before one can apply them to images, however, one must extend the one-dimensional functions into two dimensions.

The most straightforward method of two-dimensional generalization involves the separable application of band-splitting QMF's. The QMF's shown in figure 5 come in pairs, which split the frequency band into high-pass and low-pass components. In two dimensions, such filters can be applied separably in the x- and y- dimensions to produce four filters, which may be labeled low-low, low-high, high-low, and high-high. The low-high and high-low bands contain oriented information about vertical and horizontal components of the image. The low-low band contains low-passed information which can be further decomposed in the next level of the pyramid. The high-high band contains a mixture of left and right diagonal information. Figure 6 depicts this decomposition in the frequency domain. The separable decomposition retains the orthogonality of the one-dimensional transform.

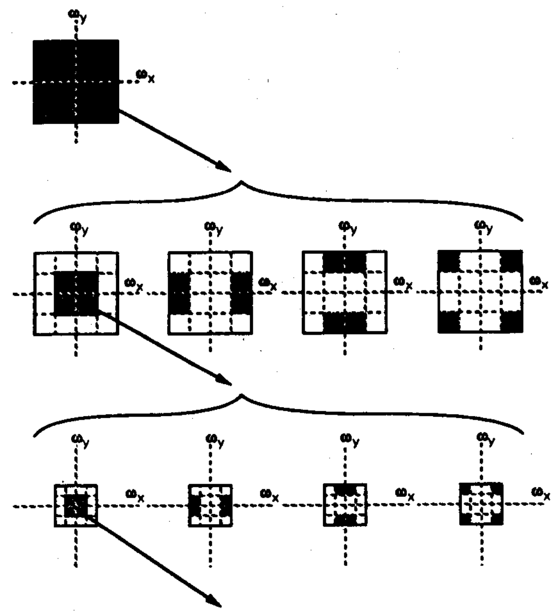


Figure 6: Idealized diagram of the partition of the frequency plane resulting from a 4-level pyramid cascade of separable 2-band filters. The top plot represents the frequency spectrum of the original image, with axes ranging from $-\pi$ to π. This is divided into four subbands at the next level. On each subsequent level, the lowpass subband (outlined in bold) is subdivided further.

A QMF pyramid built using filters of the sort described above turns out to be extremely good for image data compression. Indeed, "subband coders" based on such pyramids are among the best techniques known for efficient coding and are being widely investigated for application in image archiving and digital video transmission.

In spite of its simplicity and its success in image coding, the separable approach has some problems from the standpoint of vision systems. The high-high filter is not nicely oriented, since it contains equal contributions from the two diagonal orientations. This problem is not easily remedied; i.e. it is not easy to split the diagonal band into two oriented parts. Therefore one must seek other approaches in order to achieve orientation specificity in all of the bands.

Simoncelli and Adelson (Adelson *et al.*, 1987, Simoncelli and Adelson, 1990a) have described

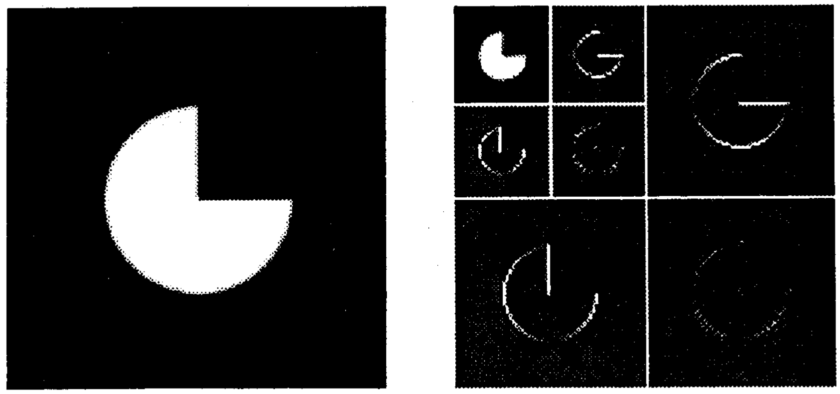


Figure 7: Construction of a 4-level QMF pyramid built on a test image.

a method of constructing QMF pyramids on a hexagonal lattice, in which the basis functions are all of the same shape and are all well-tuned in orientation and spatial frequency. An idealized frequency-domain decomposition is shown in figure 9. The actual frequency tuning of a set of such filters is shown in 10. These filters, like those of the separable pyramid, change scale by octaves from one level to the next.

The hexagonal QMF pyramid demonstrates that it is possible to capture many desirable properties in a single representation. It is an orthogonal wavelet transform; it is complete and is critically sampled, utilizing the same functions for basis functions and sampling functions. The transform is self-similar: the basis functions are all of the same shape, but appear at various sizes, positions, and rotations. The basis functions are smoothly overlapping and are well localized in both space and spatial frequency; they also bear a certain resemblance to the sort of functions that are used in modeling biological visual systems. One can use the hex pyramid for some tasks in early vision (Simoncelli and Adelson, 1990a). The hex pyramid is also a very good structure for image data compression, possibly better than the separable QMF pyramid mentioned above.

An alternate structure for building hexagonal pyramids, but with non-overlapping filters, has been described in (Crettez and Simon, 1982) and in (Watson and Ahumada, 1989). The resulting filters display an unusual blocky structure. Although we have not made a direct comparison, published results (Watson, 1990) suggest that image coding with these alternate hex pyramids requires data rate that is 2 to 4 times higher than with our hex pyramid.

Steerable Pyramids

Although the hex QMF pyramid of figure 9 manages to achieve a great many desirable properties, it does have its limitations. QMF's violate the Nyquist criterion for sampling, and are able to provide successful image representation because aliasing from adjacent bands has opposite sign and therefore cancels during reconstruction. However, there is still aliasing within

(a) (b)

Figure 8: Data compression example using a four-level pyramid. (a) Original "Lena" image at 256 x 256 pixels. (b) Compressed using 9-tap separable QMF bank. The pyramid data was compressed to a total of 16384 bits (i.e. total first-order entropy was 0.25 bit/pixel)

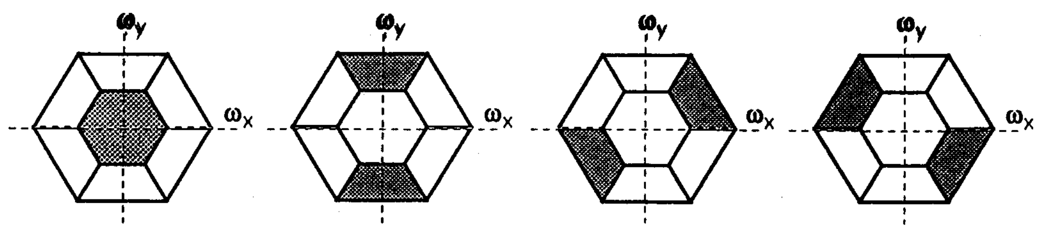


Figure 9: The low-pass and the three oriented high-pass bands of the hexagonal pyramid. Note that the high-pass subbands are not mixtures of different orientations, as in the separable decomposition. This may improve performance for coding and image analysis applications

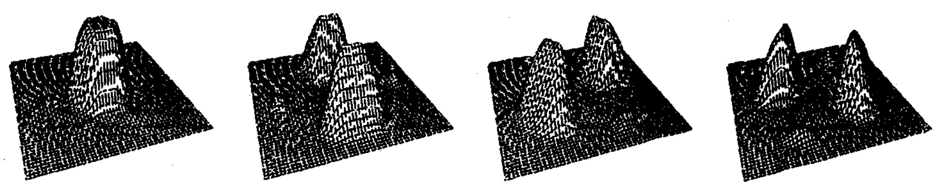


Figure 10: The power spectra for a "4-ring" set of hexagonal QMF filters.

[redacted]

a given band and this expresses itself as a lack of positional invariance in the representation. The problem can be improved if the QMF's have sharp-cut frequency responses, but then the filters lose spatial localization and produce significant "ringing" artifacts.

In fact, the very properties that tend to characterize good filters for data compression (sharp frequency cut-offs with flat responses in between) cause problems for image analysis and early vision applications. For example, in orientation analysis it is necessary that filter responses be very smooth and without flat regions, in order that the population response to different orientations should vary continuously as the orientation of the stimulus is varied.

An interesting class of filters that are well suited to orientation analysis are "steerable filters" (Freeman and Adelson, 1990). A small bank of these filters (say, 4), tuned for different orientations, can be used to analyze an image. Then if one wishes to know the response of a filter of arbitrary orientation, one can compute it as a linear combination of the responses of the original filters. Thus one can derive information about a continuum of possible oriented filters by the application of only a few. The concept can also be extended to allow steerability in phase as well as orientation: by applying the correct set of basis filters, with appropriate orientations and phases, one can synthesize the response of a filter at an arbitrary orientation and phase. One can also extract local energy measures, find the direction of maximal orientation strength, and so on, all from the same basic set of measurements.

Figure 11 (a) shows a bank of steerable filters. These particular filters were designed with one more criterion in mind, namely, that they should be useful for constructing a steerable pyramid decomposition. Indeed, these particular filters were designed to allow the construction of a self-inverting decomposition, which is to say that they were designed so that the basis functions and the sampling functions would be identical. A pyramid can be constructed from either the even or the odd phase filters.

Since the steerable filters are not orthogonal, the self-inverting property must be enforced through other means. We use a highly overcomplete set, and design the filters to "tile" in the frequency domain; i.e. the summed spectral power of the multiple bands and orientations is forced to be flat.

Figure 11 shows a steerable pyramid decomposition of a test image (which has been pre-filtered for reasons that will not be discussed here). The original image is shown in Fig. 11(b); the various levels and orientations are shown in (c), (d), and (e). Because the filters are smooth and their outputs are oversampled, the responses shown here are also quite smooth and well-behaved.

The steerable pyramid is much less efficient than the hex QMF pyramid, both from the standpoint of representation and computation. However, the filters are well-suited to such tasks orientation analysis, edge detection, and image enhancement.

Figure 12(a) shows a picture of Einstein, and figure 12(b) shows an orientation analysis applied to the same image. At each point the orientation of the line segment shows the direction of maximal orientation strength, and the length of the line segment shows the magnitude. The orientation and strength were calculated from the outputs of a pair of even and odd phase steerable filters, similar to the odd phase filters of Figure 11(a) (Freeman and Adelson, 1990) (cf. Knutsson and Granlund, 1983).

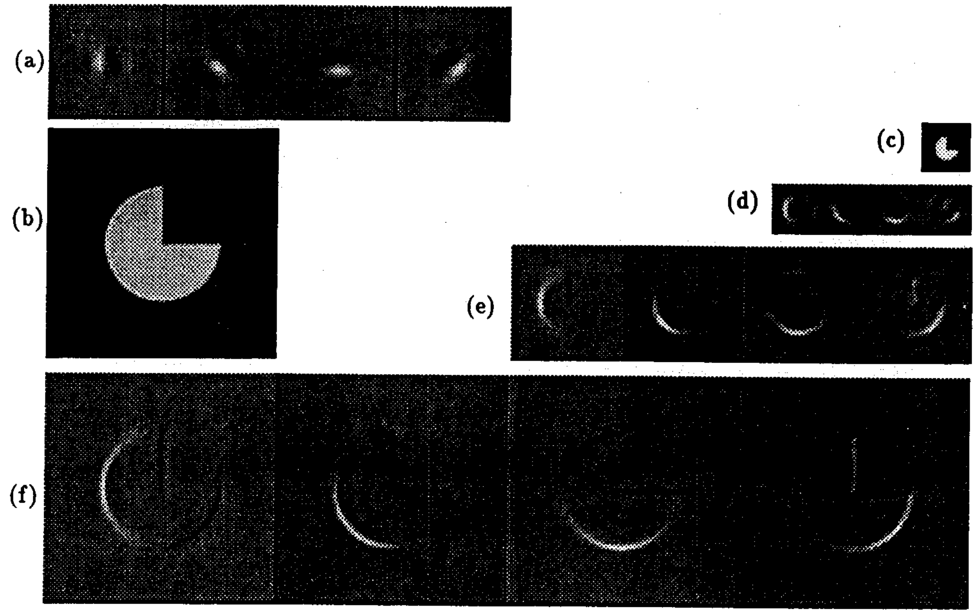
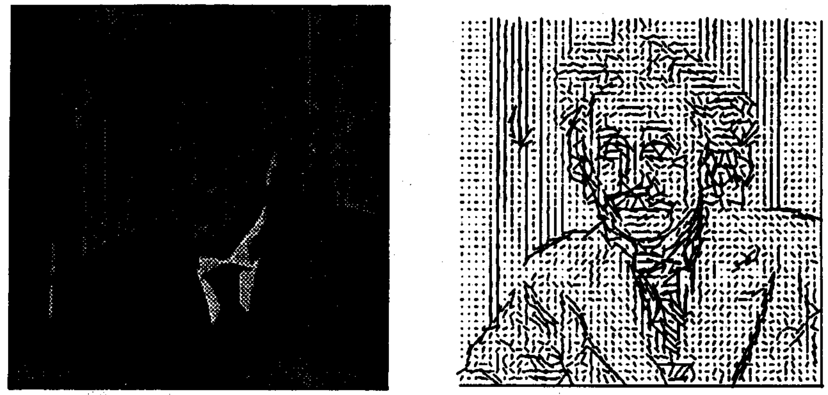


Figure 11: (a) Basis filters for steerable pyramid. Combinations of the filter at these four orientations span the space of all rotations of the filter. (b) Test image. (c) Low-pass image at top of multi-scale pyramid representation of (b). (d) - (f) Steerable, bandpass coefficients in pyramid representation. A linear combination of the transform coefficients will synthesize the response of the analyzing filter at any angle.

Figure 13 shows an example of image enhancement with steerable filters. The original image is a digital cardiac angiogram–an X-ray of a heart. The orientation was analyzed, and then a filter was applied along the direction of maximal orientation in order to enhance the oriented information. A local gain control was then applied to normalize contrast. The result is shown in Fig. 13. Linear structures, which are the ones of greatest interest here, are greatly enhanced in visibility.

Conclusions

Multiscale image representations are useful in a wide variety of vision tasks, and pyramids offer a highly convenient approach to the computation and utilization of multiscale processing. Research in pyramid image representation has revealed some of the strengths and limitations of various kinds of representaitons. Laplacian pyramids are complete and are fairly efficient for image coding, and are useful for front-end processing in various aspects of early vision. Improved coding efficiency can be achieved with QMF pyramids, which are built with orthogonal basis functions; QMF pyramids lead to discrete orthogonal wavelet transforms. By adopting a sampling structure based on a hexagonal lattice it is possible to build QMF pyramids in



(a) (b)

Figure 12: (a) Original image of Einstein. (b) Orientation map of (a).

which all of the basis functions are well-tuned in orientation and spatial frequency. We have recently explored new form of pyramid based on steerable filters, which is less efficient for coding but is well-suited to such tasks as orientation analysis, edge-detection, and image enhancement. The knowledge gained from computational experiments with pyramids may be helpful in understanding the representational issues faced by biological visual systems.

References

E. H. Adelson, E. Simoncelli, and R. Hingorani (1987): Orthogonal pyramid transforms for image coding. In *Proceedings of SPIE*, pp. 50–58, Cambridge, MA.

P. J. Burt (1981): Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16:20–51.

P. J. Burt and E. H. Adelson (1983): The Laplacian pyramid as a compact image code. *IEEE Trans. Communications*, COM-31(4):532–540.

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

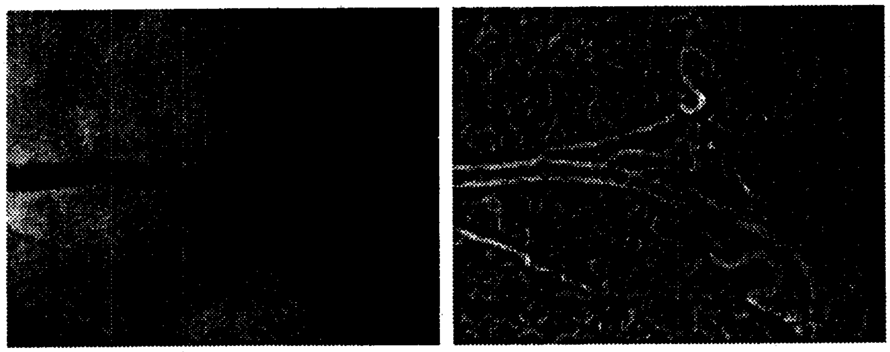
A. Croisier, D. Esteban, and C. Galand (1976): Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques. In *International Conference on Information Sciences and Systems*, pp. 443–446, Patras.

R. E. Crochiere and L. R. Rabiner (1983): *Multirate Digital Signal Processing. Signal Processing Series*, Prentice-Hall, Englewood Cliffs, NJ.

J. G. Daugman (1985): 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.

D. Esteban and C. Galand (1977): Application of quadrature mirror filters to split band voice coding schemes. In *Proceedings ICASSP*, pp. 191–195.

W. T. Freeman and E. H. Adelson (1990): Steerable filters for early vision, image analysis, and wavelet decomposition. In *Proc. Intl. Conf. Computer Vision*.



(a) (b)

Figure 13: (a) Digital cardiac angiogram. (b) Result of filtering (a) with G_2 oriented along the local direction of dominant orientation, after local contrast enhancement (division by the image's blurred absolute value). The oriented vascular structures of (a) are enhanced. (We thank Paul Granfors of G. E. Medical Systems (Milwaukee) for providing the digital angiogram.)

G. H. Granlund (1978): In search of a general picture processing operator. *Comp. Graphics, Image Proc.*, 8:155–173.

H. Knutsson and G. H. Granlund (1983): Texture analysis using two-dimensional quadrature filters. In *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, pp. 206–213.

S. G. Mallat (1989): Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. ASSP*, ASSP-37(12):2091–2110.

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

E. P. Simoncelli (1988): *Orthogonal sub-band image transforms.* Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA.

E. P. Simoncelli and E. H. Adelson (1990a): Non-separable extensions of quadrature mirror filters to multiple dimensions. In *Proceedings of the IEEE: Special Issue on Multidimensional Signal Processing.*

E. P. Simoncelli and E. H. Adelson (1990b): Subband transforms. In J. W. Woods, editor, *Subband Image Coding*, chapter 4, Kluwer Academic Publishers, Norwell, MA.

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

A. B. Watson (1989): Perceptual-components architecture for digital video. *J. Opt. Soc. Am. A*, 7:1943–1954.

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

J. W. Woods and S. D. O'Neil (1986): Subband coding of images. *IEEE Trans. ASSP*, ASSP-34(5):1278–1288.

J. W. Woods, editor (1990): *Subband Image Coding*. Kluwer Academic Publishers, Norwell, MA.