Pure orientation filtering: A scale-invariant image-processing tool for perception research and data compression

JOHN G. DAUGMAN and DANIEL M. KAMMEN
Harvard University, Cambridge, Massachusetts

A method is described for scale-invariant segregation of image structure solely on the basis of orientation content. This kind of image decomposition is an unexplored image-processing method that is complementary to the well-explored method of filtering in spatial frequency bands; the latter technique is rotation-invariant, whereas the former technique is scale-invariant. The complementarity of these two approaches is explicit in the fact that orientation and spatial frequency are orthogonal variables in the two-dimensional Fourier plane, and the filters employed in the one method depend only on the radial variable, whereas those employed in the other method depend only on the angular variable. The biological significance of multiscale (spatial frequency selective) image analysis has been well-recognized and often cited, yet orientation selectivity is a far more striking property of neural architecture in cortical visual areas. In the present paper, we begin to explore some coding properties of the scale-invariant orientation variable, paying particular attention to its perceptual significance in texture segmentation and compact image coding. Examples of orientation-coded pictures are presented with data compression to 0.3 bits per pixel.

A fertile and well-explored image-processing technique for research on spatial visual mechanisms is spatial frequency filtering. The current popularity of multiscale *pyramid* methods for image analysis and encoding (Burt, 1984; Rosenfeld & Kak, 1976), as well as for simulating human visual processes (Marr, 1982), is partially inspired by the neurobiological finding that visual neurons have receptive fields of many different sizes and spatial frequency characteristics (Blakemore & Campbell, 1969; Enroth-Cugell & Robson, 1966; Kuffler, 1953). Among the many striking demonstrations of the ability of scale-dependent (or spatial frequency selective) manipulations to simulate human visual processes have been (1) Ginsburg's (1971) suggestion that many of the classic Gestalt principles of perceptual organization, as well as the geometrical illusions, can be captured by simple lowpass spatial frequency filtering; (2) the demonstration by Harmon and Julesz (1973) that quantized pictures can be perceptually unmasked by lowpass filtering, and Julesz's (1980) demonstration of a *critical band* for such unmasking by noise of appropriate relative spectral content; and (3) Rentschler and Huebner's (1985) exploration of the processes of face recognition through the frequency-selective construction of compound images. A common technique in such investigations is to manipulate the Fourier transform of an image by multiplying it with a weighting function (a *filter*) that emphasizes some spatial frequencies at the expense of others. Taking the inverse Fourier transform of this weighted product then creates in the image domain a new version of the original image, whose spectral content has now been changed according to the specified filter function. (This operation can also be carried out entirely in the image domain by convolving the image with the desired filter's impulse response function, if it is known.) The significance of such manipulations derives in part from the fact that different kinds of information are contained in the different spatial frequency bands: highpass filtering an image extracts its edges and regions with large derivatives, whereas lowpass filtering performs a kind of averaging that blurs together details but preserves the gross image structure. Thus, different scales of spectral analysis can extract nonredundant information.

The type of spectral manipulation described above involves only one of the two polar variables spanning the two-dimensional Fourier plane, namely spatial frequency (the radial variable). Such filter functions are rotation-invariant, and hence they treat all orientations equivalently. Much less attention has been paid to the orthogonal variable, namely orientation, either in the spatial visual "channels" literature or in the image coding and data compression literatures. Our major purpose here is to begin to explore the perceptual coding properties of this neglected orientation variable, within existing paradigms of texture analysis and image data compression, by a class of image manipulations based only on the angular variable in the Fourier plane.

FORMAL METHODS

Scale-invariant orientation filtering is achieved by multiplying the Fourier transform of an image with a filter function that depends only on the angular coordinate of the Fourier plane, and then taking the inverse Fourier transform. Scale-invariance is entailed by the fact that all spatial frequencies are treated equally.

In Cartesian coordinates, the two-dimensional Fourier transform $F(u,v)$ of an image $I(x,y)$ is defined as

$$F(u,v) = \frac{1}{4\pi^2} \int\int_{-\infty}^{\infty} I(x,y)e^{-i(ux+vy)}dxdy. \quad (1)$$

The Fourier plane coordinates (u,v) can be interpreted in terms of the polar variables of spatial frequency ω and orientation θ through the coordinate transformations

$$u = \omega\cos(\theta)$$
$$v = \omega\sin(\theta), \quad (2)$$

which can be inverted to obtain

$$\omega = \sqrt{u^2+v^2}$$
$$\theta = \arctan\left[\frac{v}{u}\right], \quad (3)$$

and thus we can redefine the Fourier transform of the image $I(x,y)$ in polar coordinates as $H(\omega,\theta)$:

$$H(\omega,\theta) = \frac{1}{4\pi^2} \int\int_{-\infty}^{\infty} I(x,y)e^{-i\omega(x\cos(\theta)+y\sin(\theta))}dxdy. \quad (4)$$

The scale-invariant orientation-filtering operation consists of multiplying $H(\omega,\theta)$ by some angular filter function, $S(\theta)$, which depends only on orientation and not on spatial frequency and which is even-symmetric (in order to keep the recovered image a real rather than complex function). The orientation-filtered image, $I_\theta(x,y)$, is then the inverse Fourier transform of this product $H(\omega,\theta)S(\theta)$:

$$I_\theta(x,y) = \int_{\omega=0}^{\infty}\int_{\theta=0}^{2\pi} H(\omega,\theta)S(\theta)e^{i\omega(x\cos(\theta)+y\sin(\theta))}\omega d\omega d\theta. \quad (5)$$

The particular family of angular filter functions $S(\theta)$ that we discuss here are given by the following expression:

$$S(\theta) = [\frac{1}{2}-\frac{1}{2}\cos(2n\theta+\phi)]^m. \quad (6)$$

These filters are periodic in angle θ and parameterized with three parameters (m,n,ϕ) (m and n being positive integers) in order to provide parametric control over the following filter properties: angular frequency of the passbands (determined by n); angular bandwidth of each passband (determined by m); and angular phase of the passbands (determined by ϕ). It should also be noted that this family of filter functions are normalized to take on amplitudes only between 0 and 1. Examples of several members of this angular filter family are shown in Figure 1, giving both a bird's-eye view of the filters over the Fou-
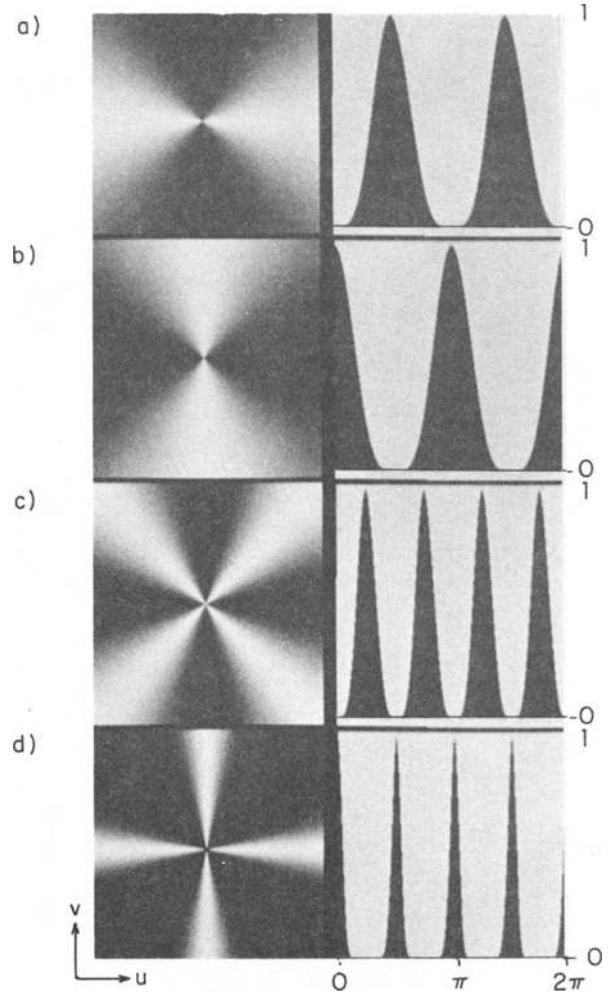

Figure 1. Examples of the angular filter function $S(\theta)$ (Equation 6). The left panels show two-dimensional filter energy distributions in the Fourier (u,v) domain, and the right panels show filter profiles as a function of orientation, independent of spatial frequency. The Fourier domain plots are defined over the 256×256 pixel grids used in the filtering operations. The parameter values are: (a) $n = 1$, $m = 2$, $\phi = 0$; (b) $n = 1$, $m = 2$, $\phi = \pi$; (c) $n = 2$, $m = 4$, $\phi = 0$; (d) $n = 2$, $m = 8$, $\phi = \pi$.

rier plane (left panels) and plots of the same filter profiles against the variable of orientation (right panels).

Finally, it should be noted that the angular filtering operations specified in Equations 5 and 6 are equivalent to certain operations carried out entirely in the space domain. For example, although we do not present the detailed derivation here, the case of angular filtering with the function $S(\theta)$ in which $m = 1$ and $n = 1$ in Equation 6, namely an hourglass filter similar to that shown in panel a of Figure 1, is equivalent to convolving an image with the space-domain singular kernel $s(x,y)$:

$$s(x,y) = \frac{4xy\cos(2\phi)\sin(2\phi)+(x^2-y^2)(\cos^2(2\phi)-\sin^2(2\phi))}{(x^2+y^2)^2},$$

$$(7)$$

and for cases in which m is greater than 1, the equivalent space-domain kernel becomes the m-fold autoconvolution of this function.

RESULTS

Texture Segmentation

A human perceptual faculty that has attracted great interest in recent years is the preattentive segregation of texture fields differing in their dipole statistics (Beck, Prazdny, & Rosenfeld, 1983; Caelli, 1982, 1985; Julesz, 1981). We apply the technique of scale-invariant angular filtering to this discrimination paradigm, and we note that the segregation of the dipole texture fields can be achieved by purely linear operations that have no sensitivity to spatial scale.

Panel a of Figure 2 illustrates the perceptual segregation of texture fields differing in their dipole statistics, as classically studied by Julesz, Beck, Caelli, and others. This texture field has been filtered in Panel b with an angular filter having parameters $m = 2$, $n = 1$, $\phi = 0$, which extracts the central texture at the expense of the surround. The converse operation is achieved in Panel c with the filter's angular phase rotated by 90°, thus
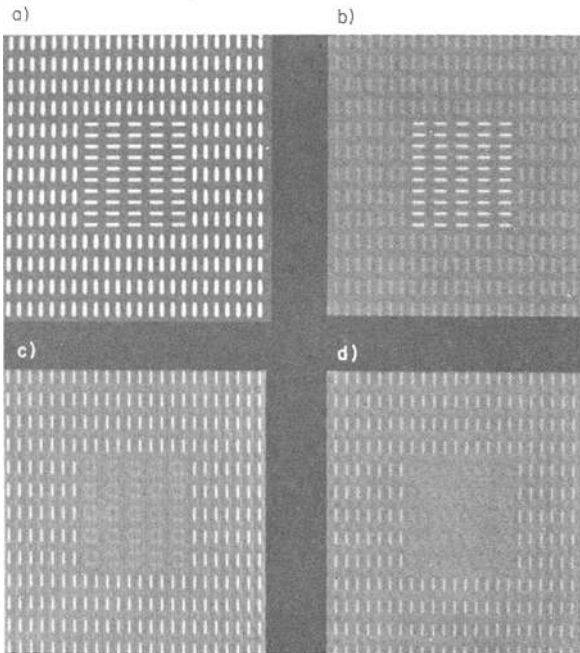

Figure 2. Application of the $S(\theta)$ angular filter to a dipole texture field. In panel a, the original texture field is presented. Panels b and c show the effect of angular filtering with various angular phases ϕ: (b) $n = 1$, $m = 2$, $\phi = 0$; (c) $n = 1$, $m = 2$, $\phi = \pi$. In each case the texture colinear with the filter orientation is passed, while the orthogonally oriented texture is strongly attenuated; some orthogonal signal is transmitted due to the wide ($\Delta\theta_{\frac{1}{2}} = 45°$) orientation passband associated with $m = 1$. In panel d, $n = 1$, $m = 4$, $\phi = \pi$. The narrower passband of this filter ($\Delta\theta_{\frac{1}{2}} = 22.5°$) almost completely attenuates the orthogonal texture field. In this respect, the human perceptual preattentive segregation of dipole texture fields can be captured by scale-invariant, linear, angular filtering, *without loss of spatial resolution.*

preferentially favoring the surround texture field. Finally, in Panel d the filter's angular passband has been further reduced (by increasing m to 8), resulting in still sharper separation of the two texture fields. It should be noted that these textural segregations have been carried out by purely *linear* filtering, and without loss of spatial resolution. If we wished to represent the segregation via a binary classification (e.g., dark center, bright surround), then a second purely linear operation, namely lowpass filtering, would result in that representation of the two texture components. It should be noted that earlier work (Caelli, 1985) has simulated such perceptual grouping of oriented textures by iterated *non*linear convolution and impletion.

Data Compression

The relationship between images and the objects of which they are projections ensures that high degrees of spatial correlation exist in images. Objects possessing some unifying physical characteristics yield images with locally homogeneous signature (such as reflectance, albedo, texture, etc.). This spatially correlated character of real images lends itself to data compression via decorrelating coding schemes. Raster-based video picture coding assumes that every pixel is different and statistically unrelated to its neighbors, consuming data bandwidths at the rate of typically 6–8 bits per pixel (requiring megahertz-range television transmission bandwidths), whereas more clever schemes exploit the intrinsic correlations in image structure across space and across time. The multiscale pyramid methods of picture coding pioneered by Burt (1984; see also Burt & Adelson, 1983) have led to representations consuming as little as 1 bit per pixel with little noticeable image degradation. (For several applications of pyramid methods, see Rosenfeld, 1984.) More extreme data compression schemes replace grayscale structure with line drawings or semantically rich primitives (for reviews, see Musmann, Pirsch, & Grallert, 1985; Pearson & Robson, 1985); such schemes are based on a psychological kind of source coding in the sense of Shannon and Weaver (1949).

Among the most interesting of recent efforts in severe data compression was the work of Sperling, Landy, Cohen, and Pavel (1985) on picture-phone image coding to allow deaf persons to converse in American Sign Language over the switched telephone network. The difficulty of such a project, as Sperling points out, is reflected in the familiar saying that "a picture is worth a thousand words": the analog bandwidth of a television channel is 4 million Hz, whereas that of a telephone channel is 3,000 Hz. Nonetheless, Sperling et al. did achieve such 1000:1 compression factors using either binarized images (cartoons, polygons) or severe lowpass filtering. The goal in such severe data compression, however, is gestural intelligibility rather than image fidelity.

The data-compression potential of the Fourier transform of an image has been widely recognized (for reviews, see Capellini, 1986; Hunt, 1983; Leger & Lee, 1982). Intui-

tively, it arises from the fact that each point in the Fourier transform contains information derived from *every* point in the image, and so even partial Fourier domain information specifies global image structure. Furthermore, images containing various kinds of coherence, such as periodicities, preferred angles, or characteristic correlation distances, will have clumped or clustered distributions of energy in the Fourier domain; thus, data compression exploits *dimensionality reduction*. (Only uncorrelated two-dimensional noise has a statistically uniform Fourier domain energy distribution.) These aspects of natural, structured images allow partial Fourier domain

information to be used for compact image representation (i.e., data compression).

The scale-invariant angular filtering approach developed in the present paper can be used for image data compression of an apparently unexplored kind. By specifying the parameters of the angular filter function $S(\theta)$, so that only that information contained in the Fourier transform along certain angles is selected for, global image structure can be recovered in a *scale-invariant* way from a fraction of the quantity of information in the full transform. A surprisingly small number of angles need be represented in the Fourier domain in order to capture image correlations
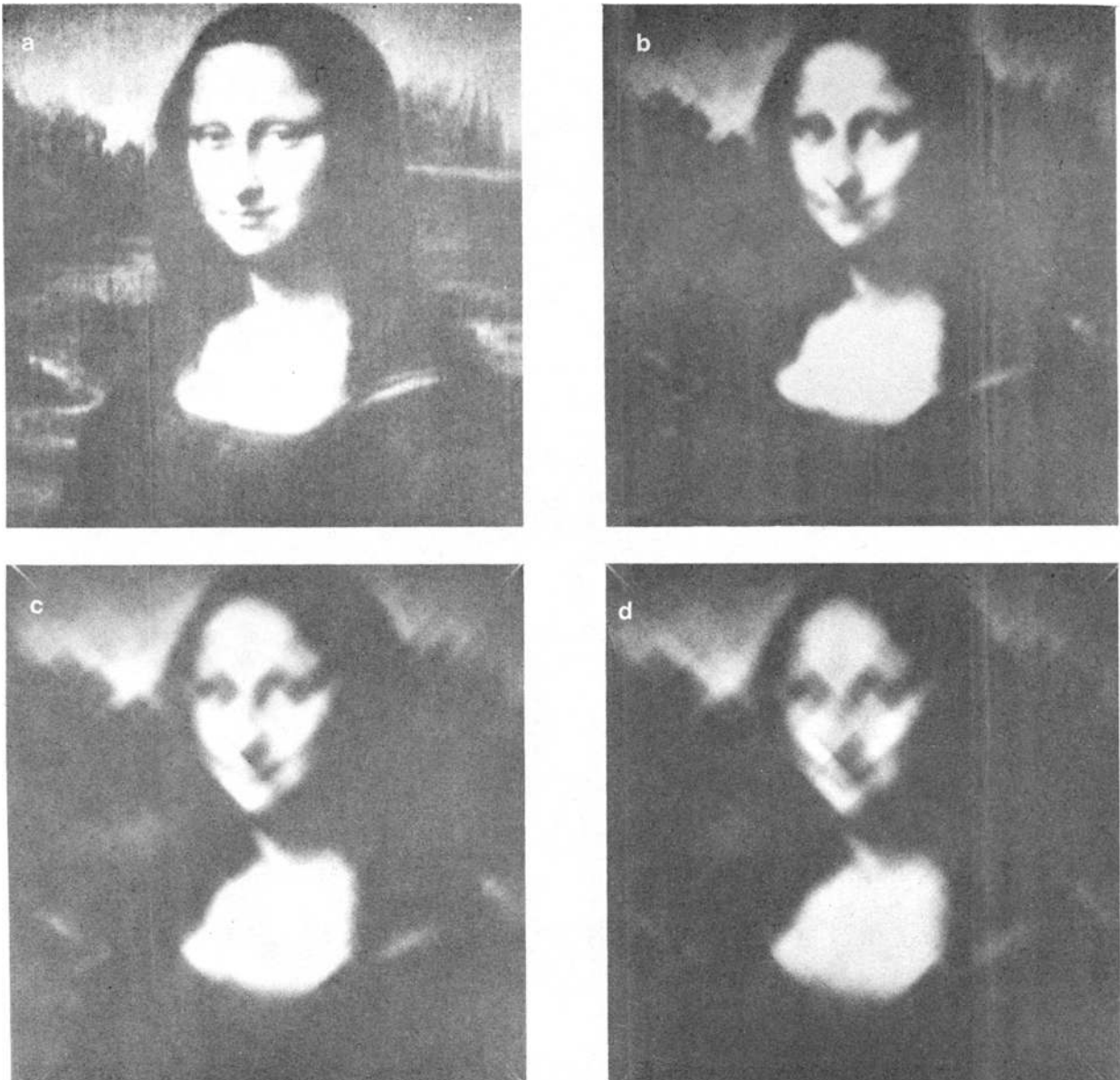

Figure 3. Examples of data compression (dimensionality reduction) with oriented filter functions. Panel a is the original picture. Panels b through d show the result of filtering the original with four-spoke filters (see panel c of Figure 1) having decreasing angular bandwidths. The parameter values are: (b) $n = 2$, $m = 16$, $\phi = 0$; (c) $n = 2$, $m = 64$, $\phi = 0$; (d) $n = 2$, $m = 128$, $\phi = 0$. The crossed-diagonal filter structure becomes increasingly apparent as m increases; in the limit $m \to \infty$, the filter approaches an "X," 1 pixel wide, in the Fourier domain. Relating the angular filter bandwidths to data compression factors, relative to the original picture's 6 bits per pixel (BPP), the data rates are: (b) $\Delta\theta_{1/2} = 5.94°$, 0.84 BPP; (c) $\Delta\theta_{1/2} = 2.98°$, 0.42 BPP; and (d) $\Delta\theta_{1/2} = 2.11°$, 0.30 BPP.

sufficient for intelligible recovery. Figure 3 illustrates this for the case of filters containing only four principal lobes, corresponding to two orientation bands (angular frequency parameter $n = 2$) and various angular bandwidths corresponding to $m = 16$ ($\Delta\theta_{1/2} = 5.94°$) in panel b; $m = 64$ ($\Delta\theta_{1/2} = 2.98°$) in panel c; and $m = 128$ ($\Delta\theta_{1/2} = 2.11°$) in panel d. Compared to the information content of the original (256×256) picture in panel a, generated with 6 bits per pixel, these compressed images correspond, respectively, to 0.84 bits per pixel, 0.42 bits per pixel, and 0.30 bits per pixel.

DISCUSSION

Among the major thrusts of image-processing research in recent years have been finding schemes for image segmentation, based directly on signal structure, and finding compact image representations for achieving data compression, based on such schemes as transform and predictive coding. In the present paper, we have sought to show that *scale-invariant orientation filtering* can be useful in both of these endeavors, as well as in studies of perceptual and biological orientation-based encoding of image structure.

Invariances have played an important role in all domains of signal processing concerned with pattern extraction and recognition. For example, speech processing requires phoneme identification independent of the pitch (and rate) of the speaker's voice; and visual pattern recognition of, say, printed letters should be size-invariant, translation-invariant, and perhaps rotation-invariant. Important efforts in optical pattern recognition (e.g., Casasent, 1977; Hsu, Arsenault, & April, 1982) have incorporated one or more of these invariances explicitly in coding schemes; familiar strategies exploit the fact that the Fourier power spectrum is image translation-invariant, and that the integrated power in annular rings around the origin of the Fourier plane is image rotation-invariant, whereas the integrated power in radial wedges is image size-invariant. Of course, the price paid when integrating spectral power for classifying a signal into one of a set of possible categories is that the original signal cannot be recovered.

Scale-invariant orientation filtering is a linear operation that segregates image structure on the basis of orientation content across scales. The operation is thus orthogonal (both geometrically and conceptually) to the more familiar methods of rotation-invariant spatial frequency filtering. Interestingly, both methods can be used for compact image coding with (qualitatively) roughly comparable compression factors; more exact comparison of these two approaches for a given standard of image quality is left to future work. A second comparison of the two methods of segregating image structure is found in texture segmentation: rotation-invariant spatial frequency filtering can segregate textures based on the sizes of texture elements (mean, or first-order statistics), but not on dipole statistics; scale-invariant orientation filtering (as illustrated in Figure 2) can segregate textures by

dipole, or second-order, statistics but not by first-order statistics. As documented compellingly by Julesz (1981), the human perceptual system is preattentively capable of both orders of pattern segregation.

Undoubtedly, one of the most striking features of mammalian neural signal processing hardware is the cortical *sequence regularity* of columns of orientation-tuned cells (Hubel & Wiesel, 1974). Clearly, orientation filtering has central importance in the logic of biological visual representation, yet the salience of this fact is underrepresented in current image-processing methods. Major genres of research that are intended to be biologically based (Marr, 1982) employ only rotation-invariant operators, such as the Laplacian of a Gaussian, followed by nonlinear zero-crossing extraction. Indeed, there is currently no significant theory, in the signal processing or information-theoretic sense (Hartley, 1928; Shannon & Weaver, 1949), of the neurophysiological orientation sampling and coding logic. It is sometimes claimed (e.g., Watt & Morgan, 1985) that the purpose of early orientation filtering in human vision is to reduce two-dimensional signal processing to one-dimensional signal processing, but this is doubtful in view of the wide angular bandwidth (typically $\pm 15°$) of spatial visual channels, not to mention the logical problem of representing two-dimensional structure with one-dimensional signals. Deeper analysis of orientation coding in human vision, in combination with spatial frequency coding for pattern analysis, has been under way in the work of Caelli and co-workers (e.g., Caelli, 1982; Caelli & Huebner, 1983; Caelli, Huebner, & Rentschler, 1985). In part we are hampered by the absence of a well-developed theory of sampling and signal processing with polar variables, despite important contributions by Stark (1979) and by Brown (1984). It is our hope that the present work may encourage research into these matters.

REFERENCES

BECK, J., PRAZDNY, K., & ROSENFELD, A. (1983). A theory of textural segmentation. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 1-38). New York: Academic Press.

BLAKEMORE, C., & CAMPBELL, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology (London)*, **203**, 237-260.

BROWN, J. L., JR. (1984). Cauchy and polar-sampling theorems. *Journal of the Optical Society of America (A)*, **1**, 1054-1056.

BURT, P. J. (1984). The pyramid as a structure for efficient computation. In A. Rosenfeld (Ed.), *Multiresolution image processing and analysis* (pp. 6-38). Berlin: Springer-Verlag.

BURT, P. J., & ADELSON, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, COM-31, 532-540.

CAELLI, T. M. (1982). On discriminating visual textures and images. *Perception & Psychophysics*, **31**, 149-159.

CAELLI, T. M. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision*, **1**, 19-30.

CAELLI, T. M., & HUEBNER, M. (1983). Coding images in the frequency domain: Filter design and energy processing characteristics of the human visual system. *IEEE Transactions: Systems, Man, & Cybernetics*, SMC-13, 1018-1020.

CAELLI, T. M., HUEBNER, M., & RENTSCHLER, I. (1985). The detection of phase-shifts in two-dimensional images: Part 1. *Biological Cybernetics*, **28**, 167-175.

CAELLI, T. M., & JULESZ, B. (1979). On psychophysical evidence for global feature processing in visual texture discrimination. *Journal of the Optical Society of America*, **69**, 675-678.

CAPELLINI, V. (1986). Two-dimensional digital filters and data compression. In P. W. Hawkes (Ed.), *Advances in electronics and electron physics* (Vol. 66, pp. 141-199). New York: Academic Press.

CASASENT, D. (1977). Optical signal processing. In D. Casasent, *Topics in applied physics* (Vol. 23, Chap. 8). Berlin: Springer-Verlag.

ENROTH-CUGELL, C., & ROBSON, J. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London)*, **341**, 279-307.

GINSBURG, A. P. (1971). *Psychological correlates of a model of the human visual system*. Unpublished master's thesis, Air Force Institute of Technology, Wright Air Development Center, Dayton, OH.

HARMON, L. D., & JULESZ, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, **180**, 1194-1197.

HARTLEY, R. L. V. (1928). Transmission of information. *Bell Systems Technical Journal*, 535-563.

HSU, Y., ARSENAULT, H., & APRIL, G. (1982). Rotation-invariant digital pattern recognition using circular harmonic expansion. *Applied Optics*, **21**, 22-28.

HUBEL, D. H., & WIESEL, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, **158**, 267-293.

HUNT, B. R. (1983). Digital image processing. In P. W. Hawkes (Ed.), *Advances in electronics and electron physics* (Vol. 60, pp. 161-221). New York: Academic Press.

JULESZ, B. (1980). Spatial-frequency channels in one-, two-, and three-dimensional vision: Variations on an auditory theme by Békésy. In C. S. Harris (Ed.), *Visual coding and adaptability* (pp. 263-317). Hillsdale, NJ: Erlbaum.

JULESZ, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, **290**, 91-97.

KUFFLER, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, **16**, 37-68.

LEGER, J. R., & LEE, S. H. (1982). Signal processing using hybrid systems. In H. Stark (Ed.), *Applications of optical Fourier transforms* (pp. 131-205). New York: Academic Press.

MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman.

MUSSMAN, H. G., PIRSCH, P., & GRALLERT, H.-J. (1985). Advances in picture coding. *Proceedings of IEEE*, **73**, 523-548.

PEARSON, D. E., & ROBSON, J. A. (1985). Visual communication at very low data rates. *Proceedings of IEEE*, **73**, 795-812.

RENTSCHLER, I., & HUEBNER, M. (1985). Hidden face perception. *Human Neurobiology* (Vol. 4). Berlin: Springer-Verlag.

ROSENFELD, A. (1984). Some useful properties of pyramids. In A. Rosenfeld (Ed.), *Multiresolution image processing and analysis* (pp. 2-6). Berlin: Springer-Verlag.

ROSENFELD, A., & KAK, A. (1976). *Digital picture processing*. New York: Academic Press.

SHANNON, C. E., & WEAVER, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

SPERLING, G., LANDY, M., COHEN, Y., & PAVEL, M. (1985). Intelligible encoding of ASL image sequences at extremely low information rates. *Computer Vision, Graphics, & Image Processing*, **31**, 335-391.

STARK, H. (1979). Sampling theorems in polar coordinates. *Journal of the Optical Society of America*, **69**, 1519-1525.

WATT, R. J., & MORGAN, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, **25**, 1661-1674.