Preserving properties of object shape by computations in primary visual cortex

Charles F. Stevens*

The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037

Contributed by Charles F. Stevens, September 9, 2004

Although our visual system is extremely good at extracting objects from the visual scene, this process involves complicated computations that are thought to require image processing by many successive cortical areas. Thus, intermediate stages in object extraction should not eliminate essential properties of the objects that are still required by later stages. A particularly important characteristic of an object is its shape, and shape has the property that it is unchanged by translations, rotations, and magnifications of the image. I show that the requirement for this property of shape to be preserved in the image, as represented by the firing of neurons in the primary visual cortex (V1), is equivalent to a particular type of computation, known as a wavelet transform, determining the firing rate of V1 neurons in response to an image on the retina. Experimental data support the conclusion that the neural representation of images in V1 is described by a wavelet transform and, therefore, that the properties of shape are preserved.

A well performed mechanical experiment, such as rolling balls down an inclined plane, gives the same result regardless of whether it is done in China or Italy and whether it was done yesterday or centuries ago. The remarkable thing, which was proved by Emmy Noether nearly a century ago, is that this simple observation (i.e., experimental results do not depend on the time or place at which the experiment is done) requires that energy and momentum must be conserved. This result is an example of what is called a “symmetry argument” because doing something, such as carrying out an experiment at a different place or time, has no effect on the result, just as rotating a 4-fold symmetric object such as a square by 90° does not change its appearance. Although arguments of this sort are very common in physics (1), they are rarely used in biology. Here, my goal is to use a symmetry argument to gain insights into how the cortex processes information. I identify abstract properties of an image that should be preserved in the cortical representation of that image (a kind of symmetry argument, because abstract image properties are preserved with cortical processing) and explore the consequences of this preservation for the types of calculations that the cortex can perform.

We see the world as filled with objects, but extracting them from the complicated pattern of light and dark that is presented to the retina requires complex computations by a sequence of cortical areas. One of the most important defining characteristics of an object is its shape. Because objects are not extracted in the primary visual cortex (V1) but rather in later areas (2), the representation of an image in V1 should not eliminate the essential properties of objects needed for computations by other areas. I show that this requirement for preserving abstract shape properties (described below) in the cortical representation of images restricts the computations that V1 can carry out to a particular class known as wavelet transforms (3–5). Indeed, analysis of experimental data reveals that V1 can be said to perform a wavelet transform of the image presented to the retina because the firing rate of each simple cell can be calculated from the image by a particular wavelet transform, and the image can, in principle, be reconstructed from the cell firing in V1 by an inverse wavelet transform.

Three-dimensional objects in the world form a two-dimensional image on the retina, and thus, I consider here only two-dimensional

shapes and not how the three-dimensional structure of an object is reconstructed from images of various views. As we move our eyes, tilt our head, and move around in the world, the image on the retina undergoes translations, rotations, and magnifications, a combination of manipulations that constitute a “similitude operation” (4). The defining property of a two-dimensional shape is that it is preserved by this similitude operation; the test for whether two shapes are the same is to determine whether they can be made to coincide by translating, rotating, and magnifying (or shrinking) one or the other of the shapes.

Each retinal image produces some particular firing pattern of neurons in the primary visual cortex, and I call the cortical firing pattern resulting from a particular image the “cortical representation” of that image. Because invariance under similitude operations is the defining property of shape, preserving shape in cortical representations means that there must be some manipulation corresponding to the similitude operation of the image that can make all cortical representations of a particular shape coincide. If the retinal image were simply copied to the cortex, or if it were filtered linearly (by blurring, for example), then just translation, rotation, and magnification could undo the application of a similitude operation to the image. (This linear filtering is how shape properties are preserved in retina and lateral geniculate.) However, as shown below, the explicit “cortical similitude operation” (required for a cortical computation beyond a simple linear filter) is not a combination of translations, rotations, and magnifications, but rather, it is another transformation that has the same role for the cortex that these operations have for the retinal image of a shape. I stress that the cortex does not necessarily use its version of the similitude operations to match shapes but, rather, that invariance under similitude operations is the abstract defining characteristic of shape to be preserved in the cortical representation of images.

Methods

A Simple Example. The purpose of this example is to illustrate, in a greatly simplified situation, the type of argument that I use here. In this illustration, the image similitude operation is reduced from the combination of translation, rotation, and magnification to a one-dimensional translation. A function $f(x)$ specifies the light intensity of an image at location x on a one-dimensional retina that stretches, for convenience, over the entire real line. This image is represented in a one-dimensional cortex by the function $\hat{f}(\omega)$, the firing rate of a neuron at cortical position $\omega \in [-\infty, \infty]$, calculated from the following inner product:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} dx \phi_{\omega}^{*}(x) f(x) \equiv (\phi_{\omega}(x), f(x)).$$

The function $\phi_{\omega}(x)$ in this integral describes the receptive field of a neuron at location ω in the cortex, with the subscript ω indicating

Freely available online through the PNAS open access option.

Abbreviation: V1, primary visual cortex.

*E-mail: stevens@salk.edu.

© 2004 by The National Academy of Sciences of the USA

the position of the neuron and the value of $\varphi_\omega(x)$ giving the response of that neuron (the firing rate) to a unit point of light at location x on the retina. Note that the cortical position ω serves as an index for the quantitative properties of the receptive field of the neurons at that position. It results that each neuron requires a parameter to specify the precise mathematical form of its receptive field, and it is convenient to require that the position variable serve also as the receptive field parameter; in general, the receptive field parameter would be some function of the position value ω , whereas here, this function is simply taken as an equality. In anticipation of arguments to follow, the receptive field function $\varphi_\omega(x)$ has been treated as possibly complex, in which case two neurons would be required for each cortical position ω , one neuron for the real part and the other neuron for the imaginary part of $\varphi_\omega(x)$.

The goal is to show that $\phi_\omega(x) = e^{i\omega x}$, if and only if the translation symmetry of the image ($f(x)$) is preserved by the inner product (described below), and the transformation is not just a linear filtering of the image.

The translation operator T_t , defined by $T_t f(x) = f(x - t) = e^{-itd/dx} f(x)$, is unitary and forms a representation of the additive group with elements x that are composed according to $x \circ y = x + y$, have the inverse $(x)^{-1} = -x$, and have the identity $e = 0$. The translation operator $T_t = e^{-itd/dx}$ can be found from a Taylor's expansion of $f(x - t)$. Suppose that the inner product of $\varphi_\omega(x)$ with the translated image $f(x - t) = T_t f(x)$ is denoted (notice subscript, which specifies the magnitude of the translation) by $\hat{f}_t(\omega) = (\phi_\omega(x), f(x - t))$. Then, "the translational symmetry is preserved by the inner product" means that there exists an operator U_t , isomorphic to the additive group, whose action is to reverse in the ω domain, the effect of T_t in the x domain, such that $U_t \hat{f}_t(\omega) = \hat{f}(\omega)$. If the retinal image, or a filtered (for example, blurred) version of it, were copied to the cortex, then translations in the cortex would correspond to translations in the retina. However, I wish to illustrate the effect of a computation that is more complex than copying or filtering. To exclude linear filtering of the image for this illustrative example, I require that U_t be some function that depends on t , which is the magnitude of the translation in visual space.

If $\phi_\omega(x) = e^{i\omega x}$, then $\hat{f}(\omega) = (e^{i\omega x}, f(x))$ is by definition the Fourier transform of $f(x)$. A well known property of Fourier transforms is that $T_t \Leftrightarrow e^{-i\omega t}$ so that the cortical operation that undoes the effects of T_t on the image is $U_t = e^{i\omega t}$. Furthermore, because $e^{i\omega a} e^{i\omega b} = e^{i\omega(a+b)}$, it is clear that the operator U is isomorphic to the additive group. Thus, if the receptive fields are $\phi_\omega(x) = e^{i\omega x}$ (with one neuron for the real part and one neuron for the imaginary part), then the translational symmetry of the image is preserved in the cortical representation.

Suppose that the unitary operator U_t exists and that, as described above, $U_t U_u = U_{t+u}$ so that U_t , like T_t , is isomorphic to the additive group. The functional equation for U implies that $U_t = \exp(u_o(\omega)t)$ for some function $u_o(\omega)$; $u_o(\omega) = dU_t/dt$ evaluated at $t = 0$. Then we have the following:

$$\begin{aligned} U_t \hat{f}_t(\omega) &= U_t(\phi_\omega(x), T_t f(x)) = U_t(T_{-t} \phi_\omega(x), f(x)) \\ &= (e^{i\omega o(\omega)t} e^{-itd/dx} \phi_\omega(x), f(x)) = (\phi_\omega(x), f(x)), \end{aligned}$$

so $e^{-itd/dx} \phi_\omega(x) = e^{-i\omega o(\omega)t} \phi_\omega(x)$, which means that $d\phi_\omega(x)/dx = u_o(\omega) \phi_\omega(x)$. If $u_o(\omega)$ is a function of ω (as I have required), then this equation indicates that $\phi_\omega(x) = \exp(i2\pi\omega x)$ is the eigen vector of d/dx and $u_o(\omega) = i2\pi\omega$ is its eigen value. Thus, I have shown that the cortical action takes a Fourier transform.

Note that the technique used here, which depends on recognizing that the eigen functions of the translation operator are $\exp(i2\pi\omega x)$, is like the one that shows translational invariance in quantum mechanics, implies momentum conservation, and identifies the momentum operator.

A Second Simple Example. The following example is used to carry out my arguments for a simplified situation with a one-dimensional similitude operation that includes both a one-dimensional translation and a magnification. This example is the same as the preceding example [one-dimensional retina with image $f(x)$], except the cortex is two-dimensional (for the reasons given below) with the coordinates ω and ξ , and the image may be scaled as well as translated. As shown above, the position coordinates will, for convenience, be used also to specify the parameters of the receptive field. Again, I seek to determine the receptive field $\varphi_{\omega,\xi}(x)$ of a neuron at location (ω, ξ) so that the firing rate of the cell is $\hat{f}(\omega, \xi) = (\phi_{\omega,\xi}(x), f(x))$ for the image $f(x)$; the parameters (ω, ξ) that give the neuron location are subscripts, emphasizing the fact that they are not variables in the inner product calculation used to compute the firing rate of the neuron from the image $f(x)$ and the receptive field $\varphi_{\omega,\xi}(x)$. If the image is translated (by t) and scaled (by s) (the unitary operator $G_{s,t} = S_s T_t$ carries out this operation), the result is $G_{s,t} f(x) = \sqrt{s} f(s(x - t))$ and the cortical representation of the scaled and translated image is $\hat{f}_{s,t}(\omega, \xi) \equiv (\phi_{\omega,\xi}(x), G_{s,t} f(x))$. The $\sqrt{s}$ appears in the effect of $G_{s,t}$ so that the integral of $f^2(x)$ is unaffected by the scaling.

The operators $G_{s,t}$ are isomorphic to the similitude group (4) with the composition $(\sigma, \tau) \circ (s, t) = (s\sigma, t + \tau/s)$, the inverse $(s, t)^{-1} = (1/s, -ts)$, and the identity $e = (1, 0)$. Considering shape properties preserved in the cortical representation is equivalent to requiring the existence of an unitary operator $U_{s,t}$, isomorphic to the similitude group, that reverses in the (ω, ξ) domain the effects of $G_{s,t}$ in the x domain as follows: $U_{s,t} \hat{f}_{s,t}(\omega, \xi) = \hat{f}(\omega, \xi)$. For $U_{s,t}$ to be isomorphic to the similitude group (necessary for preserving shape properties) and to operate on $\hat{f}(\omega, \xi)$ in the simplest possible way, I require that $U_{s,t} \hat{f}(\omega, \xi) = \hat{f}(s\omega, \xi + t/\omega)$. Note that by requiring the cortical representation to be two-dimensional (not one-dimensional, like the image), I have excluded linear filtering as the cortical computation. The goal is to show that $\hat{f}(\omega, \xi)$ is the result of a wavelet transform, with mother wavelet $\psi(x)$, if and only if shape properties are preserved in the cortical representation of the image $f(x)$.

First, suppose that $\hat{f}(\omega, \xi)$ is a wavelet transform (3, 5) of $f(x)$ so that $\hat{f}(\omega, \xi) \equiv (G_{\omega, \xi} \psi(x), f(x))$; the daughter wavelets are generated from the mother by translations and scalings (3, 5). The representation of a scaled (by s) and translated (by t) image $\hat{f}_{s,t}(\omega, \xi)$ is given by the following:

$$\begin{aligned} \hat{f}_{s,t}(\omega, \xi) &= (G_{\omega,\xi} \psi(x), G_{s,t} f(x)) = (G_{1/s, -ts} G_{\omega, \xi} \psi(x), f(x)) \\ &= \hat{f}(\omega/s, \xi - ts/\omega). \end{aligned}$$

Define the action of the operator $U_{s,t}$ by $U_{s,t} \hat{f}(\omega, \xi) = \hat{f}(s\omega, \xi + t/\omega)$. The effect of $U_{s,t}$ on the transformed image is $U_{s,t} \hat{f}(\omega, \xi) = U_{s,t} \hat{f}(\omega/s, \xi - ts/\omega) = \hat{f}(\omega, \xi)$, so $U_{s,t}$ does reverse the effect of $G_{s,t}$ if object shape properties are preserved when $\hat{f}(\omega, \xi)$ is calculated by a wavelet transform. Thus, if the cortical representation is produced by a wavelet transform, the shape symmetries exhibited by the image are preserved in the cortical representation.

Suppose that the operator $U_{s,t}$ identified above exists, and consider the following relation:

$$\begin{aligned} U_{s,t} \hat{f}_{s,t}(\omega, \xi) &= U_{s,t}(\phi_{\omega,\xi}(x), G_{s,t} f(x)) = U_{s,t}(G_{1/s, -ts} \phi_{\omega, \xi}(x), f(x)) \\ &= U_{s,t}(\sqrt{1/s} \phi_{\omega, \xi}((x + ts)/s), f(x)) \\ &= (U_{s,t} \sqrt{1/s} \phi_{\omega, \xi}((x + ts)/s), f(x)) \\ &= (\phi_{\omega, \xi}(x), f(x)) = \hat{f}(\omega, \xi), \end{aligned}$$

which means that $U_{s,t} \sqrt{1/s} \phi_{\omega, \xi}((x + ts)/s) = \phi_{\omega, \xi}(x)$. By changing variables $x \rightarrow s(x - t)$, one gets $U_{s,t} \phi_{\omega, \xi}(x) =$

$\sqrt{s}\phi_{\omega,\xi}(s(x-t))$. Recall that $U_{s,t}$ operates only on ω and ξ , not on the argument x , as described above. Thus,

$$U_{s,t}\phi_{\omega,\xi}(x) = \phi_{\omega s, \xi + t/\omega}(x) = \sqrt{s}\phi_{\omega,\xi}(s(x-t)) = G_{s,t}\phi_{\omega,\xi}(x)$$

holds, and this relation must still be satisfied for any values of ω and ξ . Set $\omega = 1$ and $\xi = 0$, and define $\phi_{1,0}(x) \equiv \psi(x)$. Then, we have the following:

$$\phi_{s,t}(x) = \sqrt{s}\psi(s(x-t)) = G_{s,t}\psi(x),$$

which means that $\hat{f}(\omega, \xi)$ is the result of a wavelet transform if $\psi(x)$ has the properties of a mother wavelet [the integral of $\psi(x)$ must vanish, and the square of the function must possess certain convergence properties (3, 5)]. Any function $\psi(x)$ is satisfactory as long as the function can be a mother wavelet so that the transformation carried out in the cortex is invertible (and, thus, does not destroy information about the object shape). Thus, if the shape symmetries of the one-dimensional image are preserved and information about the shape is not lost by the cortical processing, the two-dimensional cortical representation must be produced by a wavelet transform.

The V1 Computation. If we now consider a two-dimensional retina with an image described by $f(\mathbf{r})$, where $\mathbf{r}$ is a vector giving the x and y coordinates on the retina, and define the operator $G_{\omega,\theta,\rho} = S_\omega R_\theta T_\rho$ that acts on the image (S_ω is the scaling operator described above, R_θ rotates the image by θ , and T_ρ is the translation operator, as described), then the translation is now two-dimensional and specified by the vector ρ . Specifically, the effect of $G_{\omega,\theta,\rho}$ is as follows:

$$G_{\omega,\theta,\rho}\psi(\mathbf{r}) = \omega\psi(\omega(\mathbf{r}_\theta - \rho_\theta)),$$

where $\mathbf{r}_\theta$ is given by the following:

$$\mathbf{r}_\theta = \begin{bmatrix} x \\ y \end{bmatrix}_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \equiv R_\theta \mathbf{r},$$

with an analogous definition for ρ_θ . The firing rate of the cortical neuron at location (ω, θ, ρ) is given by $\hat{f}_{\omega,\theta,\rho}(\mathbf{r}) = (\phi_{\omega,\theta,\rho}(\mathbf{r}), f(\mathbf{r}))$, with the receptive field of the neuron being $\phi_{\omega,\theta,\rho}(\mathbf{r})$. The operator that reverses, in the cortical (ω, θ, ρ) domain, the effect of $G_{\omega,\theta,\rho}$ on the image is now given as $U_{\omega,\theta,\rho}$, and both operators are unitary.

The similitude group uses the composition (4) as follows:

$$(a, b, c) \circ (\omega, \theta, \rho) = (a\omega, \theta + b, \rho_\theta + c_\theta/\omega),$$

and it has the identity element $e = (1, 0, 0)$. As described, the shape symmetries are preserved in the V1 simple cells if and only if the V1 computation represented by these cells is a wavelet transform (linear filtering of the image as the cortical computation is excluded by requiring that the cortical representation be four-dimensional). The arguments are the same as for the preceding one-dimensional case.

Results

A Simple Example. The power of the type of argument that I use here can be illustrated by an example in which the situation is reduced to its simplest, one-dimensional form. In this example, the two-dimensional translation, rotation, and magnification is replaced by a one-dimensional translation. Picture a "Flatland" creature with a one-dimensional retina that sends information about the retinal image to a one-dimensional visual cortex. I suppose that this creature cannot move around its two-dimensional world but can move its eye so that images fall at different positions on the retina. We provide different objects for the creature to look at, with objects of one shape producing a

pattern of light intensities $f(x)$ on the creature's retina and another shape giving a pattern $g(x)$, where x specifies a particular position along the single dimension of the retina. Because the creature can move its eye, the image of the object can be translated from one retinal position to another. This possibility for eye movement means that an object producing a retinal image $f(x)$ at one time might, at another time, produce the image $f(x+t)$ (the image being translated to the left by the amount t) because the eye moved. The point of this example is to reduce the problem of comparing shapes of two objects to a single operation, translation; i.e., two images (functions) have the same shape if they can be made to coincide by translating one or the other function along the x axis.

The position of a neuron in the creature's cortex is specified by the variable ω , and the receptive field of a neuron at location ω is given by $\phi_\omega(x)$, such that the firing rate of the neuron, due to a point of light at x , is proportional to $\phi_\omega(x)$. The receptive field function $\phi_\omega(x)$ could also be written as $\phi(\omega, x)$, but I have chosen to use a subscript for the variable ω to emphasize that, when calculating the firing rate of a particular neuron, this quantity is kept constant. The effects of light at different locations are assumed to add linearly, which means that the firing rate $\hat{f}(\omega)$ of the neuron at cortical location ω , due to the retinal image $f(x)$, is the sum of all of the separate contributions of the image:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} \phi_\omega^*(x) f(x) dx \equiv (\phi_\omega(x), f(x)).$$

Here, a shorthand notation for the integral that will be used throughout appears on the right side of this equation; the asterisk denotes the complex conjugate, and the reason for its presence is given in *Methods*. Thus, the cortical representation of the image $f(x)$, is the firing rate of cortical neurons $\hat{f}(\omega)$.

The shape of an object in this simple example determines the function $f(x)$, so two objects with the same shape would give the same pattern $f(x)$ of light intensity on the retina. If the image of an object is translated to the left by t along the retina, then the retinal image would be $f(x+t)$, and this translated image could be made equal to the original image $f(x)$ by applying the translation operator T_t to $f(x+t)$ to make it equal $f(x) = T_t f(x+t)$; note that the subscript t is used to specify the size of the translation that is produced. The operator T_t acts on the function $f(x)$ to translate it to the right along the x axis by the amount t . Thus, a translation test would be the way of deciding whether two retinal images correspond to the same object shape. Because two translations in a row, one by amount t and the second by amount u , have the same effect as a single translation by $t+u$, the translation operator T must have the property $T_t T_u = T_{t+u}$.

If this translation invariance of object shape is to be preserved in the cortex, then there has to be some cortical operation U_t (corresponding to the retinal operator T_t) that makes representations of the object and its translated version equal. In other words, if $\hat{f}(\omega)$ is the cortical representation of the image of an object, and $\hat{f}_t(\omega)$ is the cortical representation of the translated image (note that the subscript t that specifies how much the retinal image had been translated), then there has to be some U_t that has the following effect:

$$\hat{f}_t(\omega) = U_t \hat{f}(\omega),$$

so the original representation and the representation of the translated image can be made to coincide. Like T , U should have the property that $U_t U_u = U_{t+u}$.

One way to preserve this shape property is for the retinal image and its cortical representation to be the same. Then, translations in the retinal image would correspond to transla-

tions of the cortical representation. However, the cortex would not have done a real computation if this were true. I show in *Methods* that, if the cortex is to carry out a more interesting computation to get the cortical representation of an image (a representation for which U does not simply produce a translation), and if the properties of shape are preserved in the representation, then that computation must be a Fourier transform. Each place ω in the cortex then requires (see *Methods*) a pair of neurons, one neuron for a cosine wave (neuron 1) and the other neuron for a sine wave (neuron 2), so that the receptive fields of the neurons at cortical location ω are as follows:

$$\varphi_{\omega,1}(x) = \cos(2\pi\omega x) \quad \text{and} \quad \varphi_{\omega,2}(x) = \sin(2\pi\omega x).$$

The firing rates $\hat{f}_1(\omega)$ and $\hat{f}_2(\omega)$ of neurons 1 and 2 at location ω are given by the following:

$$\hat{f}_1(\omega) = (\cos(2\pi\omega x), f(x)) = \int_{-\infty}^{\infty} \cos(2\pi\omega x) f(x) dx, \quad \text{and}$$

$$\hat{f}_2(\omega) = (\sin(2\pi\omega x), f(x)) = \int_{-\infty}^{\infty} \sin(2\pi\omega x) f(x) dx.$$

These integrals compute just the Fourier transform of $f(x)$, so $\hat{f}_1(\omega)$ and $\hat{f}_2(\omega)$ are the weights for the $\cos(2\pi\omega x)$ and $\sin(2\pi\omega x)$ terms in the Fourier expansion of $f(x)$. In this sense, the cortex of our Flatland creature calculates the Fourier transform of the image. All of the information about the image appears in the cortical representation because $f(x)$ can be reconstructed from $\hat{f}_1(\omega)$ and $\hat{f}_2(\omega)$ by an inverse Fourier transform.

A well known property of Fourier transforms is that a translation by amount t in the spatial (x) domain corresponds to multiplications by $\cos(2\pi\omega t)$ and $\sin(2\pi\omega t)$ in the frequency (ω) domain. Thus, the operator U_t is not just a translation of the cortical representation but, rather, a scaling of all of the cortical neuronal firing rates by appropriate sines and cosines.

A Second Simple Example. The arguments that I will apply to the actual V1 can be simplified by modifying the preceding example to consider the similitude operation consisting of a one-dimensional translation and magnification. Suppose that the Flatland creature can move around its two-dimensional world such that the image of an object not only undergoes translations, as described, but is also magnified (or shrunk) as the creature moves closer to (or farther from) the object. The shape of the object is unchanged by a size scaling operation S_s (in which the image is s times smaller on the retina if $s > 1$, unchanged if $s = 1$, and magnified by $1/s$ if $s < 1$) and a translation operation T_t . It results (see *Methods*) that permitting both scaling and translation will require a two-dimensional cortex in which cell location is specified by two variables, ξ and ω . The receptive field of the neuron at location (ω, ξ) is $\varphi_{\omega,\xi}(x)$ so that, as described, the firing rate $\hat{f}(\omega, \xi)$ of that neuron, for the retinal image $f(x)$, is as follows:

$$\hat{f}(\omega, \xi) = (\varphi_{\omega,\xi}(x), f(x));$$

the firing rates of cortical neurons $\hat{f}(\omega, \xi)$ then constitute the cortical representation of the image $f(x)$.

To extend the analysis given above, I require the following. An operator $U_{s,t}$ must exist that acts on the cortical representation to undo the effect of translation and magnification of the image. If the image $f(x)$ is translated (by t) and scaled (by s) by the operator $G_{s,t} = S_s T_t$, the cortical representation becomes $\hat{f}_{s,t}(\omega, \xi)$. That is, the firing rate $\hat{f}_{s,t}(\omega, \xi)$ of the neuron at cortical position (ω, ξ) , in response to the magnified and translated image, has to be able to be converted into the representation $\hat{f}(\omega,$

$\xi)$ for the original image by an operator U that has the following effect:

$$U_{s,t} \hat{f}_{s,t}(\omega, \xi) = \hat{f}(\omega, \xi).$$

Thus, U is an operator that can be applied to the cortical representation to reverse the effect of G acting on the image $f(x)$. The operator U has to share the property $U_{s,t} U_{\sigma,\tau} = U_{s\sigma, \tau+t/\sigma}$ (see *Methods*) with the operator G that scales and translates the retinal image.

What properties must the receptive field $\varphi_{\omega,\xi}(x)$ of the neuron at cortical location (ω, ξ) have in order for the operator U to exist? In other words, what can we say about the receptive fields if the shape properties are to be preserved in the cortical representation? In *Methods*, I show that, in order to preserve shape properties, the cortex must have neurons whose receptive fields are generated by translations and scalings according to the following:

$$\varphi_{\omega,\xi}(x) = G_{\omega,\xi} \psi(x)$$

for all possible pairs of ω and ξ and for some function (which must be determined by experiment) $\psi(x)$. A function $\psi(x)$ that can be used to generate all of the receptive fields in this way is called a “mother receptive field.” Any function ψ with certain properties (5) can be used for this argument, but a specific example of a good function is as follows:

$$\psi_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cos(2\pi x) \quad \text{and} \quad \psi_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sin(2\pi x)$$

for a pair of neurons (1 and 2) at each ω location in the cortex. Therefore, for the neuron at position (ω, ξ) in the cortex, one of the two neurons has the following receptive field:

$$\varphi_{\omega,\xi}(x) = G_{\omega,\xi} \psi(x) = \frac{\sqrt{\omega}}{2\pi} e^{-\frac{(x-\xi)^2}{2\omega^2}} \cos(2\pi\omega(x - \xi)),$$

so that the scaling by ω “compresses” the x axis and thus specifies the frequency of the cosine wave; the second neuron at this location would have a receptive field of the same form except that a sine would appear instead of the cosine. These functions are the one-dimensional Gabor functions that, in their two-dimensional versions, describe receptive fields in the actual V1 (6, 7).

The firing rate of the pair of neurons (1 and 2) at cortical location (ω, ξ) is as follows:

$$\begin{aligned} \hat{f}_1(\omega, \xi) &= (G_{\omega,\xi} \psi_1(x), f(x)) \\ &= \frac{\sqrt{\omega}}{2} \int_{-\infty}^{\infty} e^{-\omega^2(x-\xi)^2/2} \cos(2\pi\omega(x - \xi)) f(x) dx, \quad \text{and} \end{aligned}$$

$$\begin{aligned} \hat{f}_2(\omega, \xi) &= (G_{\omega,\xi} \psi_2(x), f(x)) \\ &= \frac{\sqrt{\omega}}{2} \int_{-\infty}^{\infty} e^{-\omega^2(x-\xi)^2/2} \sin(2\pi\omega(x - \xi)) f(x) dx. \end{aligned}$$

In the first example, the computation that preserved the properties of shape in the cortical representation of an image turned out to be a Fourier transform. Can a mathematical operation be identified for the present example? What I have described by the integrals given above is known as a wavelet transform, which is a mathematical operation that is in many ways similar to a Fourier transform (4, 5). For example, no information is lost when the transform is carried out because, like the Fourier transform, the wavelet transform of an image can be inverted to

give back the original image; that is, $f(x)$ can be calculated from $\hat{f}_1(\omega, \xi)$ and $\hat{f}_2(\omega, \xi)$ by an inversion equation. Because the firing rates $\hat{f}_1(\omega, \xi)$ and $\hat{f}_2(\omega, \xi)$ are the weights in an expansion that uniquely specify $f(x)$, it could be said that the creature's cortex "takes the wavelet transform of the image." The operator U described above that acts in the cortical domain to undo the effects of translations and scalings of the image does not itself translate and scale the cortical representation; its actual form is given in *Methods*.

The V1 Computation. The data above can be summarized by saying that the essential properties of shape are preserved in the two-dimensional cortical representation of a one-dimensional image when the cortex carries out a wavelet transform. Here, "carries out" means that the quantities resulting from a wavelet transform, $\hat{f}_1(\omega, \xi)$ and $\hat{f}_2(\omega, \xi)$, are the firing rates of cortical neurons and that all of the information in the image has been retained because it can, in principle, be reconstructed from these firing rates by an inverse wavelet transform. In going from a one-dimensional to two dimensional image, the only new property of shape that is added to translation and scaling invariance is the invariance with rotation. All of the arguments that led to the one-dimensional wavelet transform in the preceding example work equally well for the two-dimensional wavelet transform; i.e., the properties of object shape are preserved in the cortical representation if V1 carries out a two-dimensional wavelet transform. However, each receptive field now must be generated by an operator G that acts on a two-dimensional function $\psi(x, y)$, which is a mother receptor field, to give the following receptive field:

$$\phi_{\omega, \theta, \xi, \eta}(x, y) = G_{\omega, \theta, \xi, \eta} \psi(x, y);$$

this operator G translates (by ξ in the x direction and η in the y direction), rotates (by θ), and scales (by ω) the function $\psi(x, y)$ to specify the receptive field of the neuron indexed by the four parameters (ω, θ, ξ , and η), which are parameters that I also use to specify the location of the neuron in the cortex. As in the preceding example, the function $\psi(x, y)$ must be determined by experiment.

The Measured V1 Receptive Fields. Receptive fields of simple cells in V1 have been measured and are found to be well described by Gabor functions (6–9). That is, the receptive field function $\phi_{\theta, \xi, \eta}(x, y)$ is, up to a multiplicative constant that relates stimulus magnitude to neuron firing rate, found experimentally to be as follows:

$$\phi_{\theta, \xi, \eta}(x, y) = e^{-\frac{(x\theta - \xi\theta)^2}{2\sigma_x^2}} e^{-\frac{(y\theta - \eta\theta)^2}{2\sigma_y^2}} \cos(2\pi x\theta\omega - \alpha),$$

where the receptive field center is $(\xi\theta, \eta\theta)$, its preferred orientation is θ , and $x\theta$ and $y\theta$ indicate the x and y coordinates that have been rotated by amount θ . This experimentally determined receptive field function depends (in addition to a multiplicative constant that relates firing rate to light intensity) on seven parameters: three parameters (ξ, η , and θ) determine receptive field center and preferred orientation, and four other parameters ($\sigma_x, \sigma_y, \omega$, and α) have values that depend on which particular neuron is being studied. The parameters σ_x and σ_y determine the lengths of the receptive field in the x and y direction, respectively; the parameter ω is the spatial frequency of the sinusoidal component ($\lambda = 1/\omega$ gives the wave-length preference for the receptive field); and α determines the phase of the sinusoidal component (if $\alpha = 0$, it is a cosine wave; if $\alpha = 90^\circ$, it is a sine wave; and if α is something else, the result is a linear combination of sine and cosine waves). All of these parameters have been measured for simple cells in V1 of cat and monkey.

Fig. 1. Parameters characterizing the receptive fields of simple cells in cat V1, as measured by Jones and Palmer (8, 10). (a) A plot of the standard deviations of the Gabor Gaussians in the y direction (σ_y , length of the receptive field) vs. the x direction (σ_x = width of the receptive field), with a best-fit straight line (slope, 1.67). (b) A plot of the wave length ($\lambda = 1/\omega$) of the Gabor sinusoid vs. the larger standard deviation (σ_y) of the Gabor Gaussian (slope, 0.6).

The experimentally measured receptive field function looks like what is required to preserve the properties of shape in its V1 representation, except the descriptive equation given above provides for the possibility that the parameters ($\sigma_x, \sigma_y, \lambda = 1/\omega$) can vary independently, whereas, according to the preceding analysis, the parameters specifying receptive field size should all covary because the size scale is set by the scaling operator S_ω with a single scale factor ω . Jones and Palmer (8, 10) measured the triplet of parameters ($\sigma_x, \sigma_y, \lambda = 1/\omega$) for several dozen simple cells in cat V1. Their data, shown in Fig. 1, show that all three parameters are proportional to one another. Because of the proportionality relations among σ_x, σ_y , and $\lambda = 1/\omega$, the simple cell receptive fields can be described by the following application of the operator $G_{\omega, \theta, \xi, \eta}$ to a mother wavelet:

$$\psi(x, y) = e^{-\frac{x^2}{2} - \frac{y^2}{2(1.67)^2}} \cos(2\pi x),$$

(the factor 1.67 in Fig 1a) or to a mother wavelet with a sine function replacing the cosine. The Jones and Palmer estimates for the phase for the sinusoid are rather scattered and not very accurate; thus, it is possible that phase-shifted sines and cosines occur in the receptive field descriptions. However, Ringach (9) has confirmed the Jones and Palmer analysis for monkey V1, and he finds that sines and cosines predominate.

Thus, the receptive fields of simple cells are scaled as required for a wavelet transform and each receptive field is characterized by four parameters: two parameters to specify the receptive field

center (ξ and η), one parameter for the preferred orientation (θ), and a fourth parameter ($\lambda = 1/\omega$) for the size of the receptive field. Thus, in summary, the Jones and Palmer (8) and Ringach (9) data support the conclusions that the representation of the retinal image by simple cells in V1 is the result of a wavelet transform with a Gabor function as the mother wavelet and, therefore, that the properties of shape are preserved in the representation.

Discussion

By using a symmetry argument, I have shown that preserving properties of an object's shape in its cortical representation is equivalent to having simple cells in V1 perform a wavelet transform. Furthermore, measured receptive fields in V1 have the right characteristics for V1 simple cells to be calculating coefficients of a wavelet transform. Thus, a compact description of simple cell behavior is that the firing of these neurons specifies the magnitude of coefficients in a wavelet transform of the image, and therefore, that the properties of shape are preserved in the image as represented by the simple cell discharge.

Because Gabor functions are suitable mother wavelets, other researchers have noticed that V1 may perform a wavelet transform. For example, Mallat (11) discussed this possibility and cited appropriate literature. However, the wavelet transform was not generated in this case from a mother wavelet by a two-dimension similitude operation but rather by two one-dimensional similitude operations (two translations and two scalings rather two translations, a rotation, and a scaling).

A perfect wavelet transform requires not only that receptive fields have the correct shape but also that the size of the response varies in a particular way with the scaling factor ω . Because I have not considered the actual sizes of the response here, a more

accurate characterization of the result is that V1 simple cells represent a wavelet transform up to some multiplicative factor that depends on ω . Exactly how the relative amplitudes of the wavelet coefficients behaves was not determined.

The retina does not sample the image uniformly but, rather, samples most densely in the fovea and then progressively more sparsely with eccentricity (12). This nonuniform sampling gives rise to the well known cortical magnification factor (13, 14). Because the structure of V1 seems uniform with eccentricity, it can be noted that V1 carries out a wavelet transform on a distorted version of the image (strongly compressed towards the edges) and not on the image itself.

I have not considered which parts of the wavelet transform take place in retina, lateral geniculate, or V1, nor have I considered the cellular mechanisms that are used to generate the receptive field structure. My neglect of these issues is not intended to minimize their importance; certainly, how cortical computations are implemented is a central problem in neurobiology. Nevertheless, I believe that one can legitimately examine the question of what computation is performed independently, to some degree, of how it is implemented by the responsible neuronal circuits.

This article has been an initial attempt to use symmetry arguments to provide insight into the computations performed in the nervous system, but I anticipate that this approach could have other applications in neurobiology.

I thank H. Abarbanel, L. Abbott, D. Chklovski, V. Klyachko, and J. Snider for comments on earlier drafts. Much of the work described here was carried out at the Aspen Center for Physics (Aspen, CO) and the Santa Fe Institute (Santa Fe, NM), and I thank these institutions for their hospitality.

- Georgi, H. (1999) *Lie Algebras in Particle Physics* (Westview, Boulder, CO).
- Malach, R., Levy, I. & Hasson, U. (2002) *Trends Cogn. Sci.* **6**, 176–184.
- Grossmann, A., Morlet, J. & Paul, T. (1985) *J. Math. Phys.* **26**, 2473–2479.
- Van Den Berg, J. C. (2004) (Cambridge Univ. Press, Cambridge), pp. 453.
- Louis, A. K., Maass, P. & Rieder, A. (1997) *Wavelets, Theory and Applications* (Wiley, New York).
- Daugman, J. G. (1985) *J. Opt. Soc. Am. A* **2**, 1160–1169.
- Marcelja, S. (1980) *J. Opt. Soc. Am.* **70**, 1297–1300.
- Jones, J. P. & Palmer, L. A. (1987) *J. Neurophysiol.* **58**, 1233–1258.
- Ringach, D. L. (2002) *J. Neurophysiol.* **88**, 455–463.
- Jones, J. P. & Palmer, L. A. (1987) *J. Neurophysiol.* **58**, 1187–1211.
- Mattat, S. (1999) *A Wavelet Tour of Signal Processing* (Academic, San Diego).
- Rolls, E. T. & Cowey, A. (1970) *Exp. Brain Res.* **10**, 298–310.
- Talbot, S. A. & Marshall, W. H. (1941) *Am. J. Ophthalmol.* **24**, 1255–1264.
- Daniel, P. M. & Witteridge, D. (1961) *J. Physiol.* **159**, 203–211.