# RECEPTIVE FIELDS AND THE
# REPRESENTATION OF VISUAL INFORMATION

Steven W. Zucker

Computer Vision and Robotics Laboratory
Department of Electrical Engineering
McGill University

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, that can then be extended into a variety of representations at the cortex. These cortical representations span a range of sizes, both larger and "smaller" than those in the retina, precisely as is required for further processing. The scheme is both physiologically and psychophysically plausible, and leads to a number of predictions about receptive field size, structure, and hyperacuity that are supported. The hypothesis is supported by a formal scheme for reconstructing visual information at the cortex which proves that no information is lost. 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".

## Introduction

The structure of receptive fields constrains two aspects of early visual information processing: which measurements are taken over the raw retinal image, and how transformations of these measurements provide a representation of visual information rich enough to efficiently support subsequent processing. We shall concentrate on the X-pathway [1], along which retinal receptive fields exhibit a circular surround organization with excitatory/inhibitory interactions, while cortical (V-I) receptive fields are elongated [2]. Since this implies that the first stages of image analysis, such as orientation selection, actually take place in the cortex, the problem of how precise visual information can be communicated from the retina to the cortex arises [3]. We propose a formal solution to this problem based on a diffusion equation. Actual limits on the precision are available from the numerical analysis of the solution, and they agree with hyperacuity data. The remaining critical portions of the model are consistent with receptive field physiology. It requires, in particular, that cortical receptive fields span a wide range of sizes, and predicts, among other points, that the smaller ones will have extra side lobes [4].

The key idea behind the model is that receptive fields are composed of operators that can be modeled as convolutions with Gaussians, and that these Gaussians can be interpreted as "blurring" the image data. (Actually, receptive fields are composed of differences of Gaussians, a technical point that we shall deal with shortly.) There are then three basic questions to consider: (1) how is it possible in principle to undo this blurring [5]; (2) how good can such schemes be in practice [6]; and (3) how might such schemes be reduced to physiological terms? Our result leads to a representation of visual information distributed across receptive fields spanning a range of sizes, which is exactly the form in which subsequent processes require it [7,8]. Finally, to prove in principle that no information has been lost by this representation, we offer a scheme by which a precise visual image can be reconstructed from a combination of these multiple-size representations.

Our scheme, which supports the position that image analysis begins in the cortex, differs in basic technical ways from other schemes proposed to account for the communication gap and hyperacuity. One such class of schemes is based on $(\sin x)/x$ reconstruction filters [3,9]. The numerical analysis of such filters shows that they require too much spatial support (i.e., several lobes on either side) to function properly with the limited support apparently available. Our scheme is based on local (Hermite) polynomials [5]. A completely different theoretical position is that the receptive field operators are participating in "edge detection" [10], but both computational experiments and mathematical analysis provide evidence that this position is untenable [8].

## Blurring and De-blurring

Our scheme is derived from a diffusion in which (temporal) spread will becomes analogous to (spatial) extent of receptive fields. It is formally based on the heat equation, the simplest such diffusion which has all of the necessary mathematical properties. The basic assumption carried by the heat equation is that, for a class of functions, certain spatial derivatives (laplacians) will be formally equivalent to certain (i.e., the first) temporal derivatives. We shall begin by arguing intuitively for this assumption. Observe that, for a long conducting wire, a unit impulse of heat diffuses into increasingly larger Gaussian distrubutions as time proceeds. Mathematically, let f(x) denote the initial temperature distribution as a function of the spatial variable $x \in \Re^n$. (Clearly we are interested in the special case when $n = 2$.) Then a solution to the heat equation $u(x,t)$ giving the temperature level as a function of position $x$ and positive time $t$, satisfying

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

can be obtained from the convolution

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

where $K(x,t)$ is the "source" kernel [11]

$$K(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-|x|^2/4t}.$$

Note that this source kernel is just a Gaussian. Since it 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 (Fig. 1). In fact, assuming $f(x)$ is bounded, $u(x,t)$ as given above is entire analytic. It is the unique bounded solution to the heat equation

$$u_t = \Delta u$$

satisfying $u(x,0) = f(x)$, where $\Delta$ denotes that spatial laplacian and $u_t$ denotes $\partial u/\partial t$. Other unbounded solutions are technically possible, buth the function $u(x,t)$ given by convolution against the Gaussian kernal $K$ is the one that naturally occurs in physical systems.

Suppose we take the temperature distribution as representing the image data, but blurred by the 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 $0 < t \leq \tau$? Can $f(x) = u(x,0)$ be recovered? This is the problem of deblurring Gaussian blur.

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 sufficent conditions for the existence of a solution to the heat equation, $u(x,t), 0 < t \leq \tau$, satisfying $u(x,\tau) = g(x), x \in \Re^n$, are that $g(x)$ be analytic, and that the extension of $g(x)$ to an analytic function of several complex variables $g(z)$, $z \in C^n$, satisfy certain growth conditions. Both of these conditions, analyticity and 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 [6] 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 $\epsilon$. Then an estimate $\hat{u}(x,t)$ can be formed using discrete, local kernels such that the error $|\hat{u}(x,t) - u(x,t)|$, for $a \leq t \leq \tau$, is bounded by a constant (depending on $a$) times $\mu^{(1-\theta)}\epsilon^\theta$. Here $\theta$ 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 $\epsilon$. The coefficents of this kernel are shown in Fig. 2. This bound on the numerical stability of the backsolution is essential, because it dictates how far the process can in fact be made to work. It is the practical limit on the deblurring process.

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

## Physiological Interpretation

The physiological interpretation begins with the observation that retinal receptive fields consist of an excitatory center with an inhibitory surround. (The fact that they also consist of inhibitory centers with excitatory surrounds will be dealt with shortly.) Such local operations are useful in data coding, since they compress the required dynamic range of a channel [12, 13]. 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 to recover actual image information. However, as we now show, given the form of these operators, such recovery is straigtforward.

515

Circular surround receptive fields have been modeled by kernels given either as the difference of two Gaussians [14, 15], or as the Laplacian of a Gaussian [10]. 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

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

which is the difference of two Gaussians. We therefore take

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

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 $t \geq \tau$, the smallest receptive field (Fig. 3). Since

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

$v(x,t)$ 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 $v(x,t) = \Delta u(x,t) = \partial u(x,t)/\partial t$, 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 alternate 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 $0 < t \leq \tau$, 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 $t \geq \tau$. The data are sampled spatially, and it is likely that $v(x,t_i)$ is sampled more coarsely in $x$ for larger $t_i$, reflecting the more uniform variations of the smoothed data [14]. Values for $v(x,t_i)$, $i = 1, 2, \ldots, k-1$, with $0 < t_i < \tau$, will be obtained by backsolving. An examination of these de-blurring operators reveals the introduction of extra side lobes in a fashion strongly reminiscent of the small cortical receptive fields; see Fig. 2.

We now show that all of the original image information is available by providing a reconstruction scheme consistent with the physiological representation of the information. The integral for recovering $f(x)$ can be approximated by a weighted sum

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

thereby requiring the different size operators (or cortical receptive fields). 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 de-blurred, and it is precisely these smaller operators (receptive fields) that have been observed to have the extra side lobes [4; 16].

The last remaining issue is stability. Recall from John's results that the error is bounded, and that the accuracy is controllable, if the backsolution is non-negative. However, $v(x,t)$ can be both positive and negative, and can have a large dynamic range. But the physiology constrains neurons to only carry information within a limited range. Both of these problems can be solved by taking the measurements not as $v(x,t)$, but by splitting them in such a way that the positive and negative parts of $\Delta f(x)$ are carried separately. One way to accomplish this split is based on the observation that X-pathway neurons have a low spontaneous firing rate. We interpret this to imply that these neurons encode positive information relatively well, but negative information poorly. We accomplish this decomposition by introducing the functions

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

and

$$\Phi_-(f(x)) = -\min[f(x), 0]$$

Actually these functions are only approximations; they should really be analytic

and have a small negative excursion; see Fig. 4. We can then separate the data into

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

and

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

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 [1; 17]. Furthermore, patterns matching these receptive fields are the ones to which our visual systems are the most sensitive [18]. 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

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

## Summary and Conclusions

In summary, the reconstruction method takes as given a number of copies of the image filtered through increasingly larger Laplacian of a Gaussian masks. It is these masks that have been observed as receptive fields. The basic steps are then: (1) backsolve the data from the smallest masks to estimate data for even "smaller" masks. Again, this amounts to filtering with masks that have been observed as receptive fields, but which differ from the previous ones in that they possess additional side lobes. (2) Add the given filtered data to the backsolved data to obtain the most precise image data. Finally, for mathematical and numerical stability reasons, we proposed that the above steps are essentially carried out separately for the positive and negative parts of the filtered data, with the results recombined at the end. Physiologically this is necessary given the limited numerical range of neural coding, and gives rise to the complimentary excitatory/inhibitory organizations that have been observed.

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 to physiological variation in neuronal conduction velocities. Others have argues that it is only special features of the convolutions, such as zero crossings, that matter [21,22]. Ultimately, the physiology will decide.

## REFERENCES

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

[2] Hubel, D., and Wiesel, T., Functional architecture of macaque monkey visual cortex, Proc. Roy. Soc. (London), B, 1977, 198, 1-59.

[3] Barlow, H., Critical limiting factors in the design of the eye and the visual cortex, Proc. Roy. Soc. Lond., 1981, B 212, 1-34; Reconstructing the visual image in space and time, Nature, 1979, 279, 189-190.

[4] 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.

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

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

[7] 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.

[8] Leclerc, Y., and Zucker, S. W., The local structure of intensity discontinuities in one dimension, Proc. 7ICPR, Montreal, 1984.

[9] Crick, F., Marr, D., and Poggio, T., An information processing approach to understanding the visual cortex, in Schmitt, Warden, Adelman, and Dennis, (eds.), The Organization of the Cerebral Cortex, MIT Press, 1981.

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

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

[12] 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.

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

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

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

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

[17] 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.

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

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

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

Fig. 2. The strucure of a ($6^{th}$ order) Gaussian deblurring filter. Note the presence of the side lobes.
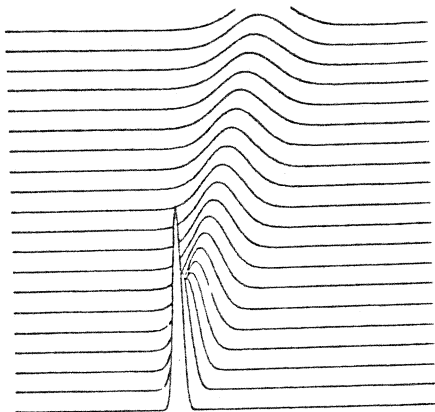


Fig.1. The Gaussian kernel for increasing time; note how, as time proceeds, the kernel spreads out, so that time here becomes analagous with the spatial coefficent (or "scale") of the Gaussian.



Fig. 3. The circular surround (Laplacian of a Gaussian) filters for increasing scale. Note that the smallest one is not infinitesimal, but is restricted by the retinal "wiring".
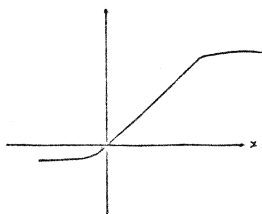


Fig. 4. A model for numerical encoding by optic tract neurons. It resembles the "positive part" function, but actually is analytic with a small negative excursion.
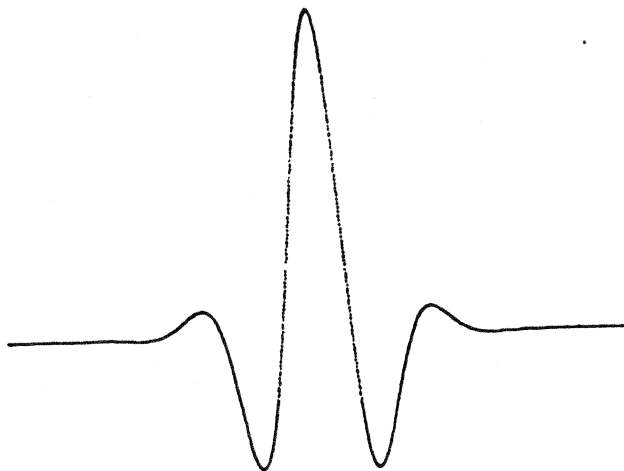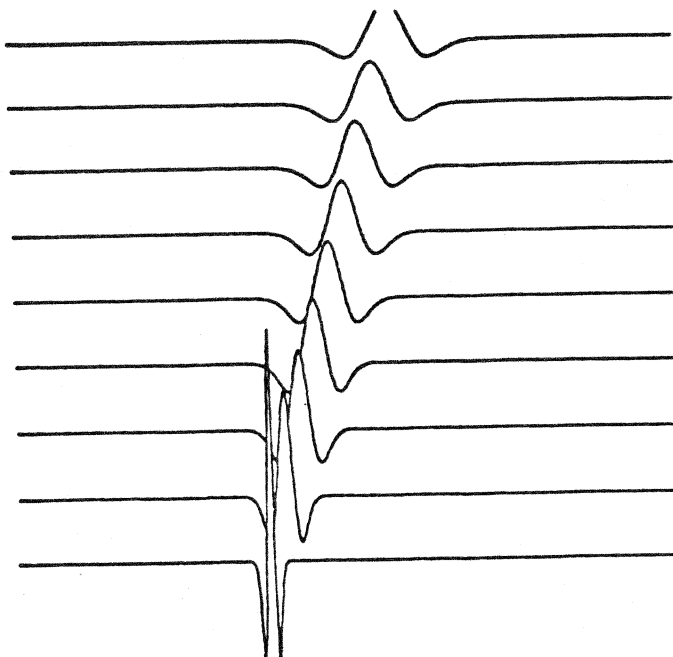
517