Infomax and Receptive Field Localization^{*1}

Kenji Okajima and Hitoshi Imaoka

Fundamental Research Laboratories, NEC Corporation^{*2}

We investigated the receptive fields of neurons in the visual cortex from a theoretical viewpoint. It was assumed that if two regions in the visual field are located far apart, local images that fall on these regions are mutually independent. An information-maximization solution was used to determine the optimal receptive field by first assuming a low signal-to-noise ratio limit. We found that the optimal receptive field is a spatially localized, Gabor-function-like receptive field, similar to that of a simple cell.

1. Introduction

Neurons in the visual cortex of cats and monkeys have spatially localized and orientation-selective receptive fields (RFs).¹⁾ A number of studies have been done to clarify the properties of these RFs from an information theoretical viewpoint.^{2–11)} Olshausen and Field²⁾ demonstrated that simple-cell-like RFs emerge by learning a "sparse code" for natural images. Bell and Sejnowski³⁾ obtained similar results by using independent component analysis (ICA) of natural images. We can consider that these studies searched for a "minimum-entropy code" for natural images⁴⁾ that represents natural images efficiently by making the cells' activities as independent as possible.

Linsker^{5,6)} suggested another approach called Infomax, in which the optimal RF is determined by maximizing the amount of information obtained from the cell's response. Okajima,^{9,10,11)} and Okajima and Imaoka¹²⁾ derived simple-cell-like¹⁰⁾ or complex-cell-like^{11,12)} RFs from Infomax by using a low signal-to-noise ratio (SNR) approximation. The results obtained by using Infomax^{9–12)} and those obtained by using sparse coding²⁾ or ICA³⁾ seem related, although the nature of the relationship is still unknown.

The Infomax approach used by Okajima^{10,11)} and by Okajima and Imaoka¹²⁾ could not derive a spa-

tially localized RF from Infomax itself. They had to assume a "localization term" in their objective functions.^{10–12)} Our objective in this paper is to refine their theory so that one can derive the RF localization from Infomax itself.

2. Infomax and the objective function

The spatially localized and orientation- (spatial-frequency-) selective RF of a simple cell can be approximated by a Gabor function,^{13–16)} which in turn is defined by a sinusoidal function localized by a Gaussian envelope:

$$\phi_{k,\theta}(x) = G_{\sigma}(x) \cos(kx + \theta) \quad (2.1)$$

where $G_{\sigma}(x)$ represents a Gaussian function, and θ represents the phase. In this paper we refer to equation (2.1) as a Gabor function and the equation $G_{\sigma}(x) \exp(i(kx + \theta))$ as a "complex Gabor function." Figure 1 shows that a two-dimensional Gabor function has an anisotropic profile. This profile is responsible for the orientation selectivity of a simple cell. The selective spatial frequency is represented by $|k|$.

A question arises as to why the visual system analyzes its input in this way. To answer this question, we modelled the output of a simple cell with the following function:

$$a = (\phi, f+n) \equiv \sum_x \phi(x) \{f(x) + n(x)\} \quad (2.2)$$

Then we applied this model to determine an optimal RF $\phi(x)$ that maximizes the amount of infor-

^{*1} Received March 4, 2002; accepted July 26, 2002

^{*2} 34, Miyukigaoka, Tsukuba, 305–8501 Japan

Fig. 1 Examples of two-dimensional Gabor functions.

mation obtained by observing output a . The expression $f(x)$ represents the input image, and $n(x)$ represents the noise, which is assumed to be uncorrelated Gaussian noise, independent of signal f . Below, we also assume $\langle f \rangle = 0$, and $\langle n \rangle = 0$. In this section, a treatment for a one-dimensional case is given, but the treatment is also applicable for a two-dimensional case if we replace one-dimensional variables by corresponding two-dimensional vectors.

To define the amount of obtained information, we made a number of assumptions. First, we assumed that all objects in the visual field could change their positions independently of each other, so their images may be shown at slightly displaced positions from time to time. In other words, we assumed that if a certain image $f_\gamma(x)$ is shown, its locally displaced images, obtained by slightly displacing each part (each object) of the image are also shown. Next, we classified all these locally displaced images into the same class, γ , as $f_\gamma(x)$. To measure the obtained information, we used the mutual information between the cell's output and the class. This mutual information represents the average amount of obtained information with regard to which class of image is shown. In the actual calculation, we used the following image model.

1) We represented each region in the visual field with a window function, $w(x-x_j)$, whose center was located at x_j and assumed the presence of a local image, $f_j(x)$, corresponding to each individual object in each region. If there is no object in the region, $f_j(x)=0$. The complete input image is the sum of these local images:

$$f(x) = \sum_j w(x-x_j) f_j(x) \quad (2.3)$$

Each local image, $f_j(x)$, is assumed to be shown independently of each other. The regions (i.e., center positions of window functions, x_j) are assumed to be distributed continuously. We also assume that $w(x)$ is an even function.

2) Next, we allowed for the fact that each local image might be in a displaced position. That is, when local image $f_{\gamma j}(x)$ is presented in the j -th region, its displaced images, $f_{\gamma j}(x-x_0)$, will have the same probability of being in that region. We then classified all these displaced images into the same class, γ , as $f_{\gamma j}(x)$. To measure the amount of the obtained information, we used the mutual information between output a and class $\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$:

$$MI = H[a] - \langle H[a|\gamma_1, \gamma_2, \dots] \rangle_{\gamma_1, \gamma_2, \dots} \quad (2.4)$$

Here, $\langle \cdot \rangle_{\gamma_1, \gamma_2, \dots}$ denotes an averaging operation over parameters $\{\gamma_1, \gamma_2, \dots, \gamma_n\}$. The expression $H[a]$ represents entropy and $\langle H[a|\gamma_1, \gamma_2, \dots] \rangle_{\gamma_1, \gamma_2, \dots}$ conditional entropy. These are defined mathematically as

$$\begin{aligned} H[a] &= - \sum_a P(a) \log P(a) \quad \text{and} \\ \langle H[a|\gamma] \rangle_\gamma &= - \langle \sum_a P(a|\gamma_1, \gamma_2, \dots) \log P(a|\gamma_1, \gamma_2, \dots) \rangle_\gamma \end{aligned} \quad (2.5)$$

using the probability $P(a)$ and the conditional probability $P(a|\gamma_1, \gamma_2, \dots)$.

The above image model is based on the following idea. Although our retinal image varies with time, we expect that within a short time, for instance, between saccades, the objects it represents remain the same. For example, suppose we observe an image (a frame of a movie) of trees. Within a short time (e.g., in the next frame) we expect to see the same objects in the image although the image itself might vary. In a different frame, each object (e.g. each leaf) might be shown at a slightly displaced position (due to its movement). The displacement vector (motion vector) at each pixel must be almost the same within a small region (corresponding to each leaf) but it will be different (and uncorrelated) in regions located far away from it (if two leaves are located far apart, their movement must be different and indepen-

dent of each other). We represent such set of images by the image model described above, where the window width represents typical size of each object. Figure 2 schematically explains the image model for a one-dimensional case; For each pixel x_i , we prepare a window, $w(x-x_i)$ (figure 2a). When local images $f_{\gamma 1}(x)$ and $f_{\gamma m}(x)$ are shown in the first and m -th windows respectively (figure 2b), the complete input image becomes like figure 2c (local images in other windows are set to zero). When $f_{\gamma 1}(x)$ and $f_{\gamma m}(x)$ are respectively displaced by $x_{0,1}$ and $x_{0,m}$, the complete input image becomes like figure 2d, and this is classified into the same class $[\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_m, \dots\}]$ as figure 2c. But when an image $f_{\gamma 1'}(x) (\neq f_{\gamma 1}(x-x_{0,1}))$ for any $x_{0,1}$ is shown in the first window, the image is classified into a different class, $\gamma' = \{\gamma'_1, \gamma_2, \dots, \gamma_m, \dots\}$. For the calculations in section 3, we adopted a rectangle window or a Gaussian window for window function $w(x)$, but the qualitative results that we obtained in this paper did not depend on the specific form of this function, provided it is a localized even function ($w(x)=w(-x)$).

Another choice for a measure of obtained information might have been mutual information $MI[a|f]$ between cell's output a and the image itself, f . But maximizing this $MI[a|f]$ cannot help us derive an orientation-selective simple-cell-like RF.¹⁰ Okajima assumed that if a certain image, $f_\beta(x)$, is present, its displaced images, $f_\beta(x-x_0)$, will have

the same probability of being present, and classified all these displaced images into the same class, β , as $f_\beta(x)$. He used the mutual information between output a and class β and showed that maximizing this mutual information leads to a simple-cell-like RF, provided we also assume a localization term in the objective function.¹⁰ We will show that maximizing MI defined by equation (2.4) leads to a spatially localized, simple-cell-like RF $\phi(x)$ without requiring us to assume a localization term.

Low SNR approximation

Unfortunately, computing the entropy or conditional entropy in equation (2.4) is in general not easy. Therefore, we evaluated mutual information MI assuming a low SNR limit^{8,10} by using the following method.

When signal f is zero, the output (see equation (2.2)) only depends on the noise and is described as

$$a_n = (\phi, n) \quad (2.6)$$

When we have a nonzero signal, the output varies slightly and described as

$$a_n \rightarrow a = (\phi, n + \varepsilon f), \quad (2.7)$$

where the signal is εf instead of f indicating that it is small. We expanded the entropy and the conditional entropy in ε ,

$$\begin{aligned} H[a] &\cong H[a_n] + \delta H(\varepsilon) \\ H[a|\gamma_1, \gamma_2, \dots] &\cong H[a_n] + \delta H'_{\gamma_1, \gamma_2, \dots}(\varepsilon) \end{aligned} \quad (2.8)$$

and obtained the mutual information by

$$MI \cong \delta H(\varepsilon) - \langle \delta H'_{\gamma_1, \gamma_2, \dots}(\varepsilon) \rangle_{\gamma_1, \gamma_2, \dots} \quad (2.9)$$

We do not need to consider all the higher-order statistics of the signal in this expansion. For example, when we evaluate MI up to the 4th order in ε , we only need to consider the statistics up to the 4th order.

When we carried out this expansion, we obtained

$$\begin{aligned} MI &\cong \frac{\varepsilon^4}{4\sigma_n^4} \sum_{k,k',j} |\tilde{\psi}(k, x_j)|^2 |\tilde{\psi}(k', x_j)|^2 \\ &\quad \times (\langle |F_j(k)|^2 |F_j(k')|^2 \rangle - \langle |F_j(k)|^2 \rangle \langle |F_j(k')|^2 \rangle) \\ &= \frac{\varepsilon^4}{4\sigma_n^4} \sum_{k,k',j} |\tilde{\psi}(k, x_j)|^2 |\tilde{\psi}(k', x_j)|^2 \\ &\quad \times (\langle |F(k)|^2 |F(k')|^2 \rangle - \langle |F(k)|^2 \rangle \langle |F(k')|^2 \rangle), \\ &\quad \|\phi\|^2 = 1 \end{aligned} \quad (2.10)$$

Fig. 2 A schematic description of the image model; (a) window functions, (b) local images, (c) an example of the complete input image, (d) another example of the complete input image.

where $F(k)$ represents the Fourier transform of the local image, and $\langle \cdot \rangle$ denotes an averaging operation over images. We have assumed that the statistical properties of all the local images are the same (i.e., we assumed that for any j and j' , $\langle |F_j(k)|^2 \rangle = \langle |F_{j'}(k)|^2 \rangle (\equiv \langle |F(k)|^2 \rangle)$ and $\langle |F_j(k)|^2 |F_{j'}(k')|^2 \rangle = \langle |F_j(k)|^2 \rangle \langle |F_{j'}(k')|^2 \rangle (\equiv \langle |F(k)|^2 \rangle \langle |F(k')|^2 \rangle)$ are satisfied). In (2.10), σ_n^2 represents the variance of the noise, and $\tilde{\psi}(k, x)$ represents the Fourier transform of $\psi(x', x) \equiv \phi(x')w(x' - x)$ with respect to x' , where $\phi(x)$ is the RF we want to determine, and $w(x)$ is the window function set in the model. Since MI does not depend on the norm of the RF $\phi(x)$, it is normalized as $\|\phi\|^2=1$. In deriving equation (2.10), we used $H[a] \equiv 1/2 \log(2\pi e \langle a^2 \rangle)$ and $H[a | \gamma] \equiv 1/2 \log(2\pi e \langle a^2 \rangle p_{a|\gamma})$, which are valid up to the 4th order in ε ,¹⁰ where $a = (\phi, n + \varepsilon f) = \sum_{x,j} \psi(x, x_j) \varepsilon f_j(x) + a_n = \sum_{k,j} \tilde{\psi}(k, x_j) \varepsilon F_j^*(k) + a_n$ is the output of the cell. Terms lower than $\mathcal{O}(\varepsilon^4)$ cancel each other out in equation (2.9). For more details, see Appendix 1.

Next, we determined the solution $\phi(x)$ which maximizes the objective function, MI .

3. Optimal receptive field

Numerical solutions

Figure 3 shows numerical solutions for $\phi(x)$ obtained by maximizing the mutual information in equation (2.10). The solutions were obtained by a gradient ascent method [$\Delta \phi(x) \propto \partial MI / \partial \phi(x)$]. We stopped the gradient ascent when MI showed no more increase. We obtained spatially localized, Gabor-like functions. The localized positions depended on the initial conditions, but all the solutions maximize MI . The localization widths of the solutions depended on the width of the window function, $w(x)$. The frequency of the Gabor function coincided with the frequency at which the power spectrum, $p(k) = \langle |F(k)|^2 \rangle$, is maximized (see the next subsection). In these calculations, we adopted a rectangle window with a width of $L=32$ (arbitrary units) for the window function, $w(x)$, and an uncorrelated Gaussian signal filtered through a DoG (difference of Gaussian) filter ($DoG(x) = G_{\sigma_{ex}}(x) - G_{\sigma_{inh}}(x)$; $\sigma_{ex}=3.8$, $\sigma_{inh}=11.4$ in the same unit as L) for each local signal $f_j(x)$. The result, however, was not sensitive to specific conditions. For example, we obtained similar results

when we used a Gaussian window for $w(x)$. But DoG filtering was necessary for obtaining a Gabor-like function of non-zero frequency. When the low frequency components were not removed by the DoG filter, we obtained a Gaussian-like non-oscillating RF.

Figure 4 shows solutions for two-dimensional cases. Again, we obtained spatially localized, 2D Gabor-like functions. The optimal RF had anisotropic profiles resulting in orientation selectivity. These results match the experimentally observed simple-cell RFs.¹³⁻¹⁷ In these numerical calculations, we adopted a Gaussian window of $\sigma=3.0$ for the window function, $w(x)$, and used an uncorrelated Gaussian signal filtered through a 2D isotropic DoG filter ($DoG(\mathbf{x}) = G_{\sigma_{ex}}(\mathbf{x}) - G_{\sigma_{inh}}(\mathbf{x})$; $\sigma_{ex}=0.96$, $\sigma_{inh}=2.86$) for each local image, $f_j(\mathbf{x})$ (for solutions in figures 4a – 4c. A sample

Fig. 3 Numerical solutions that maximize mutual information MI . These were approximated by Gabor functions, $A \exp(-(x-x_0)^2/2s^2) \cos(2\pi f(x-x_0) + \varphi)$ with $A=0.294$, $f=0.0327$, $\varphi=0.0$, $s=12.99$, $x_0=51.0$ (top), $A=0.298$, $f=0.0324$, $\varphi=1.57$, $s=12.69$, $x_0=191.0$ (middle), $A=0.298$, $f=0.0324$, $\varphi=1.57$, $s=12.69$, $x_0=111.0$ (bottom).

complete input image is shown in figure 6a). Again, the result was not sensitive to these specific conditions. For example, when we used, for $f_i(\mathbf{x})$, images taken from an image shown in figure 5, we also obtained a Gabor-like function (figure 4d); The image in figure 5 was filtered through the 2D-DoG filter described above, and partial images (32×32 pixels) were taken from it at randomly chosen positions. Then their Fourier transforms were computed and averages, $\langle |F(\mathbf{k})|^2 |F(\mathbf{k}')|^2 \rangle$ or $\langle |F(\mathbf{k})|^2 |F(\mathbf{k}')|^2 \rangle$ in equation (2.10) were calculated. Then by maximizing (2.10), the solution in figure 4d was obtained. The receptive field in figure 4e was obtained by using another set of images; First, we randomly selected two regions (i.e., two windows $w(x - x_{j1})$, $w(x - x_{j2})$). Then we generated two random patterns ($f_{j1}(x)$, $f_{j2}(x)$) for each windows (in this case, images for all other windows were set to zero). Then, f_{j1} and f_{j2} were slightly displaced (within two pixels i.e., $-2 \leq \Delta x \leq 2$, $-2 \leq$

$\Delta y \leq 2$ and independently of each other), and all these locally displaced images were defined to belong to the same class as the original one. We performed the above procedure again and again to generate a set of input images. A sample input image is shown in figure 6b. The RF in figure 5e was obtained by maximizing equation (2.4).

Uncertainty relation and the Infomax solution

A Gabor function is maximally localized in the space and frequency domains.^{18,19} Consider the relationship between this property and the Infomax solutions. For simplicity, assume that the input (local) image statistics can be described by a Gaussian. The objective function, equation (2.10), can be rewritten by using only the second-order correlation or the power spectrum of images $p(k) = \langle |F(k)|^2 \rangle$ as

$$MI = \frac{\epsilon^4}{2 \sigma_n^4} \sum_{k,x} |\tilde{\psi}(k,x)|^4 p(k)^2 \quad (3.1)$$

Fig. 4 Numerical solutions that maximize mutual information MI (two-dimensional case). These were approximated by Gabor functions, $A \exp(-|\mathbf{e}_1(\mathbf{r} - \mathbf{r}_0)|^2/2s_1^2 - |\mathbf{e}_2(\mathbf{r} - \mathbf{r}_0)|^2/2s_2^2) \cos(2\pi \mathbf{f} \mathbf{e}_1 \cdot (\mathbf{r} - \mathbf{r}_0) + \varphi)$, where $\mathbf{e}_1 = (\cos \theta, \sin \theta)$, $\mathbf{e}_2 = \delta_{i,j}$, with $A=0.290$, $f=0.119$, $\varphi=0.0$, $s_1=2.90$, $s_2=2.61$, $\mathbf{r}_0=(18.0, 18.0)$, $\theta=0.0^\circ$ (a), $A=0.290$, $f=0.119$, $\varphi=0.0$, $s_1=2.90$, $s_2=2.61$, $\mathbf{r}_0=(20.0, 11.0)$, $\theta=45.0^\circ$ (b), $A=0.290$, $f=0.119$, $\varphi=2.23$, $s_1=2.97$, $s_2=2.65$, $\mathbf{r}_0=(12.1, 20.0)$, $\theta=0.0^\circ$ (c), $A=0.295$, $f=0.125$, $\varphi=1.56$, $s_1=2.78$, $s_2=2.66$, $\mathbf{r}_0=(17.0, 14.0)$, $\theta=77.8^\circ$ (d), $A=0.271$, $f=0.135$, $\varphi=-1.30$, $s_1=3.25$, $s_2=2.65$, $\mathbf{r}_0=(12.0, 17.0)$, $\theta=46.8^\circ$ (e).

Fig. 5 An image of a tree that was used for generating the RF shown in figure 4d.

Fig. 6 Sample input images. (a) A sample input image of an image set that was used for generating the RFs in figures 4a–4c. (b) A sample input image that was used for generating the RF in figure 4e.

In deriving (3.1) from (2.10), we used

$$\begin{aligned} \langle |F(k)|^2 |F(k')|^2 \rangle &= \langle |F(k)|^2 \rangle \langle |F(k')|^2 \rangle \\ (\text{when } k \neq k') \text{ and } \langle |F(k)|^4 \rangle &= 2(\langle |F(k)|^2 \rangle)^2, \\ \langle |F(k)|^2 |F(-k)|^2 \rangle &= 2(\langle |F(k)|^2 \rangle)^2 \text{ (see Appendix 2).} \end{aligned}$$

Using this equation, we see that MI becomes large when the function $\tilde{\psi}(k, x)$ is concentrated at frequency k_0 , the frequency at which the power spectrum is at a maximum. Thus, $\tilde{\psi}(k, x)$ is localized around k_0 and also around $-k_0$ ($\psi(x, x')$ is real) in the frequency domain.

When we view $\tilde{\psi}(k, x)$ as a function of position x , (although in this case all positions are equivalent) we notice again that MI becomes large when $\tilde{\psi}(k, x)$ is localized around a certain position, x_0 , because the squared sum of $|\tilde{\psi}(k, x)|$ remains constant, irrespective of ϕ . The expression

$$\sum_{k,x} |\tilde{\psi}(k, x)|^2 = \|\phi\|^2 \|w\|^2 = \|w\|^2, \quad (3.2)$$

remains true and under this constraint, the sum of

$|\tilde{\psi}(k, x)|^4$ increases when $\tilde{\psi}(k, x)$ is localized at a certain position. Thus, MI becomes large when $\tilde{\psi}(k, x)$ is localized both in k - and x -spaces. (It should be pointed out here that we can localize the function $\tilde{\psi}(k, x)$ in x -space only when $w(x)$ is a localized function. For example, if we set $w(x) = \text{const.}$, the function $\tilde{\psi}(k, x)$ always remains delocalized in x -space no matter which function $\phi(x)$ we choose).

We can solve the maximization problem (3.1) self-consistently. First, we rewrite the objective function (3.1) as

$$\begin{aligned} MI &= \frac{\varepsilon^4}{2\sigma_n^4} \sum_{k,x} |\tilde{\psi}_{\text{old}}(k, x)|^2 |\tilde{\psi}(k, x)|^2 p(k)^2 \\ &\equiv \sum_{k,x} p'(k, x) |\tilde{\psi}(k, x)|^2 \end{aligned} \quad (3.3)$$

and solve this maximization problem self-consistently so that the solution $\tilde{\psi}(k, x)$ coincides with $\tilde{\psi}_{\text{old}}(k, x)$. Assuming that $\tilde{\psi}_{\text{old}}(k, x)$ is localized around (k_0, x_0) and $(-k_0, x_0)$, the expression $p'(k, x)$ in equation (3.3) must have maxima at $(\pm k_0, x_0)$.

Let us first consider a solution that is localized around (k_0, x_0) . We expand $p'(k, x)$ around (k_0, x_0) as

$$p'(k, x) \equiv p_0 - \alpha (k - k_0)^2 - \beta (x - x_0)^2 \quad (3.4)$$

Then we obtain

$$\begin{aligned} MI &\equiv \sum_{k,x} \{ p_0 - \alpha (k - k_0)^2 - \beta (x - x_0)^2 \} |\tilde{\psi}(k, x)|^2 \\ &= -(\alpha \Delta k^2 + \beta \Delta x^2) \|w\|^2 + \text{const.} \end{aligned} \quad (3.5)$$

for $\tilde{\psi}(k, x)$, which is localized around (k_0, x_0) . Here, Δk and Δx are respectively the localization width of the RF function in the frequency and space domains, i.e.,

$$\begin{aligned} \Delta k^2 &\equiv \sum_k (k - k_0)^2 |\tilde{\phi}(k)|^2, \\ \Delta x^2 &\equiv \sum_x (x - x_0)^2 |\phi(x)|^2. \end{aligned} \quad (3.6)$$

The uncertainty relation between them is

$$\Delta k \Delta x \geq 1/2 \quad (3.7)$$

In deriving equation (3.5), we used equalities

$\sum_x |\tilde{\psi}(k, x)|^2 = \sum_{k'} |\tilde{\phi}(k')|^2 |\tilde{w}(k' - k)|^2$ and $\sum_k |\tilde{\psi}(k, x)|^2 = \sum_{x'} |\phi(x')|^2 |w(x' - x)|^2$, where $\tilde{w}(k)$ is the Fourier transform of $w(x)$ (for more details, see Appendix 3).

From the obvious inequality $(\sqrt{\alpha} \Delta k - \sqrt{\beta} \Delta x)^2 \geq 0$ and the uncertainty relation (3.7), the inequality

$$-(\alpha \Delta k^2 + \beta \Delta x^2) \leq -2\sqrt{\alpha\beta} \Delta x \Delta k \leq -\sqrt{\alpha\beta}. \quad (3.8)$$

remains valid. Thus MI in equation (3.5) is maximized when $\phi(x)$ is a complex Gabor function that satisfies $\sqrt{\alpha} \Delta k = \sqrt{\beta} \Delta x$, i.e., when

$$\phi(x) = \phi_{k0}(x) \propto \exp(-\sqrt{\beta/\alpha}(x-x_0)^2/2 + ik_0x).$$

This is because the complex Gabor function achieves the theoretical lower limit of 1/2 for the product $\Delta k \Delta x$.

Consider the symmetry $p'(-k, x) = p'(k, x)$. If ϕ_{k0} is a solution,

$$\phi_{-k0}(x) \propto \exp(-\sqrt{\beta/\alpha}(x-x_0)^2/2 - ik_0x)$$

must also be a solution, and when the overlap between $|\phi_{k0}|$ and $|\phi_{-k0}|$ is small, any linear combination of ϕ_{k0} and ϕ_{-k0} also becomes a solution. Since $\phi(x)$ must be real, $\phi(x)$ becomes

$$\phi(x) \propto \exp(-\sqrt{\beta/\alpha}(x-x_0)^2/2) \cos(k_0x + \theta),$$

where θ is an arbitrary phase. The coefficients α and β depend on $\phi(x)$, so they must be determined self-consistently. Despite this, the solution remains a Gabor function, as shown in figure 3 and 4.

4. Summary and discussion

Assume an object is shown in a certain region within the visual field, and another object within a different region. To determine which object is shown in which region, we could design a RF that extends over both the regions, or a RF that is localized within a single region. We conclude that one localized in a single region is better since a localized RF can extract more information when both objects can change positions independently. We considered a case in which the regions were distributed continuously over the visual field and we used the model described in section 2. Locally displaced images considered to belong to the same class as the original. To measure the obtained information we used the mutual information between cell's output a and class γ to which image belonged. The optimal RF was determined by assuming a low SNR limit and maximizing this mutual information. The optimal RF was spatially localized and similar to a Gabor function, matching the experimentally observed RF of a simple

cell. We also found that the ability of the Gabor function to reach a theoretical lower limit of 1/2 for the uncertainty product $\Delta k \Delta x$ is closely related to its ability to maximize the mutual information.

We also found that the size of the optimal RF is determined by the size of the window function, $w(x)$, set to represent the size of the objects shown. In this paper we assumed that the window size was constant. However, to discuss the RF-size variation in simple cells (simple cells that are selective to lower spatial frequency tend to have larger RF size,³⁰) we need to consider possible variations in the window size (i.e., object size.) Refinements to the theory are subjects for future studies.

We believe that the approach described in this paper can also be used to derive a complex-cell-like RF^{11,12}. To do so, we would model the output of a complex cell by $a_c = |(\phi_c f + n)|^2$, where $\phi_c = \phi_1 + i\phi_2$ is a "complex RF",^{11,12} according to the energy model for a complex cell.^{21,23} By maximizing the mutual information between this output, a_c , and the class, $\gamma = \{\gamma_1, \gamma_2, \dots\}$, we should be able to obtain a spatially localized, complex-cell-like optimal RF ϕ_c .

Acknowledgements

This work was supported by the Special Coordination Funds for Promoting Science and Technology.

References

- 1) Hubel, D.H. and Wiesel, T.N. (1968): Receptive fields and functional architecture of monkey striate cortex, *J. Physiol. Lond.*, Vol. 195, pp. 215-244
- 2) Olshausen, B.A. and Field, D.J. (1996): Emergence of simple-cell receptive field properties by learning a sparse code from natural images, *Nature*, Vol. 381, pp. 607-609
- 3) Bell, A.J. and Sejnowski, T.J. (1997): The 'Independent Components' of natural scenes are edge filters, *Vision Research*, Vol. 37, pp. 3327-3338
- 4) Barlow, H.B. (1989): Unsupervised learning, *Neural Computation*, Vol. 1, pp. 295-311
- 5) Linsker, R. (1988): Self-organization in a perceptual network, *Computer*, Vol. 21, No. 3, pp. 105-117
- 6) Linsker, R. (1993): Deriving receptive fields using an optimal encoding criterion. In S.J. Hanson, J.D. Cowan and C.L. Giles (Eds.), *Advances in*

- Neural Information Processing Systems, Vol. 5. (pp. 953 – 960) San Mateo, CA :Morgan Kaufmann
- 7) Atick, J. (1992): Could information theory provide an ecological theory of sensory processing? In W. Bialek (Ed.), Princeton Lectures on Biophysics. (pp. 223 – 290) Singapore: World Scientific
- 8) Bialek, W. (1992): Optimal signal processing in the nervous system. In W. Bialek (Ed.), Princeton Lectures on Biophysics. (pp. 321 – 401) Singapore: World Scientific
- 9) Okajima, K. (1998a): The Gabor function extracts the maximum information from input local signals, Neural Networks, Vol. 11, pp. 435 – 439
- 10) Okajima, K. (1998b): Two-dimensional Gabor-type RF as derived by mutual information maximization, Neural Networks, Vol. 11, pp. 441 – 447
- 11) Okajima, K. (1998c): "Complex receptive field" obtained by Infomax coincides with the energy model for complex cells in the visual cortex, Proceedings of the fifth International Conference on Neural Information Processing, Kitakyushu, Japan, pp. 657 – 660
- 12) Okajima, K. and Imaoka, H. (2001): A complex-cell-like receptive field obtained by information-maximization, Neural Computation, Vol. 13, pp. 547 – 562
- 13) Daugman, J.G. (1980): Two-dimensional spectral analysis of cortical receptive field profiles, Vision Research, Vol. 10, pp. 847 – 856
- 14) Marcelja, S. (1980): Mathematical description of the responses of simple cortical cells, Journal of the Optical Society of America, Vol. 70, pp. 1297 – 1300
- 15) Jones, J.P. and Palmer, L.A. (1987): An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, Journal of Neurophysiology, Vol. 58, No. 6, pp. 1233 – 1258
- 16) Jones, J.P. and Palmer, L.A. (1987): The two-dimensional spatial structure of simple receptive fields in cat striate cortex, Journal of Neurophysiology, Vol. 58, No. 6, pp. 1187 – 1211
- 17) Anzai, A., Ohzawa, I. and Freeman, R.D. (1997): Neural mechanism underlying binocular fusion and stereopsis: Position vs. phase, Proc. Natl. Acad. Sci. USA, Vol. 94, pp. 5438 – 5443
- 18) Gabor, D. (1946): Theory of communication, J. Inst. Elec. Eng., Vol. 93, pp. 429 – 457
- 19) Daugman, J.D. (1985): Uncertainty relation for the resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America, Vol. A2, pp. 1160 – 1169
- 20) Maffei, L. and Fiorentini, A. (1977): Spatial frequency rows in the striate visual cortex, Vision Research, Vol. 17, pp. 257 – 264
- 21) Pollen, A.D. and Ronner, S.F. (1983): Visual cortical neurons as localized spatial frequency filters, IEEE Trans. System Man, and Cybernetics, Vol. SMC13, pp. 907 – 916
- 22) Adelson, E.H. and Bergen, J.R. (1985): Spatiotemporal energy models for the perception of motion, Journal of the Optical Society of America, Vol. A2, No. 2, pp. 294 – 299
- 23) Adelson, E.H. and Bergen, J.R. (1985): Spatiotemporal energy models for the perception of motion, Journal of the Optical Society of America, Vol. A2, No. 2, pp. 294 – 299
- 24) Kubo, R. (1962): Generalized cumulant expansion method, Journal of the Physical Society of Japan, Vol. 17, pp. 206 – 226

Appendices

1. A Derivation of Eq. (2.10)

Up to the 4th order in ε , the entropy, $H[a]$, is approximated as $H[a] \cong 1/2\log(2\pi e\langle a^2 \rangle)$,¹⁰⁾ where

$$a = (\phi, \varepsilon f + n) = \sum_{x,j} \psi(x, x_j) \varepsilon f_j(x) + a_n \\ = \sum_{k,j} \tilde{\psi}(k, x_j) \varepsilon F_j^*(k) + a_n$$

(note $f(x) = \sum w(x - x_j) f_j(x)$, and $\psi(x', x) \equiv \phi(x')w(x' - x)$) The average $\langle a^2 \rangle$ is calculated as

$$\begin{aligned} \langle a^2 \rangle &= \varepsilon^2 \sum_{j,j',k,k'} \tilde{\psi}(k, x_j) \tilde{\psi}^*(k', x_{j'}) \langle F_j(k)^* F_{j'}(k') \rangle \\ &\quad + \langle a_n^2 \rangle \quad (\text{we used } \langle n \rangle = 0) \\ &= \varepsilon^2 \sum_{j,k} |\tilde{\psi}(k, x_j)|^2 \langle |F_j(k)|^2 \rangle + \langle a_n^2 \rangle \end{aligned}$$

Here we used $\langle F_j(k)^* F_{j'}(k') \rangle = \langle F_j(k)^* \rangle \langle F_{j'}(k') \rangle = 0$ when $j \neq j'$ (note we assume that each f_j is independent of each other, and $\langle f \rangle = 0$). We also used $\langle F_j^*(k) F_{j'}(k') \rangle = 0$ when $k \neq k'$ (since $\langle F_j^*(k) F_{j'}(k') \rangle \propto \sum_{x,x'} \langle f_j(x) f_{j'}(x') \rangle e^{i(kx - k'x')} = \sum_{x,x'} c(x - x') e^{-i(kx - k'x')} \times \sum_x e^{-i(k - k')x} = 0$ when $k \neq k'$ [$c(x)$: correlation function]).

The average $\langle a_n^2 \rangle = \|\phi\|^2 \sigma_n^2$ is written as σ_n^2 (σ_n^2 : the variance of the noise n) when we normalize ϕ as $\|\phi\|^2 = 1$. Thus, $H[a]$ is calculated as

$$\begin{aligned} H[a] &\cong 1/2\log 2\pi e \left(\sigma_n^2 + \varepsilon^2 \sum_{j,k} |\tilde{\psi}(k, x_j)|^2 \langle |F_j(k)|^2 \rangle \right) \\ &\cong 1/2\log \sigma_n^2 + \varepsilon^2 / (2\sigma_n^2) \sum_{j,k} |\tilde{\psi}(k, x_j)|^2 \langle |F_j(k)|^2 \rangle \\ &\quad - \varepsilon^4 / (4\sigma_n^4) \sum_{j,j',k,k'} |\tilde{\psi}(k, x_j)|^2 |\tilde{\psi}(k', x_{j'})|^2 \\ &\quad \times \langle |F_j(k)|^2 \rangle \langle |F_{j'}(k')|^2 \rangle + 1/2\log 2\pi e \end{aligned}$$

Similarly, $\langle H[a|\gamma_1, \gamma_2, \dots] \rangle$ is expressed as

$$\begin{aligned} \langle H[a|\gamma_1, \gamma_2, \dots] \rangle &= \langle \gamma_1, \gamma_2, \dots | \\ &\equiv 1/2 \log \sigma_n^2 + \varepsilon^2 / (2 \sigma_n^2) \sum_{j,k} |\tilde{\psi}(k, x_j)|^2 \langle |F_j(k)|^2 \rangle_\gamma \\ &\quad - \varepsilon^4 / (4 \sigma_n^4) \sum_{j,j',k,k'} |\tilde{\psi}(k, x_j)|^2 |\tilde{\psi}(k', x_{j'})|^2 \\ &\quad \times \langle |F_j(k)|^2 |F_{j'}(k')|^2 \rangle_\gamma + 1/2 \log 2\pi e. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} MI &\equiv H[a] - \langle H[a|\gamma_1, \gamma_2, \dots] \rangle_\gamma \\ &\equiv \frac{\varepsilon^4}{4 \sigma_n^4} \sum_{k,k'} |\tilde{\psi}(k, x_j)|^2 |\tilde{\psi}(k', x_{j'})|^2 \\ &\quad \times (\langle |F_j(k)|^2 |F_{j'}(k')|^2 \rangle \\ &\quad - \langle |F_j(k)|^2 \rangle \langle |F_{j'}(k')|^2 \rangle), \end{aligned}$$

where we used $\langle |F_j(k)|^2 |F_{j'}(k')|^2 \rangle_\gamma = \langle |F_j(k)|^2 \rangle_\gamma \langle |F_{j'}(k')|^2 \rangle_\gamma$ (when $j \neq j'$) and $\langle |F_j(k)|^2 \rangle_\gamma = \langle |F_j(k)|^2 \rangle_\gamma, \langle |F_j(k)|^2 |F_j(k')|^2 \rangle_\gamma = \langle |F_j(k)|^2 \rangle_\gamma \langle |F_j(k')|^2 \rangle_\gamma$ (note that the variables inside the bracket are shift-invariant so that the averaging operation over γ can be replaced by that over images). Finally, by assuming that the statistical properties of all the local images are the same (i.e., by assuming $\langle |F_j(k)|^2 \rangle = \langle |F(k)|^2 \rangle$ and $\langle |F_j(k)|^2 |F_j(k')|^2 \rangle = \langle |F(k)|^2 |F(k')|^2 \rangle$ for any j) we obtain equation (2.10).

2. A derivation of Eq. (3.1)

When the input local signal, $f(x)$, is a Gaussian signal, its Fourier transform, $F(k)$, satisfies

$$\begin{aligned} \langle |F(k)|^2 |F(k')|^2 \rangle &= \langle |F(k)|^2 \rangle \langle |F(k')|^2 \rangle \\ &\text{(when } k' \neq \pm k) \text{ and} \\ \langle |F(k)|^4 \rangle &= \langle |F(k)|^2 |F(-k)|^2 \rangle \\ &= 2(\langle |F(k)|^2 \rangle)^2. \end{aligned} \quad (\text{A1}) \quad (\text{A2})$$

By using (A1) and (A2), equation (2.10) is written as

$$\begin{aligned} MI &\equiv \frac{\varepsilon^4}{4 \sigma_n^4} \sum_{k,k'} \sum_{x,x'} |\tilde{\psi}(k, x)|^2 |\tilde{\psi}(k', x')|^2 \\ &\quad \times (\langle |F(k)|^2 |F(k')|^2 \rangle - \langle |F(k)|^2 \rangle \langle |F(k')|^2 \rangle) \\ &= \frac{\varepsilon^4}{2 \sigma_n^4} \sum_{k,x} |\tilde{\psi}(k, x)|^4 \langle |F(k)|^2 \rangle^2. \end{aligned}$$

Equation (A1) is derived as follows. We define the Fourier transform of $f(x)$ as

$$F(k) = 1/\sqrt{N} \sum_x f(x) \exp(-ikx),$$

where $N \equiv \sum_x 1$. Then, the left-hand-side of (A1) is written as

$$\begin{aligned} \langle |F(k)|^2 |F(k')|^2 \rangle &= 1/N^2 \sum_{x_1, x_2, x_3, x_4} \langle f(x_1) f(x_2) f(x_3) f(x_4) \rangle e^{-ik(x_1 - x_2) - ik'(x_3 - x_4)} \end{aligned}$$

Since we assume $\langle f \rangle = 0$, the moment $\langle f(x_1) f(x_2) f(x_3) f(x_4) \rangle$ is written as (see, e.g., Ref. 24)

$$\begin{aligned} \langle f(x_1) f(x_2) f(x_3) f(x_4) \rangle &= \langle f(x_1) f(x_2) \rangle \langle f(x_3) f(x_4) \rangle \\ &\quad + \langle f(x_1) f(x_3) \rangle \langle f(x_2) f(x_4) \rangle \\ &\quad + \langle f(x_1) f(x_4) \rangle \langle f(x_2) f(x_3) \rangle. \end{aligned}$$

Then we obtain

$$\begin{aligned} \langle |F(k)|^2 |F(k')|^2 \rangle &= 1/N \sum_{x_1, x_2} \langle f(x_1) f(x_2) \rangle e^{-ik(x_1 - x_2)} \\ &\quad \times 1/N \sum_{x_3, x_4} \langle f(x_3) f(x_4) \rangle e^{-ik'(x_3 - x_4)} \\ &\quad + 1/N \sum_{x_1, x_3} \langle f(x_1) f(x_3) \rangle e^{-ikx_1 - ik'x_3} \\ &\quad \times 1/N \sum_{x_2, x_4} \langle f(x_2) f(x_4) \rangle e^{ikx_2 + ik'x_4} \\ &\quad + 1/N \sum_{x_1, x_4} \langle f(x_1) f(x_4) \rangle e^{-ikx_1 + ik'x_4} \\ &\quad \times 1/N \sum_{x_2, x_3} \langle f(x_2) f(x_3) \rangle e^{ikx_2 - ik'x_3} \end{aligned} \quad (\text{A3})$$

In (A3), the first term coincides with $\langle |F(k)|^2 \rangle \langle |F(k')|^2 \rangle$, while the second and the third term become zero. For example, in the second term, the expression

$$1/N \sum_{x_1, x_3} \langle f(x_1) f(x_3) \rangle e^{-ikx_1 - ik'x_3}$$

is calculated as

$$\begin{aligned} 1/N \sum_{x_1, x_3} \langle f(x_1) f(x_3) \rangle e^{-ikx_1 - ik'x_3} \\ = \sum_{x_1=x_3} c(x_1 - x_3) e^{-ikx_1 - ik'x_3} \times 1/N \sum_{x_3} e^{-i(k+k')x_3} = 0 \end{aligned}$$

when $k' \neq -k$ ($c(x)$: correlation function). Equation (A2) can also be derived in a similar way.

3. A derivation of Eq. (3.5)

The function

$$\begin{aligned} \tilde{\psi}(k, x) &= 1/\sqrt{N} \sum_{x'} \psi(x', x) e^{-ikx'} \\ &= 1/\sqrt{N} \sum_{x'} \phi(x') w(x' - x) e^{-ikx'} \end{aligned}$$

satisfies

$$\sum_{x'} |\tilde{\psi}(k, x)|^2 = \sum_{k'} |\tilde{\phi}(k')|^2 |\tilde{w}(k' - k)|^2 \quad \text{and} \quad (\text{A4})$$

$$\sum_{k'} |\tilde{\psi}(k, x)|^2 = \sum_{x'} |\phi(x')|^2 |w(x' - x)|^2. \quad (\text{A5})$$

Accordingly, we have

$$\sum_{k,x} |\tilde{\psi}(k, x)|^2 = \|\phi\|^2 \|w\|^2 = \|w\|^2.$$

Thus, the first term in equation (3.5)

$$(\sum_{k,x} p_0 |\tilde{\psi}(k, x)|^2)$$

becomes a constant ($= p_0 \|w\|^2$). By using (A4), the second term in (3.5) is calculated as

$$\begin{aligned}
& - \sum_{k,x} \alpha (k - k_0)^2 |\tilde{\psi}(k, x)|^2 \\
& = - \sum_{k,k'} \alpha (k - k_0)^2 |\tilde{\phi}(k')|^2 |\tilde{w}(k' - k)|^2 \\
& = - \alpha \sum_{k,k'} ((k - k') + (k' - k_0))^2 |\tilde{\phi}(k')|^2 |\tilde{w}(k' - k)|^2 \\
& = - \alpha (\sum_k k^2 |\tilde{w}(k)|^2 + \|w\|^2 \sum_{k'} (k' - k_0)^2 |\tilde{\phi}(k')|^2) \\
& - \alpha \|w\|^2 \Delta k^2 + \text{const.}
\end{aligned} \tag{A6}$$

where we used $\sum_k k |\tilde{w}(k)|^2 = 0$ which is true since $w(x)$ is real. In a similar way, by using (A5) and $\sum_x x |w(x)|^2 = 0$ (note we assume $w(x)$ an even function), we obtain

$$- \sum_{k,x} \beta (x - x_0)^2 |\tilde{\psi}(k, x)|^2 = - \beta \|w\|^2 \Delta x^2 + \text{const.} \tag{A7}$$

From (A6) and (A7) we obtain equation (3.5).