Anatomical and Physiological Correlates of Visual Computation from Striate to Infero-Temporal Cortex

ERIC L. SCHWARTZ, MEMBER, IEEE

Abstract—A review of the calculus of two-dimensional mappings is provided, and it is shown that cortical mappings are in general locally equivalent to the sum of a conformal mapping and a shear mapping. Currently available data suggests that a conformal (isotropic) mapping is sufficient to model the global topography of primate striate cortex, although the addition of a shear component to this mapping is outlined. A similar analysis is applied to the local map structure of striate cortex. Models based on a local reiteration of the complex logarithm and on Radon (and backprojection) mappings are presented. A variety of computational functions associated with the novel architectures for image processing suggested by striate cortex neuroanatomy are discussed. Applications to segmentation, perceptual invariances, shape analysis, and visual data compression are included. Finally, recent experimental results on shape analysis by neurons of infero-temporal cortex are presented.

I. INTRODUCTION

RECENT ADVANCES in understanding the anatomy and physiology of visual cortex have focused attention on the problem of visual data formatting in the central nervous system (CNS). It appears that the primate visual system may utilize a variety of novel architectures to represent visual data in the CNS. This possibility is reinforced by the use of the term “functional architecture” that Hubel and Wiesel [1] have used to describe the spatial patterns of columnar architecture in striate cortex. Here the term functional architecture will be expanded to include the global topographic structure of striate cortex, as well as the local columnar architecture represented by the orientation and ocular dominance column systems described by Hubel and Wiesel [1].

It will be shown that a conformal mapping provides a good model for global cortical topographic mapping, as well as a possible model for the local hypercolumn pattern of striate cortex. Furthermore, this mapping suggests that several integral transforms (Radon transform, backprojection) may provide workable models of the local architecture of striate cortex that are in agreement with current phenomenological descriptions.

Finally, a variety of computational applications of these anatomical models will be reviewed. These include applications to perceptual invariance, binocular segmentation, data compression, and shape analysis. Recent experimental work (in monkey infero-temporal cortex), which is related to this analysis, will be reviewed. It will be suggested that the novel architectures of image representation suggested by primate visual cortex neuroanatomy may provide an example of “computational anatomy,” which has relevance both to contemporary machine vision applications, as well as to current attempts at understanding the biological basis of visual computation.

II. Mappings of Two-Dimensional Surfaces into Two-Dimensional Surfaces

The concept of “topographic” mapping of the surface of the retina to the surface of striate cortex naturally invites an attempt at mathematical modeling. Which mapping fits the available data? In the following a complete characterization of planar mapping will be provided in an attempt to answer this question.

A. Whitney Mapping Theorem

A broad classification of two-dimensional mappings is provided by the Whitney mapping theorem, which states that topologically stable planar mappings are locally equivalent to either a “fold,” a “cusp,” or a “regular” mapping [2]. In the following, we will ignore the singularities represented by folds and cusps. Regular maps are characterized by having a Jacobian matrix of partial derivatives whose determinant is finite and nonzero. In the following section we will regroup the terms of this Jacobian in order to better understand the biological significance of the isotropic (ie, conformal) and nonisotropic maps that might occur in the nervous system.

B. Jacobian Matrix and Magnification Tensor

The (retinal) point \( (x, y) \) is taken to a cortical point \( (f, g) \) by the functions \( f(x, y) \) and \( g(x, y) \). The Jacobian of this map [2] is

\[
J = \begin{bmatrix}
\frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\
\frac{\partial g}{\partial x} & \frac{\partial g}{\partial y}
\end{bmatrix} = \begin{bmatrix} f_x & f_y \end{bmatrix}.
\]

A common manipulation from the literature of continuum mechanics (and fluid mechanics) will be useful: the Jacobian is rewritten in terms of its symmetrical and

Manuscript received August 8, 1982; revised June 1983. This work was supported in part by the AFOSR Image Understanding Program no. F49620-83-C-0108 and in part by the Systems Development Foundation.

The author is with Brain Research Laboratories, New York University Medical Center, 550 First Avenue, New York, NY 10016.
antisymmetrical components:

\[ 2J = (J + J^+) + (J - J^+) \]
\[ = \begin{pmatrix}
2f_x & f_y + g_x \\
(f_y + g_x & 2g_y)
\end{pmatrix} + \begin{pmatrix}
0 & (f_y - g_x) \\
(g_x - f_y) & 0
\end{pmatrix} \]

(2)

where the symbol *t* is used to represent the transpose of the Jacobian. The symmetrical part of the Jacobian in (2) may be rewritten in terms of its traceless and diagonal components (3):

\[ \begin{pmatrix}
2f_x & f_y + g_x \\
(f_y + g_x & 2g_y)
\end{pmatrix} = \begin{pmatrix}
2f_x - \text{tr} & f_y + g_x \\
(f_y + g_x & 2g_y - \text{tr})
\end{pmatrix} + \begin{pmatrix}
\text{tr} & 0 \\
0 & \text{tr}
\end{pmatrix} \]

(3)

where \(\text{tr}\) represents the trace of the Jacobian (the sum of its diagonal terms), such that \(\text{tr} (\text{Jacobian}) = f_x + g_y\).

The previous analysis thus shows that the most general form of the Jacobian of a regular map is the sum of a traceless symmetric matrix and two other terms, which are a diagonal matrix and an antisymmetric matrix. This is shown in the following equation.

\[ 2J = \begin{pmatrix}
\text{tr} & 0 \\
0 & \text{tr}
\end{pmatrix} + \begin{pmatrix}
0 & f_y - g_x \\
(g_x - f_y) & 0
\end{pmatrix} \]
\[ + \begin{pmatrix}
2f_x - \text{tr} & f_y + g_x \\
(g_y + f_y & 2g_y - \text{tr})
\end{pmatrix} \]

(4)

The biological significance of these terms will now be outlined by showing that the first two terms represent a conformal (i.e., isotropic) mapping, and that the last term (the traceless symmetric matrix) represents a "shear," i.e., a "squeeze" in one direction and an "expansion" in the other direction with no change in area. First the relation of this analysis to the theory of conformal mapping will be shown.

A conformal mapping is defined in a number of equivalent ways. If the mapping preserves the "shape" of infinitesimal triangles (or, equivalently, of local angles) it is conformal. Thus "isotropic" mappings are conformal. Any analytic function of a complex variable is conformal. A map is conformal if and only if it satisfies the Cauchy–Riemann equations:

\[ f_x = g_y \]
\[ f_y = -g_x. \]

(5)

The relation of the subclass of conformal maps to the class of regular maps can be made explicit: substitute the Cauchy–Riemann equation (5) into the Jacobian of (4). The last (i.e., traceless) term is identically zero. The conventional physiological magnification factor is the square root of the determinant of the Jacobian matrix, which represents the change, or "magnification" in a unit length caused by the mapping. If the map is conformal, this magnification factor is equal to \(|\nabla f|\) or to \(|\nabla g|\). Thus, conformal maps only require the specification of two num-

bers at each point: a magnitude (e.g., \(f_x\) and \(g_y\)), which determines the "magnification factor" and a direction, which is the angle specified by the rotation matrix (i.e., the antisymmetric matrix) above.

This outline represents a complete analysis of the structure of mappings from two surfaces to two surfaces. One final theorem will be mentioned that relates to the subclass of conformal mappings, which, together with "shear" mappings, represents the entire set of possible topographic map structures.

C. Riemann Mapping Theorem

The Riemann mapping theorem states that a conformal map is uniquely specified (for simply connected domains) if it is given on any finite subset (4). This has the following important experimental consequence. If the map of striate cortex were measured, say, along the horizontal meridian only, and it was determined that the map was isotropic, then the map is uniquely determined throughout the entire cortical extent (i.e., by analytic continuation (4)). This result has the somewhat surprising experimental consequence that it is not necessary to measure an isotropic map everywhere in order to determine the map function. As a matter of fact, it is not necessary to measure it by standard physiological point mapping at all: if the map is isotropic (e.g., determined by sampling, with no quantitative data), then the "shape" of the map alone (i.e., the boundary conditions imposed by the retina and cortex) uniquely determine the map, up to an orientation and a single point in both planes, as guaranteed by the Riemann mapping theorem. This fact has been made the basis of a developmental algorithm for neural maps; the connection of this algorithm to classical field theory (in general) and fluid mechanics (in particular) is further discussed in this work (5). More significantly, it indicates that a cortical map might be determined uniquely and directly from the anatomy alone, with no physiological measurement, other than the demonstration of isotropy and overall map orientation.

The results outlined in the above mathematical outline will now be applied to characterizing the global and local topographic map functions of striate cortex.

III. APPLICATIONS OF TWO DIMENSIONAL MAPPING THEORY TO NEURAL TOPOGRAPHY

A. Global Topography

1) Justification of the Present Approach: The previous analysis presented a complete categorization of planar maps, using the Whitney and Riemann map theorems, and standard manipulations of the Jacobian matrix of the mapping. This is a considerable amount of mathematical analysis to apply to an area such as visual neuroanatomy, which until quite recently appears to have received little or no mathematical attention of any kind. Is this approach justified by practical considerations, or is it merely an exercise in mathematical description? In the following dis-
cussion, evidence will be presented to support the view that a correct mathematical approach to this subject is necessary, that several substantive issues may be clarified by it, and that some confusions that have entered into this literature may be eliminated. Having demonstrated the utility of a mathematical understanding of neural maps, from a purely structural and descriptive point of view, the remainder of this paper will then discuss the computational issues that are suggested by these descriptive results.

2) Fitting Cortical Magnification Factor: Daniel and Whitteridge [6] defined "magnification factor" as the ratio of the distance (in millimeters) of movement of the locus of cortical excitation to the distance (in degrees) of movement of the corresponding visual stimuli. Since then, most studies of topographic mapping (visual, somatic) in most species have relied on measurements of magnification factor as the principle quantitative estimator of map structure.

This is inadequate, because magnification factor is a scalar (i.e., a number), but the differential quantity needed to specify a general planar map is a tensor (the Jacobian matrix). Thus, at the very least, an analysis of the structure of two-dimensional maps indicates that the quantity that has been traditionally measured (magnification factor) is not an adequate quantitative measure. This issue will emerge with some importance, below, in regard to the issue of whether or not neural maps are "isotropic."

The lack of precision in analysis, outlined above, is paralleled by a lack of precision in the data. Thus, although the topographic structure of primate striate cortex is arguably the most carefully and extensively measured aspect of visual neuroanatomy (having been studied in dozens of experiments spanning the past 40 years), there is a large amount of variance in the results. For example, most investigators have fit magnification data to a function of the form $B^1/(r + a)$, where $B$ is an overall scaling constant.

Visual eccentricity in degrees is measured by $r$ and $a$ is a constant that essentially determines the "shape" of the map. The values of the constant $a$ range from 1.7° [9] to 1° (7) to 0.3° [10]. The constant $a$ provides a one parameter fit to the entire two-dimensional structure of striate cortex topography (via the complex mapping log($z + a$)), and this critical parameter varies by over 50 percent in different experiments.

How can this variance be reconciled? One possible explanation might be differences in overall size or "shape" across individuals. However, the constant $a$ above is normalized implicitly for size. Large variations in shape or map structure do not seem likely because of Fig. 1, which shows three completely independent sets of data that are all precisely consistent with the same constant $a = 0.3$.

The most likely source of this range of variance is based on the observation of a correlation between the area of cortex studied and the magnitude of the constant $a$. Thus, Hubel and Freeman [9] found a large value for $a = 1.7$, but had only a single data point in the central 4° of field. Their data was almost entirely parafoveal. Tootel et al. [45] found a fit of $a = 1°$, based on an effective group of points at 1° (i.e., a 2D mapping of a series of rings that included one at 1° of eccentricity). The inclusion of more data towards the center of the field produced a smaller constant $a$, and hence a larger effective foveal representation. Dow et al. [10] obtained the only available data to date from the central fovea, extending to 1/6°. Their fit gave the smallest value for $a = 0.3°$, and hence the largest effective cortical foveal representation. Thus, it appears that the nonlinearity of the cortical map function may be responsible for introducing a correlation in the observed structure of the cortical map, and the extent of cortex from which data was collected.

Since Fig. 1 shows a high degree of consistency between the anatomical data of Levy et al. [12], the physiological data of Dow et al. [10], and the direct 2D mapping of Tootel et al. [45], it seems that some progress may have been made in reconciling these issues. But this consistency is only apparent via the use of the two-dimensional map function log($z + 0.3$).

3) The Relationship Between Visual Acuity, Cortical Magnification Factor, and Retinal Ganglion Cell Density: Human visual acuity decreases roughly as the inverse of eccentricity. Retinal ganglion cell density in monkeys and humans also decreases in about the same way, although there is some question as to whether this functional form is appropriate for the cell density/unit area, or the cell density/linear unit (mm) (see Romano and Virtsu [46] for discussion). These observations suggest a link between functional aspects of spatial vision (i.e., acuity) and anatomical factors and are of some importance for the understanding of human spatial vision.

Romano and Virtsu [46] have presented arguments suggesting that detailed agreement exists between retinal ganglion cell density and cortical magnification factor, and Virtsu and Romano [25] have presented data that suggests that human spatial frequency sensitivity has the same scaling behavior as retinal and cortical functions. However, the cortical data used in these analyses is based on earlier experimental work that is not consistent with the data summarized in Fig. 1. Specifically, Romano and Virtsu [46] predict that retinal ganglion cell density and cortical magnification factor are greater: along the horizontal meridian than at the same eccentricity along the vertical meridian. But it may be calculated directly from the map function log($z + 0.3$) that the reverse is true. Magnification is greater along the vertical meridian and the horizontal meridian. This result was also stated explicitly from the observations of Tootel et al. [45]. A plot of the ratio of vertical to horizontal meridian magnifications, based on the map function log($z + 0.3$) and on the fits of Romano and Virtsu [46], is shown in Fig. 2. The disagreement is maximal at the value of the constant $a$ in the fit function log($z + a$), (i.e., at 0.3°), where the discrepancy is roughly 50 percent. This is a significant disagreement. Thus, it appears from current data and analysis that the details of
for a stimulus consisting of logarithmically spaced rings and equiangular rays, as used recently by Tootle et al. [45], together with the corresponding two-deoxyglucose mapping obtained from macaque striate cortex. The operculum of the cortex is shown (as on top, in the data of Levay et al. [12]).

The 2DG map is shown in the center, bottom, and on the right (bottom) is the superposition of this data and the theoretical map function. This map function, which is consistent with the independent experimental data shown in top and middle parts of figure, clearly provided an accurate prediction for actual topographic structure of striate cortex topography, as revealed by 2DG mapping technique of Tootle et al. [45].

Fig. 1. On top left is shown a representation of the visual field (or the surface of the retina). The central 9° is shown in orthogonal projection. Polar coordinates (r, φ) are used to represent the field. On the right is shown the map of this field representation using the map function log(z + 0.3). This function is derived using the recent fit of Dow et al. [10] to cortical magnification factor, which provided a best fit of the form log(z + 0.3). The two dimensional complex log map is immediately derived from this fit as log(z + 0.3). Also shown is a superposition of the anatomical data of Levay et al. [12] showing the central 9° of striate cortex (macaque) visual representation. Clearly, the boundaries of the theoretical map provide a close fit to the boundaries of the anatomical data of Levay et al. [12]. Thus, an isotropic (i.e., conformal) map of the form log(z + 0.3) provides a good fit to the topography, anatomy, and magnification factor of macaque striate cortex. In the center is shown a reprint of Dow et al.'s [10] fit of their own and previous striate cortex magnification data (solid line). Superimposed over this plot of data is the theoretical magnification factor (large stars) calculated from the map function log(z + 0.3). Values at corresponding eccentricities along horizontal and vertical meridians are averaged, in both the theoretical and experimental data. The agreement is good between 10 min and 20° of eccentricity. The far periphery, beyond 20°, shows some disagreement. However, a function of the form log(z + 0.3)/(z + 60) (i.e., the complex logarithm of a bilinear function of visual field coordinates) is capable of fitting the entire field. On the bottom is shown a superposition of the map function log(z + 0.3)

Fig. 2. The ratio of cortical magnification at same position, on the vertical and horizontal meridia (VM/HM) is plotted on the vertical axis, against eccentricity, in degs, on the horizontal axis. The dashed lines are from Rovamo and Virsu [46] estimate of striate cortex magnification factor, based on retinal ganglion cell data. Solid line, on top, is calculated directly from the map function log(z + 0.3) (i.e., magnification is the magnitude of 1/z + 0.3). The retinal data predicts that vertical meridian magnifications are always smaller than horizontal meridian, while the map function log(z + 0.3) predicts the opposite. The effects are maximal at 0.3°, and are in the range of 50 percent at this point.
cortical magnification factor and the fit of retinal ganglion cell density of Rovamo and Virsu [46] are not precisely the same. Moreover, since Virsu and Rovamo [25] find a detailed agreement of human spatial frequency sensitivity and their fit of retinal ganglion cell density, it would appear that the “grain” of human spatial vision is determined at the retinal, rather than the cortical level.

These observations determine nontrivial statements concerning the relation of visual function and anatomical function, which would be difficult to state in the absence of a detailed understanding of the functional form of the cortical map.

4) Is the Striate Cortex Map Isotropic? Recently, Sakitt [15] has stated that “the cortical magnification factor in rhesus cannot be isotropic,” based on an analysis that claims to show that it is impossible to find an isotropic map function that fits the anatomical data (Levay et al. [10]). This claim can be simply contradicted by presenting a counter example: Fig. 1 shows an isotropic map (log (z + 0.3)) that fits the same data used by Sakitt, as well as several other independent data sets. Thus, it is, at least, possible that rhesus cortical magnification factor is isotropic.

Fig. 1 also suggests that indeed this map is approximately isotropic since there is good agreement with the details of the map function and the isotropic function log (z + 0.3).

Once again, knowledge of the detailed functional form of cortical map functions may be used to clarify basic descriptive issues.

B. Local Functional Architecture of Striate Cortex

1) Local Architecture of Striate Cortex Modeled by a Local Complex Logarithmic Mapping: The topographic mapping concept may break down at the level of scale of about 1–2 mm in striate cortex. On this scale there is a substructure evident that is no longer describable in terms of a simple point mapping. Hubel and Wiesel [1] have termed this substructure a “hypercolumn.” One salient feature of a hypercolumn is that there appears to be a mapping of the orientation of line segments in the visual field to a linear coordinate in the cortex. Specifically cortical cells that are “tuned” to lines (or gratings) of a specific orientation are grouped in roughly parallel “orientation columns.” A full 180° set of these columns occupies about 0.5 mm of rhesus cortex.

Since the retina is only one or two synapses peripheral to this orientation map, (in striate cortex) and since the retina presumably represents a more or less strict spatial (x, y) representation of visual data (spatially filtered, but not geometrically reorganized), one may ask the question: is there a simple map that takes a cartesian map to an orientation map? If one restricts the possibilities to isotropic regular mappings, then the complex logarithmic provides a prominent candidate [7], [17]. This can be seen as follows.

The mapping of rays at equal angular steps through the origin, under the map log (z + a), consists of parallel strips. This follows from the formula for the complex log: rays made up of points with constant φ are mapped to equally spaced parallel strips. (The complex log maps polar coordinates (r, φ) to cartesian coordinates (log r, φ). This point is illustrated in earlier work [7], [17] where this resemblance was used to construct a detailed model of the generation of the cortical hypercolumn pattern [18]. An underlying complex log map was summed by model cortical neurons whose dendritic fields and orientations were taken from published anatomical measurements. In addition, a lateral inhibitory operator was assumed that was unidirectional. This was modeled by a directional derivative (in contrast to the usual circularly symmetric Laplacian operator used for retinal inhibition modeling). A model of the cortical hypercolumn was developed which provided a good fit to available data concerning cortical hypercolumns [18]. In addition, this model predicted that binocular disparity tuning, in the range of 0.1° of disparity, should exist in rhesus striate cortex. Note that both binocular disparity in rhesus monkeys and ocular dominance column width in humans has been correctly estimated by the following scaling argument. A cortical cell that is located “over” a left or right eye ocular dominance column can receive information from the other eye (respectively) from an adjacent ocular dominance column. Referring this back to the retina, one obtains the estimate that (disparity) is equal to (ocular dominance column width/magnification factor). This can be used, in situations where two out of the three quantities are known, to estimate the unknown. Thus, prior to 1977, it was believed that there was no binocular disparity tuning in rhesus monkeys, based on experimental data of Hubel and Wiesel. However, the scaling argument above suggested that binocular disparity should be in the range of 0.1° (using 0.5 mm and 6 mm/deg as estimates of the other quantities. This estimate [7] was qualitatively and quantitatively confirmed by the later experimental observations of Poggio and Fischer [19]. In humans, ocular dominance columns were not yet observed, and their size was unknown. However, magnification had been estimated at 12 mm/deg foveally, and the scale of binocular disparity tuning is also known (as Panum’s fusional area), and is about 0.1°. This yielded an estimate of human ocular dominance column size as about 1 mm [13]. This prediction was confirmed by Hickey and Hitecock [47] who obtained evidence of human ocular dominance column size in the range of 0.8–1 mm.

Fig. 3 shows an illustration of this model. Parallel lines arranged at equal intervals in the visual field are shown mapped under the function log (z + 1). The reader is reminded that this local map occupies only a tiny piece of the entire cortical surface, and is repeated over and over again across the hypercolumn mosaic.

Several features are evident in this map. 1) a local complex log mapping provides a “fuzzy” set of orientation
columns that are quite overlapped. Nevertheless, the mapping of line segment orientation to a linear coordinate, in a roughly parallel fashion, is obvious from Fig. 3. Conspicuous in this figure is the fact that near the center of the hypercolumn (i.e., where \( z = 0 \)), there is increasingly greater overlap of neighboring orientations, and an increasing “fuzziness” of the column pattern. In other words, near the center (\( z = 0 \)), the orientation column pattern is “weakened.” This effect, due to the “singularity” near \( z = 0 \), was illustrated in the original discussion of this model [18], Fig. 3.

At the time that this model was proposed there was no evidence of orientation tuning “weakening.” The Hubel and Wiesel [1] data suggested continuous bands of iso-orientation columns. However, recently, Hubel and Livingstone [20] have reported that near the center of individual ocular dominance columns, there is a noticeable deficit of orientation tuning. This observation is in surprising agreement with the model of a local complex log geometry.

The previous discussion indicates a potential means of organizing local cortical mapping to provide parallel “orientation columns” (and possibly “spatial frequency” columns as well). Although originally proposed in order to account, in a relatively simple way, for orientation column geometry, this model also is consistent with later observations such as the existence of a central “unoriented” spot (cytochrome oxidase spots) and possibly existence of “spatial frequency” columns. However, one major puzzle concerning local cortical architecture is not addressed directly by this model. The “scatter” in receptive field location that was originally described by Hubel and Wiesel [1] represents an apparent “smearing” or imprecision of mapping. Specifically, neighboring cortical neurons might have receptive fields that are scattered over a distance comparable to the size of the cortical hypercolumn. Since cortical hypercolumn size is about 0.1° centrally, increasing in size roughly according to the cortical magnification factor; the scatter observed in receptive field location is about five to ten times larger than visual acuity. In other words, proof of the local cortical map appears to be about ten times too coarse to account for the known grain (i.e., acuity) of the system.

A second, related problem is due to the fact that cortical receptive fields are elongated. Integration occurs in one direction, while differentiation (i.e., lateral inhibition) occurs in the perpendicular direction. This integration is another nonlocal effect that should tend to “smear” the accuracy of spatial mapping.

In the following discussion, it will be shown that the properties of the Radon transform are capable of reconciling these apparent nonlocal effects with the known limits of spatial precision of the visual system. In particular, it will be shown that if the “scatter” of receptive field location is weighted by an inverse distance function, then this scatter may be interpreted as evidence of a Hilbert transform. The Radon transform is composed of serial application of a directional derivative, a Hilbert transform, and a back projection. The Radon transform, and its inverse, operate on data that is in the form of “line-integrals” of an image. Thus, the “scatter” of cortical receptive fields, the elongated nature of cortical receptive fields, and the existence of a directional derivative (lateral inhibition) perpendicular to the long axis of the cortical radio frequency (RF), may each be reconciled with the high spatial fidelity of primate vision via the properties of the Radon transform.

Introducing these ideas in a short article is difficult. Standard texts describing the Radon transform [21] should be consulted for a more complete outline of this material. In the following, the use of the Radon transform will be motivated by pointing out some of the similarities of the geometry associated with the Radon transform and the complex log model of cortical hypercolumns. This structural analogy will then be pursued by demonstrating the functional utility of the Radon transform, which allows the manipulation of elongated receptive fields (i.e., line integrals), together with considerable “scatter” (i.e., the Hilbert transform). Thus, if one accepts the fact that cortical receptive fields are elongated, and therefore have the nature of line integrals, the properties of the Radon transform become very obvious candidates for modeling cortical architecture.

The local complex log model outlined above provides a means of spatially organizing afferents to the cortex, so that a simple summation, by cortical cells, will provide “orientation columns.” The details of this summation, particularly with regard to the local receptive field “scatter” described by Hubel and Wiesel [1], have only been addressed in a preliminary fashion [18]. A suggestion of how to deal with this summation in detail is provided by studying the mapping of “line segments” in Fig. 2. It is clear that overlapped parallel bands of equal orientation are constructed. It is also clear that a very rough mapping of receptive field width (represented by “dot spacing” in Fig. 3) would be created i.e., “spatial frequency columns.” However, it is very suggestive that the spatial position (in the cortex) of the lines in Fig. 3 is systematically shifted, with increasing distance from the local “center” of the receptive field. If one recalls that the Radon transform [21] is a map of “line integrals” such that one coordinate \( s \) represents distance from the origin, and the other coordinate \( \theta \) represents orientation, then attention is drawn to the Radon, and related transforms as a means of modeling the details of “summation” of afferent input to a cortical hypercolumn.

For reference, the Radon transform [21], and the inverse Radon transform are stated in [8], [9]. The inverse Radon transform is equivalent to the application of three operators: a derivative in the \( s \) direction, \( D_s \); a Hilbert transform in the \( s \) direction, \( H_s \), and a back-projection \( B \). In the following discussion, we will consider the possibility that the inverse of the backprojection operator provides a

---

1 Note that the Hough transform, which is familiar in computer image processing, is a special case of the Radon transform, specialized to a binary image.
Fig. 3. (a) A series of parallel lines is shown. The "Radon" coordinates $(s, \theta)$ locate each line (or line integral) by distance $s$ from the origin and the angle $\theta$. (b) The map of this set of lines using the symmetrical model of [18]. Map function is $\log(x + 1)$. It can be seen that the image of the parallel lines on the right is entirely contained in a thin "column" about $45^\circ$. The locations of each of the imaged lines is indicated in the figure. These line images are located according to the values of coordinate $s$. Thus, a map, on the right, resembling the Radon transformation [21], is associated with local complex log model. (c) A superposition of six grids at $30^\circ$ angle steps. (d) "Orientation column" pattern. Note that angular movement in the "visual" plane (left) causes a linear shift in the "cortical" plane (center). A set of overlapped "orientation columns" is produced. (e) The center figure is shown with equal density bands outlined. Three bands shown here represent about 1.5 to 2 octaves change in density in the "cortical" plane. This is similar to "spatial frequency" columns. Clearly, the complex log mapping is associated with an angular set of parallel columns in one direction and a set of "size," "density," or spatial frequency columns in the other direction (labeled $F$). The question mark is there as a reminder that the correct coordinate for this direction in a hypercolumn is unknown. This figure suggests that both "spatial frequency" and "Radon coordinate" $s$ vary systematically in this direction.
columns that are quite overlapped. Nevertheless, the mapping of line segment orientation to a linear coordinate, in a roughly parallel fashion, is obvious from Fig. 3. Conspicuous in this figure is the fact that near the center of the hypercolumn (i.e., where $z = 0$), there is increasingly greater overlap of neighboring orientations, and an increasing “fuzziness” of the column pattern. In other words, near the center ($z = 0$), the orientation column pattern is “weakened.” This effect, due to the “singularity” near $z = 0$, was illustrated in the original discussion of this model [18], Fig. 3.

At the time that this model was proposed there was no evidence of orientation tuning “weakening.” The Hubel and Wiesel [1] data suggested continuous bands of iso-orientation columns. However, recently, Hubel and Livingstone [20] have reported that near the center of individual ocular dominance columns, there is a noticeable deficit of orientation tuning. This observation is in surprising agreement with the model of a local complex log geometry.

The previous discussion indicates a potential means of organizing local cortical mapping to provide parallel “orientation columns” (and possibly “spatial frequency” columns as well). Although originally proposed in order to account, in a relatively simple way, for orientation column geometry, this model also is consistent with later observations such as the existence of a central “unoriented” spot (cytochrome oxidase spots) and possibly existence of “spatial frequency” columns. However, one major puzzle concerning local cortical architecture is not addressed directly by this model. The “scatter” in receptive field location that was originally described by Hubel and Wiesel [1] represents an apparent “smearing” or imprecision of mapping. Specifically, neighboring cortical neurons might have receptive fields that are scattered over a distance comparable to the size of the cortical hypercolumn. Since cortical hypercolumn size is about 0.1° centrally, increasing in size roughly according to the cortical magnification factor; the scatter observed in receptive field location is about five to ten times larger than visual acuity. In other words, the precision of the local cortical map appears to be about ten times too coarse to account for the known grain (i.e., acuity) of the system.

A second, related problem is due to the fact that cortical receptive fields are elongated. Integration occurs in one direction, while differentiation (i.e., lateral inhibition) occurs in the perpendicular direction. This integration is another nonlocal effect that should tend to “smear” the accuracy of spatial mapping.

In the following discussion, it will be shown that the properties of the Radon transform are capable of reconciling these apparent nonlocal effects with the known limits of spatial precision of the visual system. In particular, it will be shown that if the “scatter” of receptive field location is weighted by an inverse distance function, then this scatter may be interpreted as evidence of a Hilbert transform. The Radon transform is composed of serial application of a directional derivative, a Hilbert transform, and a back projection. The Radon transform, and its inverse, operate on data that is in the form of “line-integrals” of an image. Thus, the “scatter” of cortical receptive fields, the elongated nature of cortical receptive fields, and the existence of a directional derivative (lateral inhibition) perpendicular to the long axis of the cortical radio frequency (RF), may each be reconciled with the high spatial fidelity of primate vision via the properties of the Radon transform.

Introducing these ideas in a short article is difficult. Standard texts describing the Radon1 transform [21] should be consulted for a more complete outline of this material. In the following, the use of the Radon transform will be motivated by pointing out some of the similarities of the geometry associated with the Radon transform and the complex log model of cortical hypercolumns. This structural analogy will then be pursued by demonstrating the functional utility of the Radon transform, which allows the manipulation of elongated receptive fields (i.e., line integrals), together with considerable “scatter” (i.e., the Hilbert transform). Thus, if one accepts the fact that cortical receptive fields are elongated, and therefore have the nature of line integrals, the properties of the Radon transform become very obvious candidates for modeling cortical architecture.

The local complex log model outlined above provides a means of spatially organizing afferents to the cortex, so that a simple summation, by cortical cells, will provide “orientation columns.” The details of this summation, particularly with regard to the local receptive field “scatter” described by Hubel and Wiesel [1], have only been addressed in a preliminary fashion (18). A suggestion of how to deal with this summation in detail is provided by studying the mapping of “line segments” in Fig. 2. It is clear that overlapped parallel bands of equal orientation are constructed. It is also clear that a very rough mapping of receptive field width (represented by “dot spacing” in Fig. 3) would be created i.e., “spatial frequency columns.” However, it is very suggestive that the spatial position (in the cortex) of the lines in Fig. 3 is systematically shifted, with increasing distance from the local “center” of the receptive field. If one recalls that the Radon transform [21] is a map of “line integrals” such that one coordinate $s$ represents distance from the origin, and the other coordinate $\theta$ represents orientation, then attention is drawn to the Radon, and related transforms as a means of modeling the details of “summation” of afferent input to a cortical hypercolumn.

For reference, the Radon transform [21], and the inverse Radon transform are stated in [8], [9]. The inverse Radon transform is equivalent to the application of three operators: a derivative in the $s$ direction, $D_s$; a Hilbert transform in the $s$ direction, $H_s$, and a back-projection $B$. In the following discussion, we will consider the possibility that the inverse of the backprojection operator provides a

1Note that the Hough transform, which is familiar in computer image processing is a special case of the Radon transform, specialized to a binary image.
Fig. 4. On top is shown a “Radon map” (Plane A of text). The line integral \( m(s, \theta) \) is represented at the point \((s, \theta)\). Convergence of this plane onto Plane B of text, with a 1/distance (i.e., \(1/s - s'\)) weighting, will result in the Hilbert transform in plane \((s', \theta)\). If, in addition, a derivative operator (i.e., via lateral inhibition) is applied prior to convergence, then plane B (lower) will represent the inverse back projection of the stimulus. Plane B has the phenomenological properties of the local architecture of a cortical hypercolumn, as described in the text.

description of hypercolumn architecture, and will outline an experimental test.

\[
m(s, \theta) = \text{[Radon} \left( g(x, y) \right) \text{]} \\
(s, \theta) = \int g(s \cos \theta - y \sin \theta, s \sin \theta + y \cos \theta) dt \\
g(x, y) = -\frac{1}{2\pi}BH_0D_s m(s, \theta) \\
D_s = \frac{\partial}{\partial s}; \quad H_0m(s, \theta) = \int \frac{m(s', \theta)}{s' - s} \, ds'. \tag{8}
\]

Suppose that, as suggested by Fig. 3, the Radon map coordinates \((s, \theta)\) represent the local architecture of striate cortex in some layer \(A\). Fig. 4 represents a mapping from this layer \(A\) to another layer \(B\), in which the inverse backprojection of the image \(g(x, y)\) is constructed. This convergence requires only that an inverse distance weighting describe the convergence from points in layer \(A\) to layer \(B\). This will provide a Hilbert transform. Specifically, if \(g(x, y)\) represents the image structure in the “retina” (i.e., the original projected stimulus structure), and if \(m(s, \theta)\) represents the structure in layer \(A\) (i.e., a Radon map), then the convergence of Fig. 3 will be represented by

\[
\rho(s', \theta) = B^{-1}g(x, y); \quad B^{-1} = H_0D_s. \tag{9}
\]

In other words, layer \(B\) will be the inverse back projection of the retinal patch represented by \(g(x, y)\).

One can now perform simulated electrophysiological recording in this layer \(B\) via a “gedanken” experiment. Passing an electrode along the \(\theta\) direction of layer \(B\) would record cells with oriented “line integral” receptive fields, which had inhibitory flanks due to the derivative operator \(D_s\) above. This corresponds to orientation columns. Passing an electrode perpendicular to layer \(B\) would encounter RF’s at the same orientation, but with scattered \(s\) coordinates. This scatter is due to the Hilbert transform. The experimental test of the existence of this Hilbert transform would be to determine if the density of the “scatter” varied as 1/distance from the average center of the receptive fields. It is interesting to note at this point that Sanderson [22] has described just this sort of “convolution” like behavior about LGN projection lines, but with a circularly symmetric rather than a linear structure. If this same result were found in striate cortex, i.e., that the apparent scatter of cortical RF’s within a hyper-column obeyed a 1/distance weighting with the \(s\) Radon coordinate (see Fig. 3), then the structure of a cortical hyper-column would resemble very much that of the “inverse backprojection operator, \(B^{-1}\) of (9).

The computational significance of this resemblance of cortical hyper-column structure to that of the inverse backprojection operator of (9) is based on the fact that the sequence of operators that construct the inverse backprojection (derivative and Hilbert transform) are required in order to retain the spatial information that is lost by having long “linear” receptive fields. This is of course the basis of computed tomography [21]. But, the visual system appears to use some sort of “line integral” representation in the form of cortical receptive fields. Thus, the operators associated with the Radon and inverse Radon transforms would be necessary, in order for the cortical map to retain the details of retinal spatial representation.

It therefore appears that the local cortical architecture outlined above, aside from being in good agreement with current phenomenology, suggests that the visual system reformat spatial data at the level of striate cortex so that fine spatial detail is retained (as an inverse backprojection), but that higher order form analysis based on a periodic orientation map is facilitated. This higher order shape analysis will be discussed below in terms of recent experimental work in infero-temporal cortex.

Thus, the local architecture of striate cortex may provide a compromise between preserving fine spatial detail (that exists at retinal levels), and providing a sophisticated map of boundary contours (required for later shape analysis). The experimental tests required to establish this hypothesis are clearly stated above.

Finally, it should be pointed out that Pollen et al. [23], in one of the first papers to discuss the relevance of spatial frequency models to cortical physiology, pointed out the resemblance of cortical RF’s to “projections,” but developed this observation in terms of a Fourier model, rather than the Radon model outlined above. The suggestion of the present paper is that the Radon (and associated) integral transforms may provide a computationally and physiologically correct alternative to the more popular Fourier transform, as a model of local cortical architecture.

IV. FUNCTIONAL CONSIDERATIONS—COMPUTATIONAL ANATOMY

A. Global Functional Architecture (Topographic Mapping)

1) Data Compression: One obvious functional advantage of the nonlinear (space variant) structure of the human visual system is that a wide range of visual resolution is
provided, without the heavy cost that would be incurred by a spatially uniform high resolution system. Since cortical magnification factor decreases by as much as 1.5–2 orders of magnitude from central fovea to far periphery, roughly 3–4 more orders of magnitude (of surface area) would have to be provided if a spatially uniform system, at foveal resolutions, were utilized.

This “data compression” aspect of human vision introduces other problems, however. Eye movement and attentional systems are required in order to utilize the small high-acuity area (the fovea) provided. In addition, the nonlinear map structure of this space variant system complicates many aspects of pattern recognition. For example, the cortical projection of a stimulus (e.g., a small square) changes as the stimulus moves (or as the eye scans) [13], (Fig. 5). Although the conformal nature of the cortical map preserves local angles, distances are distorted. This is a consequence of the nonlinear, space variant structure of the mapping. This aspect of pattern recognition, in terms of the cortical rather than the retinal map of the stimulus, will be discussed further below.

2) Psychophysical M Scaling: Many workers have pointed out that psycho-physical estimates of spatial vision, such as visual acuity and spatial frequency sensitivity, vary with visual eccentricity according to the curve of cortical magnification [1], [6], [7], [11], [24–26]. The implication is that visual processing is somehow “uniform” in the cortical plane. This idea was emphasized by Hubel and Wiesel [1], who pointed out that the primate visual system behaved as if the same “hypercolumn unit” was repeated uniformly across the surface of the cortex. Remapping this local “hypercolumn unit” back to the visual field would result in the $M$ scaling [25] of psychophysical phenomena related
to spatial vision. This scaling in terms of the cortical magnification seems to approximately summarize the scaling of acuity, spatial frequency sensitivity and human perceptive field size [26], although, as discussed earlier, this parallel may not hold in detail. Fig. 5 shows a graphic example of M scaling, based on a computer graphic simulation of the topographic structure of striate cortex.

4) Perceptual Invariances: The complex logarithmic mapping has a well known invariance property associated with it, which has been the object of extensive study. Consider an image that is described by a polar coordinate representation of dots (i.e., binary pixels) as the locus of points \((r, \phi)\). The complex log mapping of this set of points is the locus \((\log(r), \phi)\). This complex log locus has several pseudo-invariance properties under the symmetries of size and rotation changes of the original image \((r, \phi)\). Size change of the original image is equivalent to multiplication of each of the radial coordinates by a constant: \((r, \phi) \rightarrow (kr, \phi)\). The corresponding change in the complex log plane is \((\log(r), \phi) \rightarrow (\log(k) + \log(r), \phi)\). In other words, the result of a size dilation in the image plane is a linear shift, by the amount \(\log(k)\) in the complex log image [7], [27]. Similarly, the result of a rotation by some angle \(\phi\) in the image plane is a similar shift, in the perpendicular direction. An example of the size scaling properties of the complex log is illustrated in Fig. 6. A wrench was digitized with a TV camera at three successive distances. The camera was moved along its optical axis. This motion causes the projection of the image onto the camera plane (e.g., the "retina") to increase in size, and, at the same time, to move further into the peripheral "retina." The corresponding complex log images are size invariant, and shift along a straight line, as shown in Fig. 6. This property holds for all points of the image plane (outside of the central 0.3°) in other words, the streaming of the visual field, under motion that lies along the line of sight, is size invariant and rectilinear in the complex log plane, while it is radial and size variant in the image (retinal, or TV camera) plane.

Although the present illustration only holds for motion along the line of sight (since other directions of motion include components of shift, instead of pure size stream- ing), it provides a potential application of the geometric properties of the complex log mapping. Gibson [28] has extensively discussed this case of motion along the line of sight, from the point of view of the image (retinal) plane. It would appear that the complex log plane might simplify the visual streaming pattern for this particular case of motion along the visual axis.

The use of the complex logarithmic mapping in size scaling applications has been extensively discussed in both the biological and machine vision literature [7], [27], [29], [30]. Several of these applications have attempted to avoid the difficulties of the lack of shift invariance of the complex log mapping by providing a preliminary fourier transform [29], [30]. It has also been pointed out in the context of (biological) vision that a combination of a fourier transform, a complex logarithm, and a final fourier transform is shift, size, and rotation invariant [31], [32].

Cavanagh [32] calls the combination of fourier and complex log mapping a "log polar frequency mapping," and has proposed that either a global, or a local version of this mapping is a basis for human perceptual invariances. Unfortunately, globally this cannot be the case. These mappings yield a visual system which is completely invariant to shift: a small test stimulus (i.e., a Snellen letter) would be equally visible at all positions of the visual field of such a system. Clearly this is in fundamental disagreement with the space-variant aspects of human vision. Locally, the "log polar frequency mapping" is invariant only over the span of a single "hypercolumn" (which is the region of the proposed initial fourier transform). Thus, for stimuli that are larger than single hypercolumns, there is in fact no simple invariance property. On the contrary, the problem of patching together the 1000 or so individual "log polar frequency" mappings, as a stimulus changes size or position across the cortex, is extremely difficult, and has not been discussed. Further issues related to the use of fourier transforms in human vision are discussed in other
work [33]. At the present time, it is an open question whether the local hyper-column structure does, or does not suggest a "spatial frequency column" structure. As mentioned above, and in previous work ([18], Fig. 8), some systematic variation of receptive field width (or spatial frequency tuning) on the scale of a hypercolumn is expected, from the complex log model (Fig. 2). On the other hand, this paper has proposed that integral transforms other than the Fourier transform are potential candidates to describe hypercolumn structure (e.g., the Radon and backprojection transforms discussed above). At the present time, these are open experimental questions.

One final example of how the complex logarithmic mapping might provide insight into the nature of perceptual invariances is provided by the following model [34]. Consider a (three-dimensional) body in space that is projected through a point onto a plane. (The point may be taken as the nodal point of the eye, and the plane as the retina.)

Now, consider the log mappings of the two projections. If \( w = \log (z_1 - z) \) represents the log mapping about a fixation point \( z_1 \), and \( w = \log (z_2 - z) \) about a second fixation point \( z_2 \), then the difference mapping, or flow, of the two log mappings will be represented by

\[
\log (z_1 - z) - \log (z_2 - z) = \log \left( \frac{z_1 - z}{z_2 - z} \right). \tag{13}
\]

But the log of the cross ratio, which is a projective invariant, is (4):

\[
\log \text{CR}(z, z_1, z_2, z_3) = \log \left( \frac{z_1 - z}{z_2 - z} \right) + \log \left( \frac{z_2 - z_3}{z_1 - z_3} \right)
\]

\[
\text{CR}(z, z_1, z_2, z_3) = (z - z_1)(z_2 - z_3)/(z_1 - z_3)(z_2 - z). \tag{14}
\]

The log of the cross ratio and the difference map of the two log mappings differ only by terms that involve the three fixed points \( z_1, z_2, z_3 \), but no image terms \( z \). The third fixed point \( z_3 \) might be taken to be a point fixing the overall location of the eye coordinate system (e.g., locating the direction of the head). Since the complex cross ratio is a projective invariant (4), the cortical difference map changes, under projective distortions of the stimulus, by shifting, with no change in size, rotation, or "shape." It is a "pseudo-invariant" of projective transformations. Since this symmetry group includes size, rotation, and translation (and of course projection) as subgroups, the cortical difference map is invariant to all the major perceptual symmetries.

This model has been further discussed in other work [34], where it is emphasized that a variety of further aspects need to be developed before it can seriously be proposed as a model of human perceptual invariance. Specifically, the "hidden-surface" problem is not addressed (i.e., the uncovering of stimulus detail with three-dimensional rotation). Nevertheless, it does provide a possible means of integrating visual scanning with cortical topology in an invariant manner, contingent on the global topography of the cortical map.

1) Ocular Dominance Columns and Binocular Segmentation: The input (from the retina, via the LGN) to striate cortex is interlaced according to strips of tissue which represent the left and right eyes, respectively (see Fig. 1). A possible computational rationale for this interlacing of binocular input is suggested by an optical image processing procedure called difference mapping, e.g., [35]. In this procedure, two slightly different images are interlaced in strips, and the resulting interlaced image is spatially filtered. If a high pass filter is applied, whose cut-off frequency is in the vicinity of the spatial frequency of the interlacing operation, then the difference map of the two images is obtained. Applications for this technique have been mainly based on motion detection: if the two interlaced images differ slightly due to motion, or to motion of some sub-region, then this region is vividly enhanced in the difference map. Thus, a rapid segmentation of the images is performed, which is intrinsically parallel (i.e., it is performed optically). Mathematically, one can summarize this algorithm as follows. An "interlacing" of \( L(x, y) \) (left eye) and \( R(x, y) \) (right eye) results from projecting \( L(x, y) \) through a square wave grating (of period \( \omega \) and oriented in the y direction) and adding this projection to a similar projection of \( R(x, y) \) through the same grating which has been shifted by half a cycle. Using a fourier transform to represent the gratings functions, the interlaced image \( C(x, y) \) is given by

\[
C(x, y) = R(x, y) \cdot \left[ 1 + \sum_n a_n \sin \frac{2\pi}{\omega} ny \right] + L(x, y) \cdot \left[ 1 + \sum_n a_n \sin \frac{2\pi}{\omega} (ny + \pi) \right] = (R + L) + (R - L) \cdot \sum_n a_n \sin \frac{2\pi}{\omega} ny. \tag{15}
\]

Now, if \( C(x, y) \) is spatially filtered, the DC component \( (R(x, y) + L(x, y)) \) will be removed, and the "difference map," represented by the term \( L(x, y) - R(x, y) \) will appear in the high-pass output of the filter.

The advantage of an algorithm of this sort, for stereo segmentation, is speed: no computations are required, since all "computations" are performed by mapping and filtering, which are intrinsically parallel. The binocular segmentation of realistic visual images by direct correlation involves huge amounts of processing (an estimated one billion multiplications for one particular correlation algorithm [36]. Thus, if even an approximate segmentation could be effected by means of difference mapping, the ocular dominance column pattern of striate cortex might be computationally justified.

Further details of this model are discussed in other work [37], where a number of computational details and anatomical and physiological questions are discussed. It must be emphasized that current knowledge of the detailed anatomy of the ocular dominance column pattern supports only the resemblance of this feature of cortical anatomy to
the difference mapping algorithm, so that a careful examination of the details of the anatomy, and the limitations of this algorithm, is necessary. Nevertheless, it is intriguing to notice that columnar interlacing is a common aspect of cortical architecture, in both sensory and nonsensory cortical areas [38]. At the same time, extracting differences, in the form of difference mappings, is a common computational requirement. It is reasonable to hypothesize that columnar interlacing and difference map extraction may be common computational features of neural processing. As an example of the possible generality of this mechanism, an application of difference mapping to color opponent segmentation is provided in other work [37].

3) Infero-temporal (IT) Cortex and the Analysis of Curvature: The most striking spatial transformations that occur to the retinal input to striate cortex are the global topographic mapping, the interlacing of left and right eye input via the ocular dominance column system, and the regular periodic mapping of orientations which is associated with the orientation column system. In the previous discussion, a variety of computational rationales for both the global topographic mapping and the ocular dominance column systems have been investigated. The periodic mapping of contour orientations via the orientation column system is strongly suggestive of an analysis of shape in terms of some aspect of boundary orientation or curvature. In the computer literature, there is a well-known method of shape analysis known as the “fourier descriptor” approach e.g., [39]. It has been suggested, due to the orientation column mapping of striate cortex, that extra-striate cortex (and, in particular, infero-temporal cortex (IT)) may analyze the “shape” of stimuli via an analysis of the spatial texture of the striate cortex map [13]. Since spatial intervals in striate cortex correspond to differences in contour orientation (i.e., to curvature), it would be relatively simple for extra-striate visual areas (e.g., IT) to achieve a shape descriptor by “viewing” the striate cortex orientation map through a suitable spatial filtering algorithm. This hypothesis has led to the experimental test that will now be described, which used the basic visual patterns associated with the method of “fourier descriptors” in order to test the hypothesis that IT cortex may be using a periodic orientation or curvature description of “shape” [40], [41].

The method of fourier descriptors has been intensively investigated as a compact representation of shape [39], [42]. Briefly, a simple closed contour is represented by its boundary angle function. This is a representation of the orientation of the tangent to the shape at regular intervals around the boundary of the shape. Fig. 6 shows an example of this boundary angle function for a simple shape. Since the boundary angle function represents the shape as a one-dimensional function (parameterized by arc length), a one dimensional fourier analysis may be performed. Then, the shape is represented by periodic components of boundary orientation, or “fourier descriptors.” Each fourier descriptor is associated with a frequency $f_k$ and amplitude $A_k$, and a phase $\alpha_k$. If the boundary angle function is represented as $B(t)$, where $t$ represents equal steps of arc length around the perimeter of the shape, then the set of fourier descriptors (i.e., the fourier transform of $B(t)$ is given by Zahn and Roskies [39]:

$$A_k e^{i2\pi nt} = \int_0^T B(t) e^{i2\pi nt} dt. \quad (16)$$

Any (closed, simple) shape is represented by its corresponding set of fourier descriptors (FD’s). A relatively small, truncated set of FD’s provides a reasonable estimate of the “gestalt” of a shape, as shown in Fig. 7. FD’s have provided an important approach to shape description in the machine pattern recognition literature [39], [42], and represent a global approach to shape description that is contingent on a periodic representation of boundary orientation.

Since striate cortex seems to provide just such a data form (e.g., the periodic orientation column pattern), the hypothesis has been advanced that IT neurons might code shape on the basis of global shape features such as FD’s [13].

In order to explore this possibility, a set of visual stimuli was printed onto photographic slides (from a computer graphic display). The inverse transform of a single FD uniquely determines a closed boundary with a specific number of lobes (frequency), lobe indentation (amplitude) and orientation (phase), as shown in Fig. 8. These shapes are referred to in the following as FD stimuli.

If IT neurons function as band-pass filters for shape, one would expect different IT neurons to be tuned to different FD stimuli. Furthermore, the tuning should be independent of the size, contrast, and position on the retina [44]. Thus, the activity of a set of such neurons could specify or code any shape (that was simply connected).

A total of 234 visually responsive neurons in five macaques were studied with these FD stimuli. The monkeys were immobilized with pancuronium bromide (to eliminate eye movements) and anesthetized with nitrous oxide. All units were tested with slits of light, a set of complex objects (e.g., brushes, dolls, etc.), and a series of FD stimuli. The FD stimuli were white on black patterns projected onto a tangent screen at the center of gaze, and ranged in frequency from 2 to 64 cycles/perimeter in octave steps. For some units, the frequency series was repeated at different sites within the receptive field, with different FD amplitudes, with different stimulus sizes, or with reversed stimulus contrast.

Fifty-four percent of the visually responsive IT neurons were “tuned” to the frequency of FD stimuli [40], [41] with a mean bandwidth of two octaves (FWHM). Different cells had different best frequencies, and all frequencies were about equally represented (although the highest frequency used (frequency 64) had a reduced incidence). Thus, for about half of the cells in IT, the frequency of FD’s appeared to be a relevant stimulus dimension. The remaining neurons either did not respond to the FD stimuli (26 percent), responded equally well to all of the FD stimuli (13 percent), or had multimodal tuning curves (7 percent).
Fig. 7. On top left is shown a hand. The boundary angle function of this shape is plotted in the center. On the right is the power spectrum of this boundary angle function. Below is shown several “approximations” to the hand, that are obtained by truncating the set of Fourier descriptors, and then reconstructing the stimulus from the truncated set. This shows that the Fourier descriptors capture the “gestalt” of a shape, even with only a few descriptors.

Fig. 8. The shapes associated with individual Fourier descriptors are shown. (a) A set at amplitude 0.8 (units of Zahn and Roskies [39]), and frequencies 2 (bottom left), 4, 8, 16, 32, 64 (top left). (b) Amplitude 1.6, at frequencies 2 (bottom left), 4, 8, 16, 32, 64. These shapes were used as stimuli in a recent experiment studying shape sensitivity of inferotemporal cortex [40], [41].

Thus, although the use of FD methods for the physiological and psychophysical study of shape perception needs much further work, they seem to provide a promising method of bridging the gap between the local analysis of contrast (provided by conventional spatial frequency and edge detection approaches) and the global analysis of pattern and shape.

The previous discussion of boundary curvature sensitivity in IT cortex indicates that the Fourier descriptor method provides a quantitative probe of shape sensitivity in the primate (and human) visual system. It also indicates one way in which an invariant: shape description may be achieved, despite the space variant mapping that occurs early in human visual processing (i.e., at the level of striate cortex). The Fourier Descriptors are based on the intrinsic geometry of a shape: they provide a description of curvature parameterized by arc length. This is the reason that they are size and translation invariant. If a similar method of analysis is used in the human visual system, then a variable metric, with respect to visual field coordinates, clearly must be employed at some level of visual processing. Specifically, the IT cells which “view” patterns of cortical orientation columns (through intermediary extrastriate areas) could achieve this by means of space variant filtering that adjusted the measured curvature with
respect to position across the cortical surface. In other words, a given spatial interval between “lit up” orientation columns would have a curvature value that was determined by the position in the visual field indexed by that particular hypercolumn. By means of a space variant metric of this kind, the nonlinear mapping of early visual processing need not complicate later states of visual computation, could benefit from the data compression aspects of the striate cortex map, yet could provide an invariant description of shape.

VI. SUMMARY

In summary, this paper has provided a mathematical characterization of the spatial structure of striate cortex topography, and has proposed several models of the local patterns of functional architecture in striate cortex. A variety of computational applications suggested by these forms of mapping have been proposed. These include applications to visual data compression, stereo segmentation, perceptual invariances, and shape analysis. Finally, recent experimental results concerning shape analysis by high-order cortical processing (i.e., IT cortex) have been described.

The principal thrust of this analysis has been to view the intricate patterns of functional architecture of striate cortex as one key to the means by which the primate (and human) visual system achieves its impressive level of visual performance. The anatomy of the central nervous system may play a direct role in neural computation by providing novel architectures for image representation that greatly simplify the very intensive computations that must occur. In other words, it is argued that in the biology of vision, just as in many other areas of biology, form follows function.

REFERENCES


