A Hexagonal Orthogonal-Oriented Pyramid as a Model of Image Representation in Visual Cortex

Andrew B. Watson
Albert J. Ahumada, Jr.

Reprinted from
IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING
Vol. 36, No. 1, January 1989

A Hexagonal Orthogonal-Oriented Pyramid as a Model of Image Representation in Visual Cortex

ANDREW B. WATSON AND ALBERT J. AHUMADA, JR.

Abstract—**Retinal ganglion cells represent the visual image with a spatial code, in which each cell conveys information about a small region in the image. In contrast, cells of primary visual cortex employ a hybrid space-frequency code in which each cell conveys information about a region that is local in space, spatial frequency, and orientation. Despite the presumable importance of this transformation, we lack any comprehensive notion of how it occurs. Here we describe a mathematical model for this transformation. The hexagonal orthogonal-oriented quadrature pyramid (HOP) transform, which operates on a hexagonal input lattice, employs basis functions that are orthogonal, self-similar, and localized in space, spatial frequency, orientation, and phase. The basis functions, which are generated from seven basic types through a recursive process, form an image code of the pyramid type. The seven basis functions, six bandpass and one low-pass, occupy a point and a hexagon of six nearest neighbors on a hexagonal sample lattice. The six bandpass basis functions consist of three with even symmetry, and three with odd symmetry. The three even kernels are rotations of 0, 60, and 120° of a common kernel; likewise for the three odd kernels. At the lowest level, the inputs are image samples. At each higher level, the input lattice is provided by the low-pass coefficients computed at the previous level. At each level, the output is subsampled in such a way as to yield a new hexagonal lattice with a spacing $\sqrt{7}$ larger than the previous level, so that the number of coefficients is reduced by a factor of seven at each level. In the biological model, the input lattice is the retinal ganglion cell array. The resulting scheme provides a compact, efficient code of the image and generates receptive fields that resemble those of the primary visual cortex.**

Introduction

VARIOUS roles have been proposed for the neurons of primary visual cortex. A prominent idea, stated most clearly by Barlow [1], is that each cell is a *detector* of a specific image feature. A more recent, and perhaps less exciting view, is that each cell has little meaning on its own, but that the ensemble of cells serves to *represent* the visual image. However, there are on the order of a billion cells in human primary visual cortex [2], and it seems odd to devote this number to a job done quite well by the two million or so ganglion cells of the retina, so this simple representational view must be tempered by the presumption that the particular representation employed by the cortex, or by a particular cortical area [3], has some functional advantage. In this paper, we describe some of the fundamental properties of the cortical representation of visual imagery, and present a particular mathematical image transform that also exhibits many of these properties. This mathematical formulation may help in understanding the construction and utility of the representation used by primary visual cortex.

As we move from retina to primary visual cortex, a fundamental change takes place in the nature of the representation of visual information. In the retina, each ganglion cell effectively represents a small region in the image. While the receptive field surround plays an important role in adaptive gain control, it is the narrow center, often receiving input from only one cone, that effectively samples the image. In frequency terms, each cell has a broad bandwidth that is essentially equal to that of the organism as a whole. In the cortex, things are strikingly different. Receptive fields are narrow-band and oriented, and may differ markedly from one another in their orientation, size, and peak frequency. This means that we have gone from a single representation by relatively homogeneous cells, to multiple representations by distinct populations of cells differing in orientation and spatial frequency.

Despite the presumable importance of these transformations, we lack any comprehensive notion of how they occur. Here we will describe a particular scheme for achieving this sort of transformation which exhibits many of the properties of the true cortical transform. First we review in more detail the properties of cortical neurons.

Aspects of Striate Cortex

After capture by the receptors of the eye, the visual image undergoes a sequence of transformations in the retina and visual brain. These transformations take place in distinct sets of cells. Certain sets lie in a serial arrangement, so that all the cells of one set receive their input from those of the other set, in which case we may speak of a *stream*. The pathway from retina to brain, and among the various brain areas, is made up of several such streams. In the primate, one important stream is the so-called *P stream* [3], which proceeds from the β retinal ganglion cells, to the parvocellular layers of the lateral geniculate nucleus, and from there primarily to layers $4c\beta$ and $4a$ of the striate cortex.

At each stage of this stream, recordings have been made which characterize the relation between the light image

Manuscript received April 21, 1988; revised July 20, 1988.

The authors are with Vision Group, NASA Ames Research Center, Moffett Field, CA 94035.

IEEE Log Number 8824421.

0018-9294/89/0100-0097$01.00 © 1989 IEEE

and the response of the cell, typically in the form of a receptive field. In the cortex, one observes many different varieties of spatial receptive field. This suggests a branching into many streams, each specialized for some particular analysis. However, there is one large population of cells with relatively homogeneous properties. These are the oriented simple cells of $V1$ [4].

The receptive fields of these cells have a number of interesting properties. First, they exhibit linear spatial summation [5]. They are local in space, in that the receptive field covers a small, compact region of the complete visual field [6]. Since each cell captures only a small region, we suspect that the complete set includes enough cells to capture the complete visual field. The receptive fields are also local in two-dimensional (2-D) spatial frequency [7], [8]. This means that the Fourier transform of the receptive field (the *spectral receptive field*) occupies a small, compact region of the complete *spectral visual field* (the visible portion of the 2-D frequency plane). This localization in 2-D frequency means that the cell responds to a small band of radial spatial frequencies and a small band of orientations. As in the case of spatial localization, it suggests that the complete set includes enough different types to cover the entire spectral visual field. Indeed, since each different type of spectral receptive field captures a different type of information, this suggests that there must be enough *of each type* to cover the spatial visual field.

The different types of spectral receptive fields may be divided along several dimensions. The first is orientation. Each cell responds vigorously only to a range of about 45° [9], which would suggest a minimum of four different orientations. Each cell also responds only within a limited frequency bandwidth. From cell to cell, this bandwidth is most nearly constant when expressed on a logarithmic scale, with a mean value of about 1.5 octaves [7]. This constancy indicates an approximate self-similarity of the various types of receptive field, in the sense that all receptive fields are approximately magnified and rotated versions of a single canonical form. This is not exactly true, as there are rather broad distributions of both orientation and octave bandwidths, but is a useful working approximation.

Some evidence also suggests that the simple cells exist as quadrature pairs, that is, with phases 90° apart [10]. One example of a quadrature pair are cells with even and odd symmetry about a central axis of the receptive field. Quadrature pairs have proven quite useful in modeling the behavior of direction selective cells [11], [12].

In our effort to understand these biological transformations of the visual image, we have made use of concepts from the mathematics of image coding. Some of these concepts are reviewed in the following section.

Image Codes

The discipline of image coding provides some useful examples of image transforms that resemble biological image transforms in various ways. In a digital image transform, the set of pixel values are converted into a set of coefficients. In the biological analog, the response of a single cell corresponds to a single coefficient. Likewise the receptive field is analogous to the kernel of weights that specify how pixel values are combined to form coefficients.

Tanimoto and Pavlidis [13] introduced the notion of an *image pyramid*. In this sort of transform, the image is filtered into several bands of resolution, and then each band is subsampled in proportion to resolution. At the bottom of the pyramid is a high-resolution image with many coefficients, and at the top is a low resolution image with few coefficients. While Tanimoto and Pavlidis used simple averaging of adjacent pixels to reduce resolution, Burt introduced a more "biological" Gaussian filter between each level of the pyramid [14], [15], and Watson [16] proposed ideal low-pass filtering between each level. All three of the preceding transforms are constructed in such a way that they are complete (are invertible and thus permit exact reconstruction of the image), but in each there are $4/3$ more coefficients than pixels. All three produce "receptive fields" that are approximately bandpass and self-similar. None, however, partition the image by orientation.

An oriented pyramid called the CORTEX transform was introduced by Watson [17]. The receptive fields have bandwidths of 1 octave in frequency and 45° in orientation. The transform is complete, but expands the image code by $16/3$. A later version provides almost exact reconstruction, and expands the code by only $4/3$ [18].

Another set of codes that partition by orientation are quadrature mirror filter (QMF) codes [19]–[22]. These codes adopt special constraints on the sampling functions (receptive fields) and reconstruction functions to ensure that sampling artifacts generated in one band are canceled by those in the others. They are complete and produce a code the same size as the image. However, they have the distinctly nonbiological feature of partitioning orientation into three bands, horizontals, verticals, and both diagonals (oblique right and oblique left). More recently, Adelson, Simoncelli, and Hingorani [23] have derived a QMF pyramid based on a hexagonal lattice which partitions orientations into three bands of 60°.

The Gabor transform, in which each receptive field is the product of a Gaussian and a sinusoid, is a popular candidate for biological models [24]–[26], in part because its receptive fields resemble those of the cortex [27], [28], [6]. However, in its exact version it does not lead to self-similarity. Self-similiar, pyramid-style Gabor transforms could presumably be derived.

Goal of the Present Work

Our goal here is an image transform that is both mathematically coherent and consistent with the properties of primary visual cortex. Specifically, the transform should have the following properties:

pyramid structure
complete (invertible)

basis functions that are:
local in 2-D space
local in 2-D frequency
self-similar
odd and even (quadrature pairing)
built from known physiological elements
hexagonal lattice
efficient.

We are interested in transforms that are efficient in the sense of having a small number of coefficients. We are also interested in codes that can be easily constructed from the known elements of the striate pathway. In particular, this means that we must build our cortical receptive fields from the receptive fields of retinal ganglion cells. Near to the fovea, ganglion cell receptive fields form an approximately hexagonal lattice. Therefore, we also seek a transform that operates on a hexagonal lattice.

Two Stages

The transform we shall describe is most conveniently thought of as occurring in two stages: the first implemented by the retinal ganglion cells, and the second, a transform of the "neural image" supplied by the retinal ganglion cell output. The second stage can be described as a straightforward digital image transform on a hexagonal sample raster, and it may have applications independent of its role as a biological model [29], [30]. For these reasons we shall derive it first, and later explore its utility as a biological model in conjunction with the ganglion cell transformation.

Hexagonal Orthogonal-Oriented Quadrature Pyramid

We consider transformations in which each new coefficient is a linear combination of input samples. The linear combination can be defined by a kernel of weights specifying the spatial topography of the linear combination. Here we consider kernels that occupy a point and the hexagon of six nearest neighbors on a hexagonal lattice.

We derive a set of kernels under the following constraints:

1) The kernels are expressed on a hexagonal sample lattice.

2) There are seven mutually orthogonal kernels, six high-pass and one low-pass.

3) Each kernel has seven weights corresponding to a point and its six nearest neighbors in the hexagonal lattice.

4) The low-pass kernel has seven equal weights.

5) Two high-pass kernels have an axis of symmetry running through the center sample and between samples on the outer ring (at an angle of 30°).

6) Of these two kernels, one is even about the axis of symmetry, the other is odd.

7) The remaining four high-pass kernels are obtained by rotating the odd and even kernels by 60 and 120°.

8) Each kernel has a norm (square root of sum of squares of weights) of one.

Some of these constraints are suggested by the biological results cited above. For example, 1), 3), 6), and 7) induce a hexagonal structure, spatial localization, quadrature pairs, and rotational self-similarity. Orthogonality 2) is adopted to make the transform easily invertible. Constraint 8) is attractive mathematically and is irrelevant with respect to the physiological comparison. With respect to constraint 5), we have determined that there is no solution when the common axis of symmetry is at 0° (on the sample lattice of the outer ring). Note also that constraints 2) and 4) oblige the even kernels, as well as the odd, to have zero dc response (the weights sum to 0), which is approximately true of primate simple cells. Under the symmetry constraints, the kernel coefficients can be written as shown in Fig. 1.

One even and one odd kernel are shown, along with the low-pass kernel which has seven equal weights h. The variables a–g express the symmetry constraints expressed above. Note that there are two additional even kernels that are not shown, produced by rotations of 60 and 120° of the even kernel. Likewise, there are two additional odd kernels, so that there are a total of seven distinct kernels, six high-pass and one low-pass.

The remaining constraints of orthogonality and unit norm are expressed in a set of eight equations in the eight unknowns a–h. They are

$$7h^2 = 1 \qquad \text{(low-pass unit norm)} \tag{1}$$

$$a^2 + 2b^2 + 2c^2 + 2d^2 = 1 \qquad \text{(even unit norm)} \tag{2}$$

$$2e^2 + 2f^2 + 2g^2 = 1 \qquad \text{(odd unit norm)} \tag{3}$$

$$a + 2b + 2c + 2\mathrm{d} = 0 \qquad (\text{even} \perp \text{to low-pass}) \tag{4}$$

$$a^2 + b^2 + d^2 + 2bc + 2cd = 0 \qquad (\text{even} \perp \text{to } 30° \text{ self-rotation}) \tag{5}$$

$$a^2 + 2bc + 2bd + 2cd = 0 \qquad (\text{even} \perp \text{to } 60° \text{ self-rotation}) \tag{6}$$

$$e^2 + g^2 - 2ef - 2fg = 0 \qquad (\text{odd} \perp \text{to } 60° \text{ self-rotation}) \tag{7}$$

$$2eg - 2ef - 2fg = 0 \qquad (\text{odd} \perp \text{to } 120° \text{ self-rotation}). \tag{8}$$

The value of each coefficient h in the low-pass kernel is given directly by the unit norm constraint (1)

$$h = 1/\sqrt{7}. \tag{9}$$

Subtracting (5) and (6), and (7) and (8), shows that

$$b = d \tag{10}$$

$$e = g. \tag{11}$$

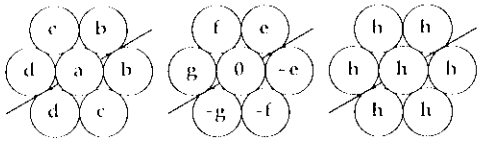


Fig. 1. Even and odd high-pass kernels, and lowpass kernel. The oblique line indicates the assumed symmetry axis at 30°.

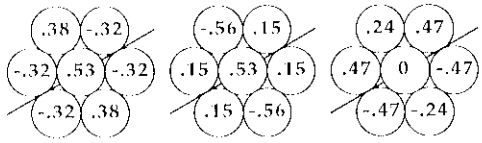


Fig. 2. Values for the two types of even kernel and the odd kernel.

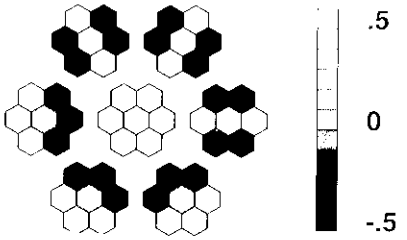


Fig. 3. Seven kernels represented by graylevels. The low-pass kernel is at the center, the three even kernels are in the upper right, and the three odd are to the lower left.

Thus, while not explicitly assumed, we see that both odd and even filters must also be symmetrical about the 120° axis.

Further simplifications lead to the following solution for the coefficients of the odd filter:

$$f = -\frac{1}{3\sqrt{2}} \tag{12}$$

$$e = 2f = -\sqrt{2}/3. \tag{13}$$

For the even filter, we find

$$a = \sqrt{2}h = \sqrt{2/7}. \tag{14}$$

But two solutions emerge for b and c:

$$b = -(1 + h)f = \frac{-(1 + 1/\sqrt{7})}{3\sqrt{2}} \tag{15}$$

$$c = (2 - h)f = \frac{(2 - 1/\sqrt{7})}{3\sqrt{2}} \tag{16}$$

and

$$b = (1 - h)f = \frac{(1 - 1/\sqrt{7})}{3\sqrt{2}} \tag{17}$$

$$c = -(2 + h)f = \frac{-(2 + 1/\sqrt{7})}{3\sqrt{2}}. \tag{18}$$

We will call the first solution the even filter of type 0, and the second solution, type 1. The three high-pass kernels are shown in Fig. 2. A complete set of kernels (using even type 0) is shown in Fig. 3. The high-pass kernels are arranged about the low-pass kernel at the center.

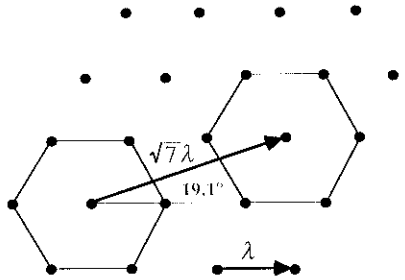


Fig. 4. Tiling the image with hexagonal neighborhoods. Two neighborhoods are shown. This tiling scheme produces a new hexagonal sample lattice that is $\sqrt{7}$ larger, and rotated by 19.1° from the original sample lattice.

Subsampling

Application of the seven kernels to a neighborhood of seven pixels will yield seven coefficients. Since the kernels are orthogonal, the seven coefficients are a complete representation of the seven pixel image neighborhood. By tiling the plane with hexagonal neighborhoods, a complete image can be transformed. The coefficients produced by each distinct type of kernel may be regarded as a subimage, and each of the seven subimages may also be regarded as a filtered and subsampled version of the original image. Since each kernel consumes seven pixels and yields one coefficient, each subimage has one seventh the number of pixels in the original image. The manner in which the image may be tiled and subsampled is illustrated in Fig. 4.

The Pyramid

One virtue of the scheme we have described is that it leads directly to a pyramid structure. The hexagonal image sample lattice is tiled with hexagons with unit sides. Each of the seven kernels is applied in each hexagon, yielding seven subimages, six high-pass and one low-pass, each with one seventh as many samples as the original. The six high-pass subimages form level 0 of the pyramid. The next level is created by applying the seven kernels to the low-pass subimage. This yields seven new subimages, six high-pass and one low-pass, each a factor of seven smaller than the subimages at level 0. This process is repeated until a level is reached at which each subimage has one sample.

This recursive process is illustrated in Fig. 5. The vertices and centers of the smallest hexagons define the input sample lattice. These smallest hexagons show a tiling of the image by the level 0 kernels. Their centers locate the samples of the level 0 subimages. The next larger hexagons tile the level 0 subimages, and their centers define the sample locations for the level 1 subimages. Their vertices (and centers) show where the weights are applied to recursively transform each low-pass subimage. Higher levels are represented by still larger hexagons.

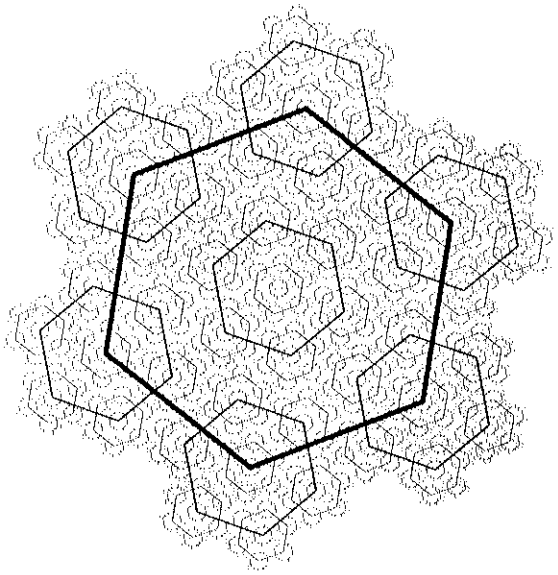


Fig. 5. Construction of the hexagonal pyramid. The image sample lattice is given by the vertices and centers of the smallest hexagons. The hexagons indicate the pooling neighborhoods for the next level. This hexagonal fractal was constructed by first creating the largest hexagon, then placing at each of its vertices a hexagon rotated by $\tan^{-1}(\sqrt{3}/5) \approx 19.1°$ and scaled by $1/\sqrt{7}$. The same procedure is then applied to each of the smaller hexagons, down to some terminating level. The image sample lattice is then a finite-extent periodic sequence with a hexagonal sample lattice defined by the vertices of the smallest hexagons. The sample lattice has 7^6 points, the same as a rectangular lattice of 343^2. The perimeter of this "Gosper flake" is a "Koch curve" with a fractal dimension of $\log 3/\log \sqrt{7} \approx 1.19$ [31]. The program used to create this image is given in Appendix 1.

While an image shape like that in Fig. 5 is very natural for this code, any shape that is one period of a hexagonally periodic sequence can be exactly encoded if the number of samples is equal to a power of seven. This includes, for example, a parallelogram with sides whose length in samples is a power of seven. Below we show how the code may be applied to a conventional rectangular image.

Pyramid Kernels

While we have described the construction of the levels of the pyramid as a recursive transformation of the low-pass subimage, we can also view each subimage as the result of direct application of a kernel to the image at appropriate sample points. As we move higher in the pyramid, the kernels are constructed from the low-pass kernels of the previous level. This yields the kernels shown in Fig. 6.

Kernel Spectra

One of our objectives was to create receptive fields that were local in frequency, that is, somewhat narrow-band and oriented. As continuous functions, the kernel spectra are easily derived. Each kernel consists of a central impulse at the origin, surrounded by three pairs of symmetric (or antisymmetric) impulses at angles of 0, 60, and 120°. Each pair of symmetric impulses transforms into a sinusoid oriented at the angle of the impulse pair, while the impulse at the center transforms into a constant. The complete transform is thus a constant plus three sinusoids at angles of 0, 60, and 120°. The constant is the value of the central coefficient, while each sinusoid has an amplitude twice that of the corresponding coefficient. For the even kernels, the sinusoids are in cosine phase; for the odd kernels, they are in sine phase.

To picture the spectra for the discrete, finite extent case we show the discrete Fourier transform (DFT) of the kernels in Fig. 6. This was done by computing each kernel in a square image, regarded as skewed coordinates of a hexagonal raster, computing the DFT of this image, and appropriately deskewing the DFT. As shown in Fig. 7, the spectra computed in this way are oriented and band-pass, as desired.

Axes of Symmetry and Orientation

We can define the *orientation* of a kernel as the orientation of the peak of the frequency spectrum, that is, the orientation of a sinusoidal input at which the kernel gives the largest response. An interesting feature of the resulting kernels is that while the axis of symmetry was fixed at 30°, the orientation of the type 0 even kernel (shown in Fig. 7) is actually orthogonal to this axis at 120°. This places its orientation axis on the hexagonal lattice. In contrast, the orientation of the type 1 even kernel and the odd kernel are equal to the initial axis of symmetry at 30°. Thus, if it is desired to have quadrature pairs with equal orientation, the type 1 even kernel must be used.

Sampling Matrices

The subsampling at each level can be formalized as follows (see [32] for a general discussion of nonrectangular sampling). The original hexagonal sampling lattice can be represented by a sampling matrix $\boldsymbol{H}$:

$$\boldsymbol{H} = \begin{vmatrix} 1 & 1/2 \\ 0 & \sqrt{3/2} \end{vmatrix}. \tag{19}$$

The column vectors of this matrix map from sample to sample, and the location of any sample can be expressed as $\boldsymbol{x} = (x, y)$,

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{r} \tag{20}$$

where $\boldsymbol{r}$ is an integer vector. Let $\boldsymbol{S}_n$ be the sampling matrix at level $\boldsymbol{n}$. Since the samples at each level must be a subset of those at the previous level, the column vectors of $\boldsymbol{S}_{n+1}$ must be integer linear combinations of the column vectors of $\boldsymbol{S}_n$. Thus,

$$\boldsymbol{S}_{n+1} = \boldsymbol{S}_n\boldsymbol{M} \tag{21}$$

where $\boldsymbol{M}$ is an integer matrix. Furthermore, the columns of $\boldsymbol{S}_{n+1}$ must be $\sqrt{7}$ longer than the columns of $\boldsymbol{S}_n$ (corresponding to the increasing radii of the hexagons at each successive level). And finally, because the determinant of a sampling matrix determines the factor by which the density of samples is reduced, we know that

$$\det(\boldsymbol{M}) = 7. \tag{22}$$

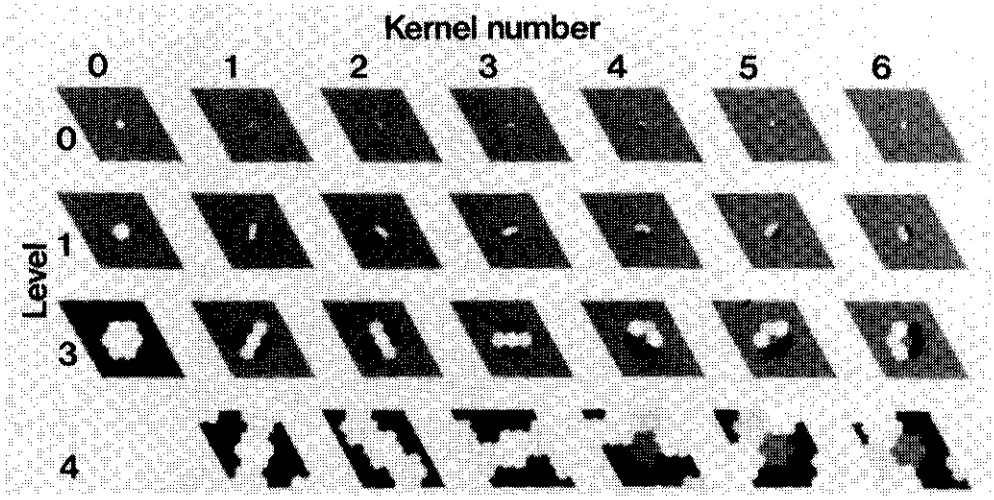


Fig. 6. Pyramid kernels. The seven kernels are numbered 0 (low-pass), 1–3 (even high-pass) and 4–6 (odd high-pass). For this figure, the kernels were first computed on a 49 × 49 pixel square, regarded as skewed coordinates of a hexagonal lattice. The square was then deskewed into a parallelogram, which is an alternate tiling of a hexagonally periodic image (see below). Even kernels are of type 0. Bright regions are positive, dark regions are negative, and gray corresponds to 0. The central kernel is shown for each type and level. At level 4 there is only one kernel of each type and it wraps around to fill the entire image.

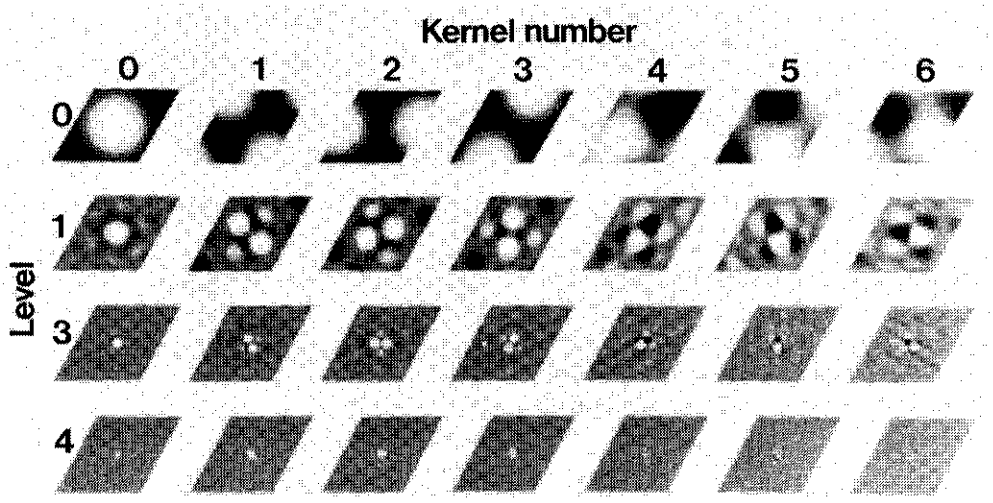


Fig. 7. Spectra of pyramid kernels. The origin is at the center of each figure. Bright regions are positive, dark regions are negative, and gray corresponds to 0.

Two matrices which satisfy these conditions are

$$M_0 = \begin{vmatrix} 2 & -1 \\ 1 & 3 \end{vmatrix} \tag{23}$$

$$M_1 = \begin{vmatrix} 1 & -2 \\ 2 & 3 \end{vmatrix}. \tag{24}$$

These generate the only two possible subsamplings from one level to the next. Then S_n can be constructed in various ways, the three most obvious being

$$S_n = HM_0^n \tag{25}$$

and

$$S_n = HM_1^n \tag{26}$$

and

$$S_n = HM_0M_1M_0M_1 \cdots (n \text{ terms}). \tag{27}$$

The first scheme (used in Fig. 5) causes a rotation of $\tan^{-1}(\sqrt{3}/5) \approx 19.1°$ in the sample lattice at each level, as does the second scheme, while the third scheme alternates between rotations of 19.1 and $-19.1°$.

Skewed Coordinates

It is well known that hexagonal samples on a Cartesian plane can also be viewed as rectangular coordinates on a coordinate frame in which one axis is skewed by 60° [Fig. 8(a)] [33], [34]. In this coordinate scheme, the sampling matrices are even simpler. They are the same as above

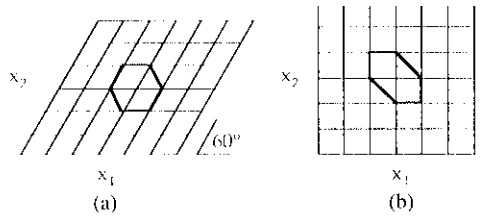


Fig. 8. (a) Hexagonal lattice represented as skewed rectangular coordinates. (b) Deskewed rectangular coordinates. The hexagon is distorted into an oblique lozenge.

[(25)–(27)] except that we drop the matrix $\boldsymbol{H}$ from each expression.

This leads to a natural method for application of this coding scheme to conventional rectangular images. When the skewed coordinates are "deskewed" [Fig. 8(b)], the hexagon is distorted into an oblique lozenge. The orthogonal pyramid may then be constructed using these lozenges as the shape for each kernel. The kernels will no longer be rotationally symmetric, but for some purposes this may be unimportant. As before, exact coding will be possible so long as the sides of the rectangle are a power of seven.

Reconstruction of an image from the transform coefficients is straightforward and is the inverse of the operations used to create the transform. First, the highest level of the pyramid (level n) is inverse transformed. Each coefficient is replaced by a set of seven pixels, produced by multiplying the coefficient and the corresponding kernel. The subimages created in this way from each of the seven subimages are added together to form a low-pass subimage at level $n - 1$. This process is then repeated for each level until the image pixels are produced.

The left side of Fig. 9 shows an original image of $7^3 \times 7^3 = 343^2$ pixels, with zero-order entropy of 7.45 bits/pixel. This image has been transformed by the preceding scheme. The coefficients of each transform subimage have been quantized using a nonlinear quantizer that exploits the masking property of human vision [18]. The severity of quantization decreases with level in the pyramid (high-spatial frequencies are more severely quantized). The image is then reconstructed from the quantized coefficients as described above. The right side of Fig. 8 shows the result when quantization is set so as to yield a coefficient entropy of 0.96 bits/pixel.

A detailed discussion of the coding efficiency and computational complexity of this transform is beyond the scope of this paper, but we note that the recursive computation of the transform is about $6N/7$ faster than direct application of the equivalent kernels where N is the number of layers, and the number of image pixels is 7^N.

Biological Image Coding

As noted earlier, the transformation from image to cortical representation may be partitioned into two stages: the first between image and retinal ganglion cell response array, and the second between this retinal neural image and the cortical neural image. In the preceding sections we developed the second stage, here we add the first stage.

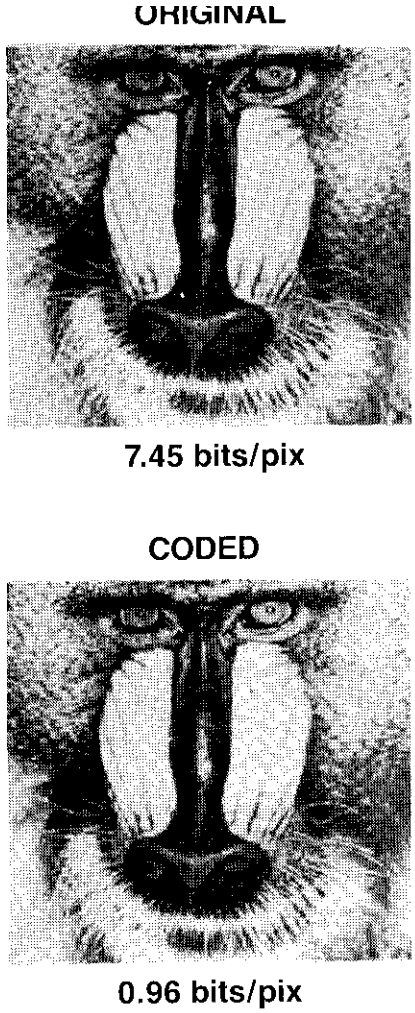


Fig. 9. Original image and image coded using the hexagonal orthogonal oriented quadrature pyramid.

Retinal ganglion cell receptive fields have a center-surround arrangement, and are described well by a difference-of-Gaussians (DOG) [35]. Near to the fovea, the center is evidently driven by a single cone. Since the cones are in an approximately hexagonal lattice, this means that ganglion cell receptive fields also form a hexagonal lattice. Each retinal ganglion cell may be represented by a DOG function with a center radius of w. The receptive fields are arranged in a hexagonal lattice with sample spacing of λ. Here we assume $w = \lambda$ (Fig. 10), which is approximately true of retinal cells [36]. The output of this hexagonal ganglion cell lattice provides the input to the orthogonal transform described above. The coefficients of the resulting transform then represent the responses of hypothetical cortical cells.

The receptive fields of the complete transform are linear combinations of the retinal DOG receptive fields. At level 0, they are produced by simply weighting each of the seven DOG's in the minimal neighborhood by the weights in the appropriate kernel. An example receptive field is shown in Fig. 11. It is a patch of elongated parallel antagonistic regions, resembling the receptive fields of striate simple cells.

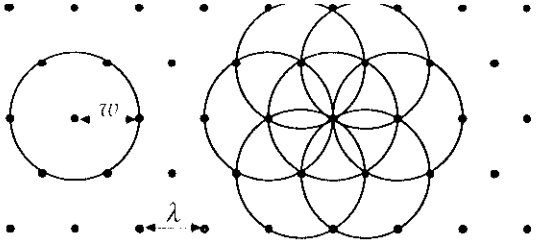


Fig. 10. Hexagonal lattice with spacing λ. Each circle represents the center Gaussian of a retinal ganglion cell receptive field. On the right, a neighborhood of seven cells is shown.

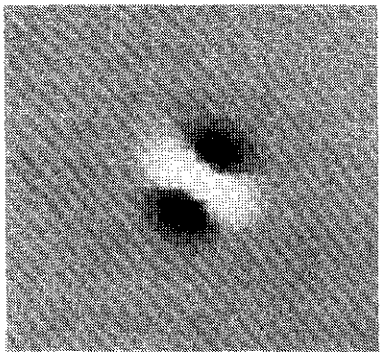


Fig. 11. Cortical receptive field produced by linear combination of seven retinal ganglion cell receptive fields.

In general, the receptive field at any level is produced by convolving the corresponding kernel (Fig. 6) with the DOG function. In the frequency domain, this corresponds to multiplying each kernel spectrum (Fig. 7) by the DOG spectrum. Examples of receptive fields and spectra are shown in Fig. 12. As we move up in the pyramid, the spatial receptive field becomes larger and the spectral receptive field smaller. Each spectral receptive field approximates a pair of blobs on either side of the origin, in agreement with the spectral receptive fields of oriented cortical simple cells. The orientation bandwidth of each receptive field is 60°, and the frequency bandwidth is about 1.5 octaves, in rough agreement with biological values.

Discussion

At this point we review our progress in attaining the goals set out at the start of this paper. We have derived an image transform which produces coefficients that are compact and local in 2-D space and in 2-D spatial frequency, with approximately correct frequency and orientation bandwidths. The receptive fields come in various sizes and orientations, and have approximate self-similarity (the self-similarity is exact in the second stage transform). The receptive fields have both odd and even forms. Because the second stage transform is orthogonal, the coefficients are a complete representation of the image provided by the retinal ganglion cells. The representation is efficient in the sense that the number of coefficients is at its theoretical minimum. And the coefficients form a pyramid structure. We have therefore satisfied all of our goals, but we conclude by noting certain defects of this scheme as a model of cortical image coding.

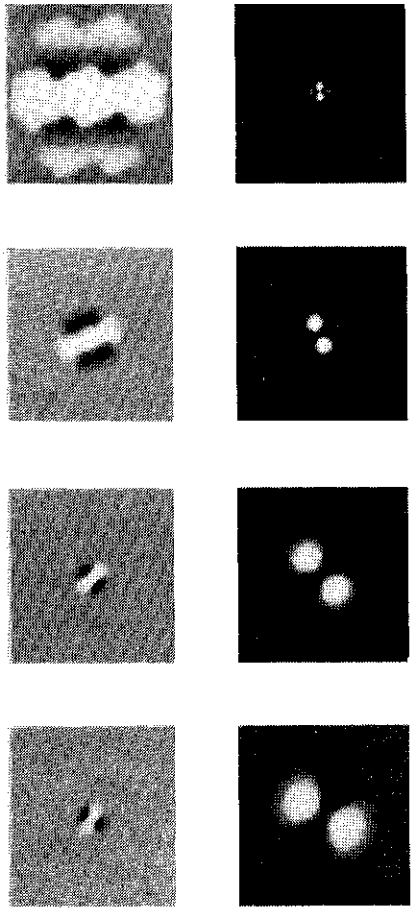


Fig. 12. Spatial and spectral receptive fields of even type 0 cells at four levels of the pyramid.

First, the frequency tuning functions of our filters are oriented in the sense of having a strongest response at one orientation, but they have a second, smaller lobe of response (of opposite sign) at the orthogonal orientation. Two-dimensional frequency tuning functions in cortical cells occasionally show such secondary lobes [9], but they do not appear to be common. Second, the units we describe change in size by $\sqrt{7}$ at each level, which might yield rather fewer different scales than are commonly supposed. Third, the 19.1° rotation of the axis of orientation at each scale reduces the degree of rotation invariance of the code, although rotational invariance is not known to hold for the cortical code. Fourth, the tuning functions produced by our scheme are broader in orientation than in spatial frequency, while cortical cells appear to have narrower tuning in orientation than in frequency [37], [6]. Fifth, tangential penetrations of striate cortex sometimes produce long sequences of cells whose orientations increase monotonically in steps considerably smaller than the 60° suggested by the present scheme [38]. Many of the apparent defects of the present scheme may be remedied by relatively minor modifications, such as larger and nonorthogonal kernels [30], which we hope to explore in the future.

Finally, the precise crystalline structure of this code is clearly different from the biological heterogeneity of vi-

visual cortex. Nonetheless, the cortex is highly regular, and a scheme like ours may be the canonical form from which the actual cortex is a developmental perturbation. These issues are discussed at greater length elsewhere [30]. Perhaps the best summary is that while this scheme may not describe exactly the cortical encoding architecture, it is an example of the form such an architecture might take. It is a means of expressing our understanding of the system, and bring into focus those aspects of the system about which we are ignorant. The particular structure we have proposed greatly simplifies the computation of the code and its inversion, providing an efficient method of image compression, progressive transmission, and a cortex-like processor for artificial vision systems.

Appendix

The following is a program in the Postscript language to draw the pyramid in Fig. 5. The number of levels drawn is determined by the variable *maxdepth*. On an Apple laser printer, a *maxdepth* of three takes about 2 min to print. Each greater depth will take a factor of seven longer.

```
/depth 0 def
/maxdepth 3 def                                    % maximum levels
/latticeRot 3 sqrt 5 atan def                      % lattice rotation angle
/root7 1 7 sqrt div def                            % scale change between levels
/negrot {/latticeRot latticeRot neg def} def
/down {/depth depth 1 add def} def                 % increments depth
/up {/depth depth 1 sub def} def                   % decrements depth
/inch {72 mul} def                                 % scale to inches

/hexside {60 rotate 1 0 lineto currentpoint translate }   % draw one side of a hexagon
   def

/drawhex                                           % draw unit hexagon
{gsave
  60 rotate 1 0 moveto 60 rotate currentpoint translate    % move to first vertex
5 { hexside } repeat                               % draw 5 sides
closepath stroke                                   % draw sixth side
grestore } def

/vertex     % angle is on stack     % go to vertex at angle, draw hexagon pyramid
{/angle exch def
gsave
angle rotate 1 0 translate angle neg rotate
fracthex
grestore
} def
/fracthex                                          % draw hexagon pyramid
{gsave
root7 dup scale                                    % reduce scale by root 7
2 72 div setlinewidth
down negrot latticeRot rotate drawhex              % move down one level, rotate
                                                   lattice, draw hex
depth maxdepth le                                  % test if at max level
    {fracthex                                      % recursive call to fracthex
    0 60 300 { vertex } for                        % call vertex at each vertex
    } if
up negrot grestore
} def

gsave                                              % main program
4.25 inch 5.5 inch moveto currentpoint translate   % set origin
6 inch 6 inch scale                                % set global scale
latticeRot neg rotate                              % set initial orientation
1 setlinejoin
fracthex                                           % do it
grestore
1 inch 1 inch moveto
/Palatino-Roman findfont 34 scalefont setfont      % label
(Hexagon Pyramid) show
showpage
```

References

[1] H. B. Barlow, "Single units and sensation: A neurone doctrine for perceptual psychology?" *Perception*, vol. 1, pp. 371–394, 1972.

[2] M. Colonnier and J. O'Kusky, "Le nombre de neurones et de synapses dans le cortex visuel de differentes especes," *Rev. Can Biol.*, vol. 40, pp. 91–99, 1981.

[3] E. A. De Yoe and D. C. Van Essen, "Concurrent processing streams in monkey visual cortex," *Trends Neurosci.*, vol. 11, pp. 219–226, 1988.

[4] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol. (London)*, vol. 195, pp. 215–243, 1968.

[5] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst, "Spatial summation in the receptive fields of simple cells in the cat's striate cortex," *J. Physiol.*, vol. 283, pp. 53–77, 1978.

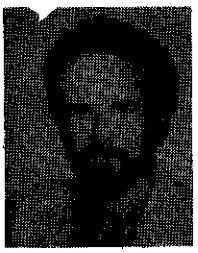
[6] J. P. Jones and L. A. Palmer, "The two-dimensional spatial structure of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, pp. 1187–1211, 1987.

[7] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Res.*, vol. 22, pp. 545–559, 1982.

[8] J. P. Jones, A. Stepnoski, and L. A. Palmer, "The two-dimensional spectral structure of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, pp. 1212–1232, 1987.

[9] R. L. De Valois, E. W. Yund, and H. Hepler, "The orientation and

direction selectivity of cells in macaque visual cortex," *Vision Res.*, vol. 22, pp. 531–544, 1982.
[10] D. A. Pollen and S. F. Ronner, "Phase relationship between adjacent simple cells in the visual cortex," *Science*, vol. 212, pp. 1409–1411, 1981.
[11] A. B. Watson and A. J. Ahumada, Jr., "A look at motion in the frequency domain," NASA Tech. Memorandum 84352, 1983.
[12] ——, "Model of human visual-motion sensing," *J. Opt. Soc. Amer.*, vol. 2, pp. 322–342, 1985.
[13] S. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing," *Comput. Graph. Image Proc.*, vol. 4, pp. 104–119, 1975.
[14] P. J. Burt, "Fast filter transforms for image processing," *Comput. Graph. Image Proc.*, vol. 16, pp. 20–51, 1981.
[15] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532–540, Apr. 1983.
[16] A. B. Watson, "Ideal shrinking and expansion of discrete sequences," NASA Tech. Memorandum 88202, Jan. 1986.
[17] ——, "The cortex transform: Rapid computation of simulated neural images," *Comput. Vision, Graph., Image Proc.*, vol. 39, pp. 311–327, 1987.
[18] ——, "Efficiency of an image code based on human vision," *J. Opt. Soc. Amer.*, vol. 4, pp. 2401–2417, 1987.
[19] M. Vetterli, "Multidimensional sub-band coding: Some theory and algorithms," *Signal Proc.*, vol. 6, pp. 97–112, 1984.
[20] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-34, pp. 1278–1288, 1986.
[21] H. Gharavi and A. Tabatabai, "Sub-band coding of digital images using two-dimensional quadrature mirror filtering," *SPIE Proc. Visual Commun. Image Proc.*, vol. 707, pp. 51–61, 1986.
[22] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," GRASP Lab Tech. Memo MS-CIS-87-22, Dep. Comput. Inform. Sci., Univ. Penn., 1987.
[23] E. H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transforms for image coding," *Proc. SPIE, Visual Commun. Image Proc. II*, 1988.
[24] A. B. Watson, "Detection and recognition of simple spatial forms, *Physical and Biological Processing of Images*, O. J. Braddick and A. C. Sleigh, Eds. Berlin: Springer-Verlag, 1983.
[25] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortex filters," *J. Opt. Soc. Amer.*, vol. 2, pp. 1160–1169, 1985.
[26] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, pp. 2379–2394, 1987.
[27] S. Marcelia, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, pp. 1297–1300, 1980.
[28] D. J. Field and D. J. Tolhurst, "The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex," *Proc. Roy. Soc. London*, vol. 228, pp. 379–400, 1986.
[29] A. B. Watson and A. J. Ahumada, Jr., "An orthogonal oriented quadrature hexagonal image pyramid," NASA Tech. Memorandum 100054, 1987.
[30] A. B. Watson, "Cortical algotecture," in *Vision: Coding and Efficiency*, C. B. Blakemore, Ed. Cambridge, England: Cambridge Univ., 1988.
[31] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman, 1983.
[32] D. A. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[33] D. P. Petersen and D. Middleton, "Sampling and reconstruction of wave-number limited functions in *N* dimensional Euclidean spaces," *Inform. Contr.*, vol. 5, pp. 279–323, 1962.
[34] R. M. Mersereau, "The processing of hexagonally sampled two-dimensional signals," *Proc. IEEE*, vol. 67, pp. 930–949, 1979.
[35] C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," *J. Physiol.*, vol. 187, pp. 517–552, 1966.
[36] V. H. Perry and A. Cowey, "The ganglion cell and cone distributions in the monkey's retina: Implications for central magnification factors," *Vision Res.*, vol. 25, pp. 1795–1810, 1985.
[37] M. A. Webster and R. L. De Valois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," *J. Opt. Soc. Amer.*, vol. 2, pp. 1124–1132, 1985.
[38] D. H. Hubel and T. N. Wiesel, "Sequence regularity and geometry of orientation columns in the monkey striate cortex," *J. Comp. Neurol.*, vol. 158, pp. 267–294, 1974.



Andrew B. Watson did undergraduate work at Columbia University, New York, and received the Ph.D. degree in psychology from the University of Pennsylvania, Philadelphia, in 1977.

He spent three years in England at the Craik Laboratory of Cambridge University, Cambridge, followed by two years at the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA. Since 1982 he has worked at NASA Ames Research Center where he is Principal Scientist of the Human Interface Research Branch. His current interest is in computational models of visual perception and their application to image coding, understanding, and display.

Dr. Watson is a member of the Association for Research in Vision and Ophthalmology, the Optical Society of America, and the American Association for Artificial Intelligence. He is also a member of the Program Committee of the Society for Information Display and the Committee on Vision of the National Research Council.

Albert J. Ahumada, Jr. received the B.S. degree in mathematics from Stanford University, Stanford, CA, in 1961 and the Ph.D. degree in psychology from the University of California, Los Angeles, in 1967.

He did research in psychoacoustics while teaching at the University of California, Irvine, and was a Research Associate in the Department of Aeronautics and Astronautics, Stanford University. Since 1980 he has been doing vision research at NASA Ames Research Center on problems of image sampling and image encoding.