

Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression

JOHN G. DAUGMAN

(Invited Paper)

Abstract—A three-layered neural network is described for transforming two-dimensional discrete signals into generalized nonorthogonal 2-D “Gabor” representations for image analysis, segmentation, and compression. These transforms are conjoint spatial/spectral representations [10], [15], which provide a complete image description in terms of locally windowed 2-D spectral coordinates embedded within global 2-D spatial coordinates. Because intrinsic redundancies within images are extracted, the resulting image codes can be very compact. However, these conjoint transforms are inherently difficult to compute because the elementary expansion functions are not orthogonal. One orthogonalizing approach developed for 1-D signals by Bastiaans [8], based on biorthonormal expansions, is restricted by constraints on the conjoint sampling rates and invariance of the windowing function, as well as by the fact that the auxiliary orthogonalizing functions are non-local infinite series. In the present “neural network” approach, based upon interlaminar interactions involving two layers with fixed weights and one layer with adjustable weights, the network finds coefficients for complete conjoint 2-D Gabor transforms without these restrictive conditions. For arbitrary noncomplete transforms, in which the coefficients might be interpreted simply as signifying the presence of certain features in the image, the network finds *optimal* coefficients in the sense of minimal mean-squared-error in representing the image. In one algebraically complete scheme permitting exact reconstruction, the network finds expansion coefficients that reduce entropy from 7.57 in the pixel representation to 2.55 in the complete 2-D Gabor transform. In “wavelet” expansions based on a biologically inspired log-polar ensemble of dilations, rotations, and translations of a single underlying 2-D Gabor wavelet template, image compression is illustrated with ratios up to 20:1. Also demonstrated is image segmentation based on the clustering of coefficients in the complete 2-D Gabor transform. This coefficient-finding network for implementing useful nonorthogonal image transforms may also have neuroscientific relevance, because the network layers with fixed weights use empirical 2-D receptive field profiles obtained from orientation-selective neurons in cat visual cortex as the weighting functions, and the resulting transform mimics the biological visual strategy of embedding angular and spectral analysis within global spatial coordinates.

I. INTRODUCTION

SEVERAL broad classes of problems for which neural networks appear to show promise involve the extraction or exploitation of redundancy. Examples include content addressable memory [1], pattern classification and learning [2], signal reconstruction from partial informa-

tion [3], separation of signals from noise [4], cooperative and fault-tolerant processing [5], estimation and prediction [6], and data compression. The last of these is perhaps both the simplest and the most generic example because it most directly depends upon the exploitation of redundancy. In principle, data compression is possible for a nonrandom signal by virtue of the fact that its value at some points can be predicted from its values at other, possibly remote, points or sequences. Correlation structure in a signal can take many forms and can involve different statistical orders, but in information-theoretic terms [7], its existence implies that the entropy or statistical complexity of the source is less than the entropy of the channel, as determined by its resolution (e.g., 8 bits/pixel). Whenever this situation exists, compression of the signal to a lower bound specified by the elimination of redundancy is in principle possible, without loss of information (cf. Theorems 4.5.1 and 4.5.2 of [7]).

Ordinary images are examples of signals having high degrees of self-correlation. Fundamentally, mutual information arises within an image because of the fact that physical objects and scenes tend to have internal morphological consistency, including first-order correlations (locally similar luminance values), second-order or dipole correlations (e.g., oriented edge continuation), as well as higher-order correlations (e.g., homogeneity of textural signature). These correlations are attributes which distinguish real images from random noise, a distinction that is not exploited in the standard pixel-by-pixel image representation. The analysis, communication, and storage of image information would benefit from an efficient means to encode image structure in ways that extracted and exploited these correlations.

A second typical goal in signal processing is to find a representation in which certain attributes of the signal are made explicit. Often this involves transformations into representations in which the attributes or features sought for in the signal are used as the expansion functions. But it is only for certain transforms that the coefficients for projecting the signal onto that chosen set of functions can be easily obtained. If the desired elementary functions are not orthogonal, for example, then simply computing their inner products with the signal will not produce the correct coefficients. A further problem may be that the primitive functions of interest for extracting certain kinds of signal

Manuscript received December 17, 1987. This work was supported by an NSF Presidential Young Investigator Award and by AFOSR U.R.I. Contract F49620-87-C-0018.

The author is with the Departments of Psychology and Electrical, Computer, and Systems Engineering, 950 William James Hall, Harvard University, Cambridge, MA 02138.

IEEE Log Number 8821369.

structure may not constitute a complete basis, or it may be difficult to establish whether or not they do except under strong constraints.

One conjoint transform which illustrates the desirability of obtaining the expansion coefficients on a set of overlapping *nonorthogonal*, yet complete, elementary functions is portrayed by Figs. 1 and 2. Displayed in Fig. 1 is a pixel histogram of the 8-bit "Lena" picture commonly used in image processing research. This gray-scale distribution of 65 536 pixels has an entropy of $S = 7.57$, where entropy is defined as average self-information of the pixel ensemble

$$S = - \sum_{i=1}^n P_i \log_2 P_i \quad (1)$$

given that

$$\sum_{i=1}^n P_i = 1 \quad (2)$$

where the P_i are the relative rates of occurrence of each of the n (in this case 256) gray levels in the picture. Characteristically, the pixel histogram is broad and multimodal, with large entropy. (Uncorrelated 8-bit white noise would have only slightly more entropy, namely, $S = 8$.) But when the Lena picture is transformed into a *complete, discrete, 2-D Gabor* representation (to be defined later), the coefficient values in the transform have the far more compact distribution shown in Fig. 2. Quantized again to 8-bit resolution, the set of 65 536 complete 2-D Gabor coefficients has an entropy of only 2.55, while capturing all of the image structure in the original picture and permitting its exact reconstruction. (The reconstruction may be seen in Fig. 8.) For data compression purposes, one consequence of this observation is that the information cost per pixel for transmitting or storing this 8-bit image could be reduced dramatically without any loss of information. By constructing a code whose word length varies inversely with the frequency distribution shown in Fig. 2, such images could in principle be encoded with a compression factor amounting to 5 fewer bits per pixel. This conjoint 2-D Gabor transform is also useful for image analysis and segmentation, since it extracts locally windowed 2-D spectral information concerning form and texture without sacrificing information about 2-D location or more global spatial relationships, as does a Fourier transform.

The problem is that the overlapping elementary functions which form the projection vectors for this transform are not orthogonal, and so finding their coefficients is difficult. In research to date, it has only been possible to find these coefficients under limiting restrictions on the relationships between the conjoint sampling rate parameters of the elementary functions, and through the use of auxiliary biorthogonal functions [8] expressed as nonlocal infinite series. The main purpose of this paper is to describe a simple neural network architecture for finding optimal coefficient values in arbitrary two-dimensional signal

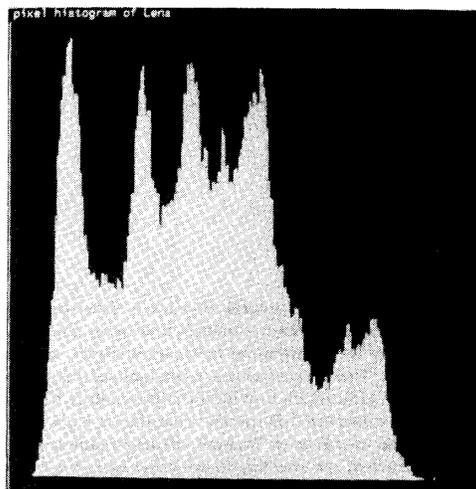


Fig. 1. Pixel histogram of the Lena image, comprising 65 536 8-bit pixels. The entropy of this pixel ensemble is 7.57, only slightly smaller than the entropy of random 8-bit noise with uniform density (namely, 8). Representing images by ensembles of independent pixels does not exploit their intrinsic correlation structure.

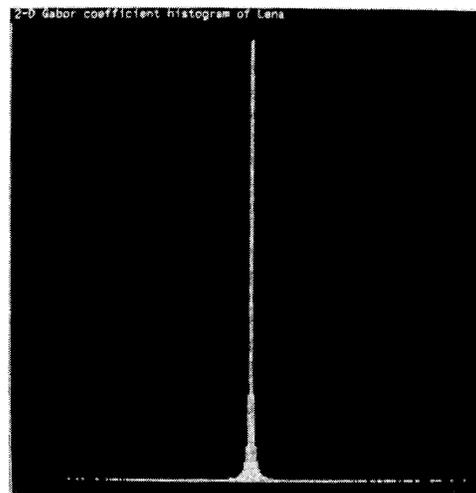


Fig. 2. Histogram of 65 536 coefficients in a complete discrete 2-D Gabor transform, quantized to 8 bits each as was the pixel histogram of Fig. 1 but obviously far more compactly distributed. The entropy of this ensemble of 2-D Gabor coefficients is only 2.55 bits, yet they completely capture the Lena image and allow its exact reconstruction (as shown in Fig. 8). The 2-D Gabor transform itself is shown in Fig. 7.

transforms which in general might be neither complete nor orthogonal. The application of this coefficient-finding scheme to generalized two-dimensional signal transforms is useful for purposes such as image analysis, feature extraction, and data compression. It also leads to an interpretation of the biologically measured two-dimensional anisotropic visual neural receptive field profiles, which have to a large extent motivated the development of the 2-D Gabor transform [10], [15].

II. NEURAL NETWORK FOR FINDING PROJECTION COEFFICIENTS

The general neural network architecture for finding the coefficients in (possibly nonorthogonal and noncomplete) signal transforms is shown in Fig. 3. We shall deal with some discrete two-dimensional signal $I[x, y]$, say, an image supported on $[256 \times 256]$ pixels in $[x, y]$, which we wish to analyze or compress by representing it as a set of expansion coefficients $\{a_i\}$ on some set of two-dimensional elementary functions $\{G_i[x, y]\}$. We may regard a given image $I[x, y]$ as a vector in a 65 536-dimensional vector space, and different representations of the image based on complete orthonormal expansions constitute different bases of this vector space. For example, the conventional pixel representation projects the image onto a set of unit basis vectors, one for each pixel, with coefficients representing lightness values. At the other extreme from the unit basis, each of the linearly independent orthonormal basis vectors might be a 2-D Fourier component, with the associated coefficient being the inner product projection of the image onto this basis vector. More generally, for certain purposes such as feature extraction, we might also wish to represent $I[x, y]$ on a set of linearly dependent vectors, which may or may not completely span the vector space; even if they are neither orthogonal nor complete, we can still find optimal projections of the image onto each one by satisfying global optimization criteria.

Thus, we wish to represent $I[x, y]$ either exactly or in some optimal sense by projecting it onto a chosen set of vectors $G_i[x, y]$. This requires finding projection coefficients $\{a_i\}$ such that the resultant vector $H[x, y]$

$$H[x, y] = \sum_{i=1}^n a_i G_i[x, y] \tag{3}$$

is either identical to $I[x, y]$ (the complete case) or generates a difference-vector $I[x, y] - H[x, y]$ of minimal length (the optimization case). If the elementary functions $\{G_i[x, y]\}$ form a complete orthogonal set, then the representation in $H[x, y]$ is exact (the difference-vector is zero) and the solution for $\{a_i\}$ is simple:

$$a_i = \frac{\sum_{x,y} (G_i[x, y] I[x, y])}{\sum_{x,y} G_i^2[x, y]} \tag{4}$$

If they do not, however, then in general the representation $H[x, y]$ will be inexact and the desired set of coefficients $\{a_i\}$ must be determined by an optimization criterion, such as minimizing the squared norm of the difference-vector:

$$E = \|I[x, y] - H[x, y]\|^2 = \sum_{x,y} (I[x, y] - H[x, y])^2 \tag{5}$$

The norm E will be minimized only when its partial derivatives with respect to all of the n coefficients $\{a_i\}$ equal zero:

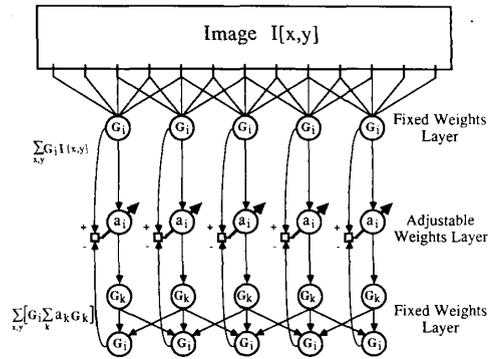


Fig. 3. A three-layered neural network for finding the optimal coefficients in arbitrary image transforms which in general may be neither orthogonal nor complete, nor limited by constraints on sampling uniformity. The first and third layers have fixed weights (in the present work taken to be 2-D Gabor elementary functions as seen in Fig. 4), while the middle layer has weights which are adjusted by interlaminar interactions. In the stable state when equilibrium is reached [see (6) and (9)], the cost function E is minimized and the weight values of the middle layer correspond to the desired transform coefficients.

$$\nabla_{a_i} \frac{\partial E}{\partial a_i} = -2 \sum_{x,y} (I[x, y] G_i[x, y]) + \sum_{x,y} \left[2 \left(\sum_{k=1}^n a_k G_k[x, y] \right) G_i[x, y] \right] = 0. \tag{6}$$

Satisfying this condition for each of the a_i then generates a system of n simultaneous equations in n unknowns:

$$\left| \sum_{x,y} (I[x, y] G_i[x, y]) = \sum_{x,y} \left[\left(\sum_{k=1}^n a_k G_k[x, y] \right) G_i[x, y] \right] \right| \tag{7}$$

Thus, the solution which minimizes the squared norm of the difference-vector [(5)] amounts to finding the set of coefficients $\{a_i\}$ such that the inner product of each vector $G_i[x, y]$ with the entire linear combination of vectors $\sum a_k G_k[x, y]$ is the same as its inner product with the original image $I[x, y]$. It should be noted that in the case when the $\{G_i[x, y]\}$ form a set of orthogonal vectors, then the inner products in the right-hand side of (7) are nonzero only for $k = i$, and so each of the n equations then has only a single unknown, and it is immediately apparent that the minimal-difference-vector solution for each a_i is identical to that given earlier in (4) as the familiar orthogonal case.

Even in the nonorthogonal case, the system of n equations [(7)] could still be exactly solved in principle by algebraic means to find the set of optimal coefficients $\{a_i\}$. But unless the enormous $(65\,536 \times 65\,536)$ matrix generated by (7) is very sparse (requiring strictly compact support for the members of $\{G_i[x, y]\}$), it would be completely impractical to solve this huge system of simultaneous equations by algebraic methods such as matrix

manipulation, since the complexity of such methods grows factorially with the number of simultaneous equations. (Using Stirling's approximation for the factorial, the general matrix solution for the system of equations in (7) would require 2.5×10^{287157} floating-point multiplications to find.) Methods based upon iterative improvement are far faster for such large n , although they converge on an exact solution only as a limit, and can become trapped in local minima. Fortunately, the difference-vector cost function (5) is quadratic in each member of $\{a_i\}$, and so a unique global minimum for E exists. The neural network architecture shown in Fig. 3 converges through iteration upon the desired image representation $\{a_i\}$ by implementing gradient descent along the $E(a_i)$ surface, which expresses the quadratic cost function's dependency on all of the $\{a_i\}$ coefficients.

A common feature of neural network architectures is the combination of layers of neurons having adjustable (or adaptive) synaptic weights, and layers with fixed weights. The present scheme begins with a layer of fixed connection strengths which are specified by an arbitrary set of (generally nonorthogonal) elementary functions $\{G_i[x, y]\}$; by summing the different image pixels through these weights, the output of the i th neuron in this layer is simply the inner product of the i th elementary function, $G_i[x, y]$, with the input image $I[x, y]$ in that region. This is precisely the neurophysiological concept of a (linear) neuron's "receptive field profile," which refers to the spatial weighting function by which a local region of the retinal image is multiplied and integrated to generate that neuron's response strength. The second layer contains adjustable weights for multiplying each of these outputs, according to a control signal which arises from interlaminar interactions. The third layer is identical to the first layer and stores the same fixed set of elementary functions. The adjustable weights of the middle layer constitute the transformed image representation as the set of coefficients $\{a_i\}$. The adaptive control signal adjusts each of the weights by an amount Δ_i , given by the difference between a feedforward signal and a feedback signal. The feedforward signal is the level of activity of the neuron from the first layer, and the feedback signal is the inner product of the weighting function of the corresponding neuron in the third layer with the weighted sum of all the other neighboring neurons in that layer with which it is connected. Thus, the weight adjustment is

$$\Delta_i = \sum_{x,y} (G_i[x, y] I[x, y]) - \sum_{x,y} \left[G_i[x, y] \left(\sum_{k=1}^n a_k G_k[x, y] \right) \right] \quad (8)$$

and the iterative rule for adjusting the value of each coefficient is $a_i \Rightarrow a_i + \Delta_i$. It should be noted that the network does not require a "teacher" that generates the weight adjustment signal by comparing the current representation with a separate copy of the desired pattern. Rather, the adaptive control signal Δ_i arises only from interlaminar network interactions.

It can be seen by inspecting [6] and [8] that the weight adjustment rule is equivalent to

$$\Delta_i = -\frac{1}{2} \frac{\partial E}{\partial a_i}.$$

It should be noted that the minus sign implies that the weight adjustment is always in the downhill direction of the cost surface $E(a_i)$, and that the adjustment is proportional to the slope of the cost surface at this point. A fuller discussion of gradient descent methods may be found in [4, ch. 4]. The equilibrium state of the network that is reached when all $\Delta_i = 0$ is the state in which the cost function E representing the difference-vector squared norm $\|I[x, y] - H[x, y]\|^2$ has reached its minimum; this is the point at which the partial derivative of E with respect to all of the adjustable weights is nil:

$$\mathbf{v}_i, \quad \Delta_i = 0 \Leftrightarrow \frac{\partial E}{\partial a_i} = 0. \quad (9)$$

Thus, in the stable state, the middle layer of the network has weights which represent the optimal coefficients $\{a_i\}$ for the projection of the signal $I[x, y]$ onto any set of elementary functions $\{G_i[x, y]\}$ which, as noted earlier, need be neither orthogonal nor complete.

III. 2-D GABOR ELEMENTARY FUNCTIONS AND BIOLOGICAL VISION

The particular choice of nonorthogonal elementary functions which will be used in the remainder of this paper for the fixed-weight layers of the network are taken from actual neurophysiological measurements of the two-dimensional anisotropic receptive field profiles describing single neurons in mammalian visual cortex [9], [10], [15]. A scientific topic of great interest to neural network researchers is the investigation of the properties and functioning of "real" (biological) neural networks. In the case of the mammalian visual nervous system, a great deal is now known about neural signal processing strategies for the extraction and representation of image structure, at least in the earlier levels of visual processing (retina, lateral geniculate, and primary visual cortex). Among the many questions which can fruitfully be studied regarding signal processing strategies in biological visual systems are the following: how image structure is encoded at various levels; the efficiency of these codes in terms such as dynamic range compression, entropy, noise characteristics, and invariances; the interweaving of multiple coding dimensions within single channel firing rates and across separate channels; the roles of spatiotemporal filtering and of nonlinear operations; and the transformations of image information which support higher level visual cognition. For all of these questions, a potential dialogue between neural network theory, signal processing theory, and experimental neurobiology is an exciting prospect, and the potential mutual benefits for all three disciplines could be high.

The several cortical visual areas of mammals contain

many populations of neurons, some linear and many non-linear, with selectivities for a variety of stimulus attributes. These include location in 2-D visual space, orientation, motion, color, stereoscopic depth, size or spatial frequency, symmetry, and others [11]. In the primary visual cortex (Area 17), perhaps the most striking of these is orientation selectivity [12], which imparts to individual neurons a pronounced dependency between their firing rate and the planar orientation of a stimulus such as an edge or bar. Moreover, assemblies of neurons are organized into "columns" which share the same orientation preference, and on a larger scale, these columns reveal a functional "sequence regularity" of systematic shifts in their preferred orientation [13]. The sequence regularity of columnar orientation preference is one of the most crystalline features of visual cortical architecture now known, and it clearly plays a crucial role, although an as yet unspecified role from a signal processing viewpoint, in the logic of the brain's representation of the visual world. A second striking feature, although true only of the linear class of neurons (so-called "simple cells"), is their pairing by symmetry into quadrature phase pairs: adjacent simple cells have spatial receptive field profiles which share the same location in space and the same orientation preference but differ by 90° in their phase [14]. This quadrature phase relation in neural receptive field pairs is suggestive of a kind of local harmonic expansion of image structure.

One suitable model of the two-dimensional receptive field profiles encountered experimentally in cortical simple cells, which captures their salient tuning properties of spatial localization, orientation selectivity, spatial frequency selectivity, and quadrature phase relationship, is the parameterized family of "2-D Gabor filters," as seen in Fig. 4. This neural model was originally proposed in 1980 simultaneously by Daugman [15] in two-dimensional form and by Marcelja [16] in one-dimensional form. The 2-D form has the virtue of capturing explicitly the critical neurobiological variables of a given neuron's orientation and spatial frequency preference, the tuning bandwidths for these variables, the receptive field dimensions, and the relationships among all of these parameters as captured by generalized uncertainty relationships [10] which the 2-D filter family (in complex form) optimizes.

The general functional form of the 2-D Gabor filter family is specified in (10) and (11), in terms of the space-domain impulse response function $G(x, y)$ and its associated 2-D Fourier transform $F(u, v)$:

$$G(x, y) = \exp \left(-\pi \left[(x - x_0)^2 \alpha^2 + (y - y_0)^2 \beta^2 \right] \right) \cdot \exp \left(-2\pi i \left[u_0(x - x_0) + v_0(y - y_0) \right] \right) \quad (10)$$

$$F(u, v) = \exp \left(-\pi \left[\frac{(u - u_0)^2}{\alpha^2} + \frac{(v - v_0)^2}{\beta^2} \right] \right) \cdot \exp \left(-2\pi i \left[x_0(u - u_0) + y_0(v - v_0) \right] \right). \quad (11)$$

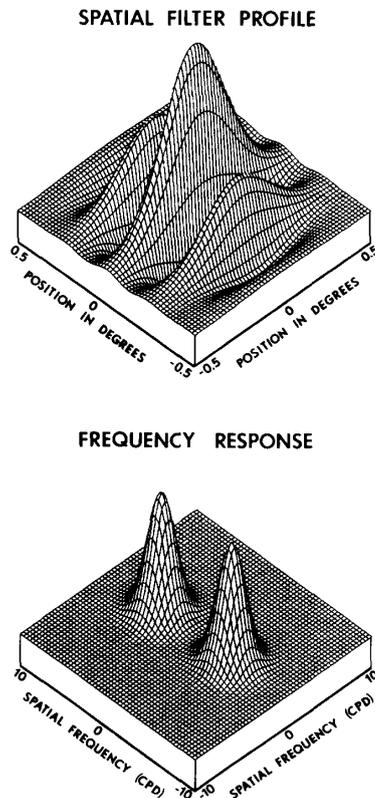


Fig. 4. Example of a 2-D Gabor elementary function (real part) and its 2-D Fourier transform, as originally proposed by Daugman in 1980 [15]. These functions have optimally compact support in conjoint 2-D spatial/2-D spectral representation, and they achieve the lower bound in the general uncertainty relation (12). In the present network (Fig. 3), they provide the weighting functions $\{G_i[x, y]\}$ for the first and third layers.

This family of 2-D elementary functions constitutes a generalization of the 1-D elementary functions proposed in 1946 by Gabor [17] in his famous monograph, "Theory of communication." It should be noted that the 2-D Gabor filter impulse response function $G(x, y)$ and its 2-D Fourier transform $F(u, v)$ have identical functional form; the 2-D Fourier transform theorems for shift, similarity, and modulation are reflected in the position parameters (x_0, y_0) , the modulation parameters (u_0, v_0) , and the two scale parameters (α, β) . If $\alpha \neq \beta$, then a further degree of freedom [for simplicity not included in (10) and (11)] is coordinate rotation of (x, y) out of the principal axes corresponding to (α, β) , which in the Fourier domain results in the same coordinate rotation of (u, v) . These "noncanonical" members of the 2-D Gabor family simply have additional cross terms in xy in (10) and in uv in (11).

An important property of the family of 2-D Gabor filters is their achievement of the theoretical lower bound of joint uncertainty in the two conjoint domains of (x, y) visual space and (u, v) spatial frequency variables. Defining uncertainty in each of the four variables by the nor-

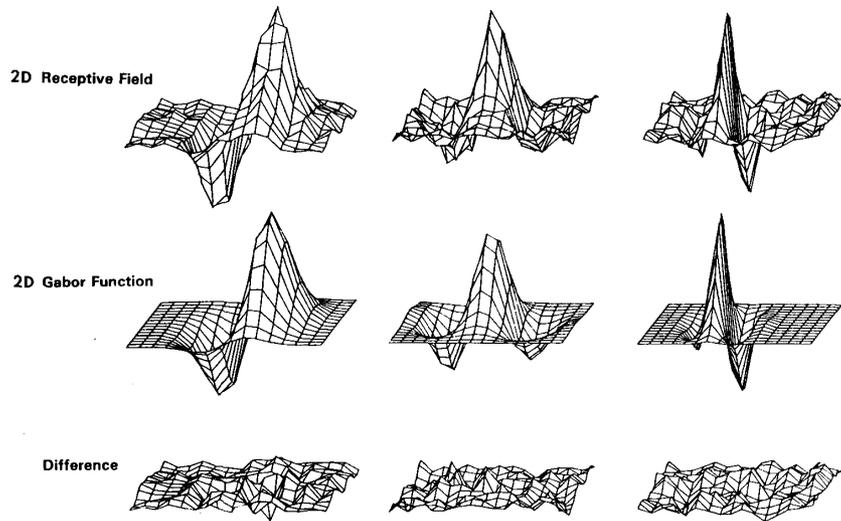


Fig. 5. Top row: illustrations of empirical 2-D receptive field profiles measured by J. P. Jones and L. A. Palmer (personal communication) in simple cells of the cat visual cortex. Middle row: best-fitting 2-D Gabor elementary function for each neuron, described by (10). Bottom row: residual error of the fit, indistinguishable from random error in the Chi-squared sense for 97 percent of the cells studied.

malized second moments (Δx) , (Δy) , (Δu) , (Δv) about the principle axes (see Daugman [10] for details), it may be shown that a fundamental uncertainty principle exists:

$$(\Delta x)(\Delta y)(\Delta u)(\Delta v) \geq 1/16\pi^2 \quad (12)$$

and that the lower bound of the inequality is achieved by the family of 2-D Gabor filters [(10) and (11)]. In this sense, these filters achieve the maximal possible joint resolution in the conjoint 2-D visual space and 2-D Fourier domains. These elementary functions also can be regarded as forming a continuum between the opposite extremes of either Kronecker delta functions in the space domain (inherent in the pixel representation of an image) or Kronecker delta functions in the frequency domain (inherent in the 2-D Fourier representation of an image). These limiting cases arise when the parameters (α, β) in (10) and (11) become either very large or very small; in the mixed case when one is very large and the other very small, the representation corresponds to taking 1-D Fourier transforms on each raster line in a rastered image. In general, we will work with intermediate values of (α, β) in self-similar conjoint representations, because this situation appears to have great neurobiological significance.

It is interesting that the great majority of mammalian cortical simple cells (97 percent in the studies described in [9] and [10]) have 2-D receptive field profiles which can be well fit, in the sense of satisfying statistical chi-squared tests, by members of the family of 2-D Gabor elementary functions. Three examples of such empirical studies by J. Jones and L. Palmer (personal communication) are presented in Fig. 5. The top row shows the empirical 2-D receptive field profiles measured with small

spots of light spanning a (16×16) position grid, plotted as the excitatory or inhibitory effect of the stimulus on the neuron's firing rate. The middle row shows the best-fitting 2-D Gabor elementary function for each cell; and the bottom row shows the residual error of the fit. Extensive discussions of the experimental and analytic methods are provided in [9].

Clearly, the parameters in the 2-D Gabor family of elementary functions directly capture the chief neurophysiological properties of localization in visual space (x_o, y_o) , spatial dimensions (α, β) , preferred orientation and spatial frequency (captured by converting the Cartesian (u_o, v_o) parameters into polar coordinates), and the tuning bandwidths for orientation and spatial frequency (determined jointly by u_o, v_o, α , and β). To this extent, because the neural receptive field profiles $G(x, y)$ are localized *both* in (x, y) visual space *and* in (u, v) 2-D spectral coordinates, we can describe the biological early visual cortical analysis of image structure as forming a conjoint spatial/spatial frequency signal representation with optimized joint resolution, subject to the 4-D uncertainty principle of (12). Roughly speaking, such a representation facilitates the extraction of local 2-D spectral information (texture, scale, axes of modulation) without sacrificing concurrent extraction of information about 2-D location and metrical relationships. For example, the textural structure of a given image region can be separated into its identifying 2-D spectral constituents, while in the same representation, the global spatial structure of the image can be separated into the distinct regions in which a given 2-D spectral structure appears. This scheme of image representation might be considered analogous to a

speech spectrogram, generalized to four dimensions; separate signal components having conjoint support in one domain can be given disjoint support in the other domain, a strategy of proven utility in statistical pattern recognition [19]. Further discussion about conjoint 2-D/2-D anisotropic filter representations and neurobiological mechanisms may be found in [10], [15], and [18].

IV. COMPLETE DISCRETE 2-D GABOR TRANSFORMS

For machine vision, the utility of representing image structure in terms of 2-D Gabor elementary functions is complicated by the fact that they do not constitute an orthogonal basis. The inner-product of two members of the set specified by (10), in the same location (x_o, y_o) but parameterized differently by i and j , is nonzero:

$$\begin{aligned} & \langle G_i(x, y); G_j(x, y) \rangle \\ &= \exp \left(-\pi \left[\frac{(u_i - u_j)^2}{(\alpha_i^2 + \alpha_j^2)} + \frac{(v_i - v_j)^2}{(\beta_i^2 + \beta_j^2)} \right] \right). \end{aligned} \quad (13)$$

One solution to this problem, developed by Bastiaans [8], is to introduce an auxiliary biorthogonal function $\gamma[x, y]$ which allows one to find the correct coefficients by the usual inner product rule for projecting the signal onto the elementary functions. Thus, in the discrete case, if the elementary functions $\{G_i[x, y]\}$ form a complete but nonorthogonal set on which the image $I[x, y]$ can be exactly represented as

$$I[x, y] = \sum_{i=1}^n a_i G_i[x, y], \quad (14)$$

then it may be possible under specific restrictions on $\{G_i[x, y]\}$ to find an auxiliary function $\gamma[x, y]$ such that the desired coefficients $\{a_i\}$ can be found directly by the rule

$$\begin{aligned} a_i &= \sum_{x,y} \gamma[x - x_i, y - y_i] \\ & \cdot \exp[-2\pi i(u_i x + v_i y)] I[x, y]. \end{aligned} \quad (15)$$

Thus, Bastiaans' auxiliary function $\gamma[x, y]$ is biorthogonal to the (invariant) Gaussian window of the chosen elementary functions $\{G_i[x, y]\}$, and it is derived by demanding that the Kronecker delta inner product rule for orthogonal basis functions be satisfied. Although Bastiaans' 1-D solution can be readily generalized to the 2-D case as a Cartesian product, it is expressed only as an infinite series [8], and so in practice an approximation must be found. More importantly, its derivation depends upon certain severe restrictions on the elementary functions $\{G_i[x, y]\}$; in particular, they must all share the same windowing function. This entails that the spatial frequency bandwidths (in octave terms) and orientation bandwidths of the elementary functions will both be inversely proportional to their center frequencies. We would prefer to relax this requirement, in part because the biological 2-D receptive field profiles tend to have a roughly invariant template shape across scales as illustrated by the



Fig. 6. Five examples of 2-D Gabor elementary functions displayed as luminance primitives. These biologically modeled "wavelets" can all be generated from a single complex member by dilations, rotations, and translations, as specified by (22).

luminance profiles in Fig. 6, lending them constant log-polar bandwidths, rather than having a window of constant size which would entail constant linear bandwidths. A further motivation for averting the requirements of the Bastiaans' biorthogonal approach is that we would also like to be able to find optimal conjoint coefficients $\{a_i\}$ even when the elementary functions do *not* form a complete set, as arises from irregular sampling rules. In these cases, the auxiliary $\gamma[x, y]$ biorthogonal function approach to obtaining the coefficients is not helpful, but the approach based on the neural network architecture illustrated in Fig. 3 is.

Before applying the network to the general (nonorthogonal and noncomplete) case, we first demonstrate its ability to accomplish the same goal as the Bastiaans method for regular sampling with invariant window function (the nonorthogonal yet complete case). Here the 2-D Gabor elementary functions are parameterized for an invariant Gaussian window which is positioned on (fully overlapping) Cartesian lattice locations

$$\{x_m, y_n\} = \{mM, nN\} \quad (16)$$

for integers (m, n) and corresponding lattice cell dimensions M, N . The complex exponentials which modulate these overlapping Gaussians are accordingly parameterized for a Cartesian lattice of 2-D spatial frequencies $\{u_r, v_s\}$ appropriate to the M, N spatial lattice:

$$\{u_r, v_s\} = \left\{ \frac{r}{M}, \frac{s}{N} \right\} \quad (17)$$

for integer increments of (r, s) spanning $\{-(M - 1/2), (M - 1/2)\}$ and $\{-(N - 1/2), (N - 1/2)\}$, respectively. Thus, for the neural network shown in Fig. 3, we use for the fixed weighting functions of the first and third

layers the 2-D Gabor elementary functions

$$G_{mnr s}[x, y] = \exp\left(-\pi\alpha^2[(x - mM)^2 + (y - nN)^2]\right) \cdot \exp\left(-2\pi i\left[r\frac{x}{M} + s\frac{y}{N}\right]\right) \quad (18)$$

and allow the network to converge to its stable state, when (9) is satisfied, at which point we may read out the desired coefficients $a_{mnr s}$ from the adjustable weights of the middle layer.

These obtained coefficients $a_{mnr s}$ constitute a complete 2-D Gabor transform of the input image. Each coefficient is complex, but because the input image is real, there is conjugate symmetry among the coefficients: over both parameters r and s , the real part of $a_{mnr s}$ has even symmetry and its imaginary part has odd symmetry. Fig. 7 displays the nonredundant halves of the complete set of real and imaginary coefficients $a_{mnr s}$ as a (256×256) image, giving a complete 2-D Gabor transform of the Lena picture. It is noteworthy that the fundamental uncertainty principle expressed in (12) is implicit in the space/spectral sampling rules expressed in (16) and (17). The larger the size of each spatial lattice cell M or N , which means the fewer the number of spatial sampling positions, the larger is the number of spatial frequency components required in each patch in the corresponding dimension, as expressed above by the ranges of the indexes r and s . Thus, the product of the ranges of the four indexes m, n, r, s is a constant, and in the complete case, is equal to the number of pixels in the image.

The (m, n) lattice that was used in constructing the complete 2-D Gabor transform shown in Fig. 7 is apparent by the periodic clusters of points, which correspond to the centers of the overlapping Gaussian envelopes. Although the size of each $(M \times N)$ lattice cell here was (16×16) pixels, each of the overlapping elementary functions in this transform is fully supported on (32×32) pixels, with Gaussian space constant $(1/\alpha\sqrt{\pi})$ equal to ± 9 pixels at the $1/e$ points; thus, the value at which the overlapping Gaussians are finally truncated and equated to zero is 0.05. Although the value of the Gaussian scale constant α in (18) is arbitrary from the standpoint of completeness and only affects the amount of effective overlap of the 2-D elementary functions across neighboring m, n lattice locations, it does determine the required support size (number of pixels) of each elementary function so that the truncation of the Gaussian tails is negligible. Since the degree of effective overlap of the Gaussians is a free parameter, as was the particular tradeoff between the m, n spatial sampling density and the number of r, s spatial frequency components per patch, these can be manipulated in a signal-dependent fashion without affecting the completeness of the representation. These are signal-dependent flexibilities of the present neural network approach, which are not possible in the biorthogonalizing approach that requires uniform sampling rules and an invariant Gaussian window throughout the image.

Within each of the (m, n) lattice cells apparent in Fig.

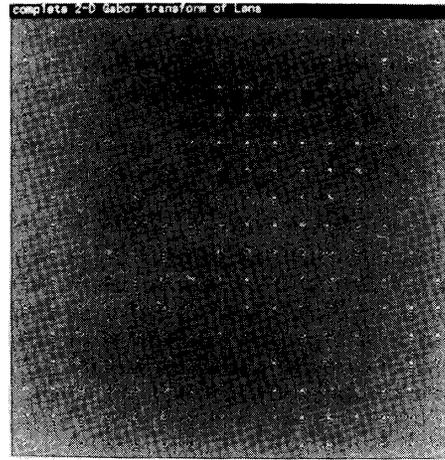


Fig. 7. Complete 2-D Gabor transform of Lena computed by the network of Fig. 3. The amplitude coefficients $\{a_{m,n,r,s}\}$ are quantized to 8-bits and plotted as pixel values (gray being zero), with the spatial center positions m, n of the overlapping elementary functions constituting the global (16×16) lattice centers, and with their 2-D spectral parameters r, s mapped out within each of these local lattice regions. Coefficient histogram shown in Fig. 2; complete reconstruction of Lena from this transform shown in Fig. 8.

7 are embedded the coefficient values $a_{mnr s}$ as (r, s) span their ranges. Thus, the conjoint character of the 2-D Gabor transform is made clear by the way in which local spectral variables (r, s) are embedded within the global spatial image variables (m, n) , for representing the image as the set of coefficients $a_{mnr s}$ on the overlapping, non-orthogonal, elementary functions $G_{mnr s}[x, y]$.

Finally, the completeness of the representation found by the neural network is demonstrated in Fig. 8, which shows the exact reconstruction of the Lena picture from the 2-D Gabor transform of Fig. 7. Each of the transform coefficients was quantized to 8 bits (as in the original pixel image), and the reconstructed picture in Fig. 8 was simply created by the sum of all of the 2-D Gabor elementary functions weighted by their coefficients:

$$H[x, y] = \sum_{m,n,r,s} a_{mnr s} G_{mnr s}[x, y]. \quad (19)$$

The dark points specify the (m, n) lattice locations, and the mean-squared-error of the recovered image is close to zero. Recalling the original entropy comparisons of Figs. 1 and 2, it is striking that all of the image structure seen in Fig. 8 was recovered from the seemingly very impoverished image in Fig. 7, whose histogram has an entropy of only 2.55 bits. Indeed, with the complete 2-D Gabor transform of Fig. 7 quantized to 8 bits, so that each coefficient becomes an integer between -127 and $+128$, about 75 percent of all the coefficients fall within 3 bins of zero. (See Fig. 2.) This means that nearly all the image structure that was recovered in Fig. 8 was contained in just a small subset of the complete 2-D Gabor transform coefficients. For this reason, dramatic factors of data compression are possible by representing images in terms



Fig. 8. Reconstruction of the Lena picture from the complete 2-D Gabor transform displayed in Fig. 7, at only 2.55 bits/pixel. Dark points represent lattice centers for the overlapping 2-D Gabor elementary functions.

of these nonorthogonal elementary functions, whose coefficients can be found by the neural network.

V. IMAGE REPRESENTATION IN SELF-SIMILAR 2-D GABOR “WAVELET” SETS

By eliminating degrees of freedom in the family of 2-D Gabor elementary functions so that they all are dilations, rotations, and translations of each other, with the spectral parameters of the set distributed in a 2-D log-polar lattice, it is possible to represent images on a sparse self-similar family of primitives with advantageous reductions in complexity. In this more biologically inspired scheme as was illustrated in Fig. 6, the different 2-D Gabor elementary functions G_{mnr} have sizes distributed in octave steps (and hence, preferred frequencies also changing in octave steps). In (10), this corresponds to setting α and β proportional to u_o and v_o , thus eliminating two degrees of freedom which correspond to orientation bandwidth and spatial frequency bandwidth. (See [10, Fig. 2] for clarification.) The orientations of the elementary functions, given by

$$\theta_o = \tan^{-1} \left(\frac{v_o}{u_o} \right), \quad (20)$$

are chosen from a fixed set of angles (e.g., six distinct orientations differing in 30° steps). The spectral characteristics of one such set of *log-polar* parameterized 2-D Gabor elementary functions are illustrated in Fig. 9. All the elementary functions in this example have spectral envelopes with a 2:1 aspect ratio (a reflection of their 30° orientation bandwidth and 1.5-octave spatial frequency bandwidth), with center frequencies distributed on a log-polar radial octave grid (the defining 2-D spectra sampling rule), and with self-similarity across all scales, reflecting the invariant shape of the image-domain templates.

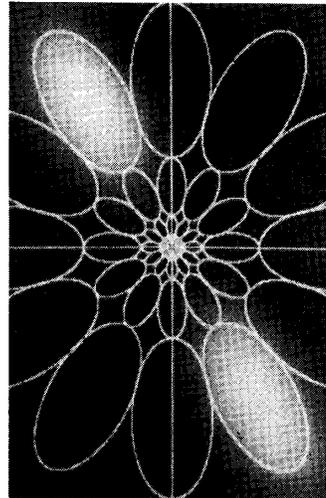


Fig. 9. 2-D Fourier transforms of the Gabor elementary functions employed in one log-polar radial octave “wavelet” scheme. Following physiological data [9], [10], these primitives have logarithmically dispersed center frequencies, $\pm 15^\circ$ orientation bandwidths, 1.5 octave spatial frequency bandwidths, and hence a constant template shape and a 2:1 bandwidth aspect ratio.

In certain of these respects, this set of elementary functions resembles the “wavelet” expansions developed recently by Meyer, Daubechies, Grossmann, Morlet, and Mallat (see [20]–[25]) for analyzing 1-D signals into a self-similar family of wavelets, all of which can be generated by dilations and shifts of a single basic wavelet. Families of wavelets have been recently developed which have strictly compact support and which constitute complete orthonormal bases for $L^2(\mathbf{R})$ functions ([20]). All wavelet schemes, including the present nonorthogonal one, are parameterized by a geometric scale parameter m and position parameter n which relate members of the family to each other:

$$\Psi_{mn}(x) = 2^{-m/2} \Psi(2^{-m}x - n). \quad (21)$$

Generalizing to two dimensions and incorporating discrete rotations θ into the generating function (21) together with shifts p , q and dilations m , the present 2-D Gabor “wavelet” set can be generated from any given member by

$$\Psi_{mpq\theta}(x, y) = 2^{-m} \Psi(x', y') \quad (22)$$

where

$$x' = 2^{-m} [x \cos(\theta) + y \sin(\theta)] - p \quad (23)$$

$$y' = 2^{-m} [-x \sin(\theta) + y \cos(\theta)] - q. \quad (24)$$

By using the network of Fig. 3 to find optimal coefficients on this self-similar multiresolution wavelet scheme in which 2-D Gabor elementary functions serve as the $\Psi_{mpq\theta}(x, y)$, significant further factors of code compression may be achieved as illustrated in Fig. 10. Each column of Fig. 10 corresponds to a different choice for the number of distinct orientations in the wavelet set, and the



Fig. 10. Image compression achieved by the 2-D Gabor "wavelet" transform. Columns: different numbers of distinct wavelet orientations, ranging from six to two. Rows: different quantization depths for each Gabor coefficient, ranging from 8 bits to 5 bits. Overall bit/pixel rates as indicated.

different rows reflect different degrees of quantization of the computed coefficients ranging from 8 bits to 5 bits per coefficient, with the coarsest level always having 2 bits higher quantization accuracy than the finest level. There are 6 distinct values of the scale parameter m of (22)–(24) employed in each decomposition scheme, producing a five-octave range of resolution scales in one-octave steps. Thus, for example, the image in Fig. 10, marked "3 orientations, 1.03 bit/pixel" was reconstructed from 2-D Gabor wavelets present in 3 orientations (changing in 60° steps), 2 quadrature phases, and a total of 2610 positions spanning 5 levels of resolution with variable quantization depth. It is remarkable that rather high image quality is achieved here at only 1 bit/pixel using the coefficients found by the network, even though as few as 3 distinct orientations are represented by the elementary function wavelets.

VI. IMAGE SEGMENTATION

Finally, by examining the distributions of the 2-D Gabor coefficients found by the network in different image regions, it is possible to achieve image segmentation on the basis of spectral signature [26] as demonstrated in Fig. 11. Here the input image to the network (top left panel) is texture consisting of a collage of anisotropically filtered white noise fields, with the noise in different regions of the image having different 2-D bandpass principal orientations.

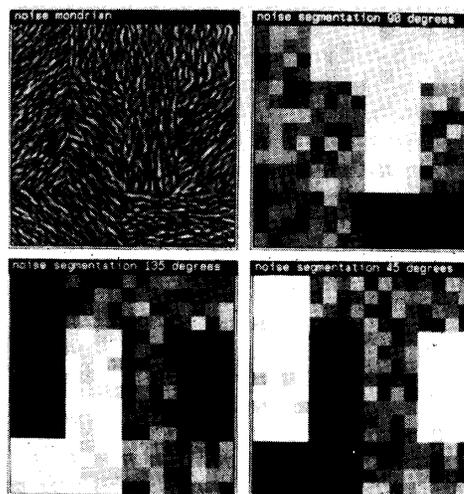


Fig. 11. Image segmentation of anisotropic white noise texture collage (upper left), by the dipole clustering of coefficients in the complete 2-D Gabor transform displayed in Fig. 12.

tations. The complete 2-D Gabor transform of this texture image is displayed in Fig. 12. Close inspection of the transform reveals that associated with each local image region, the 2-D Gabor coefficients a_{mnr} have significant amplitudes that tend to form dipoles of distinct orienta-

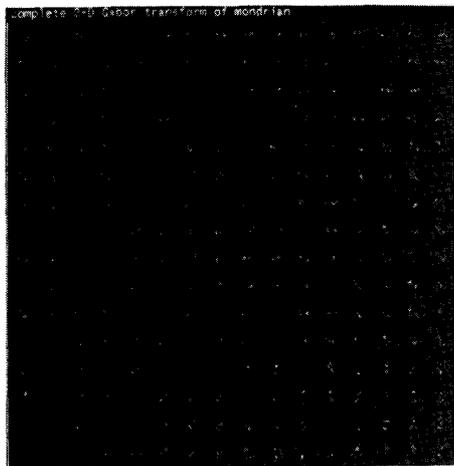


Fig. 12. Complete 2-D Gabor transform of the anisotropic white noise Mondrian displayed in Fig. 11. Different local spectral dipoles are apparent in regions of the transform corresponding to regions of the image described by different anisotropic texture moments.

tions. These orientations correspond to the predominant anisotropic texture moment in that region of the image. On this basis, the original textured image was segmented into distinct regions characterized by a certain spectral signature, as demonstrated in the other three panels. Since the 2-D Gabor coefficients which the network generated as shown in Fig. 12 constitute a conjoint space-spectral representation, spectral information remains localized in the image; hence, it can be associated with particular regions of the image having a given textural signature. Many studies [26]–[33] have confirmed the utility of deriving such regional spectral measures for various signal processing applications. We have seen that the neural network of Fig. 3 for computing the transform coefficients on nonorthogonal 2-D Gabor elementary functions can also be used for texture-based image segmentations.

REFERENCES

- [1] T. Kohonen, *Associative Memory—A System-Theoretical Approach*. New York: Springer-Verlag, 1977.
- [2] K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 826–834, 1983.
- [3] D. Psaltis and N. Farhat, "Optical information processing based on an associative-memory model of neural nets with thresholding and feedback," *Opt. Lett.*, vol. 10, pp. 98–100, 1985.
- [4] B. Widrow and S. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [5] J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proc. Nat. Acad. Sci. USA*, vol. 79, pp. 2554–2558, 1982.
- [6] A. Lapedes and R. Farber, "Nonlinear signal processing using neural networks: Prediction and system modelling," Los Alamos Nat. Lab., preprint LA-UR-87-2662, 1987. (Submitted to *Proc. IEEE*.)
- [7] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [8] M. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc. IEEE*, vol. 68, pp. 538–539, 1980.
- [9] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, pp. 1233–1258, 1987.
- [10] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer.*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [11] D. Van Essen, "Hierarchical organization and functional streams in the visual cortex," *Annu. Rev. Neurosci.*, vol. 2, pp. 227–263, 1979.
- [12] D. Hubel and T. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," *J. Physiol. (London)*, vol. 160, pp. 106–154, 1962.
- [13] —, "Sequence regularity and geometry of orientation columns in the monkey striate cortex," *J. Comput. Neurol.*, vol. 158, pp. 267–293, 1974.
- [14] D. Pollen and S. Ronner, "Phase relationships between adjacent simple cells in the visual cortex," *Science*, vol. 212, pp. 1409–1411, 1981.
- [15] J. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, pp. 847–856, 1980.
- [16] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, pp. 1297–1300, 1980.
- [17] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.*, vol. 93, pp. 429–457, 1946.
- [18] J. Daugman, "Six formal properties of two-dimensional anisotropic visual filters: Structural principles and frequency/orientation selectivity," *IEEE Trans. Syst., Man, Cybern.*, vol. 13, pp. 882–887, 1983.
- [19] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [20] Y. Meyer, "Principe d'incertitude, bases hilbertiennes, et algèbres d'opérateurs," *Seminaire Bourbaki*, 1985–1986, no. 662.
- [21] J. Morlet, G. Arens, I. Fourgeau, and D. Giard, "Wave propagation and sampling theory," *Geophysics*, vol. 47, pp. 203–236, 1982.
- [22] A. Grossman and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, pp. 723–736, 1984.
- [23] A. Grossman, J. Morlet, and T. Paul, "Transforms associated to square integrable group representations. I. General results," *J. Math. Phys.*, vol. 26, pp. 2473–2479, 1985.
- [24] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math. Phys.*, vol. 27, pp. 1271–1283, 1986.
- [25] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal., Machine Intell.*, vol. 10, 1988, in press. (Univ. Pennsylvania GRASP LAB 103, MS-CIS-87-22.)
- [26] J. Daugman, "Image analysis by local 2-D spectral signatures," *J. Opt. Soc. Amer. (A)*, vol. 2, p. P74, 1985.
- [27] Y. Zeevi and M. Porat, "Combined frequency-position scheme of image representation in vision," *J. Opt. Soc. Amer. (A)*, vol. 1, p. 1248, 1984.
- [28] M. Turner, "Texture discrimination by Gabor functions," *Biol. Cybern.*, vol. 55, pp. 71–82, 1986.
- [29] M. Clark, A. Bovik, and W. Geisler, "Texture segmentation using a class of narrowband filters," in *Proc. Int. Conf. Acoust., Speech, Signal Processing 87*, 1987, pp. 571–574.
- [30] R. Hecht-Nielsen, "Nearest matched filter classification of spatio-temporal patterns," *Appl. Opt.*, vol. 26, pp. 1892–1899, 1987.
- [31] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 610–621, 1973.
- [32] H. Szu, "Two-dimensional optical processing of one-dimensional acoustic data," *Opt. Eng.*, vol. 21, no. 5, pp. 804–813, 1982.
- [33] H. Szu and H. Caulfield, "The mutual time-frequency content of two signals," *Proc. IEEE*, vol. 72, pp. 902–908, 1984.



John G. Daugman was born in 1954. He received the B.A. degree in physics in 1976 and the Ph.D. degree in psychology in 1983, both from Harvard University, Cambridge, MA.

His main research interests include neural networks, multidimensional signal processing, and neurophysiological and psychophysical studies of biological visual systems. He serves as a consultant to Lincoln Laboratories and as Scientific Advisor for the Hecht-Nielsen Neurocomputer Corporation. He belongs to the Editorial Board of *Neural Networks*. Since 1985–1986 he has been an Assistant Professor of Psychology and of Electrical, Computer, and Systems Engineering at Harvard University.

Dr. Daugman is the recipient of a 1988 NSF Presidential Young Investigator Award.