Visual Communications and Image Processing IV

Part 3

PROCEEDINGS
SPIE—The International Society for Optical Engineering

Visual Communications and Image Processing IV

William A. Pearlman
Chair/Editor

8–10 November 1989
Philadelphia, Pennsylvania

Sponsored by
SPIE—The International Society for Optical Engineering

Cooperating Organizations
IEEE Circuits and Systems Society
IEEE Acoustics, Speech, and Signal Processing Society
European Association for Signal Processing (EURASIP)

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 206/676-3290 (Pacific Time) • Telex 46-7053

SPIE PROCEEDINGS SERIES

Volume 1199
Part Three of Three Parts

SPIE (The Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to advancing engineering and scientific applications of optical, electro-optical, and optoelectronic instrumentation, systems, and technology.

The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:
Author(s), "Title of Paper," *Visual Communications and Image Processing IV*, William A. Pearlman, Proc. SPIE 1199, page numbers (1989).

Library of Congress Catalog Card No. 89-63405
ISBN 0-8194-0238-9

Address inquiries and notices to Director of Publications, SPIE, P.O. Box 10, Bellingham, WA 98227-0010 USA.

VISUAL COMMUNICATIONS AND IMAGE PROCESSING IV

Volume 1199

CONTENTS

(continued)

(continued)

Part Two

(continued)

(continued)

VISUAL COMMUNICATIONS AND IMAGE PROCESSING IV

Volume 1199

SESSION 5A

Hierarchical Image/Video Coding II

Chair

Mark J. Smith
Georgia Institute of Technology

Nonseparable QMF pyramids

Eero P. Simoncelli[†] and Edward H. Adelson[‡]

The Media Laboratory
[†]Department of Electrical Engineering and Computer Science
[‡]Department of Brain and Cognitive Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

It is widely recognized that effective image processing and machine vision must involve the use of information at multiple scales, and that models of human vision must be multi-scale as well. The most commonly used image representations are linear transforms, in which an image is decomposed into a sum of elementary basis functions. Besides being well understood, linear transformations which can be expressed in terms of convolutions provide a useful model of early processing in the human visual system. The following properties are valuable for linear transforms that are to be used in image processing and vision modelling:

Scale Specificity (Radial Spatial Frequency): As noted above, one property that is considered to be important in image processing is an explicit representation of *scale*. Basis functions which represent a particular scale are restricted to localized annular region in the frequency domain.

Orientation tuning (Angular Spatial Frequency): Orientation tuning is an important property of most cortical cells and psychophysically inferred channels. In machine vision it is important in the detection of lines and edges. Basis functions which are oriented are restricted to angularly localized regions in the frequency domain.

Spatial localization: In addition to localization in the frequency domain, it is often important for a representation to retain some locational information; thus the basis functions (and sampling functions) should be restricted to some spatial region. This may also improve computational efficiency.

Completeness (Invertibility): Invertibility guarantees that the transformation does not discard any information about the image. In the standard linear algebra terminology, the basis set of an invertible transformation is said to be *complete*.

Orthogonality (Self-inversion): In many cases it is advantageous to work with representations that are orthogonal, or more generally, self-inverting. A self-inverting transform is one whose basis functions and inverse or *sampling* functions are identical. In the case of a linearly independent basis set, this condition is equivalent to orthogonality.

Computational Efficiency: A final property worth considering is that of computational efficiency. There are two ways to achieve efficiency in convolution-based linear transforms: one can use basis functions which have a small region of support, or one can use basis functions which are computable as cascades of simple operations. Both are typically employed in the construction of pyramids.

The first three requirements given above are conflicting: localization in one domain leads to a lack of localization in the other. This joint localization issue was characterized by Gabor [1], who described a set of functions which were optimally localized in both domains. Unfortunately, the corresponding sampling functions for his basis set are very poorly localized in both domains, limiting the use of the transform in praotical situations.

We have developed an image transform that captures all of the properties mentioned above. It is computed using quadrature mirror filters (QMF's) with hexagonal symmetry. QMF's form an orthogonal basis set that is well

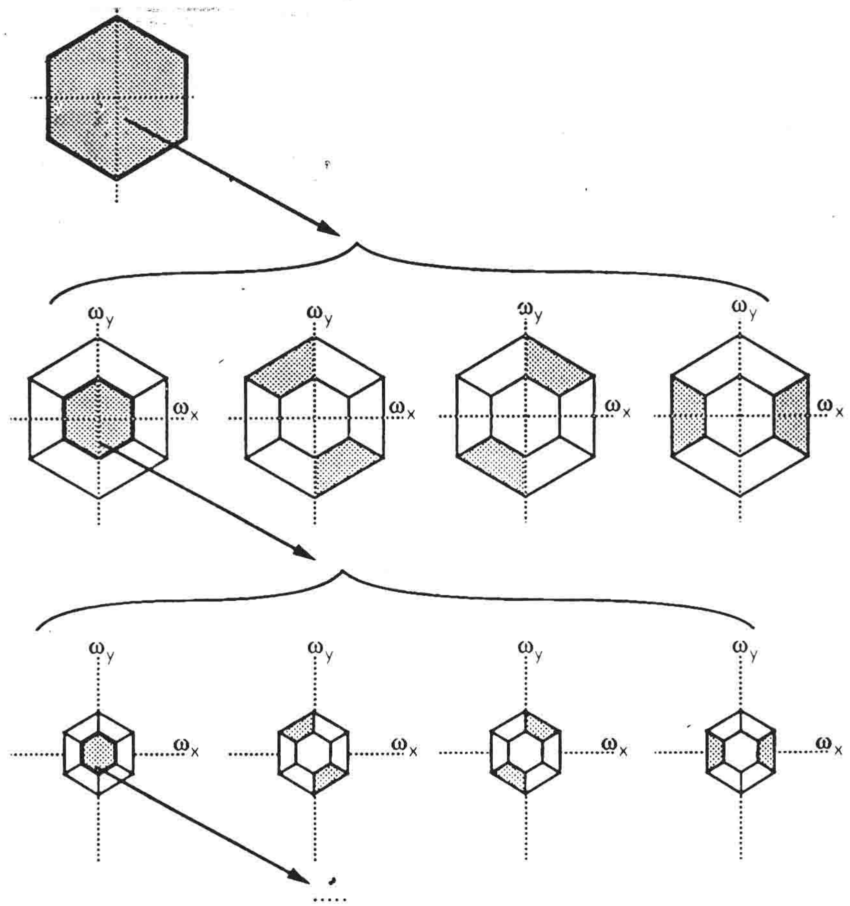

Figure 1: Idealized diagram of the partition of the frequency plane resulting from a four-level pyramid cascade of hexagonal filters. The top plot represents the frequency spectrum of the original image. This is divided into four sub-bands at the next level. On each subsequent level, the lowpass sub-band (outlined in bold) is sub-divided further.

localized in space and spatial frequency [2, 3]. Most research with two dimensional QMF's has involved separable filters, in which one of the subbands contains mixed orientations [4, 5, 6], or non-separable filters which are not orientation-specific [4]. We have previously described a hexagonal QMF pyramid in which all bands are oriented [2]. In the present work we apply such a pyramid to several tasks. We also note that the same concept can be extended to three dimensions; if the third dimension is taken to be time, the resulting filters will be tuned for motion. A more detailed discussion may be found in [7].

Figure 1 illustrates the partition of the frequency domain produced by the hex QMF pyramid transform. The image must be sampled on a hexagonal grid, and its spectrum is assumed to be band-limited to a hexagonal region of the Fourier plane. At each stage the pyramid subdivides the spectrum into four subbands; one is low-pass and the other three are oriented and band-pass. The subbands are subsampled by a factor of four, and the transformation is recursively applied to the low-pass subband. The resulting pyramid constitutes a self-similar orthogonal image transform. To illustrate the decomposition into oriented subbands, we have applied the hex QMF pyramid to a disk image; this is shown in figure 2.

We have also used the pyramid for image data compression, as shown in figure 3. The pyramid coefficients were quantized and entropy coded, to reduce the data rate from 8 bit/pixel to 0.25 bit/pixel.

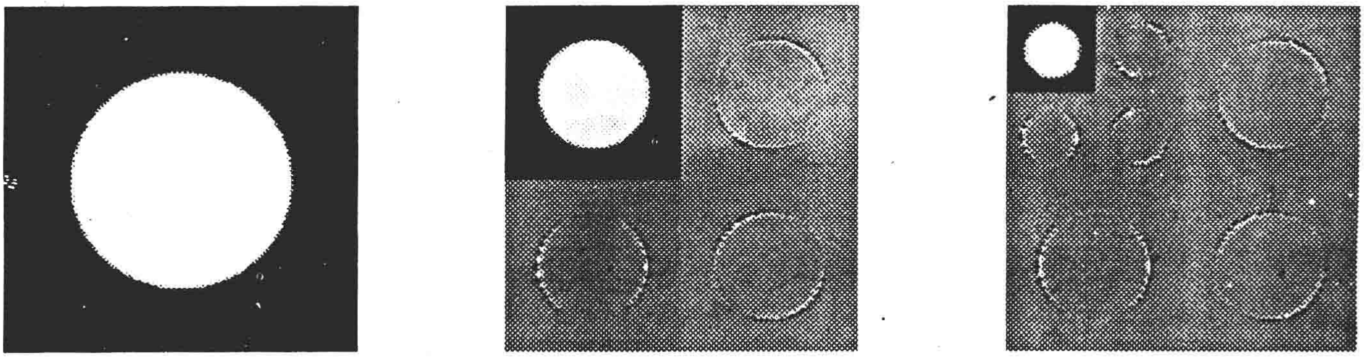

Figure 2: Results of applying a hexagonal QMF bank to an image of a disk. On the left is the original image. In the center is the result after one application of the analysis section of the filter bank. The image has been decomposed into a low-pass and three oriented high-pass images at 1/4 density. On the right, we have applied the filter bank recursively to the low-pass image to produce a two-level pyramid decomposition.

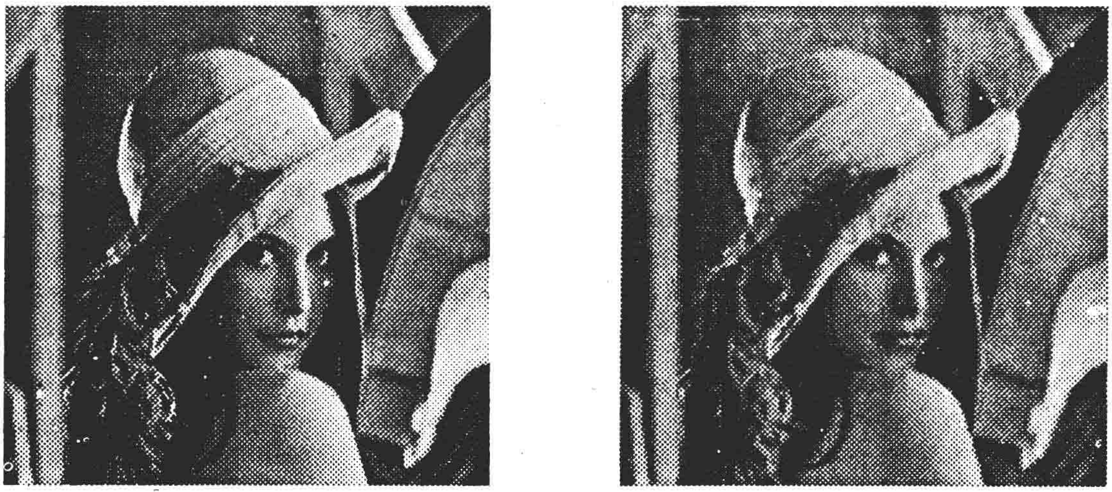

Figure 3: Data compression using the 4-ring hexagonal filter bank. On the left is the original "Lena" image at 256×256 pixels. On the right is a the compressed image. The entropy of the quantized pyramid was 0.25 bits per pixel for a total of 16384 bits.

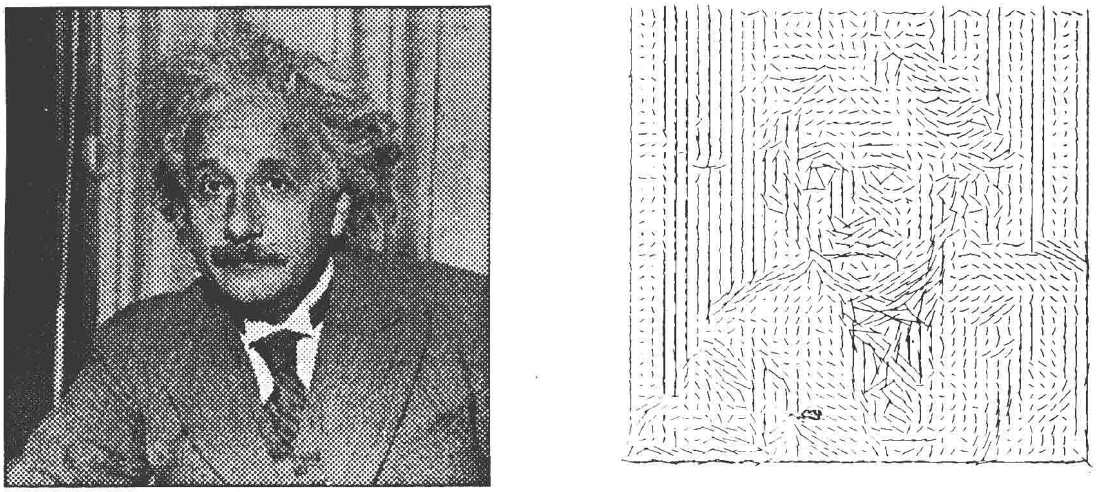

Figure 4: On the left is the original image of Albert Einstein. On the right is an oriented line drawing produced by using linear combinations of the oriented hexagonal filter outputs to measure the strength and orientation of the local image anisotropy.

Because all of the high-pass filters are oriented, they offer a useful set of measurements of local image properties. In figure 4 we illustrate their use in analyzing local orientation of an image, using a technique developed by Freeman and Adelson [8]. Oriented energy measures were taken over the Einstein image, and these were used to derive the both the strength and angle of the local orientation.

The concepts used in the design of the hexagonal QMF's may be extended to three dimensions. Just as a hexagonal lattice is tightest packing of disks in two dimensions, a rhombic dodecaheral lattice is the tightest packing of spheres in three dimensions. It is possible to design three dimensional QMF pyramids, in which the basis functions are tuned for various three dimensional orientations. If the third dimension is time, the filters will isolate local motion information [7].

References

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

[2] Edward H. Adelson, Eero Simoncelli, and Rajesh Hingorani. Orthogonal pyramid transforms for image coding. In *Proceedings of SPIE*, October 1987.

[3] Eero P. Simoncelli. *Orthogonal sub-band image transforms.* Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, May 1988.

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

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

[6] 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.