Biol. Cybern. 44, 47–58 (1982)

Biological Cybernetics

A Model of the Early Stages of the Human Visual System: Functional and Topological Transformations Performed in the Peripheral Visual Field

C. Braccini, G. Gambardella, G. Sandini, and V. Tagliasco

Istituto di Elettrotecnica, University of Genova, Genova, Italy

Abstract. A model of the early stages of the visual system is presented, with particular reference to the region of the visual field outside the fovea and to the class of retinal and lateral geniculate nucleus cells which are most active in the processing of pattern information (*X*-cells). The main neuroanatomical and neurophysiological properties taken into account are: the linear increase of the receptive fields diameter with eccentricity, the constancy of the overlap factor and the topological transformation operated upon the retinal image by the retino-cortical connection. The type of filtering taking place between the retina and the visual cortex is analyzed and some simulations are presented. It is shown that such a filtering is of a band-pass space variant type, with center frequencies that decrease from the center (i.e. the fovea) toward the periphery of the visual field. This processing is "form invariant" under linear scaling of the input. Moreover, considering the properties of the retino-cortical connection, it is shown that the "cortical images" undergo simple shifts whenever the retinal images are scaled or rotated.

Introduction

The aim of this work is to present a contribution to the modeling of the human visual system, by giving a functional interpretation of the neuroanatomy and neurophysiology of the retina and the lateral geniculate nucleus.

Under a few simplifying hypotheses, we will show how the structure of the peripheral visual system is such to imply an image processing (or transformation) whose result (or output) is pseudoinvariant with respect to linear scalings and rotations of the stimulus (or input). Here, the term pseudoinvariant refers to the property that the output of a two-dimensional processing is simply shifted in its representation domain when the input is linearly scaled or rotated. This property will be also referred to in the sequel as "form-invariance".

The idea that the peripheral visual system operates in a form-invariant fashion on the retinal input has already been introduced, in its most explicit form, by Schwartz (1977, 1980a, b). His approach, however, assumes a direct representation (i.e. a simple projection) of the image from the retina into the cortex, without any intermediate processing. The present work shows that the processing performed by the peripheral nervous cells belongs to that particular class of linear signal transformations such that a linear scaling of the input is mapped into a linear scaling of the output, thus yielding the pseudoinvariance property.

The organization of the paper is as follows. In Sect. 1 we present an analysis of the cells and the structure of the peripheral visual system together with the related simplifying hypotheses and assumptions, that represent the physical base of our mathematical formalization. The latter is described and discussed in detail in Sect. 2, while some simulation examples are presented in Sect. 3. In the fourth and last section, we will discuss the limits of our work and compare it with recent related works by other authors who have pursued the same line of thought and reached similar results, as in the case of the already mentioned Schwartz's paper.

1. Properties of the Peripheral Visual System

In this section, we review some properties of the peripheral stages of the visual system, i.e. the retinal ganglion cells (RGC) and the lateral geniculate nucleus (LGN). In doing so, we emphasize the signal processing aspects, that represent the base of the mathematical model introduced in Sect. 2.

0340-1200/82/0044/0047/$02.40

1.1. Retina and Lateral Geniculate Nucleus

The visual information, processed by the retinal receptors and the first layers of retinal cells, is coded, in terms of nervous impulses, at the level of RGCs. After the pioneering work of Granit (1947), the first description of RGCs in term of receptive field (RF) properties was made by Kuffler (1953). He described RGC's RF as being formed by two mechanisms, the center, roughly circular in shape, which produces a response either to the introduction (ON response) or to the withdrawal (OFF response) of the stimulus, and the surround, annular in shape, which produces a response opposite to that observed in the center. In this way RGCs can be subdivided into two classes, one characterized by an ON center and an OFF surround, the other by an OFF center and an ON surround; the percentage of cells belonging to each class is almost equal.

Starting from these basic studies, Enroth-Cugell and Robson (1956) and Cleland et al. (1971) distinguished two further classes: the classes *X* and *Y*. In particular, for *X*-cells the spatial summation over their RF is approximately linear, while for *Y*-cells it is nonlinear. Only recently (Stone and Hoffmann, 1972) a third class of RGC has been described: the *W*-cell class, characterized by the slow conduction velocity of their axons.

Anatomically, three different classes of neural cells were described in the retina: ALPHA, BETA, and GAMMA cells, and their relation with the functional classes *X*, *Y*, and *W*-cells has been clearly established, at least for what concerns *X* and *Y*-cells, which are shown to correspond respectively to ALPHA (Cleland et al., 1975) (large sized cell body) and BETA (medium sized cell body) (Cleland et al., 1975; Boycott and Wässle, 1974) populations. The remaining class, that of *W*-cells, having slow conduction velocity, is thought to belong to the anatomical GAMMA class (small sized cell body).

Coming back now to the properties of the receptive field of RGCs, a further specification has to be considered, namely the interaction between the center and surround mechanisms previously introduced. Rodieck and Stone (1965) suggested that a center-surround RF might be produced by two independent mechanisms: the center, summing the activity of all the elementary areas that exhibit a center-type response, and the surround, which conversely represents the activity of all the elementary areas of surround-type response; the two mechanisms are spatially described by two bell shaped weighting functions, both centered on the geometric center of the RF and acting antagonistically upon the ganglion cell. It should be noticed that, in this model, the term surround loses its topological meaning as well as the annular shape. Furthermore, Enroth-Cugell and Robson (1966) found that, in the case of *X*-cells and stationary stimuli, signals from photoreceptors in the center and surround regions add linearly within each region before affecting the response of the ganglion cell. We will further assume that, in the case of *X*-cells, the signals coming from the center and the surround regions subtract arithmetically under any type of luminous stimulation and, in accordance with other authors (Enroth-Cugell and Robson, 1966; Rose, 1979), that the sensitivity profiles of the center and surround regions are gaussian. This simple model does not hold in describing the activity of *Y*-cells (Hockstein and Shapley, 1976a, b).

Besides the retina, a further processing level must be considered in our model, namely, the LGN level. The role of the LGN in the transformation of visual information has been investigated by many authors. Hubel and Wiesel (1961) first showed that the LGN cells exhibit characteristics similar to those of the RGCs (Bullier and Norton, 1979). The basic differences are: a considerable reduction in the spontaneous activity of *X*-cells at the LGN level, and an increase in the inhibition exerted by the surround on the center response of *X*-cells. Stated in terms of spatial filtering, the band-pass characteristic exhibited by *X*-cells and schematized by the difference of gaussians (DOG) sensitivity profile, still describes the characteristics of LGN *X*-cells, with the only modification that the bandwidth is somehow decreased by the stronger inhibition of the surround mechanism (Maffei and Fiorentini, 1973).

As to their functional role, even if the experimental data are far from being exhaustive, the fact that the *X* population projects to the LGN only and is not very selective for target movement, whereas the *Y* population projects also to the superior colliculus (SC) (where sensory-motor integration takes place) and shows peculiar response to the movement of stimuli over its RF, leads to the hypothesis that, while the role of *X*-cells is predominantly that of pattern analysis, for *Y*-cells this role is qualitatively of minor importance.

1.2. Distribution and Shape of the Receptive Fields

We come now to the description of ganglion cell RFs, in relation to eccentricity, size and coverage of centers. It has been known for a long time that the spatial distribution of photoreceptors over the human retina is nonuniform (Osterberg, 1935; Pirenne, 1967). Moreover, an increase in cone density is associated with an increase in GC density. As the increase in GC density is associated with an increase in the amount of visual information (for unit visual angle) transmitted to

the higher levels of visual processing, it is clear that those regions of higher GC density correspond to the regions of the retinal image which are more densely sampled (or, in other words, the regions of higher visual acuity). As there is a point to point projection of the retinal image into the visual cortex, the resulting mapping is topologically distorted, in the sense that the regions with higher GC density correspond to larger areas in the cerebral cortex.

The first quantitative analysis of GC distribution was made by Stone (1965) in the cat's retina. The estimated range of cell's density goes from about $6000/mm^2$ in the center to less than $100/mm^2$ in the far periphery. Even if these data are derived from the cat's retina, other studies in primates and humans confirm, at least qualitatively, the conclusions presented above. Also interesting is the observation that the ganglion cell density gradient is very similar to the gradient of visual acuity across the human visual field and to the gradient of cortical representation of the visual field in monkeys (Stone, 1965). On the other hand, the size of RGC RF, obtained by measuring the radius of the central region of the RF, is an increasing function of retinal eccentricity (Enroth-Cugell and Robson, 1966; Hubel and Wiesel, 1960; Fischer, 1973). Therefore, we are dealing with two, possibly uncorrelated, parameters: ganglion cell density, which decreases with eccentricity, and RF size, which, on the contrary, increases with eccentricity. Fischer (1973), measuring the mean value of RF centers of RGC in cat at different eccentricities, found a nearly linear increase of center radius. This finding is confirmed by other authors both in cat and monkey (Ikeda and Wright, 1972; Hubel and Wiesel, 1960). It should be noted that the linear increase mentioned above refers to the *mean* RF center at different eccentricities, and not to the absolute dimension, which, on the contrary, is scattered around the mean in a way that will be described later.

A further characteristic of RGC RF, which could vary with eccentricity, is related to the fact that, at each retinal location, the receptive fields of several ganglion cells overlap. This is due to the fact that photoreceptors converging to neighboring ganglion cells are intermixed, so that the same spatial area is covered by the RFs of different ganglion cells. Fischer (1973), starting from the data reported by Stone (1965) about the spatial distribution of RGC, measured the RF centers of the same class of cells at different eccentricities and found that an overlap factor of 35 fits well with the physiological and anatomical data, independently from the visual angle. This finding can be restated by saying that, if one takes as the unit of measure of spatial extent the mean RF center area, the density of RGC is constant at any eccentricity.

The size of the RF centers is a very important factor, especially with respect to the information transfer to the visual pathway (Enroth-Cugell and Robson, 1966; Maffei and Fiorentini, 1973). As it has been stated before, RGCs can be described as band-pass filters operating on the visual information falling over their RF, the bandwidth of the filter being related to the RF center diameter. This means that, if the RFs of a particular class of RGCs at one retinal location have a wide range of sizes, this would have the effect that the spatial information is transferred within different spatial frequency bands, whereas, if the scatter of RF center size is small, at any retinal location, only a limited frequency band will be present at the RGC output. In order to clarify this matter, a combined measure of RGC density and RF center size must be performed at different retinal locations and, if possible, for the three different classes of RGCs. Peichl and Wässle (1979) performed the measurement on *X* and *Y*-cells and found that *X*-cells have smaller RF centers than *Y*-cells, and that the RF centers of the two populations increase with the distance from the area centralis. More quantitatively, they found at each retinal location that the diameter of the largest RF (in each class) is about twice the smallest. Due to the fact that the two ranges ore only slightly overlapped, the overall RF range at any retinal location is about four.

Concerning the overlap of RGC RFs, two parameters must be considered: the mean RF center size and the density of RGCs at each retinal eccentricity. For *X*-cells, the "equivalent center size" (Peichl and Wässle, 1979) (that is the radius where the sensitivity of the center mechanism has dropped to $1/e$) goes from about $20'$ in the central area to about $70'$ at 27 deg eccentricity; in the same range of eccentricities, the density of BETA cells goes from $6500/mm^2$ to about $100/mm^2$ (Peichl and Wässle, 1979). Therefore, the decrease in density between the area centralis and the periphery is balanced by a simultaneous increase in the RF center dimensions. This produces an overlap factor of 7 to 10 for *X*-cells, except for the central 1 deg of visual field, where the overlap factor for *X*-cells goes up to 30. These data suggest that the RFs of about 8 *X*-cells overlap at each retinal location; moreover, taking into account that the percentage of ON-center units and OFF-center units is equal, we can conclude that, at each retinal location, about 4 ON-center and 4 OFF-center *X*-cells overlap. These data refer to the "equivalent center size" of GC RFs, which is (Peichl and Wässle, 1979) an underestimation of about 1.6 of the radius of the center mechanism, causing an underestimation of the overlap factor of about 2.25.

Concerning the shape of the RGC and LGN receptive fields, for *X*-cells many experimental data show that it has a circularly symmetrical profile that

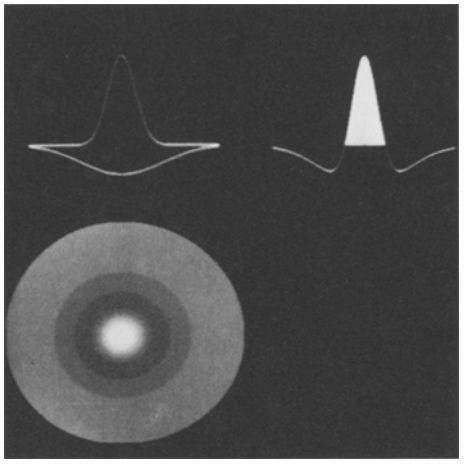

Fig. 1. Example of DOG weighting function: the two gaussian functions (top left); the resulting profile (top right; the white area corresponds to the center mechanism) and its two-dimensional representation (bottom)

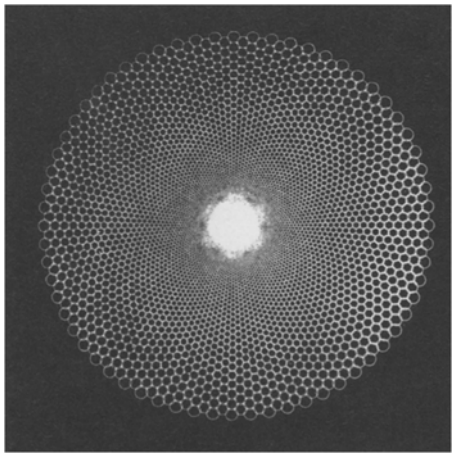

Fig. 2. Schema of the geometric distribution of the RFs of RG and LGN cells. The geometric centers of the circles represent the sampling points and form a triangular lattice. The diameter of the circles is proportional to the RF size. Both the sampling distance and the RF size increase linearly with eccentricity. The parameters of the figure correspond to the high frequency filtering plane, where the overlap between contiguous RFs is minimum (see text)

can be modeled, as already mentioned, by the difference of two gaussian functions:

$$w(\varrho, \sigma_i, \sigma_e) = \frac{A}{\sqrt{2\pi}\,\sigma_e}\exp(-\varrho^2/2\sigma_e^2) - \frac{1}{\sqrt{2\pi}\,\sigma_i}\exp(-\varrho^2/2\sigma_i^2). \quad (1)$$

The two terms of (1) represent the center and surround mechanisms, respectively. Equation (1) can be used to model both RGC and LGN cell behaviour, by introducing different values for A and for the ratio σ_i/σ_e. Specifically, it turns out that values of $\sigma_i/\sigma_e = 2$ and $A = 5$ are appropriate for RGCs, while for LGN RF profiles σ_i/σ_e should be set equal to 3 and A equal to 1.

The previously discussed property that the RF size varies with eccentricity r can be expressed analitically by introducing in (1) a linear dependence of the standard deviations σ_e and σ_i on r. In the case of LGN cells, (1) becomes

$$w(k, r, \varrho, \sigma_i, \sigma_e) = \frac{1}{\sqrt{2\pi}\,kr\sigma_e}\exp(-\varrho^2/2k^2r^2\sigma_e^2) - \frac{1}{\sqrt{2\pi}\,kr\sigma_i}\exp(-\varrho^2/2k^2r^2\sigma_i^2), \quad (2)$$

where different values of the parameter k, representing the linear dependence of the standard deviations on eccentricity, define different classes of RFs (for each cell population). As we will discuss in Sect. 2.3, the number of these classes (or "processing planes") will be determined on the base of some assumptions concerning the overlap factor.

It should be pointed out here that the kind of processing represented by (1) and (2) (i.e. by means of oscillating weighting functions) is a band-pass filtering in the domain of the spatial frequencies ω_x and ω_y. Due to the symmetry of the RF profiles, such a filtering is performed along the radial frequency $\omega_r = (\omega_x^2 + \omega_y^2)^{\frac{1}{2}}$ independently from the orientation. According to (2), when r varies, the center frequency of the band-pass filters varies too, as well as their bandwidth. It can be verified that the latter is proportional to the center frequency at any r, so that the overall filtering performed on the input image is of the constant-Q type. The filter bandwidth is, in the case of (2) and for $\sigma_i/\sigma_e = 3$, of about 1.5 octaves. An example of a DOG profile of this kind is shown in Fig. 1.

A further parameter to be considered is the spacing between the centers of the RFs so far described, that can be referred to as the "sampling distance". This parameter is related to the overlap of RF centers at a given eccentricity: the smaller the sampling distance, the higher the overlap. We know from the literature that, at eccentricities greater than 1 deg, the overlap of RF centers is quasi-constant, and moreover the size of the RFs is uniformly distributed in a range of 1 octave (the smallest being about 1/2 the largest). For this reason, we can say that the sampling distance is a linear function of eccentricity regardless to the size of the RF itself (Fig. 2).

1.3. Retino-cortical Projection

As it has been known for a long time, the retinal input is orderly mapped in diverse neural structures. The particular mapping of the retinal image, as seen at layer IVc of the cat's striate cortex, can be understood

starting from two distinct observations. The first one is that the radius of RGC RFs increases almost linearly as a function of eccentricity; or, stated in other words, the RGC axons transfer information falling on a visual angle which is wider going from the center toward the periphery. The second observation is that the part of the visual system after the retina is homogeneous, meaning that the information carried by each ganglion cell's axon is processed uniformly, with no regard to the eccentricity in the visual field. Topologically, this means that the same amount of processing is devoted to the analysis of narrow angles of the visual field in the central retina and to relatively wider angles in the periphery.

Daniel and Whitteridge (1961) and Hubel and Wiesel (1974), measuring the cortical magnification factor (the extent of visual cortex corresponding to 1 deg in the visual field), found a linear dependence of the inverse of the cortical magnification factor on eccentricity. Starting from this observation, Fischer (1973) and Schwartz (1978) found that the retino-cortical projection can be described as a logarithmic conformal transformation. Later works (Rovamo et al., 1978; Drasdo, 1977) describe a psychophysical correlate of the cortical magnification factor, demonstrating the functional relevance of this mechanism for visual information processing.

2. Mathematical Representation of the Peripheral Visual Processing

In this second section we will present a mathematical model of the processing performed by the cell populations previously analyzed (i.e. RGC and LGN). The model is limited to the X-cells and to the region of the visual field where the overlap factor is constant and the RF size depends linearly on eccentricity.

2.1. Weighting Functions of the RFs *and Form-Invariance*

Letting $f(x, y)$ be the image which is formed in the retina (considered as a planar surface with its center, or origin 0, in the fovea), the processing performed by the RGC, or LGN, considered before, can be approximately represented, in continuous form, by the integral (Braccini et al., 1981)

$$z(r,\theta)=\int_0^{\infty} w_0(\varrho/r)\varrho\, d\varrho \cdot \int_0^{2\pi} f(r\cos\theta-\varrho\cos\varphi, r\sin\theta-\varrho\sin\varphi)\, d\varphi \qquad r>0, \tag{3}$$

where the polar coordinates, $r=(x^2+y^2)^{\frac{1}{2}}$ and $\theta=\operatorname{artan} y/x$, have been chosen since they are more suitable for representing the properties of the weighting functions, namely the circular symmetry and the size growing with r. In (3), (r, θ) is the "running point" around which the image $f(\ )$ is being analyzed [through a receptive field whose center is assumed to coincide with (r, θ)], while $w_0(\ )$ is the set of weighting functions identifying a particular class of RFs. For the time being, we do not further specify the shape of the w_0's and proceed to analyze the properties of (3) in general. In this way, in fact, we will be able to check which properties derive only from the two hypotheses on the RFs already embodied in (3), i.e. the circular symmetry and the size growing linearly with r. An example of how the RF sizes can vary according to this law is shown in Fig. 2, where the circles represent the center region of the RFs.

In signal theory terminology, (3) represents the linear filtering of the signal $f(\ )$ performed by means of a system characterized by the space-variant point-spread function $w_0(\ )$ and it is called a "space-variant convolution".

It has been shown (Braccini et al., 1981; Braccini and Gambardella, 1982) that the basic property of the class of space-variant convolutions defined by (3) is the form-invariance of the output with respect to linear scalings of the input. In formal terms, this can be expressed by saying that, if the input $f(x, y)$ is replaced by its scaled version $f(ax, ay)$, with the parameter a real, then the output $z(r, \theta)$ becomes $1/a^2\ z(ar, \theta)$, that is a linearly scaled version of the previous output (except for the coefficient $1/a^2$), with the same scaling factor of the input signal.

It has been shown (Braccini and Gambardella, 1982) that this property could *not* be satisfied if the input-output relationship, i.e. the convolution, were of the space-invariant type; on the other hand, if a dependence of the weighting functions on r is allowed, the only dependence compatible with the form-invariance property is the one implying that the size of $w(\ )$ linearly grows with r. In this sense we can say that the form-invariance is a basic property of (3).

Let us now see how this property looks like if the output plane (r, θ) of (3) is mapped into the new plane (r', θ), being $r'=\log r$. If $v(r', \theta)=z(\exp(r'), \theta)$ is the output of (3) in the new plane, the form invariance implies now that the output corresponding to the scaled input $f(ax, ay)$ is $1/a^2 v(r'+a', \theta)$ with $a'=\log a$. In other words, the output due to a scaled input, turns out to be a simply shifted replica (along the r' axis) of the previous output.

It is worth noticing that if the output space (r', θ) is a cartesian space in the coordinates r' and θ, then also the input rotations around the origin are mapped, in the output domain, into simple shifts along the θ

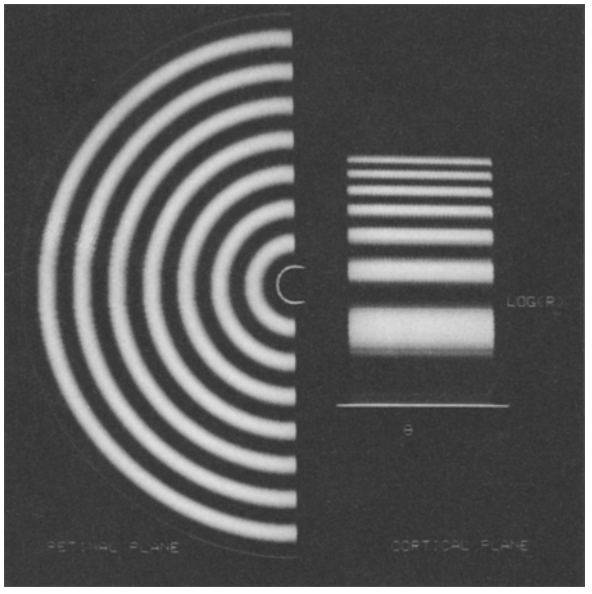

Fig. 3. The complex logarithm mapping, showing how a set of constant width annuli in the retinal plane (left) is transformed into a set of unequal width rectangles in the cortical plane (right)

coordinate. For this to be true, however, the allowable range of θ must be extended outside the 0–2π interval, or, alternatively, it must be represented along the circular section of a cylinder, r' being one of the cylinder generatrixes.

We may then conclude that the use of the space-variant convolution described by (3) and the mapping of the output into the (r', θ) plane is very convenient whenever a scale and rotation independent processing is desired (e.g. in template matching operations for pattern recognition, where a single reference template for any scaled and rotated input can thus be used). Within our work, the main interest of such a conclusion stems from the fact that in the human visual system the same type of processing and mapping does actually take place. In fact, the result of the peripheral processing [represented by (3)] is mapped into the cortex by means of the retino-cortical projection, which is modeled, as already mentioned, by the logarithmic conformal mapping. The cortical representation in the new (cartesian) coordinates (r', θ) is discussed in detail by Schwartz (1977, 1980a, b). Here we simply recall that the transformation from r to $r' = \log r$ is essentially due to the fact that, while the density of the RGC and LGN RFs in non uniform (namely proportional to $1/r$), the density of the corresponding cortical cells (that is the cells receiving a direct input from the LGN) is uniform.

This transformation [from (x, y) to (r', θ)] is such that the topological attributes (like connectivity) of the image are preserved, while other properties (like parallelism/perpendicularity relations) and the metric (area, perimeter) are not preserved. More specifically, the logarithmic conformal mapping from the retinal coordinates (x, y) to the cortical coordinates (r', θ) is synthetically expressed as $z = \log w$, with $w = x + jy$ and $z = r' e^{j\theta}$. Some effects of this mapping are pictorially shown in Figs. 3 and 4. In particular, Fig. 4a and b illustrate the dependence of the mapping on the fixation point (black and white circles in the checkerboard-like input image) for a given image.

Before leaving this subject, it is important to emphasize that the association of the processing modeled by (3) and the representation of its output in the (r', θ) coordinates must not be considered as the (temporal and spatial) sequence of the two events. In fact, what our model implies is simply that the processing performed by the LGN on the input image can be described by (3), and that the output of such a processing is represented in the cortex in the (r', θ) coordinates. In other words, we introduced a *formal* distinction (consistent with the experimental data and convenient from a theoretical viewpoint) between the two operations, processing and representation, that need not to be necessarily distinct in the visual system.

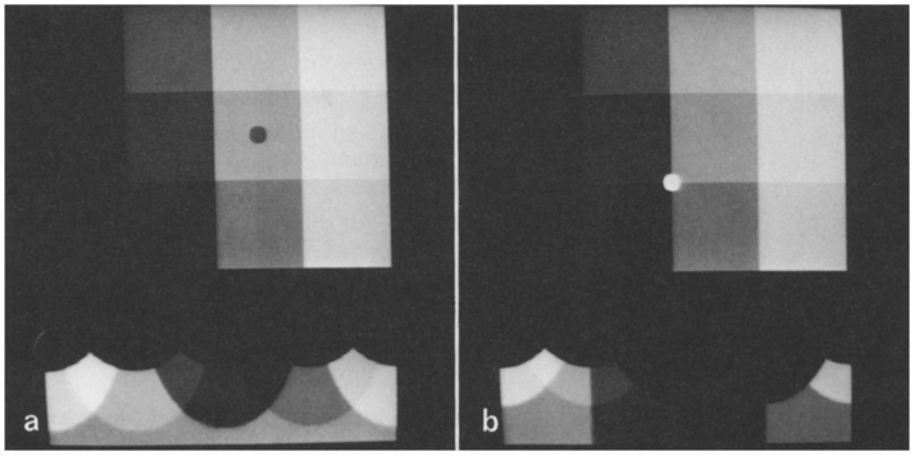

Fig. 4a and b. Another example of the complex logarithm mapping from the retinal (top) to the cortical (bottom) plane, with two different fixation points (corresponding to the small disks): **a** in the center, **b** shifted toward the bottom left. Notice that the shape of the squares in the cortical representation varies with the fixation point

2.2. *Band-Pass Filtering by Circularly Symmetrical Weighting Functions*

We make now a first step in specifying (3), by introducing a peculiar attribute of the RF weighting functions, namely their oscillating behaviour. It should be emphasized, once more, that we are interested mainly in the RF of the LGN cells, since introducing their shape in (3) is equivalent to describing the global processing performed at retinal and LGN levels.

It is known from signal theory that any given oscillating function can be written as the product of an envelope $w_e(\ )$ and the cosine function, the latter having the same period of the oscillating function. Therefore, we can write

$$w_0(\varrho/r) = w_e(\varrho/r)\cos(\omega_r \varrho) = w_e(\varrho/r)\cos(2\pi\varrho/R_r), \qquad (4)$$

where R_r is the oscillation period and $\omega_r/2\pi$ the corresponding frequency. The period R_r depends linearly on the eccentricity, corresponding to the fact that the number of oscillations within the extent of w_0 (or w_e) is constant, independent of r. Of course, the dependence of ω_r on the eccentricity is of the kind $\omega_r = \beta/r$.

Once the expression (4) is introduced in (3), it becomes apparent that, due to the oscillating nature of $w_0(\ )$, (3) represents a band-pass filtering operation. The domain where such a filtering is performed is defined by the spatial frequencies ω_r, which are not associated with a particular orientation in the input image, being the weighting functions, and therefore their Fourier transforms, not dependent on the orientation. In other words, for any pair (r, θ), the portion of the image around the point (r, θ) is processed by a band-pass filter defined over an annular region centered at ω_r in the $(\omega_\xi, \omega_\eta)$ plane. The filter center frequency is $\omega_r = \beta/r$ and its bandwidth corresponds to the frequency extent of the envelope $w_e(\varrho/r)$, whose bandwidth is, in turn, inversely proportional to r and of the same order of $\omega_r = \beta/r$. The result of this particular frequency analysis, that is the output $z(r, \theta)$ of (3), can be interpreted as the spatial filtering of the input around radial frequencies that are higher near the origin and decrease linearly with eccentricity. Since the associated bandwidth too varies according to the same law, a constant-Q filtering is implemented, where the selection of the β coefficient in $\omega_r = \beta/r$ affects, along with the bandwidth of $w_e(\varrho/r)$, the energy distribution of the filtered image $z(r, \theta)$ in the radial frequency domain.

2.3. DOG *Profiles and Overlap Factor*

The mathematical model described up to this point has taken into account the linear dependence of the RF size on eccentricity and the circularly symmetrical and oscillating shape of their profiles. In order to completely specify the parameters of the model, we now exploit other properties introduced in Sect. 1, concerning the RF profile and the overlap factor.

The chosen approximation of the RF profiles for LGN X-cells is the DOG function, expressed by (2), with a ratio $\sigma_i/\sigma_e = 3$ (where σ_i and σ_e are proportional, respectively, to the diameters of the surround and center regions). The parameter k appearing in (2) represents the linear dependence of the standard deviations on eccentricity. Therefore, (3) becomes

$$z_G(r, \theta) = \int_0^\infty \left[\frac{1}{\sqrt{2\pi} k r \sigma_e} \exp(-\varrho^2/2k^2r^2\sigma_e^2) - \frac{1}{\sqrt{2\pi} k r \sigma_i} \exp(-\varrho^2/2k^2r^2\sigma_i^2) \right] \varrho \, d\varrho \cdot \int_{-\pi}^{\pi} f(r\cos\theta - \varrho\cos\varphi, r\sin\theta - \varrho\sin\varphi) \, d\varphi . \qquad (5)$$

Coming now to the overlap factor, this parameter is strictly related to the sampling distance introduced at the end of Sect. 1.2. However, there is a certain arbitrariety in the definition of the sampling distance, which must be related to the band-pass characteristic of the RFs, or, alternatively, to their size. The optimal choice, on engineering considerations and assuming the structure as being locally homogeneous, would be such that the highest spatial frequency that is "passed" by the convolution process at a given eccentricity corresponds to at least $1/\sqrt{3}$ of the sampling frequency, according to the sampling theorem and to the exagonal sampling used in the processing examples (Sandini and Tagliasco, 1980). We arbitrarily define the smallest RF (to which corresponds the highest center frequency in the frequency domain) at any eccentricity as being the one with a CENTER SIZE covering the whole visual field with minimum overlapping (Fig. 2). Due to this assumption, the radius of the center mechanism of the smallest RF at each eccentricity, r_{min}, is expressed as a function of the sampling distance d as

$$r_{min} = d/\sqrt{3}.$$

We call the resulting structure the "high frequency filtering plane".

The overlap of different RFs at each eccentricity can be introduced in the model by superimposing other filtering planes to the above defined high frequency plane. With reference to (5), different filtering planes can be specified by choosing different values of the parameter k. The number of filtering planes can be determined taking into account that, at each eccentricity, the size of the largest RF is twice the smallest and assuming a discrete number of RF sizes uniformly distributed in this range. This assumption corresponds

to using everywhere in the (r, θ) plane the same sampling schema used for the high frequency filtering plane.

The overlap factor derived from physiological data is obtained by multiplying the density of RGCs by the mean area of the center mechanism at a given eccentricity. If we consider the sampling schema of Fig. 2, we can express the cell density D as a function of the sampling distance d:

$$D = 2N/d^2 \sqrt{3},$$

where N is the parameter to be determined, that is the number of filtering planes. The mean RF center size can also be expressed as a function of d:

$$r_{\text{mean}} = d\sqrt{3}/2.$$

Since the overlap at the LGN level can be assumed approximately equal to the overlap of RGC (due to the fact that only a little convergence of RGCs into LGN cells is usually found), we will consider, as a tentative value, the overlap factor determined for X-type RGCs in the cat's retina, that is a value of 15 to 22. This value refers to the center size (and not to the "equivalent center size") and includes both the ON-center and the OFF-center units. The resulting number of filtering planes is approximately between 6 and 8. It should be noticed that the number N of filtering planes derived in this way is dependent, besides the value of the overlap factor, on the arbitrary choice of the ratio between the sampling distance d and the smallest RF size; however, this arbitrary choice is not a peculiar characteristic of the model we propose, since changing it would simply cause a different number of filtering planes to be considered. Finally, the number N derived above (from 6 to 8) must be divided by two, if pairs of corresponding ON-center and OFF center filtering planes are grouped together.

We wish to point out that the resulting organization in terms of filtering planes does not have any known anatomical correspondence in the nervous structures we are modeling [although a subdivision into layered structures, processing different ranges of spatial frequencies, has been found in the visual cortex of cats (Maffei and Fiorentini, 1977)], but represents a way of organizing the processing performed by the early stages of the visual system.

To conclude the presentation of our model, we notice that it satisfies an important normalization constraint, i.e. it implicitly takes into account the fact that, regardless of RF size or eccentricity, the maximum cell output in terms of spikes/s is about the same. To achieve this result in terms of model output $z_G(r, \theta)$, a normalization factor $1/z_M$ ought to be used, being z_M the maximum output value

$$z_M = \int_0^{\infty} \int_0^{2\pi} w^2(\varrho/r)\varrho \, d\varrho \, d\varphi .$$

In the case of the DOG profiles used in (5), z_M turns out to be independent of r, due to the amplitude scale factor $1/r$ introduced to make the standard deviations of the gaussian curves be linearly increasing with eccentricity. Therefore, the outputs of (5) are self-normalized.

3. Processing Examples

In this section we show some examples of image processing that implement (5), and therefore simulate the processing of the LGN cells. We also present some examples of the $(\log r, \theta)$ representation of the output of (5), as opposed to the polar representation in (r, θ). As already pointed out, this transforms a linear scaling of the input $f(r, \theta)$ into a shift of the output along the axis $r' = \log r$.

All processings have been performed digitally. The input image $f(x, y)$ is acquired in cartesian coordinates, on a suitable sampling grid, with eight bit quantization.

To implement (5), the structure schematized in Fig. 2 has been used, consisting of a set of 2-D DOG profiles (see Fig. 1), whose geometric centers form a triangular lattice in the (r, θ) plane. The sampling distance, as well as the RF size, increases linearly with eccentricity, so that the percentage overlap of contiguous DOG functions is constant. The smallest eccentricity where the processing is performed is chosen on the base of both physiological and numerical considerations. In fact, the linear decrease of RF size toward the center of the visual field is limited to the region outside the fovea and the numerical implementation of RFs with radius less than 4 elements is not reasonable. For the same physical and numerical reasons, the logarithmic conformal mapping cannot be

Fig. 5. Original image, digitized on a 512×512 cartesian grid, with 8 bit quantization

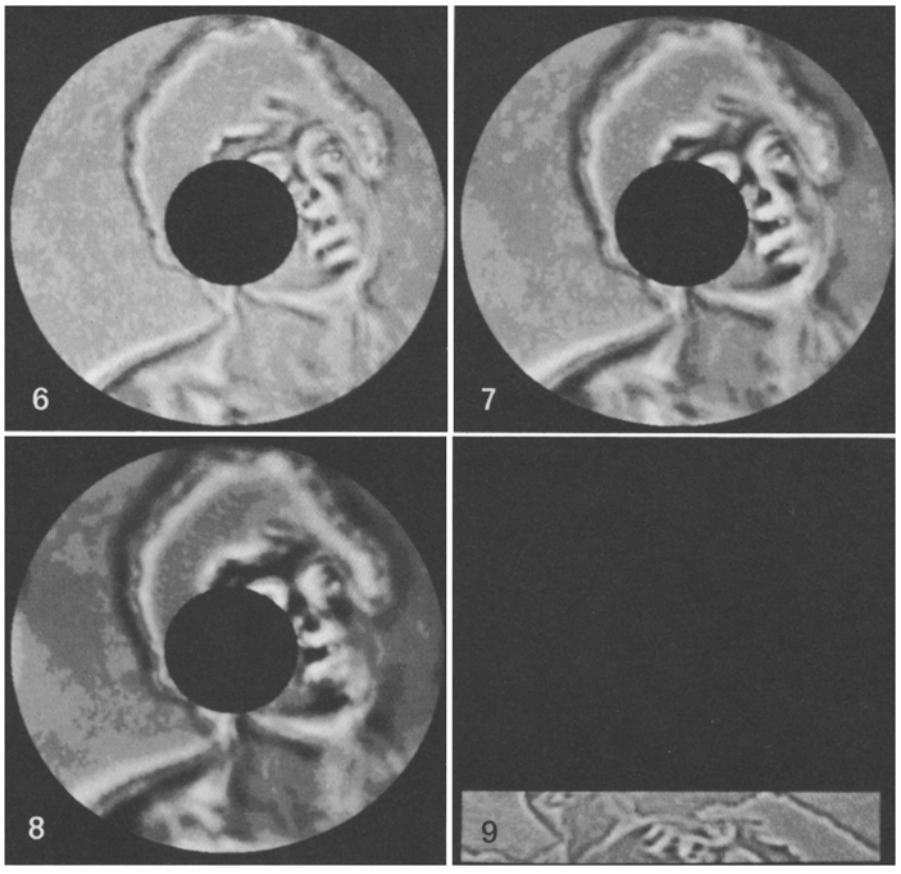

Fig. 6. Result of the processing represented by Eq. (5) applied to the original image of Fig. 5. The value of the parameter k appearing in (5) is chosen here to be 1.38×10^{-2} and corresponds to the highest frequency processing plane discussed in the text. Notice that the small details are retained in the processed image. The central black circle represents the fovea, where no processing is performed

Fig. 7. Same as Fig. 7, with $k = 2 \times 10^{-2}$, corresponding to the medium frequency plane

Fig. 8. Same as Fig. 6, with $k = 2.76 \times 10^{-2}$, corresponding to a low frequency filtering plane: the processed image has lost many details, while retaining the information on the bright and dark masses distribution

Fig. 9. Cortical representation of Fig. 6, being θ the horizontal axis and $\log r$ the vertical axis

implemented near the origin, which represents a singular point of the transformation.

Figure 5 represents the original (digitized) image that we have chosen in order to perform the processing examples that we are going to show.

The Figs. 6–8 represent the processing, according to (5), of the original image of Fig. 5, for k equal respectively to 1.38×10^{-2}, 2.0×10^{-2}, 2.76×10^{-2}. These three values of k correspond to three different filtering planes, each one representing the superposition of an ON-center and an OFF-center filtering plane. In these figures, in fact, the intermediate gray level corresponds to zero-output of the convolution, while brighter and darker gray levels correspond, respectively, to positive and negative outputs (approximating the output of ON-center and OFF-center units). The central region of Figs. 6–8 is filled by the black circle corresponding to the fovea, i.e. to that part of image where, as stated before, our processing stops.

From a qualitative point of view, we can say that the outputs of Figs. 6–8 correspond to a band-pass filtering around progressively lower (radial) frequencies. Hence Fig. 6 corresponds to the filtering at the higher frequencies (in the average, since the frequencies filtered around the fovea are always higher than those filtered at the periphery). The processing of Fig. 6 implements, in fact, a sort of "contour extraction" on the original image of Fig. 5. On the contrary, the output of Fig. 8, which corresponds to the filtering at the lower frequencies, shows quite blurred contours and rather gives an indication of the distribution of the bright and dark masses in the original image of Fig. 5. In the case of Fig. 7, the effects of the processing are obviously in between those of Figs. 6 and 8.

The images of Figs. 6–8 do not yet take into account the retino-cortical mapping. In fact, in order to evidence the space-variant characteristic of the processing, the value of the output of the convolution performed by each RF is extended over areas increasing with eccentricity according to the structure of Fig. 2, as a sort of back projection to the retinal plane.

Figure 9 shows the "cortical representation" of the output of Fig. 6, with the abscissa corresponding to θ and the ordinate corresponding to $\log r$. As we know, this type of representation allows to convert scalings and rotations of the original image into simple shifts, along respectively the $\log r$ and the θ axis, of the processed image.

We shall give now an example of this property through the next three figures. Figure 10 represents a

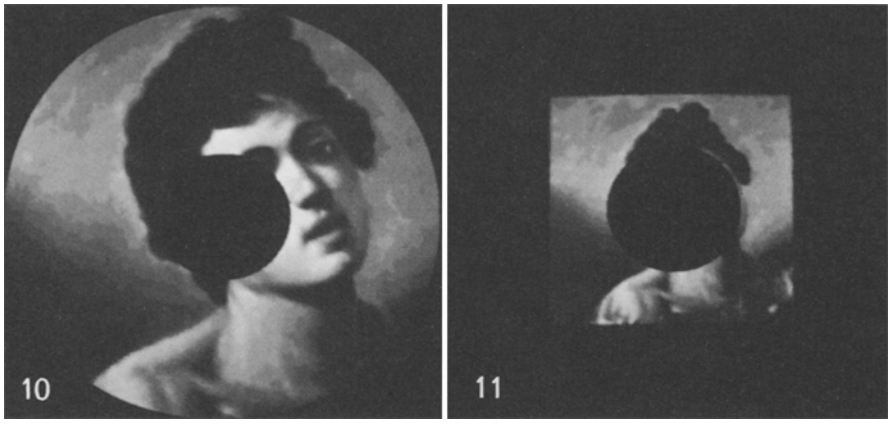

Fig. 10. Smoothed version of the original image of Fig. 5, obtained by a space-variant averaging over circular areas distributed according to the schema of Fig. 2

Fig. 11. Same smoothing as in Fig. 10, performed on a 2 : 1 scaled version of the original image (see text)

Fig. 12. Cortical representation of the two results of Fig. 10 (bottom) and Fig. 11 (top), showing that they are a shifted version of each other along the log$r$ (vertical) axis (see text)

smoothed version of the original image of Fig. 5, obtained through simple local averages within circular areas whose sizes vary according to the schema of Fig. 2. From the mathematical point of view, the processing of Fig. 10 corresponds to (3), that is to a $z(r, \theta)$ characterized by weighting functions that are constant over a circular area and zero elsewhere.

Figure 11 represents the output of the same type of processing applied to a scaled version (2 : 1) of the original image (Fig. 5). Notice that the size of the black hole in the center is the same as before, implying that the occluded part of the image is now relatively bigger. By comparing Figs. 10 and 11, we can recognize that the two outputs, apart from the already mentioned fovea effect, are different from each other simply because of a scale factor (the same factor 2 affecting the original inputs). A minor and not meaningful difference is actually represented by the two margins, since the one of Fig. 10 is circular while that of Fig. 11 is rectangular, due to the fact that the scaling brings the whole original within the circular processing area.

The cortical representations of the two outputs of Figs. 10 and 11 are shown in Fig. 12, respectively at bottom and top positions. The comparison of the two representations is immediate and shows that they are a (shifted) replica of each other (apart from the fovea effect and the margin effect mentioned above). This pictorial result is just what one would have expected on the base of the form-invariance property.

4. Comments and Conclusions

In the previous sections of this work, we have presented and illustrated by means of examples a mathematical model of the processing performed by the peripheral visual system, based on the properties discussed in Sect. 1. In particular, the spatial distribution of the RFs and of their size has been taken into account, as well as the retino-cortical projection. By following a signal theory approach, we have shown that the processing performed by the peripheral visual cells is form-invariant under linear scalings and rotations of the input, since the latters simply map into shifts of the cortical representation. Such features have been derived in general, i.e. with no assumption on the specific shape of the RF weighting function, like the DOG response used in the processing examples.

In this last section we further discuss the previous results and stress analogies and differences with respect to some related works.

Let us start with two works that exploit, in a way similar to our approach, the idea of decomposing the original image through a set of processing planes, each providing a simplified, or less redundant, description of the image, oriented to specific processing and analysis operations (e.g. contour extraction or texture segmentation). Such operations are implemented by parallel processing algorithms and structures, similarly, again, to what takes place in the visual system.

We first mention Marr (1976), Marr et al. (1979) and Marr and Hildreth (1980) who introduce the

concept of "raw primal sketch" (i.e. a "primitive but rich description of the image"), that is the output of an image processing system consisting of a set of processing planes, or channels. The output of each plane is the zero-crossing pattern of the convolution between the input image and the weighting function $\nabla^2 G$, which is the laplacian of $G(r)=(\pi\sigma^2/2)\exp(-r^2/2\sigma^2)$. The operation corresponding to such a response is basically a band-pass frequency analysis in the radial frequencies ω_r. The value of the variance of the gaussian profiles (that define the different processing planes and the filter center frequencies) are chosen in such a way that the resulting frequency analysis is constant-Q. The outputs of the different channels are essentially the contours of the original image, extracted at different degrees of spatial (and therefore frequential) resolution. These outputs are then compared to each other, to build up the "true contours" of the image and to perform further processing functions.

Second, we quote the work (though not specifically oriented to the modeling of the visual system) by Granlund (1978), where a system is presented, based on a single picture operator iteratively applied through a hierarchy of processing levels, differing from each other essentially for the degree of spatial and frequency resolution of the analysis. The operator, based on the Fourier transform, performs a short-space (i.e. local) frequency analysis, which is also constant-Q, in the radial frequencies, that we have seen to correspond to the processing of the early stages of the visual system.

We remark once again the connection of our work with the property of a size invariant cortical representation of the retinal inputs, introduced by Schwartz (1977, 1980a, b) through his model of a complex logarithm mapping. However, this model is not completely satisfactory, since it does not include any processing of the retinal input associated with the mapping. On the other hand, the cortical representation of a retinal input is obtained by the combination of a topological transformation *and* a local processing performed by the peripheral cells. Clearly, in order for the output of the overall process to be invariant (except for a shift) with respect to linear scalings of the retinal input, the local processing mentioned above must be form-invariant in the sense specified in Sect. 2.1. Our work shows that this particular processing, along with the logarithmic conformal mapping, adequately models the early stages of the human visual system, and in this sense it represents a useful clarification and a further support to the hypothesis of Schwartz.

We further mention a work dealing with the modeling of the peripheral visual system (Crettez, 1980), where a model is proposed lacking in consistency under several respects, when critically evaluated in the light of the experimental data. For example, this model allows an increasing range od receptive fields as eccentricity decreases, up to the point that all possible sizes of RF would coexist at minimal eccentricity.

We wish to end this discussion by mentioning some works where image acquisition and processing techniques are proposed, oriented to automatic recognition tasks and based, in a way or another, on the vision mechanisms (and therefore exhibiting properties similar to the ones discussed in this paper).

Weiman and Chaikin (1979) have proposed an acquisition scheme exploiting a digitization grid based on logarithmic spirals rather than cartesian coordinates. Such a scheme implies a mapping from the original picture space to a computation space where a variety of computer graphics and image processing operations can be performed in a simpler and more efficient way. One of the computational advantages stems from the property that both magnifications and rotations in the "picture space" map into simple shifts in the "computation space".

Processing techniques yielding shift, rotation and magnification invariant results are proposed by Casasent and Psaltis (1976) for recognition purposes and by Altes (1978) to model the peripheral auditory system in mammals. These methods are based on the combined use of Fourier transforms and a logarithmic mapping or of Fourier and Mellin transforms. A consequence of this kind of processing, however, is to destroy some geometric information on the patterns, like their size, that might often be useful to retain, suitably translated. e.g. into shifts.

Acknowledgements. We wish to thank Mr. Buffoni for the processing of the photographic material and Miss Bice Del Papa for typing the manuscript. This work has been partially supported by the "Consiglio Nazionale delle Ricerche". A fellowship from the "Scuola Normale Superiore – Pisa" is gratefully acknowledged by Dr. Sandini.

References

Altes, R.A.: The Fourier-Mellin transform and mammalian hearing. J. Acoust. Soc. Am. **63**, 174–183 (1978)

Boycott, B.B., Wässle, M.: The morphological types of ganglion cells of the domestic cat's retina. J. Physiol. **240**, 397–419 (1974)

Braccini, C., Gambardella, G.: Linear shift-variant filtering for form-invariant processing of linearly scaled signals. Signal Processing, April 1982

Braccini, C., Gambardella, G., Sandini, G.: A signal theory approach to the space and frequency variant filtering performed by the human visual system. Signal Processing **3**, 231–240 (1981)

Bullier, J., Norton, T.T.: Comparison of receptive field properties of X and Y ganglion cells with X and Y lateral geniculate cells in the cat. J. Neurophysiol. **42**, 274–291 (1979)

Casasent, D., Psaltis, D.: Position, rotation and scale invariant optical correlation. Appl. Opt. **15**, 1795–1799 (1976)

Cleland, B.G., Dubin, M.W., Levick, W.R.: Sustained and transient neurons in the cat's retina and lateral geniculate nucleus. J. Physiol. **217**, 473–496 (1971)

Cleland, B.G., Levick, W.R., Wässle, H.: Physiological identification of a morphological class of cat retinal ganglion cells. J. Physiol. **248**, 151–171 (1975)

Crettez, J.P.: A visual model based on the overlapping of the cell receptive fields in the visual pathway. In: Proc. of the First Scandinavian Conf. on Image Analysis, Linkoeping, Sweden 1980

Daniel, P.M., Whitteridge, D.: The representation of the visual field on the cerebral cortex in monkeys. J. Physiol. **159**, 203–221 (1961)

Drasdo, N.: The neural representation of visual space. Nature **266**, 554–556 (1977)

Enroth-Cugell, C., Robson, J.G.: The contrast sensitivity of retinal ganglion cells of the cat. J. Physiol. **187**, 517–552 (1966)

Fischer, B.: Overlap of receptive field centers and representation of the visual field in the cat's optic tract. Vision Res. **13**, 2113–2120 (1973)

Granit, R.: Sensory mechanisms of the retina. London: Oxford University Press 1947

Granlund, G.H.: In search of a general picture processing operator. Comput. Graph. Im. Proc. **8**, 155–173 (1978)

Hochstein, S., Shapley, R.M.: Quantitative analysis of retinal ganglion cell classifications. J. Physiol. **262**, 237–264 (1976a)

Hochstein, S., Shapley, R.M.: Linear and non-linear spatial subunits in *Y* cat retinal ganglion cells. J. Physiol. **262**, 265–284 (1976b)

Hubel, D.M., Wiesel, T.N.: Receptive field of optic nerve fibers in the spider monkey. J. Physiol. **154**, 572–580 (1960)

Hubel, D.M., Wiesel, T.N.: Integrative action in the cat's lateral geniculate body. J. Physiol. **155**, 385–398 (1961)

Hubel, D.M., Wiesel, T.N.: Uniformity of monkey striate cortex: a parallel relationship between field size, scatter and magnification factor. J. Comp. Neurol. **158**, 295–302 (1974)

Ikeda, M., Wright, M.J.: Differential effects of refractive errors and receptive field organization of central and peripheral ganglion cells. Vision Res. **12**, 1465–1476 (1972)

Kuffler, S.W.: Discharge patterns and functional organization of mammalian retina. J. Neurophysiol. **16**, 37–63 (1953)

Maffei, L., Fiorentini, A.: The visual cortex as a spatial frequency analyzer. Vision Res. **13**, 1255–1267 (1973)

Maffei, L., Fiorentini, A.: Spatial frequency rows in the striate visual cortex. Vision Res. **17**, 257–264 (1977)

Marr, D.: Early processing of visual information. Philos. Trans. R. Soc. London B **275**, 483–524 (1976)

Marr, D., Poggio, T., Ulman, S.: Bandpass channels, zero-crossings, and early visual information processing. J. Opt. Soc. Am. **69**, 914–916 (1979)

Marr, D., Hildreth, E.: Theory of edge detection. Proc. R. Soc. Lond. B **207**, 187–217 (1980)

Osterberg, G.: Topography of the layer of rods and cones in the human retina. Acta Ophthal. Suppl. **6**, 1–103 (1935)

Peichl, L., Wässle, M.: Size, scatter and coverage of ganglion cell receptive field centers in the cat retina. J. Physiol. **291**, 117–141 (1979)

Pirenne, M.H.: Vision and the eye. London: Chapman and Hall 1967

Rodieck, R.W., Stone, J.: Analysis of receptive fields of cats retinal ganglion cells. J. Neurophysiol. **28**, 833–849 (1965)

Rose, D.: Mechanisms underlying the receptive field properties of neurons in the cat visual cortex. Vision Res. **19**, 533–544 (1979)

Rovamo, J., Virsu, V., Näsänen, R.: Cortical magnification factor predicts the photonic contrast sensitivity of peripheral vision. Nature **271**, 54–56 (1978)

Sandini, G., Tagliasco, V.: An anthropomorphic retina-like structure for scene analysis. Comput. Graph. Im. Proc. **14**, 365–372 (1980)

Schwartz, E.L.: Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. Biol. Cybern. **25**, 181–194 (1977)

Schwartz, E.L.: A quantitative model of the functional architecture of human striate cortex with application to visual illusion and cortical texture analysis. Biol. Cybern. **37**, 63–76 (1980a)

Schwartz, E.L.: Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. Vision Res. **20**, 645–669 (1980b)

Stone, J.: A quantitative analysis of the distribution of ganglion cells in the cat's retina. J. Comp. Neurol. **124**, 337–352 (1965)

Stone, J., Hoffmann, K.P.: Very slow-conduction ganglion cells in the cat's retina: a major new functional type? Brain Res. **43**, 610–616 (1972)

Weiman, C.F.R., Chaikin, G.: Logarithmic spiral grids for image processing and display. Comput. Graph. Im. Proc. **11**, 197–226 (1979)

Received: November 18, 1981

Prof. Carlo Braccini
Istituto di Elettrotecnica
dell'Università di Genova
Viale Francesco Causa 13
I-16145 Genova
Italy