Efficiency of a model human image code

Andrew B. Watson

Perception and Cognition Group, MS 239-3, NASA Ames Research Center, Moffett Field, California 94035

Received May 26, 1987; accepted August 13, 1987

Hypothetical schemes for neural representation of visual information can be expressed as explicit image codes. We may test whether a given code is sufficient, in the sense of retaining all the information that the human perceives, and necessary, in the sense of retaining only that information. The latter is a test of efficiency. Here, we explore a code modeled on the simple cells of the primate striate cortex. The Cortex transform maps a digital image into a set of subimages (layers) that are bandpass in spatial frequency and orientation. The layers are sampled so as to minimize the number of samples and still avoid aliasing. Samples are quantized in a manner that exploits the bandpass contrast-masking properties of human vision. The entropy of the samples is computed to provide a lower bound on the code size. Finally, the image is reconstructed from the code. We devise psychophysical methods for comparing the original and reconstructed images to evaluate the sufficiency of the code. When each resolution is coded at the threshold for detection artifacts, the image-code size is about 1 bit/pixel.

1. INTRODUCTION

The purpose of this research is to explore the efficiency of an image code based on a simple model of spatial processing in the early stages of human vision. There are two reasons for pursuing this issue. First, progress in image compression has been hampered by the lack of an explicit model for the visibility of coding artifacts. As will become evident, the compression scheme developed here is a direct outgrowth of a model of visibility and may thus offer some practical improvement on existing image-coding schemes.

Second, explicit image codes provide a language in which to express conjectures about how visual information is represented in the brain. One test of a hypothetical code is that it must retain all the information to which the human observer is sensitive. (Or, if the code simulates a distinct module of vision, then the code must retain all the information to which the observer is sensitive in the domain of that module.) This test is accomplished be comparing the original image and an image reconstructed from the code. This tests whether the code is sufficient.

Furthermore, given two codes that produce perceptually perfect reconstructions, the one of larger size contains information that is not retained by human vision. Hence among all perceptually lossless codes, the one of smallest size (highest efficiency) is the one closest to containing only the information contained in the true visual code. This is a test of the necessity of the code.

2. IMAGE CODING AND COMPRESSION

The scheme described here can be divided into the steps shown in Fig. 1. The image is transformed into a set of coefficients, and the coefficients are quantized to a discrete number of levels. A reconstruction process then transforms the quantized coefficients back into an image. A lossless coding scheme can be devised to convert the quantized coefficients into a code of size approximately equal to the first-order entropy of the quantized coefficients. Thus, although we do not implement this step, we express code size as this first-order entropy.

3. CORTEX TRANSFORM

The first step in our coding process is a linear transformation of the image into a set of coefficients. In general, a transform of this sort assumes a particular set of component images, or features, and discovers how much of each component is present in the image. Here we use the Cortex transform,[1] in which the components are small patches of oriented sinusoid. A brief review of the Cortex transform is given below.

A. Objective

The area of the visual brain known as V1 receives essentially all the visual projection to the cortex.[2] Approximately one half of the neurons in V1 are so-called simple cells, which sum light approximately linearly over their receptive field and are typically composed of parallel elongated excitatory and inhibitory regions. In the two-dimensional spatial-frequency domain, these cells have passbands that are localized and much smaller than the passband of the observer as a whole.[3] Median frequency and orientation bandwidths are about 1.4 octaves[4,5] and 40 deg.[6] Because of their evident importance in visual processing, these cells have formed the basis of a number of models of early spatial vision.[7-9] The Cortex transform was designed as a means of rapidly computing the responses of arrays of model cells to an image.

B. Implementation

In the Cortex transform, all cells are divided into types according to spatial frequency and orientation. Each type is associated with a particular receptive field shape, or impulse response. The receptive field may also be characterized by the discrete Fourier transform (DFT) of the impulse response (a filter). The cells of one type form an array sampling the visual field, which we call a layer. A given layer may be computed by discrete circular convolution of the image and the given impulse response or, equivalently, by multiplying the image DFT and the filter and computing the inverse DFT of the result.

Each cortex filter has a passband located in a particular region of the two-dimensional spatial-frequency plane, with

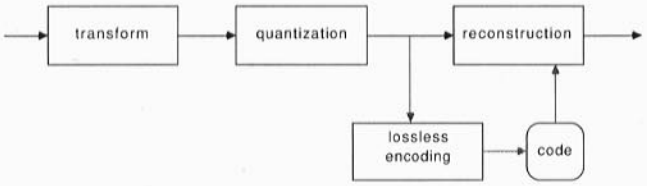

Fig. 1. Stages in the compression and reconstruction of an image.

a radial-frequency bandwidth of 1 octave and an orientation bandwidth of 45 deg. The borders of each filter are Gaussian. The filters are designed to be rotated and magnified versions of one another and to subdivide the Fourier space exhaustively. The complete set of filters is shown in Fig. 2. Figure 3 shows an impulse response obtained as the inverse DFT of one filter. This represents the receptive-field shape of the cell type corresponding to the filter. Details of filter construction are given elsewhere.[1,10]

The layer is given by the inverse DFT of the product of the filter DFT and the image DFT. It is a bandpass image, each pixel of which corresponds to the reponse of a cell whose receptive field is centered on the pixel. Within a layer, pixels correspond to cells tuned to the same spatial frequency and orientation. The various layers that result from one image[11] are shown in Fig. 4.

As the radial frequency (resolution) of the filter declines, the frequency support of the filter becomes smaller. Multiplication and the inverse DFT are applied only over this smaller support. Thus the size of each layer is inversely proportional to the resolution. This is equivalent to (but more efficient than) multiplication and inverse DFT over the entire DFT, followed by subsampling at an appropriate rate.[1,12] This decline in layer size with resolution results in a pyramid structure, similar to those devised by Watson,[12] Tanimoto and Pavlidis,[13] Burt and Adelson,[14] and Crowley and Stern,[15] except that we have a separate pyramid for each orientation.

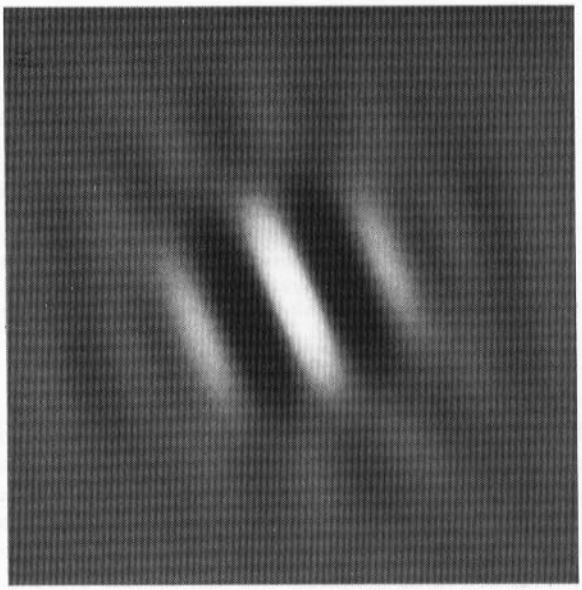

Fig. 3. Impulse response of the cortex filter at one frequency and orientation.

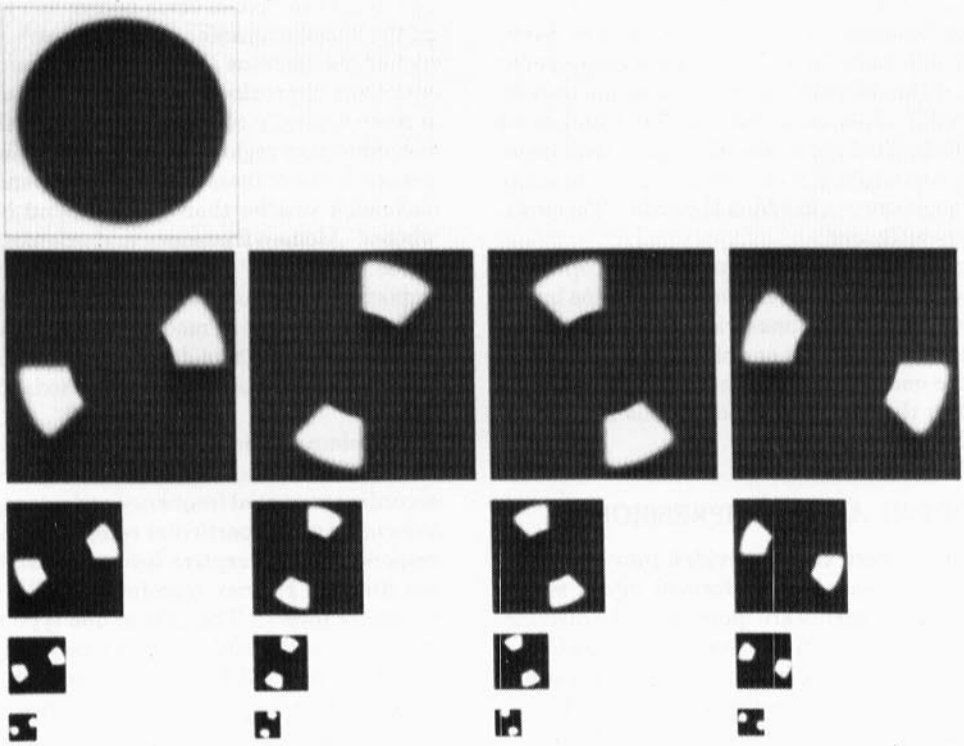

Fig. 2. Frequency spectra of the cortex filters. Filters differ in resolution down each column and in orientation across each row. The high-residue filter is shown in the upper left, and the low-residue filter is in the lower left. For each filter, the frequency origin is at the center. All filters are purely real.

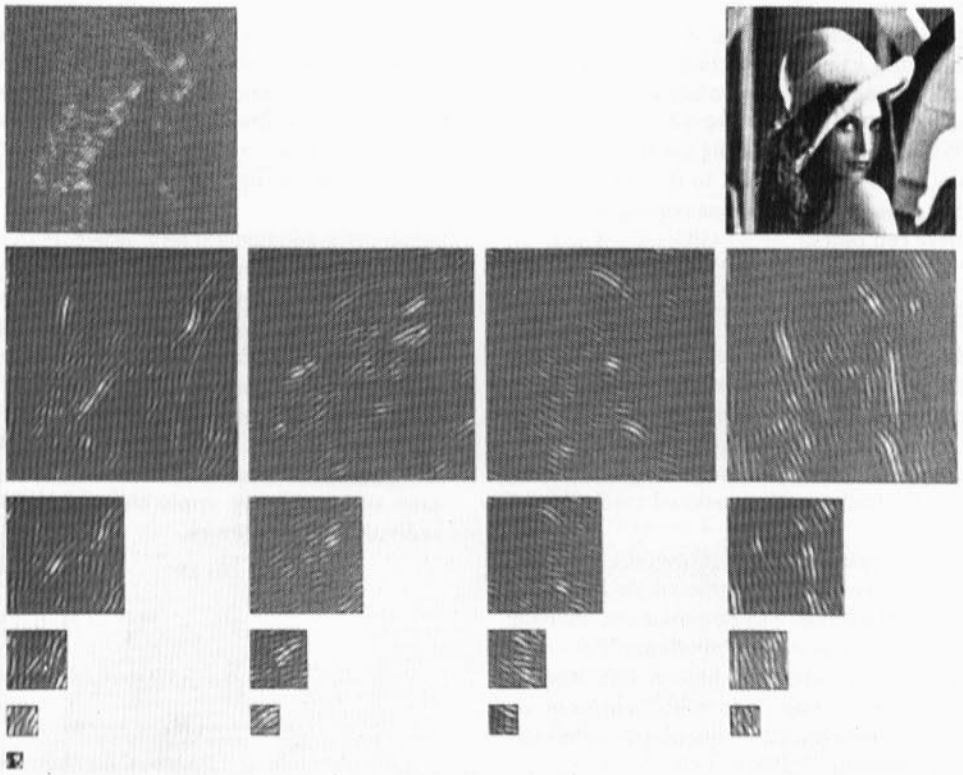

Fig. 4. Cortex transform layers for the image of the woman. The original is at the upper right. The high and low residues are at the upper left and lower left, respectively. The layers vary in orientation across a row and in resolution down a column. All layers are shown at maximum contrast.

The transform includes two special layers, the high and low residues. They are what is left over at the high and low frequencies after the various bandpass layers have been subtracted from the image. The low residue has few pixels but conveys potentially important information (such as the mean intensity of the image) and is therefore retained. The high residue, on the other hand, often has little information and can be discarded without altering the appearance of the image. In what follows we generally neglect the high residue.

Since the highest-resolution layers are equal in size to the original image, and since each lower-resolution layer is smaller by a factor of 4, the total number of coefficients in the transform (excluding the high residue) is $4N(1 + 1/4 + 1/16 + \ldots) \approx 16N/3$, where N is the number of pixels in the original image.

In what follows, we will index filters and layers by an orientation number $O = 0, 1, 2, 3$ (proceeding counterclockwise from the band at 0–45 deg) and a resolution number R, equal to the $\log_2$ of the width of the layer (or the filter DFT). Thus, for an image of 256×256 pixels, the highest-resolution layer is 8. The filter of resolution R and orientation O is written $f_{R,O}(\mathbf{x})$.

A useful feature of the cortex transform is that it is invertible through simple expansion and addition operations. Each layer is expanded to the size of the original by ideal interpolation and then added to the other layers. This can be accomplished easily in the DFT domain by simply adding the various layer DFT's, ensuring that DFT's for different sizes are properly registered.

C. Analytic Filters

The filters described above have even symmetry, and the DFT of the real image has conjugate symmetry. Thus their product (the DFT of one layer) also has conjugate symmetry. This means that the two lobes of each layer DFT contain the same information, and one can be dropped without loss. The layer that results will, of course, be complex.

A simpler view of the situation is to construct an analytic filter by dropping one lobe of each cortex filter DFT (Fig. 5). Application of this filter to the image will yield a complex layer; that is, each pixel in a layer will have both a real part and an imaginary part. Alternatively, each layer may be regarded as two layers, one real and one imaginary. The real layer contains the responses of cells with even symmetry; the imaginary contains the responses of cells with odd symmetry.[16]

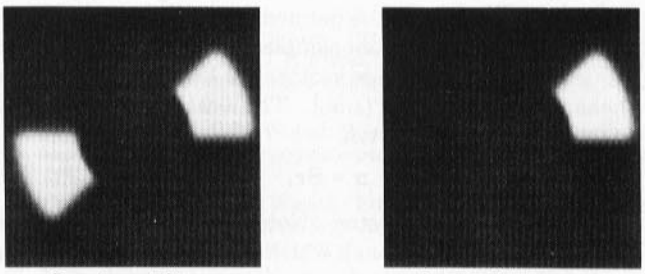

Fig. 5. Creation of an analytic cortex filter by removal of one lobe of a cortex filter.

Is there a biological analog to this complex version of the cortex filter? There is some evidence that the oriented simple cells of the cortex exist in matched pairs, with the same location, frequency, and orientation but with phases 90 deg apart.[17] Paired cells with a 90-deg phase difference have been found to be useful in extracting estimates of local motion.[18-21] It may therefore be useful to think of the real and imaginary parts of each layer pixel as corresponding to the responses of these cell pairs.

4. SAMPLING

It was shown in Section 3 how the number of pixels in the output layers could be reduced by subsampling in proportion to resolution. Here, we take that process one step further. In effect, the previous subsampling took advantage only of the reduced one-dimensional resolution of each layer. Here, we exploit the limited two-dimensional resolution of each layer.

The two principles at work here are that periodic sampling of an image induces a periodic replication of its DFT and that the information in the image will be preserved (aliasing will be avoided) if the replicas do not overlap. These notions are well known for one dimension but are less familiar for two dimensions. Therefore we first review the essentials of sampling theory for general periodic two-dimensional sampling.[22]

A. General Periodic Two-Dimensional Sampling

Consider an image as a finite-extent discrete sequence $x(\mathbf{n})$ with a rectangular support of width N_1 and height N_2. This finite-extent sequence can be regarded as one period of the periodic extension of $x(\mathbf{n})$, written as $\tilde{x}(\mathbf{n})$. The extension is created by replicating the image at horizontal intervals of N_1 and vertical intervals of N_2 (Fig. 6). It has two periodicities, horizontal (N_1) and vertical (N_2). The mapping from one replica to another can be associated with two periodicity vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ that map between corresponding samples. For a rectangular image, a natural (but not unique) set of vectors is $\mathbf{n}_1 = [N_1, 0]$ and $\mathbf{n}_2 = [0, N_2]$. The two vectors can be combined to form a periodicity matrix $\mathbf{N} = [\mathbf{n}_1|\mathbf{n}_2]$. The periodicity of the sequence can then be expressed by the relation

$$\tilde{x}(\mathbf{n} + \mathbf{N}\mathbf{r}) = \tilde{x}(\mathbf{n}), \tag{1}$$

where $\mathbf{r}$ is any integer vector. A useful feature of $\mathbf{N}$ is that $|\det \mathbf{N}|$ is equal to the number of samples in one period, that is, in our finite extent sequence (image) $x(\mathbf{n})$.

To subsample the image, we select a subset of the pixels in the image and set the others to zero. For general periodic subsampling, the subset is defined by two sampling vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ that map from one sample to the next, as illustrated in Fig. 7. As before, these vectors can be combined to form the sampling matrix $\mathbf{S} = [\mathbf{s}_1|\mathbf{s}_2]$. The location of any selected sample can then be written:

$$\mathbf{n} = \mathbf{S}\mathbf{r}, \tag{2}$$

where $\mathbf{r}$ is an integer vector. Note that the matrix $\mathbf{S}$ is not constrained to be diagonal, which corresponds to sampling vectors that are horizontal and vertical.

The factor by which the number of samples is reduced is $|\det \mathbf{S}|$. If we ensure that $\det \mathbf{N}/\det \mathbf{S}$ is an integer, then it is simple to show that the subsampled sequence will continue to have a periodicity matrix $\mathbf{N}$, and the number of selected samples in the finite extent sequence will be $|\det \mathbf{N}/\det \mathbf{S}|$.

The subsampling operation can be represented by multiplication of the image by a periodic array of impulses. The transform of this array is also an array of impulses (see Fig. 7), with a sampling matrix

$$\mathbf{F} = (\mathbf{S}^T)^{-1}\,\mathbf{N}. \tag{3}$$

By the convolution theorem, the DFT of the sampled image is equal to the circular convolution of the DFT of the image with $\mathbf{F}$. This corresponds to periodic replication of the DFT at the locations defined by the replicating matrix $\mathbf{F}$. In short, sampling the image by $\mathbf{S}$ replicates the spectrum by $\mathbf{F}$. Note that the number of replicas is $|\det \mathbf{N}/\det \mathbf{F}| = |\det \mathbf{S}|$. Thus, as expected, there is an inverse relationship between the number of samples and the number of replicas. In the next subsection we apply these principles to sampling of individual cortex layers.

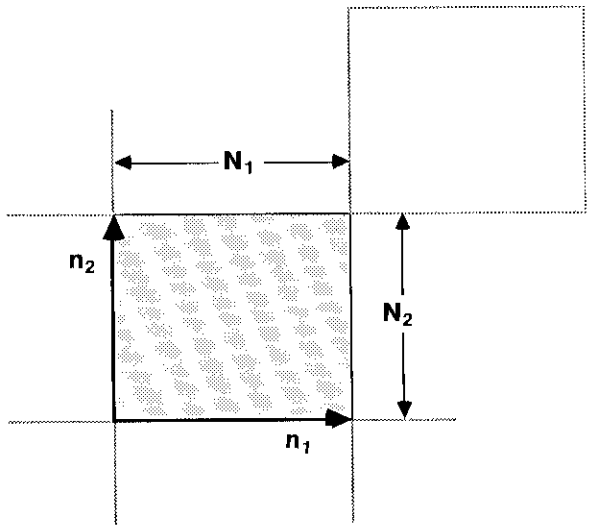

Fig. 6. Periodic extension of a rectangular finite-extent discrete sequence of size N_1 by N_2. The periodicity vectors are $\mathbf{n}_1 = [N_1,0]$ and $\mathbf{n}_2 = [0, N_2]$.

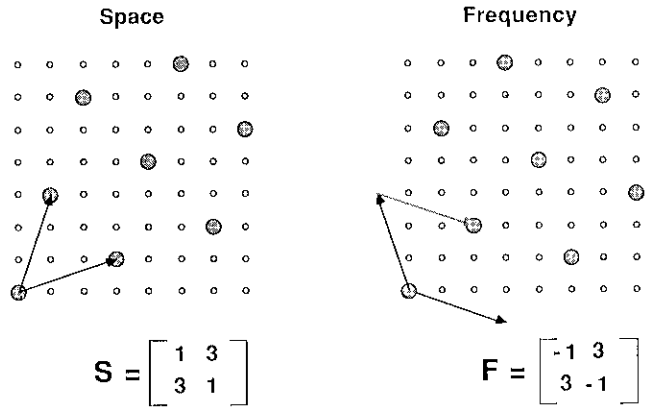

Fig. 7. A sampling matrix $\mathbf{S}$ and its corresponding replication matrix $\mathbf{F}$. The two sampling vectors (the column vectors of the matrix) are shown. In the picture, sampling is applied to an 8×8 discrete sequence, reducing the number of samples from 64 to 8 and producing 8 spectrum replicas in the Fourier domain at the locations specified by the replication matrix $\mathbf{F}$. Samples are numbered from 0 from the lower left.

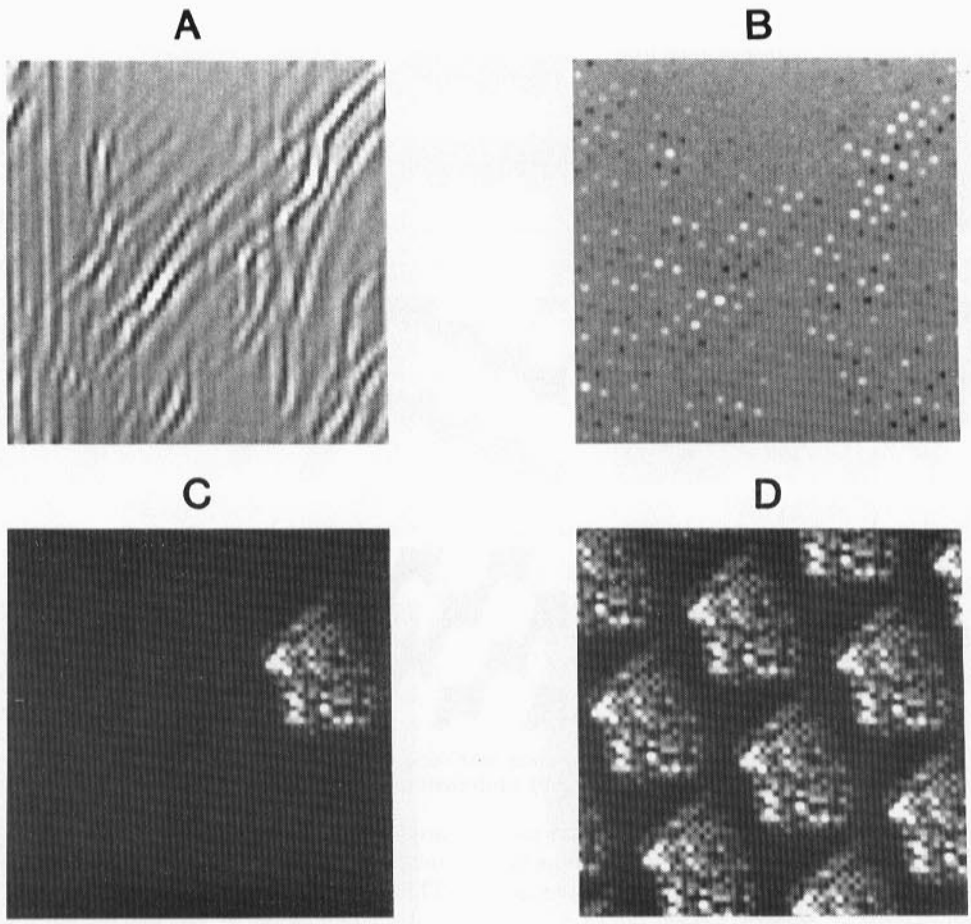

Fig. 8. Periodic sampling of a cortex layer. The real component of a cortex layer is shown A, before and B, after subsampling by means of the matrix $\mathbf{S} = [1\ 3\ 3\ 1]$. The amplitude spectrum of the layer is shown in C, and that of the sampled layer is in D. The replicas in D are placed according to the replicating matrix $\mathbf{F} = [-1\ 3\ 3\ -1]$.

B. Subsampling Cortex Layers

Sampling the image replicates the spectrum. As in the one-dimensional case, no aliasing will occur if the spectral replicas do not overlap. To illustrate how this can be arranged, we introduce a concrete example. Figure 8A shows one layer of the Cortex transform. The corresponding DFT is shown in Fig. 8C. Subsampling the layer by using the matrix $\mathbf{S} = [1\ 3\ 3\ 1]$ is shown in Fig. 8B, and the resulting replicated spectrum is shown in Fig. 8D. (For convenience, we write the elements of the matrix in the order of the column vectors, $[s11\ s21\ s12\ s22]$.) Note that the replicas are placed according to the replicating matrix $\mathbf{F} = [-1\ 3\ 3\ -1]$. In this case, the replicas do not overlap; hence aliasing will not occur.

Recall that det $\mathbf{S}$ is the factor by which the number of samples is reduced. Thus we want det $\mathbf{S}$ to be as large as possible. However, det $\mathbf{S}$ is also the number of replicas, which must be kept small if the replicas are not to overlap. Since in the example in Fig. 8 each replica occupies about one tenth of the area of the DFT, det $\mathbf{S}$ must clearly be less than about 10. Finally, we are constrained that det $\mathbf{N}$/det $\mathbf{S}$ must be an integer, which in the present case means that det $\mathbf{S}$ must be a power of 2. Thus we are led to a value of 8 for det $\mathbf{S}$. At this point, we would like to evaluate all possible sampling matrices with det = 8. But the number of such matrices is infinite. In particular, any matrix $\mathbf{SK}$, where $\mathbf{K}$ is an integer matrix with det = 1, will generate the same samples as $\mathbf{S}$. Using this rule to eliminate redundancies, we can generate a set of possible sampling matrices with det = 8. The replications induced by these matrices are shown for a number of examples in Fig. 9.

Visual examination suggests that the matrix $[1\ 3\ 2\ -2]$ (functionally equivalent to $[1\ 3\ 3\ 1]$) exhibits the least spectral overlap. This can be confirmed by sampling and reconstructing an image that is zero everywhere but at the origin and then computing the rms error.

C. Filter Orientation and Sampling Matrices

We have seen that the matrix $[1\ 3\ 3\ 1]$ is optimal for $f_{R,0}$, the filter with orientation 0 (0–45 deg). Because the other three oriented filters are various reflections of $f_{R,0}$, it is simple to show that their optimal sampling matrices are simple transformations of the matrix $[1\ 3\ 3\ 1]$. In particular, since $f_{R,1}$ (45–90 deg) is obtained from filter $f_{R,0}$ by exchanging x and y axes, we exchange rows of the sampling matrix to give $[3\ 1\ 1\ 3]$. But this yields an equivalent sampling pattern, since columns of the sampling matrix are the sampling vectors, which can be exchanged without effect. $f_{R,2}$ (90–135 deg) is obtained from $f_{R,0}$ by exchanging x and y axes and negating the y axis (reflection about the positive diagonal and about the y axis). Applying the same transformation to the matrix yields $[-1\ 3\ -3\ 1]$. $f_{R,3}$ is obtained from $f_{R,2}$ by reflection

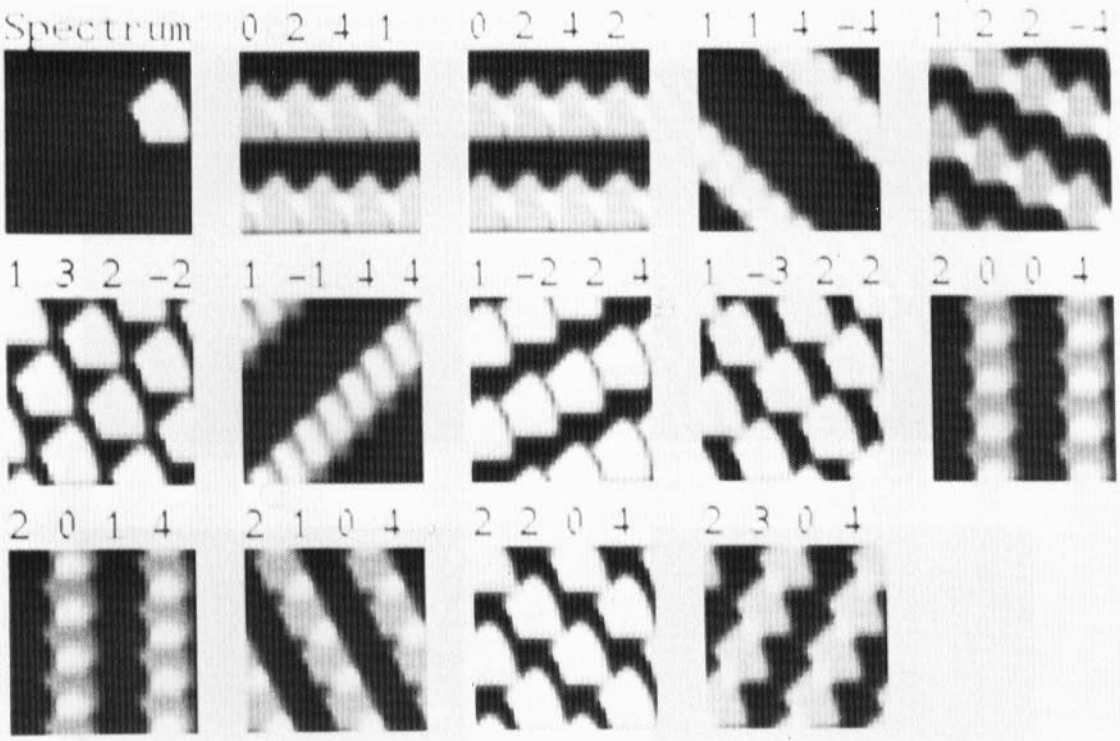

Fig. 9. Replications of cortex filter spectrum obtained by sampling with various matrices with determinant 8. The sampling matrix is shown above each spectrum. Note that matrix [1 3 2 −2] is functionally equivalent to the matrix [1 3 3 1] used in Fig. 8.

through the negative diagonal (exchanging x and y and negating both), which yields [−3 1 −1 3]. Again, since columns may be exchanged without effect, this is the same matrix as for $f_{R,2}$.

In summary, we arrive at two matrices: [1 3 3 1] for orientations 0 and 1 and [−1 3 −3 1] for orientations 2 and 3. Since these matrices determine the positions of the samples taken from each layer, this means that the samples in layers 0 and 1 will be located differently from those in layers 2 and 3.

D. Sampled Layers

The subsampling matrix that we have derived is appropriate for the highest-resolution layers. At the lower-resolution layers, the sample density can be reduced by a factor of 2 in each dimension for each step lower in resolution. This is equivalent to multiplying the sampling matrix by the scalar 2. Alternatively, since the Cortex transform already incorporates a subsampling by a factor of 2 in each dimension at each layer, we can apply the same subsampling matrix to these already-subsampled layers.

The determinant of the sampling matrix is 8; hence the number of samples is reduced by that factor. Recall that the original Cortex transform increases the number of samples by a factor of 16/3. The use of complex filters doubles that number, since each pixel has a real part and an imaginary part. The subsampling divides the factor by 8. Thus the end result is a number of samples approximately 4/3 as large as the number of pixels in the original image. This is equivalent to the number of samples in the Laplacian pyramid.[14]

F. Aliasing

Because the replicated spectra overlap slightly, an image reconstructed from subsampled layers will exhibit a small amount of aliasing. Thus, whereas the straightforward cortex transform is invertible, the subsampled transform is not. This residual aliasing could be removed through orthogonalization techniques,[23] but since this would complicate the description of the coding process and would likely lead to only small improvements in visual fidelity, we have not yet taken this step.

F. Summary

The result of the preceding steps is a set of coefficients, each corresponding to the response of a hypothetical visual cell at a particular location, tuned to a particular band of spatial frequency and orientation. The coefficients may be organized into layers by frequency and orientation. This representation is similar to the Laplacian pyramid developed by Burt and Adelson.[14] Both representations decompose the image into subsampled bandpass layers. Both the Laplacian pyramid and the Cortex transform are complete, in the sense of permitting exact reconstruction of the image. The Cortex transform differs in several respects. First, the Cortex transform further subdivides the image by orientation. Second, the radial-frequency passband of the Laplacian pyramid is approximately a difference of Gaussians, whereas the cortex passband is a more sharply tuned difference of mesa filters.[1] Third, the Laplacian pyramid is constructed by hierarchial discrete correlation,[24] which results in somewhat different filters at each resolution, whereas the filters used in the Cortex transform are all identical except for scaling and rotation.

5. LAYER ENTROPIES

The process of coding an image may often be divided into two parts (Fig. 1). The first is a transform from pixels into

some new set of quantized coefficients, and the second is an information-preserving encoding process, for example, assignment of variable-length code words to coefficients. If no use is made of possible correlations among coefficients, then the lower bound on the size of the code will be given by the first-order entropy of the coefficients, defined as

$$H = -\sum_{v=\min}^{\max} p(v)\log_2 p(v), \tag{4}$$

where $p(v)$ is the frequency histogram of coefficient values v. In practice, this bound can be approached quite closely by using techniques such as Huffman coding.[25] Since we are interested in the earlier, lossy stages in the coding process, we express the code size as the first-order entropy.

The entropy H is in units of bits per coefficient. It is traditional in image coding to express the code size in units of bits per pixel of the original image. This conversion can be made by scaling the entropy of each layer by the ratio of coefficients to pixels:

$$|\det \mathbf{S}|^{-1} 2^{2(R_{\max}-R)}, \tag{5}$$

where det S is the determinant of the subsampling matrix discussed in Section 4, R is the layer resolution, and $R_{\max}$ is the resolution of the original image. The total entropy of the image in bits per pixel is the sum of all of the scaled layer entropies.

6. QUANTIZATION

A. Terms and Notation

Quantization reduces a potentially continuous distribution of levels to a finite number. Typically any input level lying in a bin between the thresholds T_n and T_{n+1} is assigned the output level L_n, as illustrated in Fig. 10. A particular quantizer is defined by a particular set of input thresholds $\{T_i; i = 0, \ldots, n\}$ and output levels $\{L_i; i = 0, \ldots, n-1\}$. Note that the top threshold T_n must be greater than or equal to the maximum of the input, and the bottom threshold T_0 must be less than or equal to the minimum input.

B. Contrast Masking

The goal of quantizer design is to select thresholds and levels that minimize entropy while also minimizing some measure of distortion. Max[26] has derived a quantizer that is optimal when the distribution of input levels is Gaussian and the distortion is measured as rms error. However, we are concerned here with visual distortion, which may not be measured well by rms error. In particular, rms error equates all errors of a given size, regardless of the pixel value from which the deviation occurs. To the human viewer, however, errors are less readily detectable when they are deviations from a large value than when they are deviations from a small value.

The heavy line in Fig. 11 is a diagram of this contrast-masking effect. It represents the results of an experiment in which an observer discriminates between two images that differ only in contrast. One image (the background) has a contrast of c; the other (the background plus the increment) has a contrast of $c + \Delta c$. The increment Δc is varied to find the threshold at which c + Δc can just be discriminated from c. When the background contrast is zero ($c = 0$), one measures a contrast-detection threshold C. In plotting the experimental results it is convenient to normalize both the background contrast and the increment threshold by the detection threshold C. The graph plots the normalized increment threshold $\Delta c/C$ as a function of c/C.[27-33] Between normalized backgrounds of 0 and 1, the increment threshold declines by as much as a factor of 2 or 3.[28,30] This decline is not pictured in Fig. 11 and is not used in our analysis, although it may prove to be useful in the future. At higher background contrasts, the increment threshold grows, generally following a power function of background contrast.[13] We model the increment threshold function (ignoring the initial dip) as

$$\Delta c(c) = C \max[1, (c/C)^W], \tag{6}$$

where W is the exponent of the power function. A survey of the literature suggests an average value of W of about 0.7.[29,31-33] The form of the increment threshold function and the value of the exponent have been found to vary little with mean luminance, spatial frequency, target size, retinal location, and psychophysical method.

C. Quantizer Design

An effective quantizer should minimize the visibility of quantization errors. The contrast-masking function provides a description of the visibility of errors and may be applied directly to the design of a quantizer. Our method is derived in Fig. 11 and is a modification of a method described by Sharma and Netravali.[34] Within a layer, the

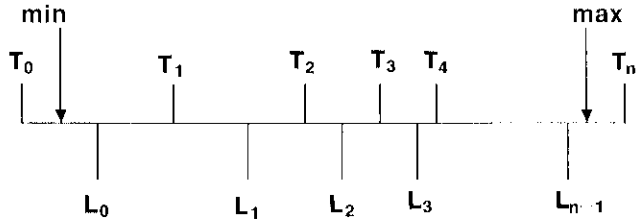

Fig. 10. Quantization terminology. Max and min are the largest and smallest input values. T's and L's are thresholds and levels. All input values between T_i and T_{i+1} are replaced by L_i.

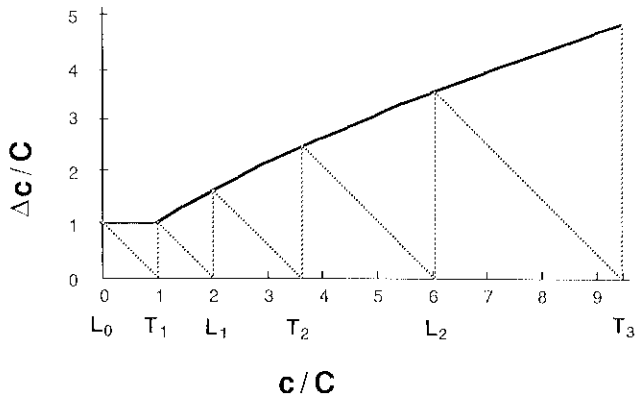

Fig. 11. Contrast-masking function and quantizer design. The heavy line represents the hypothetical contrast increment threshold function. The abscissa is the contrast of a signal relative to the threshold; the ordinate is the increment threshold, relative to the threshold. The dashed line illustrates the construction of quantizer levels and thresholds (see the text).

distribution of coefficients is nearly symmetrical about zero, with a large proportion of values near zero; therefore it is reasonable to place the first output level at zero.

We must next select a threshold T_1. The quantization error for a value at T_1 will be simply equal to the distance from T_1 to L_0. For this error to be at the threshold of visibility, this distance must equal the height of the masking function at L_0. Thus the first threshold T_1 is placed a distance $\Delta c(L_0)$ from L_0, as indicated by the first segment of dashed line in Fig. 11. Input levels between T_1 and L_1 will be quantized to L_1, resulting in increment errors. The worst case will be for an input just greater than T_1, which will suffer an increment error of $L_1 - T_1$. For this error to be at the detection threshold, it must be equal to $\Delta c(T_1)$, the height of the masking function at T_1. Thus the second level is placed a distance $\Delta c(T_1)$ from T_1, indicated by the second downward segment of dashed line. An input level between L_1 and T_2 will suffer a decrement error equal to the distance of the level from L_1. The worst case will be for a level at T_2, which will have a decrement error of $T_2 - L_1$. The masking function, as typically plotted, indicates the visibility of increments. However, a discrimination between a and $a + b$ may be regarded as either an increment of b on a or a decrement of b on $a + b$. Thus the visibility of a decrement of size $T_2 - L_1$ on T_2 is equivalent to the visibility of an increment of the same size on L_1. The threshold error at L_1 is given by $\Delta c(L_1)$, so T_2 is placed this distance from L_1, as indicated by the third downward segment of dashed line. This process is iterated until the next threshold is greater than the largest input value.

Formally, the levels and thresholds are given by

$$\begin{aligned} L_0 &= 0, \\ T_i &= L_{i-1} + \Delta c(L_{i-1}), \\ L_i &= T_i + \Delta c(T_i). \end{aligned} \tag{7}$$

The same set of thresholds and levels is reflected about zero to produce the bins for negative input values. Note that a complete quantizer depends on only the two parameters C and W of the masking function. In what follows, we fix W at a value of 0.7 and explore the effect of varying C. Thus C, the hypothetical contrast threshold, will control the severity of quantization for each layer.

Implicit in our use of the masking function in quantizer design is the notion that each quantized coefficient represents the contrast of an image component for which the masking function in Fig. 11 is accurate. As will be shown, in the process of reconstruction each coefficient represents the contrast of a small patch of sinusoidal grating (Fig. 3). Thus the masking function pictured (which has been measured for both large and small patching of grating[29]) is in fact an appropriate description.

Note also that if, in the contrast-discrimination experiment, the background and the increment are of different spatial frequencies or orientations, then much less masking occurs.[27,29,35] It is therefore essential that the masking-based quantization be applied to coefficients of a frequency and orientation band-limited transform like the one used here. In principle, a quantizer designed in this way will ensure that all quantization errors are below the threshold of visibility, provided that C and W are chosen correctly and that the masking function for each reconstruction component does indeed have the form of Eq. (6), independent of the value of all other coefficients.

In the interest of simplicity, in the preceding discussion we neglected the probabilistic nature of visual thresholds and considered only perceptually perfect coding. Relaxation of either of these constraints would require attention to the actual distribution of quantization errors as well as a theory for how multiple errors combine to determine visibility. Some preliminary discussion of these issues is given in Subsection 10.D.

We have chosen to control the degree of quantization with a parameter Q (the quantization strength), related to C by

$$Q = \log_2 C + 10.9. \tag{8}$$

For reference, 0.5% contrast corresponds to $Q = 3.26$.

7. ENTROPY OF CODED IMAGE

The entropy of the coded image is the sum of the entropies of the individual layers. To express layer entropies in units of bits per pixel of the original image, the entropy of each layer must be scaled by the ratio of the numbers of samples in the layer and to that the image, as was described in Section 5.

Calculation of the total code size must also include any additional information required for reconstruction that changes from image to image, such as possibly the value of C for each resolution. However, at present we treat this as a nonadaptive code and do not envision alteration of C for each image. To determine whether adaptive setting of C is worthwhile would require study of a larger class of images. Note also that the addition of four values of C would not materially change the code size. Figure 12 illustrates the entropy of each resolution as a function of the quantization strength Q. Each resolution entropy is the sum of the entropies for the four orientations. The total entropy (less that low residue entropy of 0.027 bits/pixel) is also shown. Note that about 3/4 of the total entropy is consumed by the highest-resolution band, primarily because it has about 3/4 of the coefficients. It is also clear that to achieve substantial compression at least the high-resolution band must be quantized with a strength of around 5 (1.67% contrast). For comparison, the minimum human contrast threshold is about 0.5%

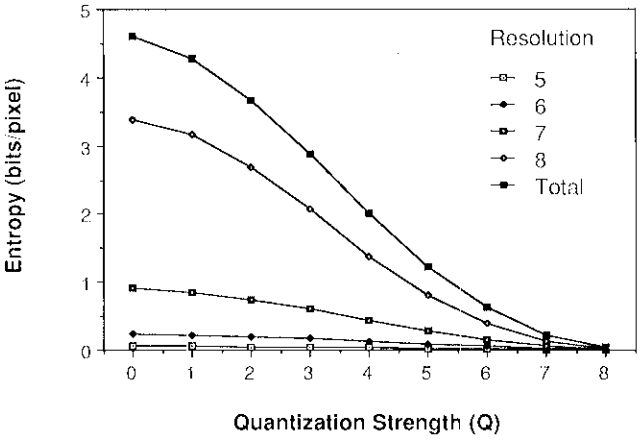

Fig. 12. Entropy versus quantization strength Q. Separate curves are shown for the four resolutions and the total.

($Q = 3.26$). Whereas, this is correct only for the most visible spatial frequencies under optimal conditions, if all resolutions were quantized with this threshold the resulting entropy would be about 2.5 bits/pixel.

8. RECONSTRUCTION

The set of sampled, quantized layers at various resolutions and orientations constitutes the information we have about the original image. In a general sense, the process of reconstruction is an attempt to estimate the image that was most likely to have given rise to the code in hand. In the perceptual analog, the code may be likened to the small number (relative to the total) of visual neurons active enough to contribute to the perceptual process. The reconstruction is the analog of inferring the retinal image from these responses.

To reconstruct the image from the sampled, quantized coefficients, we first insert the zeros deleted from the sampled layer (upsample) to recreate a rectangular sequence. This layer is then filtered to remove the sampling replicas as well as some of the artifacts of quantization. The layer is then combined with the layers at other resolutions and orientations. The resulting image may be compared with the original to evaluate the fidelity of the coding and reconstruction processes.

The complete signal flow is shown for one layer in Fig. 13. The images and spectra that result at various stages in the process are shown in Fig. 14. In Subsection 12.A.4 below, it is suggested how the two sequences enclosed in dashed boxes might be converted to single steps.

To compare sensitivity to quantization errors to contrast sensitivity, it is useful to note that the reconstruction process may also be viewed as placing into the reconstruction image, at the location of each coefficient, a reconstruction signal with a contrast proportional to the quantized coefficient. The reconstruction signal is the inverse DFT of the cortex filter. A quantization error therefore corresponds directly to the addition of a reconstruction signal with a contrast proportional to the size of the quantization error.

A. Spectral Effects of Sampling and Quantization

Before subsampling and quantization, the spectrum of the layer is confined to a small region of the DFT. Subsampling replicates the spectrum, as described above (Fig. 8). Quantization also scatters energy outside the original passband of the layer. Figure 14 shows the effect of quantization and sampling on the layer spectrum.

B. Reconstruction Filter

In reconstruction we wish to remove all the energy lying outside the passband for the layer. This is done by means of a reconstruction filter that is a copy of the filter used to create the layer. Applying this to the sampled, quantized layer confines the signal once again to the passband of the original layer, as shown in Fig. 14.

If no quantization or subsampling occurs, the signal passes twice through the same filter, once to create the layer and once during reconstruction. This has the same effect as passing the image through a filter equal to the square of the previous filter. To ensure that the latter filter subdivides the image without remainder, one of the cortex filters discussed above is used. Thus the actual coding filter and the reconstruction filter are equal to the square root of a cortex filter. When neither sampling nor quantization occurs, this will lead to exact reconstruction of the image.

C. Building the Spectrum

After the application of the reconstruction filter, the spectrum is added to the portions of the spectrum contributed by the other layers. The final reconstructed image is obtained by an inverse DFT of the reconstructed spectrum. This method of combining spectra automatically incorporates expansion of each layer to the size of the original by ideal interpolation.[12]

As noted above, reconstruction may also be described as replacing each coefficient by a reconstruction signal of the corresponding frequency, orientation, phase, and location, whose amplitude is the value of the coefficient.

9. EVALUATION OF IMAGE FIDELITY

We are now at a point where we may compress and reconstruct images and would like a method of evaluating the degree to which they visually resemble the original. This is a sufficiently important subject that it warrants a separate discussion.

It is clear that progress in this area has been hampered by a lack of objective techniques for evaluating image fidelity. Although a number of computational techniques have been proposed,[36] none is generally accepted as a substitute for human evaluation. On the other hand, subjective rating scales have not been shown to provide accurate, robust, or repeatable measures. What are required, at least until a computational technique of sufficient power is developed, are objective measures of human preceived fidelity.

There are many criteria on which an original image and a compressed version may be compared, and these criteria will depend strongly on the application. If the application is known, for example, transmission of sign language, then direct measures of success in the application can be used as a criterion.[37,38] The criterion that we will consider is near-perfect perceptual fidelity. This may also be described as perceptually lossless encoding. This criterion might be used in applications, such as visual arts and photography, in which essentially no artifacts are tolerable.

How do we implement this criterion? There are many

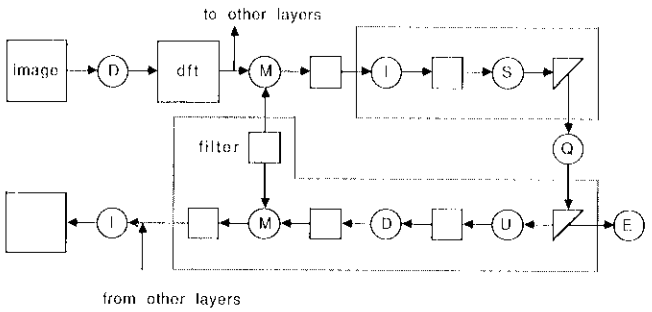

Fig. 13. The sequence of operations used to code and reconstruct an image, shown for one layer only. The triangle representing the result of subsampling indicates that the image is no longer represented in a square sample lattice. The circles indicate operations: D, DFT; M, complex multiplication; S, subsampling; Q, quantization; U, upsampling; I, inverse DFT; E, entropy computation. Sections enclosed in dashed lines are discussed in Subsection 12.A.4.

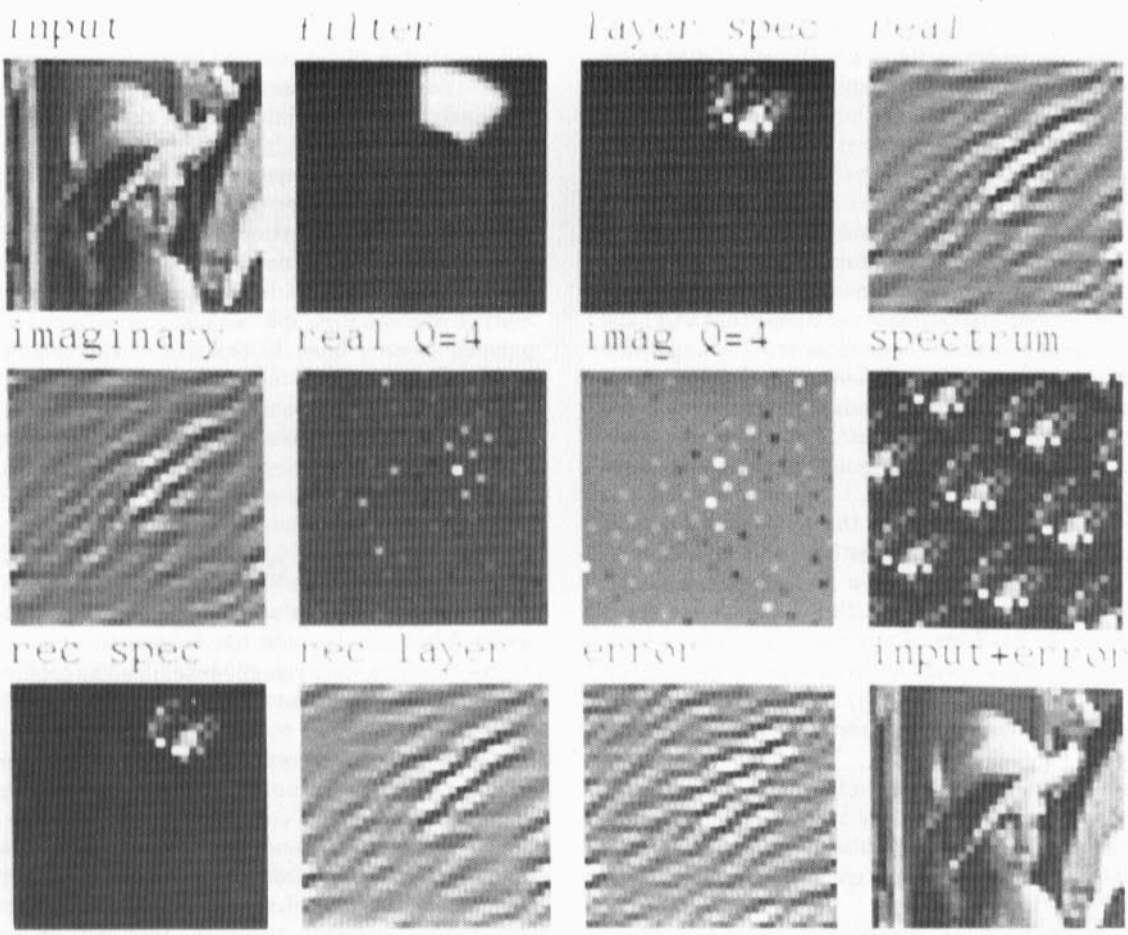

Fig. 14. Process of coding and reconstruction for one layer. From left to right, top to bottom, the images are the original, the analytic cortex filter, the spectrum of the layer response, the real layer, the imaginary layer, the real layer after sampling by the matrix [1 3 3 1] and quantization at $Q = 4$, the sampled and quantized imaginary layer, the spectrum of the sampled and quantized layers, the spectrum after application of the reconstruction filter, the reconstructed layer, the error owing to sampling and quantization (shown at full contrast), and the sum of the original and the error.

possible implementations yielding possibly different estimates. The one that we have adopted has the virtues of being objective, reasonably accurate, reasonably consistent across subjects, and portable to any laboratory. We use a forced-choice discrimination between the original and the compressed image. We estimate the degree of compression that yields 82% correct discriminations.[39] Details of our method are given below, but we draw attention to several aspects of the viewing conditions that we believe are particularly significant.

First, an image size must be specified. We chose 256^2 as a compromise between high resolution and manageable computations and experiments. There may be reasons for the selection of other sizes, but citations of particular compressions achieved should never be given without an indication of the image size. In general, large compressions are easier to obtain for images with many pixels, because when the image is displayed within a constant visual angle, more of the image is necessarily pushed into regions of spatial frequency at which the human observer is less sensitive.

Second, for similar reasons, the viewing distance will have powerful effects on obtainable compression and must be given whenever compression figures are cited. We used a viewing distance of 91.2 cm, so that the highest image frequency was 16 cycles/deg, still within the visible range for human observers. This distance corresponds to 7.125 times the image height.

Another important consideration is the gamma function of the display, which relates pixel values to luminance on the screen. It is difficult to specify a correct gamma function without reference to the ultimate purpose of the image. Nevertheless, the gamma function can certainly obscure or reveal detail, and it is therefore essential that testing be done under specified display conditions. Since our image transformations are defined on a luminance signal, we used a carefully linearized display.[40]

Finally, it is important that the display monitor itself does not obscure image information excessively. This can be ensured by using a monitor of sufficient bandwidth and small enough spot size (as we have done) or by expanding the image to bypass the monitor limitations.

10. ESTIMATION OF QUANTIZATION THRESHOLDS

Quantization of each layer is governed by a parameter C that is in theory given by the contrast-detection threshold for the corresponding resolution. In principle, then, we could set C

Fig. 15. Images in which one resolution band (6) has been quantized by varying amounts, as indicated by the quantization strength Q.

for each resolution equal to an estimate of contrast sensitivity. Instead, we have estimated the value of C for each resolution that yields just detectable quantization.

A. Experimental Design

At each of four resolutions (8, 7, 6, and 5), the four layers corresponding to the four orientations were extracted from the image and individually quantized with strengths Q ranging from 0 to 8 (corresponding to contrast thresholds C ranging from 0.05 to 13.4%). The individual layers were then reconstructed and combined with the (uncompressed) remainder of the image. This results in an image in which only one resolution band has been compressed, as shown in Fig. 15. This permits us to assess the effects of quantization separately at each resolution.

Each of the four sets of nine images was then used in a two-

alternative forced-choice staircase experiment. In each trial, the observer was presented with a pair of images, displayed side by side for 2 sec. One was the original, and the other was a compressed image. The compressed image appeared randomly on the left or right. The observer's task was to report on which side the compressed image appeared.

The quantization strength used in the next trial was determined by a QUEST staircase.[41] In this method previous correct responses lead to less quantization, and incorrect responses lead to more quantization, so that the sequence of trials converged on a quantization yielding 82% correct. Separate staircases for each resolution were interleaved, so that the observer did not know what resolution or quantization would be viewed next. After each trial, the observer was told whether he or she was correct or incorrect.

B. Methods

Images were displayed on a Mitsubishi M-6950 monochrome monitor driven by an Adage RDS-3000 raster graphics system. The display was calibrated and linearized so that pixel values were proportional to the luminance. The contrast of the images (defined as peak minus mean divided by mean luminance) was 100%. The mean luminance was 100 cd/m^2. Further details on the method of image display are available elsewhere.[40] The viewing distance was 91.2 cm, and the display resolution was 20 pixels/cm, so that the image Nyquist frequency was 16 cycles/deg.

C. Results

Data from one staircase for one observer at one resolution are shown in Fig. 16. Each data point is the proportion correct at a particular quantization strength. The proportion correct rises from chance at slight quantization to perfect performance at more severe levels.

To estimate a quantization threshold, we fitted a Weibull function to the data and extracted the quantization strength at which the curve is at 82% correct.[42] This technique was used to estimate thresholds for each of the different resolutions for six observers. These thresholds are shown in Fig. 17, expressed as contrast sensitivities by means of Eq. (8).

There is considerable variability among observers. We suspect that this is due to different strategies employed by different observers in searching for the artifacts that betray the compressed image. Some observers quickly discover an effective cue, whereas others do not. This problem would no doubt be ameliorated by the use of a different image on each trial, but this is beyond our present capability. There is little doubt that observers learn the artifacts over the course of the session, so that our method overestimates the visibility of artifacts in images viewed infrequently and hence underestimates the potential for perceptually lossless compression.

D. Relation to the Contrast-Sensitivity Function

The general shape of the mean curve resembles that of a contrast-sensitivity function, as it would if the quantization threshold C were in fact equal to the contrast threshold for the corresponding resolution. Recall that the quantization error of a coefficient results in a contrast increment or decrement in a reconstruction signal at a particular frequency, orientation, phase, and location. For comparison, we collected contrast thresholds for the individual reconstruction signals themselves. We used a forced-choice method with 2-sec side-by-side viewing of a blank field and a field containing a reconstruction signal. Ten thresholds, each based on 64 trials, were collected at each of the four resolutions for observer WCP. All orientations were used at least twice, and no systematic differences among orientations were noted. These thresholds are shown as the lower line in Fig. 17.

The mean quantization sensitivity (upper line) lies considerably above the mean detection sensitivity (lower line). The discrepancy may be due to spatial integration among artifacts: the detection thresholds are for single reconstruction signals, whereas the quantization error for a given image may consist of many such signals of varying contrast distributed over space and orientation. Pursuing this explanation, we considered the following prediction. Assume that all

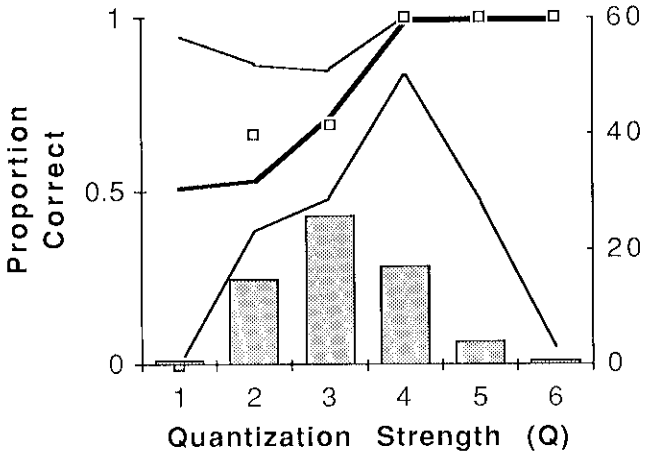

Fig. 16. Quantization detection as a function of quantization strength Q for one observer at one resolution. The square symbols represent percent correct, the number of trials at each strength is indicated by the histogram, and the heavy line is the best-fitting Weibull function used to estimate a value of quantization threshold. The upper and lower thin lines are binomial confidence limits about the data.

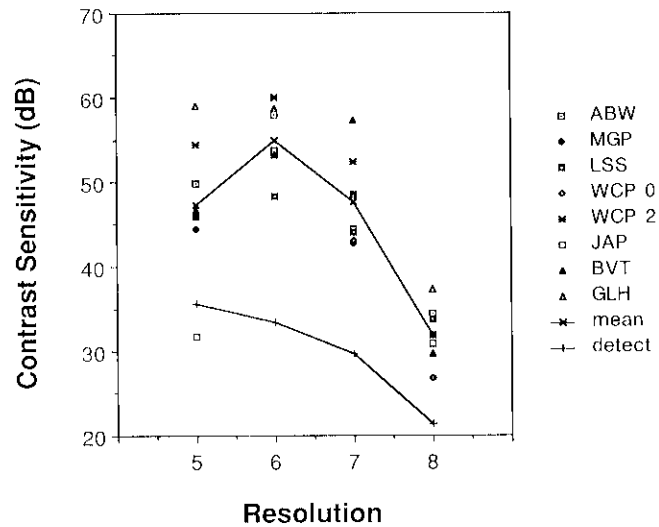

Fig. 17. Quantization sensitivity (1/threshold) at each resolution for eight observers for the image of the woman (two replications for subject WCP). The upper line shows the mean. The lower line shows contrast-detection sensitivities for single reconstruction signals for observer WCP.

artifacts, after scaling by the appropriate value of the contrast increment function, are combined through probability summation.[42,43] If each contrast artifact is q_i, we compute the scaled contrast artifact $q_i/\Delta c(c_i)$, where c_i is either the true contrast (if $q_i > 0$) or the quantized contrast (if $q_i < 0$). These scaled artifacts are combined by means of the formula

$$\sum_i |q_i/\Delta c(c_i)|^\beta. \tag{9}$$

The function Δc was given by Eq. (6), with $W = 0.7$ and C given by the detection threshold for the corresponding resolution (lower line in Fig. 17). This sum should equal 1 when the compound of many artifacts is at the threshold. For detection thresholds, the parameter β is typically 3.5. With this value, the predicted quantization sensitivities are far too high. By using the actual $\{q_i\}$ resulting from the mean quantization thresholds in Fig. 17, I estimated the value of β required for correct prediction. For the four resolutions used (5, 6, 7, 8), the values required were 8.2, 6.6, 7.5, and 10.4. These large values suggest little probability summation among artifacts. This is consistent with the data of Legge and Foley,[29] who found little spatial integration for increment thresholds, in contrast to that for detection thresholds.

Other factors that must be considered in predicting the quantization threshold are masking between coefficients and the effects of spatial inhomogeneity of visual acuity. The latter will certainly result in less spatial integration at the highest resolutions, consistent with the high value of β noted above.

11. COMPRESSION RESULTS

A. Image at the Threshold Quantization

Figure 18 shows an image in which each resolution was quantized to its threshold, as determined in the previous section. The entropy of this image is 1.11 bits/pixel. Figure 19 shows progressively more severe compression of the same image. Figure 20 shows results for a second image. In summary, for these two images, visually lossless compression, by the somewhat arbitrary standard adopted here, requires about 1 bit/pixel. Subjectively high fidelity is still possible at as low as 0.5 bit/pixel.

B. Comparison with Other Schemes

Unfortunately, there are few established schemes for quantitative comparison of image coding schemes. This is in part why we developed the psychophysical evaluation method described above. But, before applying our method to other schemes, we must rely on qualitative reports from other authors. In this context, it appears that our method is competitive with other existing methods, such as discrete cosine transform,[44,45] block truncation coding,[46] Laplacian pyramid,[14] and subband coding.[47]

12. DISCUSSION

A. Image Compression

The results obtained here, although preliminary, suggest the promise of coding schemes based explicitly on visual models. In this section we draw attention to several of the possibly novel features of this effort.

1. *Model as Code*

Although visual models were previously applied to the design of image codes,[48] here we have used the model itself as a coding device. In other words, we have coded the image in the same way that it may be coded by the visual system. Each coefficient of the code is a simulated response of a single hypothetical visual cell, quantized so as to reflect the ability of the rest of the brain to resolve different levels of response. There is no guarantee that this will result in the most efficient code (cells could be highly correlated), but it seems likely to approach this goal, since the visual system also contends with limited information capacity.

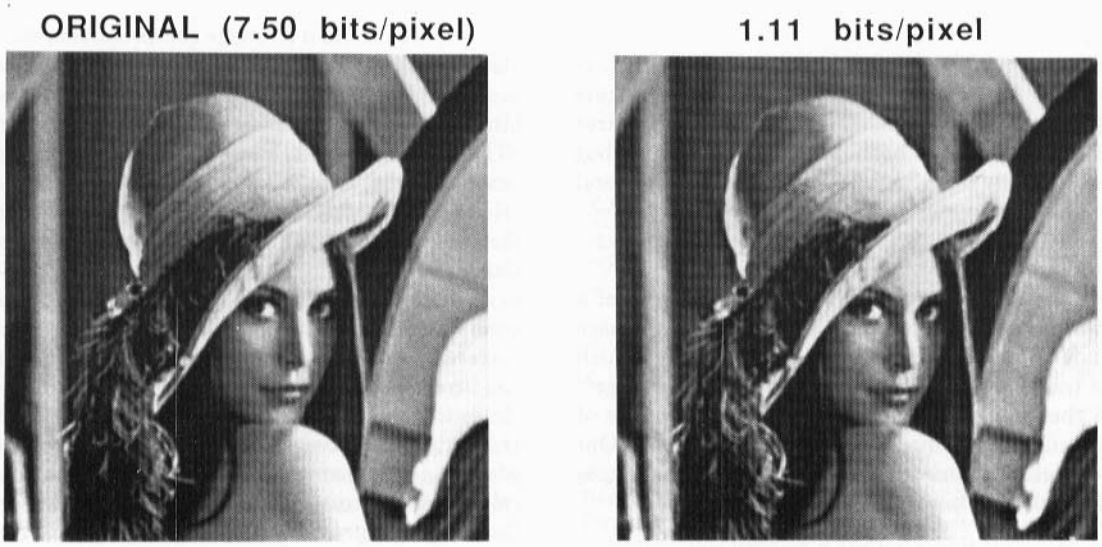

Fig. 18. Original image of the woman and reconstruction at 1.11 bits/pixel. Quantization strengths of 3, 2, 4, and 6 were used at resolutions of 5, 6, 7, and 8.

Fig. 19. Image of woman reconstructed from codes of various sizes. Entropies and Q at resolutions 5, 6, 7, and 8, respectively: 0.66 bit/pixel, 4, 3, 5, 7; 0.45 bit/pixel, 4, 5, 6, 7; 0.36 bit/pixel, 5, 4, 6, 8; 0.23 bit/pixel, 5, 5, 7, 8.

2. *Masking*

A second feature of our scheme is the use of the contrast-masking function for two-dimensional band-limited targets as the basis of quantizer design. The notion of a quantizer based on visual masking has been discussed widely,[49–51] but masking has typically been based on edges or some general measure of image busyness.

3. *Frequency and Orientation*

Another fundamental aspect of this scheme is the use of a transform whose basis functions are localized in both space and frequency domains. This resembles schemes in which Fourier-like transforms are applied to blocks of the image[44] or in which the image is separated into different bands of resolution for subsequent quantization and coding.[14,52] Our scheme also partitions orientation into separate bands, as was recently explored elsewhere.[53]

4. *Future Developments*

We have so far explored only the most straightforward application of the Cortex transform and masking-based quantization to the problem of image compression. We suspect that there are substantial additional economies to be exploited.

For example, the only masking so far implemented is contrast masking occurring within each hypothetical cell. There are, however, masking effects of some power that extend some distance over space, frequency, and orientation. Fortunately, the coefficients generated by the Cortex transform are ideally suited to simple implementation of such effects.

A second possible improvement involves a more efficient method of generating and reconstructing the subsampled layers. Since only certain samples of the layer are required, an inverse transform can be designed that in effect encapsulates the coding steps enclosed in the dashed box in Fig. 13. Likewise, the enclosed reconstruction steps can be condensed. These amendments largely affect the computational cost of the transform (each DFT is one quarter as large) but may also provide some improvement in compression by alteration of the prefiltering and postfiltering.

Next, we suspect that more-detailed optimization of the quantization may be worthwhile. Particularly when quanti-

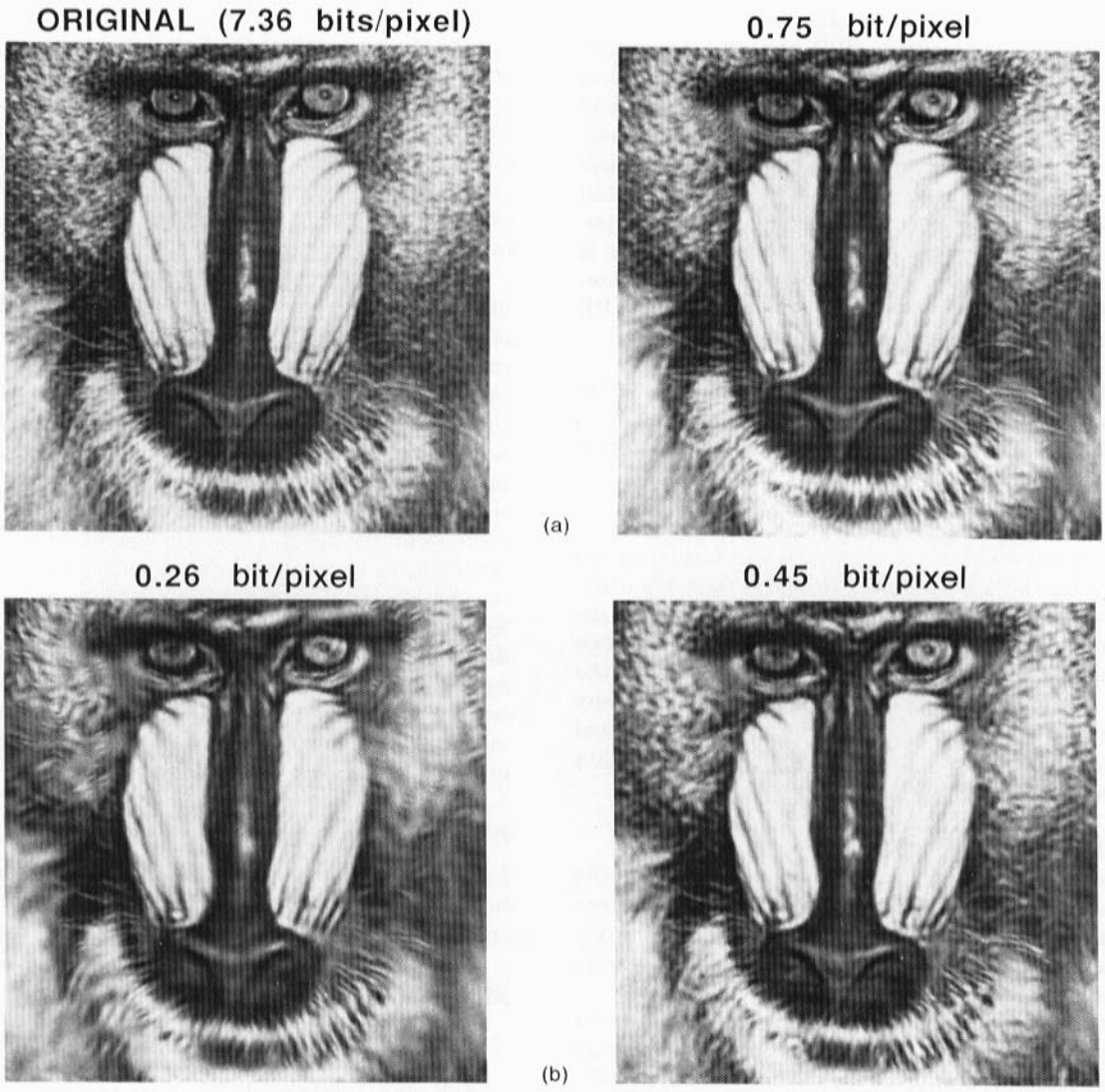

Fig. 20. Mandrill image reconstructed from codes of various sizes. Entropies and Q at resolutions 5, 6, 7, and 8, respectively: 0.75 bit/pixel, 4, 5, 6, 7; 0.45 bit/pixel, 5, 4, 6, 8; 0.26 bit/pixel, 5, 5, 7, 8.

zation is severe (C is large), careful placement of levels may be critical. Also, the masking function on which we have based our quantization was obtained on a uniform background, whereas our quantization errors are viewed on a background consisting of the rest of the image. Contrast-discrimination functions collected on nonuniform backgrounds may differ.

Finally, whereas the Cortex transform serves in part to decorrelate the coefficients of the image, the coefficients may still retain some correlation, both within and between layers, that can usefully be removed. Masking between frequency and resolution bands, mentioned above, may provide a way of removing residual correlations between these layers.

B. Visual Representation

As was noted in Section 1, a second goal of this project was to explore the representational structure underlying early spatial vision. The implicit notion is that the visual input undergoes various transformations yielding various intermediate representations. We assume a visual pathway consisting of multiple nodes and branching and converging paths. We may regard the retinal image as the root node, the receptor response arrays of the rods and cones as four branches from this node, each type of retinal cell as another node, and so on. Why, in this context, might perceptual judgments reflect the information in only a single node? Because either all distal branches of the network eventually converge to this node or pathways that do not converge to this node convey information that is useless for the task at hand. We assume that our psychophysical judgments and estimates of image fidelity are based on a particular representation, which we may call the spatial node. Whereas we may speculate that this node has some distinct physiological basis, at this point it is defined only by a particular set of perceptual tasks, such as the detection of compression artifacts.

Our goal is to discover the code at the level of the spatial node. The method used here is to conjecture a specific code, based on general principles of signal processing as well as specific facts about primate neurophysiology, and to test whether it retains the information to which the human observer is demonstrably sensible.

The logical basis of this test is as follows. The universe of

images is exhaustively partitioned into a number of nonoverlapping perceptual equivalence (PE) classes. All images within a class appear equivalent to the human observer and hence map to the same perceptual code. Likewise many images will yield the same instance of our hypothetical code. Thus the image set can also be partitioned into hypothetical equivalence (HE) classes based on our hypothetical spatial code. Ultimately, we would like to show that some hypothetical code is identical to the perceptual code. This is equivalent (modulo information preserving transformations) to showing that the HE classes are equal to the PE classes. This, in turn, is equivalent to showing that

(1) Any two images with the same code should appear identical, and

(2) Any two images with different codes should appear different.

Our psychophysical evaluation of compressed images is a crude test of proposition (1). Reconstruction from the code creates an image that is physically different from the original yet results in the same code. If the two images are indiscriminable, then we have at least one instance of two different images that map both to the same PE and to the same HE. A more elaborate test would be to generate many samples from the equivalence class for the hypothetical code. This could be done by varying each coefficient within the bounds of its quantization bin.

1. *Spatial Inhomogeneity*

A possible confusion is present in our use of a spatially homogeneous image representation. It is well known that human spatial resolution declines precipitously with the distance from the point of fixation, and a code attempting to mirror human visual representation should do likewise.

If such a code were developed, it would be appropriate for only one point of fixation. Psychophysical testing would require that fixation be appropriate, either by brief presentation or by stabilized viewing. While this code might in fact bear a closer correspondence to the true code than the one we have explored, it would be less useful in practical image-compression applications. Our code may be regarded as a homogenized version of the human spatial code, extending the resolution of the fovea throughout the visual field.

2. *Other Visual Dimensions*

The methods used here can be extended to other dimensions of visual representation. For example, the coding scheme used here is readily generalized to color, and recent advances in our understanding of the physiology of color coding provide strong suggestions about how luminance and color codes might differ.[54] This attack on the representational problem posed by color vision is a necessary complement to the more-traditional study of single isolated mechanisms (analogous to one coefficient of the code) or of purely empirical properties of color perception.

3. *Functional Architecture of the Visual Cortex*

It is interesting to speculate on how a representational scheme such as the one we have described might relate to particular physiological structures. First, as is shown by our choice of the name Cortex transform, we believe that our scheme most resembles the functional behavior of portions of the primate striate cortex. This is reasonable in the context of the network notions discussed earlier, since essentially all input from the two eyes that is destined for the cortex converges to this area before distribution to other centers and since the information conveyed by these cells is of the sort that is appropriate for the spatial discriminations involved.

It is now believed that the striate cortex consists of modules, each a 2 mm × 1.5 mm column extending the full depth of the cortex.[54,56] Each module corresponds to a particular small region of visual space. Within a module there is a full set of cells tuned to various orientations and spatial frequencies (as well as color and ocular dominance). There is as yet no clear statement of how these cells are arranged within a module, although tangential penetrations often show long sequences of regularly changing orientation.[54,57] In our representation, a cortical module corresponds to all the coefficients from the various layers related to a particular small region of the image, equal in size to the receptive field of the lowest-resolution layer. The number of cells in each module devoted to each spatial frequency would increase as the square of the spatial frequency. More-detailed hypotheses regarding the layout of cells within a hypothetical module must await further physiological results or further insight into the signal-processing architecture within a module.

ACKNOWLEDGMENTS

I thank Jeffrey Mulligan, Leland Stone, and especially Albert J. Ahumada, Jr., for useful discussions and Kia Soroosshian and Khanh Nguyen for software assistance.

REFERENCES AND NOTES

1. A. B. Watson, "The Cortex transform: rapid computation of simulated neural images," Comput. Vision Graph. Image Process. **39,** 311–327 (1987).
2. P. Lennie, "Parallel visual pathways: a review," Vision Res. **20,** 561–594 (1980).
3. R. M. Shapley and P. Lennie, "Spatial frequency analysis in the visual system," Annu. Rev. Neurosci. **8,** 547–583 (1985).
4. R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," Vision Res. **22,** 545–559 (1982).
5. M. A. Webster and R. L. De Valois, "Relationship between spatial frequency and orientation tuning of striate-cortex cells," J. Opt. Soc. Am. A **2,** 1124–1132 (1985).
6. R. L. De Valois, E. W. Yund, and H. Hepler, "The orientation and direction selectivity of cells in macaque visual cortex," Vision Res. **22,** 531–544 (1982).
7. B. Sakitt and H. B. Barlow, "A model for the economical encoding of the visual image in cerebral cortex," Biol. Cybern. **43,** 97–108 (1982).
8. A. B. Watson, "Detection and recognition of simple spatial forms," in *Physical and Biological Processing of Images,* O. J. Braddick and A. C. Sleigh, eds. (Springer-Verlag, New York, 1983), pp. 100–114.
9. S. A. Klein and D. M. Levi, "Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation," J. Opt. Soc. Am. A **2,** 1170–1190 (1985).
10. The parameters of the filters used here, in the terms defined in Ref. 1, are $\beta = 0.9$, $\gamma = 8$, $s = 2$. The filters have center orientations of 22.5, 67.5, 112.5, and 157.5.
11. A. Weber, "Image data base," USCIPI Rep. 1070 (Image Processing Institute, University of Southern California, Los Angeles, Calif., 1983).

12. A. B. Watson, "Ideal shrinking and expansion of discrete sequences," NASA Technical Memorandum 88202 (National Aeronautics and Space Adminstration, Moffett Field, Calif., 1986).
13. S. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing," Comput. Graph. Image Process. **4,** 104–119 (1975).
14. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **COM-31,** 532–540 (1983).
15. J. L. Crowley and R. M. Stern, "Fast computation of the difference of low-pass transform," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6,** 212–222 (1984).
16. Consider a filter $f(\mathbf{x})$ at orientation θ. By analogy to analytic signals in one dimension [R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1978), p. 267], we define the analytic filter

$$f_a(\mathbf{x}) = {}^1/_2|f(\mathbf{x}) - if_h(\mathbf{x})|,$$

where $f_h(\mathbf{x})$ is the Hilbert transform of $f(\mathbf{x})$ taken along the direction vector $\mathbf{d} = [\cos\theta, \sin\theta]$,

$$f_h(\mathbf{x}) = f(\mathbf{x}) * \frac{-1}{\pi\mathbf{x}\cdot\mathbf{d}}.$$

The Fourier transform of $f_a(\mathbf{x})$ is

$$F_a(\mathbf{u}) = \frac{1}{2}[F(\mathbf{u}) + \mathrm{sign}(\mathbf{u}\cdot\mathbf{d})F(\mathbf{u})] = \begin{cases} F(\mathbf{u}), & \mathbf{u}\cdot\mathbf{d} > 0 \\ 0, & \mathbf{u}\cdot\mathbf{d} < 0 \end{cases}.$$

Since the filter transform consists of two identical lobes symmetrically placed on either side of the line $\mathbf{u}\cdot\mathbf{d} = 0$, the analytic filter discards one of these lobes. Since $f(\mathbf{x})$ is even along $\mathbf{d}$, $f_h(\mathbf{x})$ will be odd along $\mathbf{d}$. Thus the analytic filter consists of the original real even filter plus the corresponding complex odd filter, all divided by 2. These two components of the analytic filter correspond to odd and even receptive fields.
17. D. A. Pollen and S. F. Ronner, "Phase relationship between adjacent simple cells in the visual cortex," Science **212,** 1409–1411 (1981).
18. A. B. Watson and A. J. Ahumada, Jr., "A look at motion in the frequency domain," in *Motion: Perception and Representation,* J. K. Tsotsos, ed. (Association for Computing Machinery, New York, 1983), pp. 1–10.
19. A. B. Watson and A. J. Ahumada, Jr., "Model of human visual-motion sensing," J. Opt. Soc. Am. A **2,** 322–342 (1985).
20. J. P. H. van Santen and G. Sperling, "Elaborated Reichardt detectors," J. Opt. Soc. Am. A **2,** 300–321 (1985).
21. E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," J. Opt. Soc. Am. A **2,** 284–299 (1985).
22. D. A. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1984).
23. R. E. Williamson and H. F. Trotter, *Multivariate Mathematics* (Prentice-Hall, Englewood Cliffs, N.J., 1974).
24. P. J. Burt, "Fast filter transforms for image processing," Comput. Graph. Image Process. **16,** 20–51 (1981).
25. A. Rosenfeld and A. C. Kak, *Digital Picture Processing* (Academic, New York, 1982), p. 191.
26. J. Max, "Quantizing for minimum distortion," IRE Trans. Inf. Theory **IT-6,** 7–12 (1960).
27. F. W. Campbell and J. J. Kulikowski, "Orientation selectivity of the human visual system," J. Physiol. **187,** 437–445 (1966).
28. J. Nachmias and R. Sansbury, "Grating contrast: discrimination may be better than detection," Vision Res. **14,** 1039–1042 (1974).
29. G. E. Legge and J. M. Foley, "Contrast making in human vision," J. Opt. Soc. Am. **70,** 1458–1471 (1980).
30. J. M. Foley and G. E. Legge, "Contrast detection and near-threshold discrimination in human vision," Vision Res. **21,** 1041–1053 (1981).
31. G. E. Legge, "A power law for contrast discrimination," Vision Res. **21,** 457–467 (1981).
32. H. R. Wilson, D. K. McFarlane, and G. C. Phillips, "Spatial frequency tuning of orientation selective units estimated by oblique masking," Vision Res. **23,** 873–882 (1983).
33. A. Bradley and I. Ohzawa, "A comparison of contrast detection and discrimination," Vision Res. **26,** 991–997 (1986).
34. D. J. Sharma and A. N. Netravali, "Design of quantizers for dpcm coding of picture signals," IEEE Trans. Commun. **COM-25,** 1267–1274 (1977).
35. G. C. Phillips and H. R. Wilson, "Orientation bandwidths of spatial mechanisms measured by masking," J. Opt. Soc. Am. A **1,** 226–232 (1984).
36. W. K. Pratt, *Digital Image Processing* (Wiley, New York, 1978).
37. G. Sperling, M. Landy, Y. Cohen, and M. Pavel, "Intelligible coding of ASL image sequences at extremely low information rates," Comput. Vision Graph. Image Process. **31,** 335–391 (1985).
38. D. E. Pearson and J. A. Robinson, "Visual communication at very low data rates," Proc. IEEE **73,** 795–812 (1985).
39. The 82% value is chosen for mathematical convenience (it is the probability reached when the Weibull exponent is 1). Any percentage point that can be accurately estimated would do as well.
40. A. B. Watson, K. R. K. Nielsen, A. Poirson, A. Fitzhugh, A. Bilson, K. Nguyen and A. J. Ahumada, Jr., "Use of a raster framebuffer in vision research," Behav. Res. Methods Instrum. **18,** 587–594 (1986).
41. A. B. Watson and D. G. Pelli, "QUEST: a Bayesian adaptive psychometric method," Percept. Psychophys. **33,** 113–120 (1983).
42. A. B. Watson, "Probability summation over time," Vision Res. **19,** 515–522 (1979).
43. J. G. Robson and N. G. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," Vision Res. **21,** 409–418 (1981).
44. N. Ahmed, T. Nataragan, and K. R. Rao, "Discrete cosine transform," IEEE Trans. Comput. **C-23,** 90–93 (1974).
45. J. W. Modestino, N. Farvardin, and M. A. Ogrinc, "Performance of block cosine image coding with adaptive quantization," IEEE Trans. Commun. **COM-33,** 210–217 (1985).
46. E. J. Delp and O. R. Mitchell, "Image compression using block truncation coding," IEEE Trans. Commun. **COM-27,** 1335–1342 (1979).
47. J. W. Woods and S. D. O'Neil, "Subband coding of images," IEEE Trans. Acoust. Speech Signal Process. **ASSP-43,** 1278–1288 (1986).
48. D. J. Sakrison, "Image coding applications of vision models," in *Image Transmission Techniques,* W. K. Pratt, ed. (Academic, New York, 1979).
49. J. O. Limb and C. B. Rubinstein, "On the design of quantizers for dpcm coders: a functional relationship between visibility, probability, and masking," IEEE Trans. Commun. **COM-26,** 573–578 (1978).
50. J. C. Candy and R. H. Bosworth, "Methods of designing differential quantizers based on subjective evaluations of edge busyness," Bell Sys. Tech. J. **51,** 1495–1516 (1972).
51. A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," Proc. IEEE **65,** 536–548 (1977).
52. W. F. Schreiber, C. F. Knapp, and N. D. Kay, "Synthetic highs—an experimental TV bandwidth reduction system," J. Soc. Motion Pict. Eng. **68,** 525–537 (1959).
53. M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second-generation image-coding techniques," Proc. IEEE **73,** 549–574 (1985).
54. M. S. Livingstone and D. H Hubel, "Anatomy and physiology of a color system in the primate visual cortex," J. Neurosci. **4,** 309–356 (1984).
55. D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," J. Physiol. **195,** 215–243 (1968).
56. D. H. Hubel and T. N. Wiesel, "Functional architecture of the macaque monkey visual cortex. Ferrier lecture," Proc. R. Soc. London Ser. B **198,** 1–59 (1977).
57. D. H. Hubel and T. N. Wiesel, "Sequence regularity and geometry of orientation columns in the monkey striate cortex," J. Comp. Neurol. **158,** 267–294 (1974).