Proceedings CSCSI/SCEIO Conference 1984
The University of Western Ontario
London, Ontario, Canada, 15, 16, 17 May 1984

RECEPTIVE FIELDS AND THE RECONSTRUCTION OF VISUAL INFORMATION

Steven W. Zucker
Computer Vision and Robotics Laboratory
Department of Electrical Engineering
McGill University
Montreal, Quebec, Canada

Robert A. Hummel
Courant Institute of Mathematical Sciences
New York University
New York, N.Y.

ABSTRACT

Receptive fields in the retina indicate the first measurements taken over the (discrete) visual image. Why are they circular surround with an excitatory/inhibitory structure? We hypothesize that they provide a representation of the visual information in a form suitable for transmission over the optic nerve, a rather limited channel. The hypothesis is supported by a formal scheme for reconstructing visual information at the cortex. The scheme is both physiologically plausible, and leads to a number of predictions about receptive field size, structure and hyperacuity that are supported. The existence of such a scheme suggests that the analysis of visual information really begins in the cortex, a suggestion that stands in strong opposition to many current beliefs about "edge detection".

We have developed a theory of image reconstruction which explains precisely how it is possible that detailed visual information can be made available to the cortex. The theory provides a single, consistent role for much of the spatial processing along the X-pathway from the retina through the lateral geniculate nucleus to the visual cortex. The need for this information follows, in principle, from the precision with which we can interact with our visual environment [1]. Under the assumption that receptive fields carry measurements that can be represented by operator convolutions, the theory accounts for the following physiological observations: (i) retinal receptive fields have a center surround organization, with an excitatory center and inhibitory surround; as well as (ii) with an inhibitory center and excitatory surround [2]. We are using the terms as defined in [3], to stress the antagonistic manner in which information is combined within a receptive field. (iii) At the cortex, receptive fields span a range of sizes [4]. The theory takes into account that (iv) neurons have a limited capacity to carry information, and that (v) neurons along the X-pathway have a rather low spontaneous firing rate [5]. Finally, for truly accurate reconstructions the theory requires (vi) the presence of additional side lobes in certain cortical receptive fields [6;7]. Since these facts summarize most of what is known about the basic X-pathway, the suggestion that the analysis of visual information begins in the cortex becomes very much more plausible.

The reconstruction scheme is based on a restoration of data obtained from receptive field measurements. An essential feature of these measurements is the blurring or diffusion of information, which we shall model according to the heat equation. While the formal role of the heat equation will be introduced shortly, the intuition comes from the observation that a unit impulse of heat diffuses into increasingly larger Gaussian distributions as time proceeds (Fig. 1). Such a (temporal) spread will become analgous to the (spatial) extent of a receptive field.

Mathematically, let $f(x)$ denote the initial temperature distribution as a function of the spatial variable $x \in \mathbb{R}^n$. Then a solution to the heat equation $u(x,t)$, $t \geq 0$, satisfying

$$u(x,0) = f(x)$$

can be obtained from the convolution

$$u(x,t) = \int_{\mathbb{R}^n} K(x-x',t)\, f(x')\, dx'$$

where $K(x,t)$ is the "source" kernel [8]

$$K(x,t) = (2\pi t)^{-n/2}\, e^{-|x|^2/4t}.$$

Since the kernel acts as a blurring operator, we can regard the distributions $u(x,t)$ as representing continuously coarser representations of the original data $f(x)$ as t increases. In fact, assuming $f(x)$ is bounded, $u(x,t)$ as given above is analytic. It is the unique bounded solution to the heat equation

$$u_t = \Delta u$$

satisfying $u(x,0) = f(x)$. Other (bounded) solutions are technically possible, but the function $u(x,t)$ given by convolution against the Gaussian kernel K is the one that naturally occurs in physical systems.

Suppose initial image data were blurred by a Gaussian kernel. Is it possible to reproduce the original data? Specifically, given $g(x) = u(x,\tau)$, for some fixed $\tau > 0$, is it possible to solve the heat equation backwards to recover $u(x,t)$ for $0 < t < \tau$?

There are two separate aspects to the answer: whether recovery is possible in principle and whether it is possible in practice. In principle it can be shown that necessary and sufficient conditions for the existence of a solution to the heat equation, $u(x,t)$, $0 \leq t \leq \tau$, satisfying $u(x,\tau) = g(x)$, $x \in \mathbb{R}^n$, are that $g(x)$ be analytic, and that the extension of $g(x)$ to an analytic function of several complex variables $g(z)$, $z \in C^n$, satisfies certain growth conditions. Both of these conditions, analyticity are bounded growth, fit smoothly into the vision context.

Given the existence of a backsolution, calculating it may still be impractical. The difficulty is that the backward heat problem is unstable; that is, a small change in the initial data $g(x)$ can lead to an arbitrarily large change in $u(x,t)$. No matter how well $g(x)$ is approximated numerically, there will always exist examples where the resulting calculation for $f(x)$ is arbitrarily far off.

Nevertheless, John [9] has shown that, if a non-negative backsolution exists, then stable reconstruction of $u(x,t)$ is possible for $a \leq t \leq \tau$, where $a > 0$ is fixed. Specifically, suppose that $0 \leq g(x) \leq \mu$, and that $g(x)$ is sampled so that it can be reconstructed to within accuracy ε. Then an estimate $u(x,t)$ can be formed using discrete kernels such that the error $\bar{u}(x,t) - u(x,t)$, for $a \leq t \leq \tau$, is bounded by a constant (depending on a) times $\mu^{1-\Theta}\varepsilon^{\Theta}$. Here Θ is a number between 0 and 1, which implies that, e.g., a twofold increase in the accuracy of reproducing $u(x,t)$ requires more than a twofold increase in the representational accuracy. The coefficients of this kernel are shown in Fig. 2.

In summary, then, the essential restrictions amount to requiring that $g(x)$ be bounded, that a non-negative back solution exists, and that a solution is sought only back as far as some positive time $t \geq a \geq 0$. And, for numerical stability, we must represent $g(x)$ as accurately as possible.

The physiological interpretation begins with the earlier points (i) and (ii): that receptive fields have an antagonistic center/surround organization. Such local operations are useful in data coding, since they compress the required dynamic range of a channel [10,11]. They may also exist in primate visual systems for evolutionary reasons. But, since they are modifying the initial light measurements, they would seem to make reconstruction more difficult. Somehow their effects need to be undone. However, as we now show, given the form of these operators, we just need to add an extra step onto the reconstruction scheme.

Circular surround receptive fields have been modeled by kernels given either as the difference of two Gaussians [12, 13], or as the Laplacian of a Gaussian [14]. Although these kernels are distinct, we can interpret the former as a discrete analog of the latter. This follows since the heat kernel $K(x,t)$, a solution of the heat equation, satisfies

$$\Delta K(x,t) = \frac{\partial}{\partial t} K(x,t) \approx [K(x,t_1)-K(x,t_2)]/(t_1 - t_2)$$

which is the difference of two Gaussians. We therefore take

$$v(x,t) = \int_{\mathbb{R}^n} \Delta K(x - x', t)\, f(x')\, dx'$$

as a continuous version of the initial measurements, where t parameterizes the effective size of the receptive field. Note that $v(x,t)$ is only available for $t \geq \tau$, where τ corresponds to the size of the smallest receptive field. Since

$$v(x,t) = \Delta \int_{\mathbb{R}^n} K(x - x', t)\, f(x')\, dx'$$

$$= \int K(x - x', t)\, (\Delta f)\, (x')\, dx'$$

$v(x,t)$ can be interpreted either as the Laplacian of the blurred intensity data, i.e., as $\Delta u(x,t)$, or as the bounded solution to the heat equation using the initial data $\Delta f(x)$. From the former interpretation and the fact that $u(x,t)$ satisfies the heat equation, we have that $v(x,t) = \Delta u(x,t) = \frac{\partial}{\partial t} u(x,t)$, so

$$-\int_0^T v(x,t)dt = u(x,0) - u(x,T).$$

Now, $u(x,T)$ is nearly constant for large T, so the above integral can be used to recover $f(x)$ modulo an additive constant. From the other interpretation of $v(x,t)$, we see that $v(x,t)$ is itself a solution to the heat equation. Thus the values of $v(x,t)$, for $0 \leq t \leq \tau$, will have to be obtained by backsolving the heat equation using $v(x,\tau)$ as initial data.

To discuss physiological realizations of these formulas, we must confront problems of discretization and stability. The receptive field measurements $v(x,t)$ are not continuous in t, but rather are given by the discrete approximation $t = t_k, t_{k+1}, \ldots, t_m$, with each $t_i \geq \tau$. The data are sampled spatially, and it is likely that $v(x,t)$ is sampled more coarsely in x for larger t, reflecting the more uniform variations of the smoothed data [15]. Values for $v(x,t_i)$, $t = t_1, t_2, \ldots, t_{k-1}$, with $0 < t_i < \tau$, will be obtained by backsolving. An examination of John's coefficients for the backsolution kernel reveals the addition of decreasing side lobes for higher orders of approximation, as would be expected from the de-blurring of Gaussians [16]. The third order approximation agrees nicely with physiological observation (vi); see [6, fig. 9b]. The integral for recovering $f(x)$ can be approximated by a weighted sum

$$-\int_0 v(x,t)\ dt = \sum_{i=1}^{m} (t_i - t_{i-1})\ v(x,t_i)$$

thereby requiring the different size operators (physiological observation (iii)). It is especially interesting to note, with regard to this sum, that it is only the data for t near 0 that needs to be reconstructed, and it is precisely these smaller operators (receptive fields) that have been observed to have the extra side lobes [6]; see also [17] for psychophysical support.

The last remaining issue is stability. John's results require that the backsolution be non-negative, but $v(x,t)$ can be both positive and negative. Therefore we shall split the manner in which $v(x,t)$ is represented, separating the positive and negative parts. Recalling physiological observation (v), that X-pathway neurons have a low spontaneous firing rate, we introduce a smooth approximation to the "positive part" function:

$$\phi_+(x) = \max\,[x, 0]$$

for $x < A$, a saturation level; see Fig. 3. Similarly, set $\phi_-(x) = \phi_+(-x)$, and note that

$$\phi_+(x) - \phi_-(x) \approx x \qquad \text{for } |x| < A$$

since ϕ is linear in the range in which it is used. We can then separate the receptive-field convolution data into

$$v_+(x,t_i) = \int K(x-x', t_i)\ \phi_+(\Delta f(x'))\ dx'$$

and

$$v_-(x,t_i) = K(x-x',t_i)\ \phi_-(\Delta f(x'))\ dx'$$

Note that this functional form implies that it is only the small convolutions that are sent back from the retina, which is in agreement with the physiology [5; 18]. This formulation also suggests that there should only be a limited range over which X-cells behave linearly, which is also well known. Patterns qualitatively matching these receptive fields are, it is further interesting to note, the ones to which our visual systems are the most sensitive [19]. Since v_+ and v_- are both solutions to the heat equation with (essentially) non-negative data, both can be backsolved from $t = \tau$ to $t = t_1$, with $t_1 > 0$, in a stable fashion. The desired data $v(x,t_i)$, as used before, can then be recovered from the difference

$$v(x,t) = v_+(x,t) - v_-(x,t).$$

In summary, the reconstruction method involves sampling and transmitting the receptive field convolutions $v_+(x,t_i)$ and $v_-(x,t_i)$ for small $t_i \geq \tau$, backsolution of $v_+(x,t_i)$ and $v_-(x,t_i)$ for $t_i < \tau$, and (weighted) summing of all these measurements with the larger, smoothed convolutions. Note that these larger convolutions are obtained just by Gaussian blurring of small center surround receptive fields. The backsolution is stable for both v_+ and v_- provided reconstruction is attempted only as far as some resolution level $t > 0$, as would be expected from hyperacuity data.

Our proposal differs fundamentally from those that implicate the different size operators with notions of "edge detection" [14]. Qualitatively, this other approach asserts that physical events such as reflectance, depth, or illumination changes take place over different spatial "scales", so that different size operators are necessary to capture them. However, this approach suffers from several problems. First, from photometric observations, it can be shown that the physical processes responsible for generating the intensity changes operate over many scales simultaneously [20] and non-linearly. Second, there is the problem of how to combine the information at the different scales. Finally, there is increasingly more psychophysical and computational evidence that early vision in general, and anything like edge detection in particular, requires very precise information [1; 21; 22]. Nevertheless, attempts have been made within this approach to use the zero-crossings of the different size operators as "edge" locations, and diffusion equations have been used to study their migration across scales [20; 23].

While our method has some qualitative similarity to others based on the Shannon sampling theorem (e.g., [24], and references cited therein), there are fundamental differences. The spatial support required for sin x/x reconstruction is larger than the local polynomials that we incorporated, raising serious questions about accuracy [25]. Also, the idea of backsolving to obtain the highest frequency data is not there. Finally, one of the principle advantages of that scheme - - noise averaging in the larger receptive fields -- is present in our scheme as well.

While our reconstruction method demonstrates the possibility of precise image reconstruction in principle, it does not imply that it is necessarily taking place in practice. Perhaps only part of the scheme is utilized, such as just step (1) above, since this also amounts to an effective Gaussian de-blurring strategy. Many sources of such blur exist early in vision, from receptive field convolutions to motion smear [26] to physiological variation in neuronal conduction velocities. Or perhaps the reconstruction takes place only implicitly, within a subsequent level of processing such as orientation selection [21]. Ultimately, the physiology will decide.

REFERENCES

[1] Barlow, H., Reconstructing the visual image in space and time, Nature, 1979, 279, 189-190.

[2] Kuffler, S., and Nichols, J., From Neuron to Brain, Sunderland, Ma: Sinauer Associates, Inc., 1976.

[3] Stein, A., Mullikin, W., and Stevens, J., The spatiotemporal building blocks of X-, Y-, and W-ganglion cell receptive fields of the cat's retina.

[4] Hubel, D., and Wiesel, T., Functional architecture of macaque monkey visual cortex, PROC. ROY. SOC. (LONDON), B, 1977, 198, 1-59.

[5] Stone, J., Dreher, B., and Leventhal, A., Hierarchical and parallel mechanisms in the organization of visual cortex, Brain Res. Rev., 1979, 1, 345-394.

[6] Movshon, J., Thompson, I., and Tolhurst, D., Spatial summation in the receptive fields of simple cells in the cat's striate cortex, J. Physiol. (London), 1978, 283, 53-77.

[7] DeValois, Albrecht, D., Thorell, L., Cortical cells: Bar and edge detectors or spatial frequency filters, in Frontiers in Visual Science, S. Cool and E. Smith (eds.), New York, Springer, 1978.

[8] Widder, D., The Heat Equation, New York, Academic Press, 1975.

[9] John, F., Numerical solution of the equation of heat conduction for preceeding times, Annali de Matematica, 1955, 17, 129-142.

[10] Barlow, H., Critical limiting factors in the design of the eye and the visual cortex, Proc. Roy. Soc. Lond., 1981, B 212, 1-34.

[11] Srinivasan, M., Laughlin, S., and Dubs, A., Predictive coding: A fresh view of inhibition in the retina, Proc. Roy. Soc. Lond., 1982, B, 427-459.

[12] Rodieck, R., Quantitative analysis of cat retinal ganglion cell response to visual stimuli, Vis. Res., 1965, 5, 583-601.

[13] Enroth-Cugell, C., and Robson, J., The contrast sensitivity of retinal ganglion cells of the cat, J. Physiol. (Lond.), 1966, 187, 517-552.

[14] Marr, D., and Hildreth, E., Theory of edge detection, Proc. Roy. Soc. (London), B, 1980, 207, 187-217.

[15] Burt, P., and Adelson, E., The laplacian pyramid as a compact image code, IEEE Trans. Comm., 1983, COM-31, 532-540.

[16] Kimia, B., and Zucker, S.W., Deblurring gaussian blur, IEEE Conf. Computer Vision and Image Proc., Washington, D.C., June, 1983.

[17] Wilson, H., McFarlane, D., and Phillips, G., Spatial frequency tuning of orientation selective units estimated by oblique masking, Vis. Res., 1983, 9, 873-882.

[18] Schiller, P., Finlay, B., and Volman, S., Quantative studies of single-cell properties of monkey straite cortex. I. Spatiotemporal organization of receptive fields, J. Neurophys., 1976, 6, 1288-1319.

[19] Watson, A., Barlow, H., and Robson, J., What does the eye see best?, Nature, 1983, 302, 419-422.

[20] Witkin, A., Scale space filtering, Proc. 8th Int. Joint Conf. on Artificial Intelligence, Karlsruhe, W. Germany, 1983, 1019-1022.

[21] Zucker, S.W., Cooperative grouping and early orientation selection, in O. Braddick and A. Sleigh (eds.), Physical and Biological Processing of Images, Springer, New York, 1983.

[22] Leclerc, Y., Ph.D. Thesis, McGill University, in preparation.

[23] Yuille, A., and Poggio, T., Scaling theorems for zero-crossings, A.I. Memo 722, Cambridge: M.I.T., 1983.

[24] Sakitt, B., and Barlow, H., A model for the economical encoding of the visual image in cerebral cortex, Biol. Cybern., 1982, 43, 97-108.

[25] Hummel, R., Sampling for spline reconstruction, SIAM J. Appl. Math., 1983, 43, 278-288.

[26] Burr, D., Motion smear, Nature, 1980, 284, 164-165.

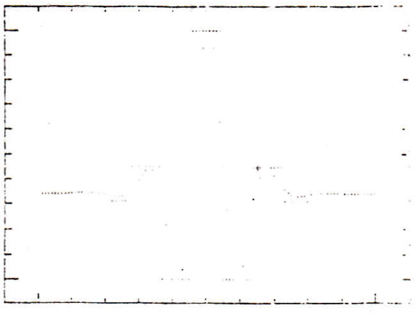

Fig. 2. A comparison of [illegible] notification [illegible] necessary practically or [illegible] physiologically

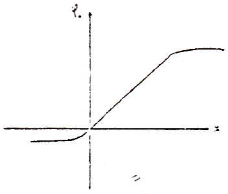

Fig. 3. A sketch of the "positive part" function [illegible] to encode neural [illegible] the linear range between the [illegible] positive and [illegible] negative saturation levels.

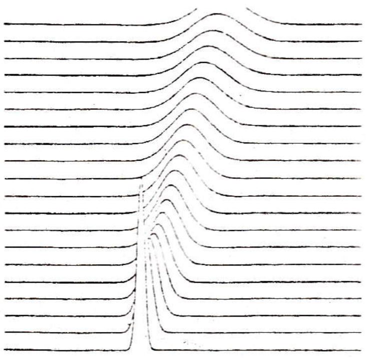

Fig.1. The Gaussian kernel for increasing [illegible] time proceeds, the kernel [illegible] time here becomes analogous with the spatial coefficient (or "scale") of the Gaussian.