Statistical Mechanics and Invariant Perception

William Bialek and A. Zee
Institute for Theoretical Physics, University of California, Santa Barbara, Santa Barbara, California 93106
(Received 26 June 1986)

We consider the problem of discrimination among ensembles of images generated by distortions of a prototype and the addition of noise. As the noise level is increased the discrimination task becomes qualitatively more difficult in that optimal discrimination requires the computation of increasingly longer-ranged correlations or the solution of increasingly difficult optimization problems. These results suggest the use of such image ensembles in probing the computational abilities of the human visual system, and possible theoretical implications of such experiments are discussed.

PACS numbers: 87.30.Ew, 05.20.−y, 87.10.+e, 89.70.+c

A basic question about the central nervous system concerns its capacity to solve problems of great computational complexity. There is at present substantial interest in the connections between complexity theory and statistical mechanics on the one hand,[1] and between the dynamics of various statistical systems and neural networks on the other.[2] What is missing is some quantitative or even qualitative characterization of the computational abilities of *real* nervous systems which could be used as a guide in formulating theoretical issues.

To make progress we choose a very specific issue, the question of local versus nonlocal computation in the visual system. Most recent theoretical attempts to understand visual information processing have relied on the "feature detector" ideas which originated in the physiological experiments of the 1950's[3]: Neurons in the visual system are assumed to compute nonlinear functionals of the image intensity which signal the presence of local features in the image,[4] converting a continuous pattern into a set of discrete "feature tokens" which can be processed by subsequent layers of neurons. An alternative view is that of the *Gestalt* physchologists, who maintain that the perceptual content of an image cannot be reduced to the sum of local features, particularly in view of the fact that our perceptions are invariant under a very large group of transformations.[5]

There are some problems, such as the detection and identification of small-amplitude movements, for which local computations are not only a possible solution,[6] but in fact the optimal solution in a well-defined sense.[7] Here the structure of the optimal computation maps cleanly onto the parallel architecture of the visual system, and so it is (qualitatively) easy to understand how the biological system achieves its impressive speed. If, on the other hand, we have a problem whose solution requires the computation of strongly nonlocal correlations among the outputs of the photoreceptor array, then the mapping of the problem onto the system architecture becomes nontrivial, and feature-detector neurons of the usual type are rather useless.

Our goal in this Letter is to outline a strategy for construction of nonlocal problems which exploits the established invariances of our perceptions. The tasks we consider involve discrimination among *ensembles of images*, an idea which has its roots in Julesz's studies of discrimination among stochastic "textures."[8] Statistical-mechanics methods provide a powerful analytic approach to an understanding of the kinds of computations that are required for optimal discrimination among such ensembles. We emphasize that our goal is to use these methods in the design of new experiments which can probe the computational abilities of the visual system in a theoretically well-motivated manner; it is not our intention to present a theory of perception.

An image (or more precisely a picture) may be described by a field $\phi(x)$, scalar for black-and-white pictures and vector for color. Imagine starting with some prototype image $\phi_0(x)$, distorting this image, and adding noise to obtain the image $\phi(x)$ which the observer actually sees; distortions are summarized by another field $\chi(x)$. In the absence of noise there is some deterministic procedure which allows us to apply a particular distortion χ to the prototype ϕ_0 and obtain the image ϕ. In the presence of noise, ϕ is only probabilistically related to a distorted version of the prototype, and so we define some probability distribution functional $P[\phi(x)\,|\,\phi_0(x);\chi(x)]$. To be precise we also asssume that the distortion is randomly chosen from some known probability distribution $P[\chi(x)]=\exp\{-W[\chi(x)]\}$. Then a single prototype image $\phi_0(x)$ generates an *ensemble of images* defined by the conditional probability of our observing a particular $\phi(x)$ given $\phi_0(x)$:

$$P[\phi(x)\,|\,\phi_0(x)]=\{Z[\phi_0(x)]\}^{-1}\int D\chi(x)e^{-W[\chi(x)]}P[\phi(x)\,|\,\phi_0(x);\chi(x)]. \quad (1)$$

The interesting and experimentally accessible quantities are the reliabilities with which two or more such ensembles can be discriminated from one another. Optimal unbiased discrimination is accomplished by maximum likelihood[9]: If

© 1987 The American Physical Society

we are forced to choose between prototypes $\phi_0(x)$ and $\phi_1(x)$, for example, we calculate

$$\lambda[\phi(x)] = \ln\left[\frac{P[\phi(x)\,|\,\phi_0(x)]}{P[\phi(x)\,|\,\phi_1(x)]}\right] \tag{2}$$

and guess that $\phi(x)$ derived from $\phi_0(x)$ if $\lambda[\phi(x)] > 0$ and conversely. The probability of correctly identifying the ensemble generated by $\phi_0(x)$ is

$$P_c[\phi_0(x) \text{ vs } \phi_1(x)] = \int D\phi(x) P[\phi(x)\,|\,\phi_0(x)]\Theta\{\lambda[\phi(x)]\}. \tag{3}$$

Here Θ denotes a step function. The difficulty of computing $\lambda[\phi(x)]$ and thus reaching optimal performance may be studied by our taking Eq. (1) seriously as a statistical mechanics for the scalar field $\phi(x)$. Thus we write

$$P[\phi(x)\,|\,\phi_0(x)] = \{Z[\phi_0(x)]\}^{-1}\exp\{-S_{\text{eff}}[\phi(x);\phi_0(x)]\}, \tag{4}$$

and

$$Z[\phi_0(x)] = e^{-F[\phi_0(x)]} = \int D\phi(x)\exp\{-S_{\text{eff}}[\phi(x);\phi_0(x)]\}, \tag{5}$$

where the effective action $S_{\text{eff}}[\phi(x);\phi_0(x)]$ is obtained by our integrating out the distortion field and the free energy $F[\phi_0(x)]$ is defined in the usual way as a function of the external field $\phi_0(x)$. We see that

$$\lambda[\phi(x)] = S_{\text{eff}}[\phi(x);\phi_1(x)] - S_{\text{eff}}[\phi(x);\phi_0(x)] + F[\phi_0(x)] - F[\phi_1(x)], \tag{6}$$

so that the complexity of computing $\lambda[\phi(x)]$ is determined by the structure of the effective action. If this action is local in $\phi(x)$ then for an image sampled at N points a serial computer can evaluate $\lambda[\phi(x)]$ in $O(N)$ computational steps, and we expect that the nervous system can solve the problem with $O(1)$ layer of neurons in $O(1)$ time step. Following the discussion above, any task with such a local effective action cannot be used to look for computational abilities beyond those available in conventional feature detectors.

A basic fact about human vision[5] is that it admits, at least approximately, a large group of invariances or "perceptual constancies." Thus we recognize triangles as triangles despite changes in size, orientation, and position in the visual field, we recognize colored objects despite changes in the spectral content of the illumination, and so forth. Evidently certain distortions $\chi(x)$ are considered nearly as plausible as no distortion at all. This presumably reflects the relative likelihood of such distortions occurring in the natural environment, so that if we want to simulate "natural" tasks we must choose a weighting function $W[\chi(x)]$ which assigns minimal weight to distortions that respect the perceptual invariances. To make things simpler we promote these approximate invariances of perception to exact symmetries of our artificial ensembles. Space permits us to discuss only one example; details and more examples are given elsewhere.[10]

The geometric invariances of perception are perhaps the most obvious. To a first approximation we can identify any rigid Euclidean transformation of an image, and indeed these transformations are constantly occurring as we move through the world. Geometric distortions of a two-dimensional image are defined by the function $y(x)$ which maps points in one image to points in the other, and if we add spatially white contrast noise we have

$$\exp\{S_{\text{eff}}[\phi(x);\phi_0(x)]\} = \int Dy(x)\exp\left[-W[y(x)] - \frac{1}{2C}\int d^2x[\phi(x) - \phi_0(y(x))]^2\right]. \tag{7}$$

The weighting function $W[y(x)]$ must be chosen so that the rigid translations, rotations, and dilations are given minimal weight if we are to respect the Euclidean invariance of our perception, and if we promote this invariance to an exact symmetry then W itself must be a scale-invariant scalar quantity. This symmetry of course provides severe constraints; the simplest allowed possibility is to write $y(x) = x + A(x)$, with the components $A_\mu = \partial_\mu\eta + \varepsilon_{\mu\nu}\partial_\nu\chi$, and choose

$$W = -\int d^2x[(1/g_1^2)\eta(x)\partial^6\eta(x) + (1/g_2^2)\chi(x)\partial^6\chi(x)].$$

The functional integral in Eq. (7) can be viewed as a statistical-mechanics problem for the field $y(x)$, for which several approaches are possible. The first is perturbative, where we write

$$\phi_0(y(x)) = \phi_0(x + A(x)) \sim \phi_0(x) + A_\mu(x)\partial_\mu\phi_0(x) + \cdots$$

and perform a double expansion in A and $\Delta\phi = \phi - \phi_0$. Carrying through the algebra one finds a masslike term

$$\frac{1}{2C}\int d^2x\, A_\mu(x)A_\nu(x)[\partial_\mu\phi_0(x)\partial_\nu\phi_0(x)]$$

which regulates the infrared singularities of the free action $W[\eta(x);\chi(x)]$ and thus must be included in the zero-order action around which one does perturbation theory. To do this one replaces $\partial_\mu\phi_0(x)\partial_\nu\phi_0(x)\rightarrow\delta_{\mu\nu}\phi_0^2/l^2$, where ϕ_0 and l^{-1} are typical values of $\phi(x)$ and its logarithmic derivative, respectively. Although this seems like a drastic approximation, it captures the analytic structure introduced by the mass term and can be improved by a convergent perturbation theory provided that the $\phi_0(x)$ is relatively smooth. In the opposite limit of a "bloblike" image we have an alternative approach, as described below.

Expansion in A is sensible only if typical values of A which enter the integration are smaller than l, and this allows us to identify the perturbation parameter $\varepsilon=(g^2C/\phi_0^2l^2)^{1/4}$, with g some combination of $g_{1,2}$. We find the leading nonlocal contribution to the effective action to be

$$\sim\frac{\varepsilon^2l^2}{2C^2}\int d^2x\,d^2y[\partial_\mu\phi_0(x)\partial_\mu\phi_0(y)]\Delta\phi(x)\Delta\phi(y)e^{-|x-y|/\xi}F(|x-y|/\xi), \tag{8}$$

where F is an oscillatory function and the correlation length is $\xi=s^{-1}l\varepsilon$. In this expression we assume that g_1/g_2 is of order 1; if we vary the coupling ratio over a wide range we can achieve the same effective correlation length in many different ways. We have analyzed higher-order terms in the perturbation series[10] and find that the qualitative long-distance behavior is unchanged, although the correlation length acquires finite corrections and there are terms in the effective action involving multipoint correlations of $\Delta\phi$. Thus there exists a well-defined limit, $g\rightarrow 0$ at fixed ε, in which arbitrarily long-ranged correlations make a finite contribution to the effective action and thus must be computed if one is to reach optimum discrimination performance. If we also allow ε to become of order 1 then arbitrarily many-point nonlocal correlations become important in the computation of $\lambda[\phi(x)]$.

An alternative approach to evaluation of $S_{\rm eff}[\phi(x);\phi_0(x)]$ is in mean-field theory. This is equivalent to our finding a particular distortion $y(x)$ which brings $\phi(x)$ and $\phi_0(y(x))$ into correspondence —template matching— and which at the same time is plausible as measured by $W[y(x)]$. In mean-field theory we approximate

$$-S_{\rm eff}^{\rm mft}[\phi(x);\phi_0(x)]\sim W[y_*(x)]+\frac{1}{2C}\int d^2x[\phi(x)-\phi_0(y_*(x))]^2, \tag{9}$$

where $y_*(x)$ is the solution of the variational equation for stationary points of the exponent in Eq. (7). The difficulty in computation of this approximation to $S_{\rm eff}$ and hence λ is generally dominated by the complexity in the solving of the variational problem. While variational problems do not fit into the classification of local and nonlocal problems introduced at the outset, the structure of the variational calculation confirms our conclusions about the inherent difficulty in the solving of a nonlocal problem.

Suppose that $\phi(x)$ is chosen out of the ensemble defined by $\phi_0(x)$, so that $\phi(x)=\phi_0(y_0(x))+\psi(x)$, with $y_0(x)$ the "correct" distortion and ψ some instance of the added noise. We need to compute objects like

$$E[y(x);y_0(x)]=\int d^2x\,\phi_0(y(x))\phi_0(y_0(x)),$$

and

$$\delta E[y(x)]=\int d^2x\,\phi_0(y(x))\psi(x).$$

For bloblike images of the form $\phi_0(x)=\sum_m B(x-b_m)$, with B a narrowly peaked function centered on the origin, each of the terms E_{mn} in the expansion

$$E[y(x);y_0(x)]=\sum_{mn}\int d^2x\,B(y(x)-b_n)B(y_0(x)-b_m)=\sum_{mn}E_{mn}[y;y_0]$$

vanishes unless $y^{-1}(b_n)\approx y_0^{-1}(b_m)$. Schematically we can write

$$E[y(x);y_0(x)]\sim l_0^2\sum_{mn}\tilde{B}(y^{-1}(b_n)-y_0^{-1}(b_m)), \tag{10}$$

where $\tilde{B}$ is some new sharply peaked (dimensionless) function and l_0 is the blob width which must appear to keep the dimensions right. In this approximation the "potential" $(2C)^{-1}\int d^2x[\phi(x)-\phi_0(y(x))]^2$ depends only on y^{-1} at a discrete set of points corresponding to the blob positions. This potential consists of a deterministic piece which favors y being close to y_0, with the depth of this minimum of order of the number of blobs N, and a fluctuating piece of order $\sqrt{N}$ which arises from coupling to ψ. There are "false minima" of the deterministic potential at values of y which bring some but not all of the blobs into correspondence, and there are completely erroneous minima generated by the nosiy background potential. These latter effects can be calculated by well-known techniques for study of the extrema

of random Gaussian fields,[11] here in $2N$ dimensions. The result[10] is that erroneous minima of depth $E \gg \sqrt{N}$ occur with a probability density (at large N) $P(E) \sim E^{2N}\exp(-E^2/\alpha N)$, where α is a parameter proportional to the noise level. This implies that typical erroneous minima of depth $E^* \sim N\sqrt{\alpha}$ and correspondingly the barriers between false and true minima are of order N, not $\sqrt{N}$. This is similar to the structure found in the "energy landscape" of NP-hard optimization problems such as the traveling salesman,[1] suggesting (although by no means proving) that as the noise level α becomes large this problem also becomes difficult.

We have given an example of a rather natural discrimination task in which arbitrarily long-ranged and multipoint correlations must be computed if optimum performance is to be reached at least in certain limits which are controllable as the different image ensembles are generated; this same structure is found in other tasks as well.[10] It is known that, in discrimination among simpler image ensembles, human observers can approach optimum performance in the sense defined here.[12] This suggests experiments in which the performance of humans is measured as a function of the parameters which control the correlation lengths defined above: If the visual system can only compute local functionals, as with feature detectors, performance (percent correct discrimination) should follow the optimum only for a restricted range of correlation lengths and then fall away dramatically. If, on the other hand, the system can adapt to compute strongly nonlocal functions of image intensity or solve hard optimization problems, no such abrupt drop will be observed. These experiments will be difficult, but they have the potential to provide serious challenges to our understanding of computation in the nervous system. As the reader may have guessed, our suspicion is that the system *can* solve nonlocal problems, and that there are interesting theoretical questions to be answered about the algorithms and hardware responsible for such computations. Suspicions aside, the approach described here provides the tools for asking very definite questions about the computational abilities of the brain.

We thank J. Langer, J. Moody, and S. Palmer for helpful discussions. This work was supported by the National Science Foundation under Grant No. PHY82-17852, supplemented by funds from the National Aeronautics and Space Administration.

[1]F. Barahona, J. Phys. A **15**, 3241 (1982); S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983); C. P. Bachas, J. Phys. A **17**, L709 (1984); J. Vanninmenus and M. Mezard, J. Phys. (Paris), Lett. **45**, 1145 (1984); H. Orland, J. Phys. (Paris), Lett. **46**, 763 (1985); M. Mezard and G. Parisi, J. Phys. (Paris), Lett. **46**, 771 (1985); S. Kirkpatrick and G. Toulouse, to be published.

[2]J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982), and **81**, 3088 (1984); D. J. Amit, H. Gutfruend, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985), and Phys. Rev. Lett. **55**, 1530 (1985); J. J. Hopfield and D. W. Tank, Biol. Cybern. **52**, 141 (1985).

[3]H. B. Barlow, J. Physiol. (London) **119**, 69 (1953); S. W. Kuffler, J. Neurophysiol. **16**, 37 (1953); J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, Proc. IRE **47**, 1940 (1959).

[4]For a formalization of feature-detector ideas in these terms see T. Poggio and W. Reichardt, Q. Rev. Biophys. **9**, 377 (1976).

[5]For a recent review see S. Palmer, in *Human and Machine Vision,* edited by J. Beck, B. Hope, and A. Rosenfeld (Academic, New York, 1983), p. 269.

[6]W. Reichardt and T. Poggio, Q. Rev. Biophys. **9**, 311 (1976).

[7]R. R. de Ruyter van Steveninck, W. H. Zaagman, and W. Bialek, to be published; R. R. de Ruyter van Steveninck, Academisch Proefschrift, Rijksunviersitiet Groningen, 1986 (unpublished).

[8]B. Julesz, IRE Trans. Inf. Theory **8**, 84 (1962).

[9]J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals* (McGraw-Hill, New York, 1950); D. M. Green and J. A. Swets, *Signal Detection Theory in Pyschophysics* (Krieger, New York, 1966).

[10]W. Bialek and A. Zee, to be published.

[11]S. O. Rice, in *Selected Papers on Noise and Stochastic Processes* edited by N. Wax (Dover, New York, 1954), p. 133.

[12]H. B. Barlow, Philos. Trans. Roy. Soc. London, Ser. B **290**, 71 (1980).