

Stereo, Shading, and Surfaces: Curvature Constraints Couple Neural Computations

This paper discusses computational problems faced by the mammalian visual system, articulates theoretical models of its solution methods, and outlines the implications for computer vision applications.

By STEVEN W. ZUCKER, Fellow IEEE

ABSTRACT | Vision problems are inherently ambiguous: Do abrupt brightness changes correspond to object boundaries? Are smooth intensity changes due to shading or material properties? For stereo: Which point in the left image corresponds to which point in the right one? What is the role of color in visual information processing? To answer these (seemingly different) questions we develop an analogy between the role of orientation in organizing visual cortex and tangents in differential geometry. Machine learning experiments suggest using geometry as a surrogate for high-order statistical interactions. The cortical columnar architecture becomes a bundle structure in geometry. Connection forms within these bundles suggest answers to the above questions, and curvatures emerge in key roles. More generally, our path through these questions suggests an overall strategy for solving the inverse problems of vision: decompose the global problems into networks of smaller ones and then seek constraints from these coupled problems to reduce ambiguity. Neural computations thus amount to satisfying constraints rather than seeking uniform approximations. Even when no global formulation exists one may be able to find localized structures on which ambiguity is minimal; these can then anchor an overall approximation.

KEYWORDS | Boundary detection; computational vision; constraint satisfaction; neural computation; shading analysis; stereo

I. INTRODUCTION

Cortex consists of billions of neurons and trillions of synapses, all in support of various neural computations. Key to understanding these computations is building a proper abstraction. While one routinely thinks of neurons as decision-making units, it is most important to understand which questions they are attempting to answer. Knowing the answers could suggest insights from neuroscience to guide engineering theories and applications; at the same time, practical considerations can provide insight into neural computations.

Our focus is on problems of early and intermediate-level vision. These problems are difficult for applications (and for brains) because they are inverse problems [94]. Computer graphics, by contrast, is a forward problem: shading can be calculated directly given models of surfaces, viewing geometry, and lighting [24]. Going the other way there are (in general) many different surfaces and lighting combinations that could account for a given shading distribution. Structuring these inverse choices is what makes vision an inference problem.

Big data and machine learning define, to some extent, our intellectual environment. It is already the case that solutions to certain classification problems, such as reading zip codes, can be learned automatically [70]. But how far can one go: is it possible to learn how to infer surfaces from shading in an unconstrained, unsupervised fashion? We maintain that there are deep insights into

Manuscript received November 6, 2013; revised February 11, 2014; accepted March 10, 2014. Date of publication April 23, 2014; date of current version April 28, 2014. This work was supported by the U.S. Air Force Office of Scientific Research (AFOSR), the National Institutes of Health (NIH), the National Science Foundation (NSF), and The Paul G. Allen Family Foundation.

The author is with the Departments of Computer Science and Biomedical Engineering, Yale University, New Haven, CT 06520-8285 USA (e-mail: steven.zucker@yale.edu).

Digital Object Identifier: 10.1109/JPROC.2014.2314723

0018-9219 © 2014 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

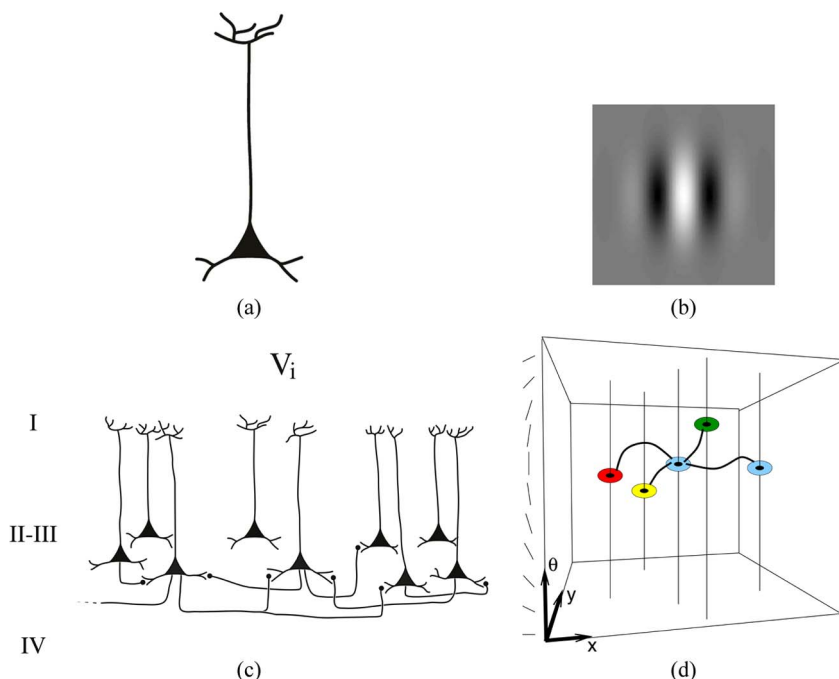


Fig. 1. Function of an individual neuron (a) in visual cortex is classically summarized by its receptive field (b). Shown is a Gabor filter tuned to the vertical orientation. (c) Connections between such neurons define networks and (d) different abstractions of these networks lead to different theoretical ideas. One focus of this paper is to understand an abstraction based on geometric principles.

these problems that are geometric in nature and that could provide novel constraint. And as we will show, the geometry is also reflected in neurobiology. The lesson, in short, is that geometry serves (at least) as a surrogate for higher order statistical analysis. A concrete example in edge statistics supports this claim, and a surprising result about the role of color reinforces its usefulness.

A. From Neural Connections to Distributed Models

The selectivity of individual neurons to patterns of light has strongly influenced ideas about neural computation in the visual system (Fig. 1). Receptive fields, or the pattern of light to which a neuron responds, can be related to the statistics of natural images by independent component analysis [7] and sparse coding [86]. At a larger scale there are about 50 anatomically distinct visual areas [31], each of which consists of elaborated networks of neurons. For nearly every feedforward connection from neurons in one area to the next, there is a feedback projection from the higher area.

Since receptive fields can be built up from earlier projections, they have been taken as a proxy for feedforward connections between neurons in different areas. Repeated across several “hidden layers” we obtain a model for cortical architecture (Fig. 2, middle row, right). Such deep network models began with the neocognitron [34]; modern extensions [108] have different nonlinearities imposed between the feedforward convolutions. Passing the output layer into a

classifier leads to recognition systems [30], [70]. Popular algorithms exist for both supervised and unsupervised learning of network parameters [41].

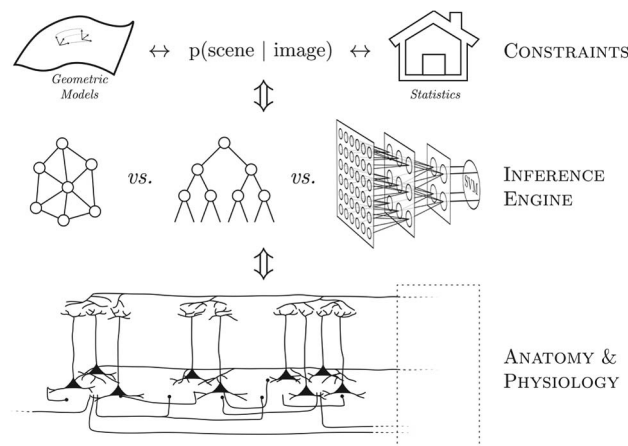


Fig. 2. Levels of explanation are grounded in neurobiology and include both the inference engine and the constraints on which it operates. At the inference engine level we show (right) a deep convolutional network, with many “hidden” layers that is, in effect, equivalent to a specialized computation on (middle) directed acyclic graphs; such graphs are a special case of general graphical models (left). At the constraint level, which provides the “edges” in the graphical models, are (left) statistics derived from the world and those derived from models (right). We will concentrate on geometric models in this paper.

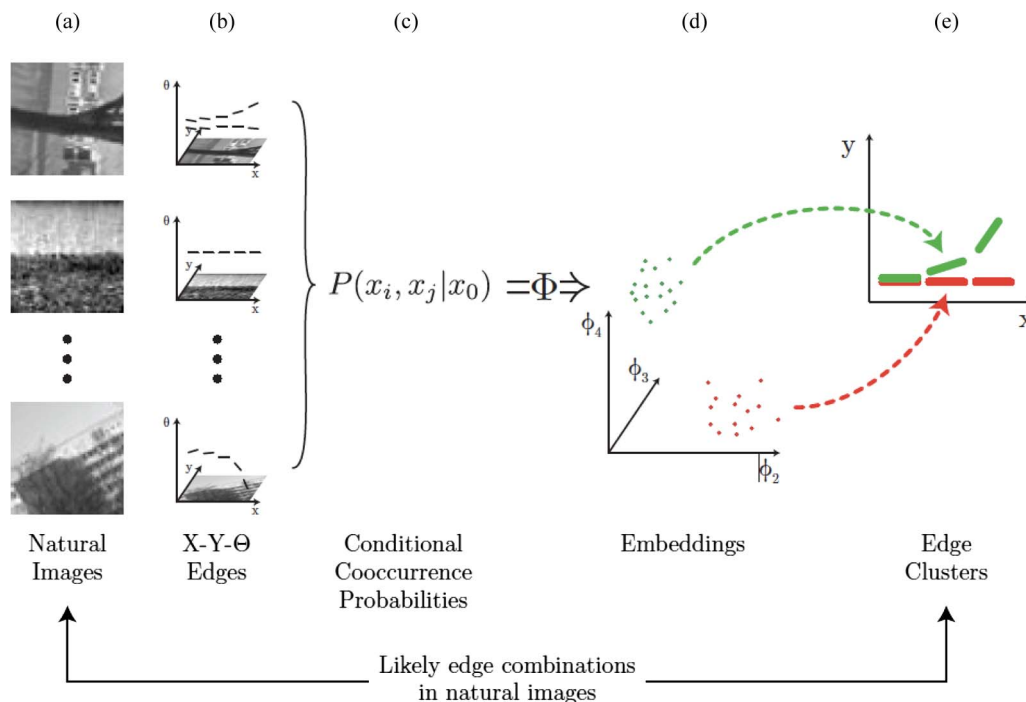


Fig. 3. From image statistics to constraints. Edges in natural images (a) can be represented as points in (position, orientation) space (b). The joint probability of multiple edges co-occurring in a large corpus of image patches can be estimated (c). Since this probability “matrix” is positive semidefinite, its eigenvectors can provide an embedding (d) in which those edge triples likely to co-occur appear as clusters [see (3)]. Mapping these clusters back to a position representation reveals the geometry of curves (e). Figure after [69].

But there is much more to cortical anatomy. There are several interconnected pathways in the ventral stream implicated in object representation [61], and neurons within each of the areas participate in elaborate networks involving both short-range and long-range connections. Fig. 2 (bottom) shows a cartoon elaboration of this intraarea network, the output of which projects to the next area. The recurrent backprojection is shown arriving in the superficial (top) layers. Deep convolutional networks are essentially directed acyclic graphs [Fig. 2 (middle)]; more realistic functionality requires a graph with cycles [Fig. 2 (middle, left)]. How might this more elaborate function be described computationally? Again, there are many possibilities.

In some deep networks, the feedforward projections specify activity, and the feedback modifies synaptic weights by error signal backpropagation. Richer classes of graphical models [60] have been suggested for computational reasons. Hierarchical Bayesian networks [71] postulate inferences supported by a combination of feedforward observations and feedback priors. For a problem such as shape from shading, for example, feedforward data about image intensity might be interpreted with regard to feedback involving surface and light source priors. (We discuss this further in Section III-B.) In computer vision terms, such inverse problems are often formulated as finding a (latent) parameter vector that best describes given (e.g., image) data according to a model [122]. The model

is realized as an energy function, and the model parameters are learned from training data. A practical consideration is that there are fast algorithms to guide the search for interpretation parameters, but only for certain graphs [16], [116].

Bayesian networks [14] and Markov random fields (MRFs) are related realizations [1], [76], [115]. A popular form resembles statistical mechanics [44] and motivates a connection to regularization terms in MRFs [95]. Boltzman machines [42] exploit the underlying probability distribution for sampling.

In a simple sense, neurons can be viewed as decision makers, firing an action potential when they receive sufficient support (ionic current) from other neurons projecting to them. Considering the set of “neurons” as nodes in a graph, we obtain a very simple form for such networks. Let the edges specify which neurons are connected in the graph and, leaving technical considerations aside, we obtain a natural quadratic “energy” form relevant to Hopfield networks and (symmetric) relaxation labeling [50]. In symbols, if p_i denotes the probability that neuron i fires and $c_{i,j}$ denotes the synaptic coupling from neuron j to i , then summing over all interacting neighbors for each node in the graph yields

$$\text{Energy} = \sum_{i,j} p_i c_{i,j} p_j. \quad (1)$$

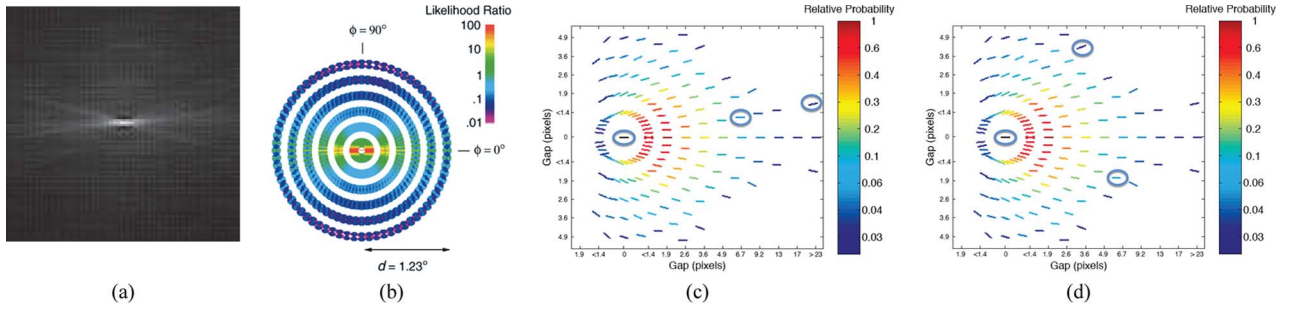


Fig. 4. Association fields derive from pairwise co-occurrence statistics and illustrate the probability (likelihood) of a particular edge near a horizontal edge at the center position. Two equally likely pairs of edge pairs are shown in (c) and (d); higher order co-occurrence probabilities are necessary to determine which of these is more likely. (a) After [4]. (b) After [35]. (c) and (d) Data from [28].

Constrained gradient ascent provides an approach to determine which neurons should be active to maximize energy from an initial distribution p_i . More generally, when neurons are viewed as coupled decision makers, a more subtle connection to polymatrix games arises [80], [81]; in this case, the optimal payoff is given by the Nash equilibrium, and the constraints no longer need to be symmetric. For general MRFs, the constraints $c_{i,j}$ are embedded in clique potentials. Constraints are thus the “guts” of the models: there are several different types of machines—inference engines—within which to use them. The question, then, is how to find these constraints.

B. From Image Statistics to Abstract Constraints

Statistical regularities underlie many models of machine and biological learning. For example, objects in our visual world are coherent, and this coherence is reflected in the probabilities that edge elements (or image intensities or other features) co-occur [111]. The famous Hebb synapse [18] is often summarized by the phrase: *cells that fire together wire together*. Since many cells respond to edges, it is natural to start with those statistics (Fig. 3).

Let E_i denote an edge at position, orientation $r_i = (x_i, y_i, \theta_i)$. Viewing this as a $\{0, 1\}$ -valued random variable \mathcal{E}_i , the joint distribution $P(\mathcal{E}_i, \mathcal{E}_j)$ is well studied [4], [27], [35], [62]. It is convenient to view this distribution around a horizontal edge at the center of an image patch (Fig. 4). Such “association field” [32] models of continuation are prominent in psychophysical research [26], [39]. While pairwise information is useful, higher order structure could be even more useful. Thus far, such higher order information has been developed through models tied to applications [36], [54]. As we now show, following [69], it is possible to infer higher order statistical information directly.

The association field is a representation of pairwise information: it displays roughly the probability that edge E_i is present given a horizontal edge at the center. Now consider triples of edges. These could derive from edge

pairs that are equally likely to occur but not likely to occur together [Fig. 4(c) and (d)]; or from pairs that are likely to occur together. Statistically, such third-order questions are complex to answer (but see [119]).

Denote positive edge triple co-occurrences by $P(\mathcal{E}_i = 1, \mathcal{E}_j = 1, \mathcal{E}_k = 1) = P(i, j, k)$. This matrix can be estimated from natural image edge patches by finding a strong edge, moving it to the center of the patch (20×20 pixels; ten orientations/position) and then rotating so that it is horizontal

$$P(i, j|0) = P(\mathcal{E}_i = 1, \mathcal{E}_j = 1 | \mathcal{E}_0 = 1) \quad (2)$$

where $P(\mathcal{E}_0) = 1$ denotes a horizontal edge at the origin. (Edges are isolated by enforcing local nonmaxima suppression and inhibiting lateral spread.) Since $P(i, j|0)$ is positive semidefinite, edge triples can be visualized by forming an embedding based on the eigenvectors that diagonalize the matrix [38]

$$P(i, j|0) = \sum_{l=1}^n \lambda_l \phi_l(i) \phi_l(j)$$

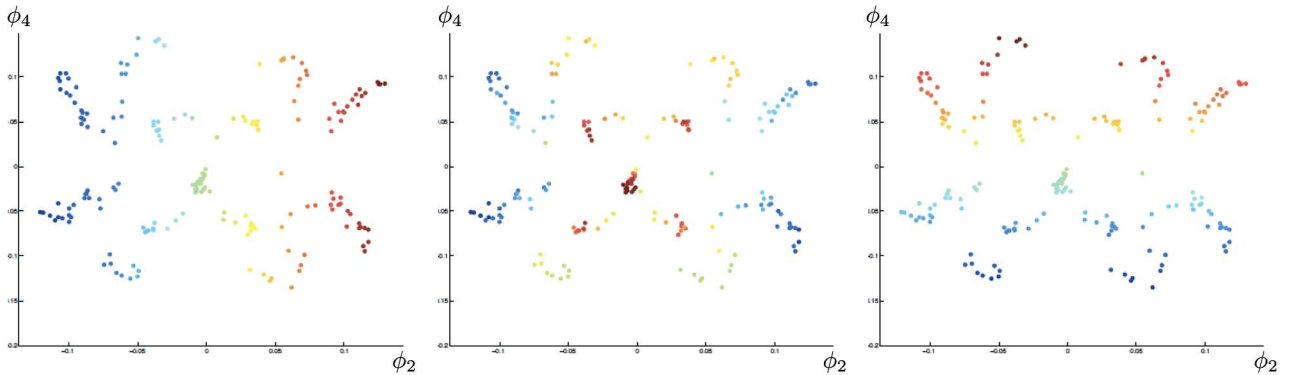
where the eigenvectors ϕ_l allow a spectral embedding

$$\Phi : r_i = \begin{pmatrix} x_i \\ y_i \\ \theta_i \end{pmatrix} \rightarrow \mathbb{R}^n.$$

Φ maps edges to points in an embedded space where squared distance is equal to relative probability

$$\Phi(r_i) = \left\{ \sqrt{\lambda_1} \phi_1(i), \sqrt{\lambda_2} \phi_2(i), \dots, \sqrt{\lambda_n} \phi_n(i) \right\}. \quad (3)$$

Spectral embedding colored by embedding coordinates



Edge map colored by embedding coordinates

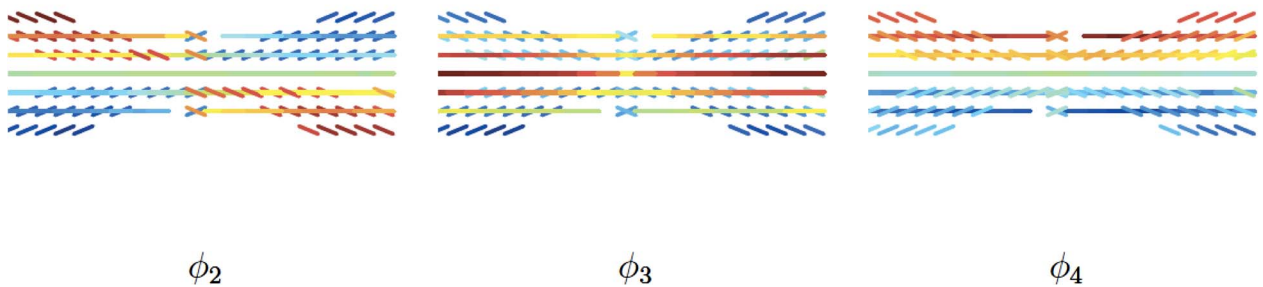


Fig. 5. Display of third-order edge structure showing how oriented edges are related to their spectral embeddings. (Top) Spectral embeddings. Since the spectrum of $P(i, j|O)$ decays rapidly, edges (points in illustration) that are likely to co-occur with E_0 can be visualized as clusters (small diffusion distance). Embedded edges are plotted in (ϕ_2, ϕ_4) coordinates and colored by the value of ϕ_2, ϕ_3, ϕ_4 as shown. (Bottom) Edge distributions mapped back into (x, y) and again colored by eigenfunctions. ϕ_2 shows linear organization and ϕ_4 shows a curvature organization. Compare with Fig. 4 where red edges all have high probability of occurring with the center, but no information is known about their co-occurrence probability. Figure after [69].

In this space, the Euclidean distance between embedded points is given by (see also [21])

$$\begin{aligned} \|\Phi(r_i) - \Phi(r_j)\|^2 &= \langle \Phi(r_i), \Phi(r_i) \rangle - 2\langle \Phi(r_i), \Phi(r_j) \rangle \\ &\quad + \langle \Phi(r_j), \Phi(r_j) \rangle \\ &= E\left[\mathcal{E}_{r_i}^2 | \mathcal{E}_{r_0} = 1\right] - 2E\left[\mathcal{E}_{r_i} \mathcal{E}_{r_j} | \mathcal{E}_{r_0} = 1\right] \\ &\quad + E\left[\mathcal{E}_{r_j}^2 | \mathcal{E}_{r_0} = 1\right]. \end{aligned}$$

The first and last terms in this embedding are basically the association field: the edges likely to occur with the center, horizontal edge. The middle term measures the co-occurrence of the other pairs; in other words, edges E_i and E_j that both frequently co-occur with a horizontal edge at the center (see Fig. 5). These include straight continuations and curves with positive and negative curvatures. In

other words, high-order edge statistics reflect the natural geometry of contours.

In summary, whether we are using hidden variables, priors, or synaptic connections is determined by the inference engine employed. In all cases, these variables represent constraints: constraints between neurons at the physiological level or constraints between tokens at the scene level. Here we showed that there is significant higher order statistical structure to edge elements, but we had to develop a special technique to reveal it. This can be viewed as a learning strategy. Most importantly, it revealed an identification with geometrical ideas, which we take as a surrogate to working with very high-order statistics.

C. Overview of the Paper

Lighting and material properties combine in the image formation process: even simple photometric models involve a product of lighting and surface albedo. Such coupling between problems has been addressed in

computer vision as a series of coordinated intrinsic images [6], [118], [124]. The intrinsic images model has its roots in Land and McCann's retinex theory, which was developed to explain color constancy. Retinex is based on the idea that sharp (high-frequency) variations denote material changes and slow (low-frequency) variations denote cast shadows. The modern version [118] keeps the idea that properties (e.g., albedo) are scalar fields over the image and also characterizes points of change (large image derivatives). As we will show, there is significantly more to differential structure than that involved in boundaries, and significantly more to color variations than abrupt material versus gradual shadow edges. This will elaborate the notion of geometrical models introduced above, and will take the form of different flows—essentially vector fields—defined along curves and within shaded or colored regions.

Flows are related to differential equations. Shape from shading typically involves a partial differential equation (PDE) to be solved over a global surface from boundary information [45] using smoothness constraints; similar smoothness constraints have been postulated for stereo (e.g., [20], [79], and [96]). We develop a different path: biology suggests looking “locally,” and we show that some parts of the shape-from-shading problem are inherently less ambiguous than others. Hence, there could be a real advantage to “locking down” certain parts of the solution and then interpolating others. It is a little like doing a puzzle: start with those pieces about which you are certain and then use constraints to fit nearby pieces together with them. Just as neurons are connected into networks, problems such as these (and their decompositions) imply networks of local problems that can be fitted together [78], [83], [103].

II. EARLY INFERENCE PROBLEMS

Early biological vision often connotes boundary detection and segmentation to computer vision researchers, because the first cortical visual area, V1, contains neurons selective for (a sampling of) all orientations at every retinotopic position [Fig. 6(a) and (b)]. It is thought that these are local edge detectors. Taken together we have a columnar model [48] that suggests an identification with the geometry of fiber bundles. We start with such models to set the stage, and then move to stereo and shading analysis.

A. Contour Geometry

Visual cortex in primates provides a rich substrate for realizing networks of orientationally selective neurons that could implement the high-order statistical constraints just described (Fig. 2, bottom). Orientation selectivity begins in layer 4 [82], [113]; there is a substantial projection to the upper levels [19], [25] that is associated with boundary processing [2]. Anatomical studies reveal that these intrinsic connections are clustered [37] and orientation dependent [15], leading many to believe that consistent

firing among neurons in such circuits specifies the orientations along a putative contour [32], [52], [128]. Random fields and neural networks are all about using context (e.g., along the contour) to remove noisy responses that are inconsistent with their neighbors' responses or to reinforce weak or missing responses. How might constraints $c_{i,j}$ be designed for such a task? Do they resemble third-order edge statistics?

We apply this machinery to contour detection in Fig. 6 following [9]. Fig. 6(b) shows how neurons form circuits with long-range horizontal connections [3], [15], [100]. Activity in such circuits can be interpreted geometrically [Fig. 6(c)]: viewing orientationally selective responses as signaling local, linear approximations to a contour, suggests interpreting them as signaling tangents to contours. Mathematically, a tangent can be transported along an approximation to the curve (indicated as the osculating circle) to a nearby position. Compatible tangents are those that agree with sufficient accuracy in position and orientation following transport; this is the cocircularity approximation [89]. In (position, orientation) space [Fig. 6(d)], a length of circle in the image lifts to a length of helix in (x, y, θ) . Identifying this diagram with the one above it shows that the transport operation need not be carried out mathematically; it can be embedded in the long-range connections. Projection into the image plane of these connections indicates either straight [Fig. 6(e)] or curved [Fig. 6(f)] patterns. In biology, such connections are called projective fields [72]. Returning to (1), these are the $c_{i,j}^{\kappa}$, for i denoting diagonal in the center and j denoting another edge. The superscript κ indicates that these are a function of the curvature; cf., the clusters of third-order edge structure (Fig. 5).

Algorithmically, we can use these connections by elaborating the index in (1) to include curvature: $i = (x_i, y_i, \theta_i, \kappa_i)$. The gradient ascent in energy is then as follows.

Given: connections $\{c_{i,j}\}$ and initial probability estimates $\{p_i^0\}$ for each discretized position, orientation, and curvature. Update: the probability estimates (until convergence) by

$$p_i^{n+1} \leftarrow p_i^n + \eta \frac{\partial}{\partial p_i} [\text{Energy}] \quad (4)$$

$$= p_i^n + \eta \prod \left[\sum c_{i,j} p_j^n \right] \quad (5)$$

where η is a step size and \prod is a projection operator onto the probability simplex (necessary to keep $0 \leq p_i^n \leq 1$ and appropriate $\sum_i p_i^n = 1$). Consistency in firing according to patterns would, of course, reduce noisy responses implying an increase in firing sparsity [120].

In addition to the connections intrinsic to V1, there are feedforward projections from layers 2/3 to higher visual

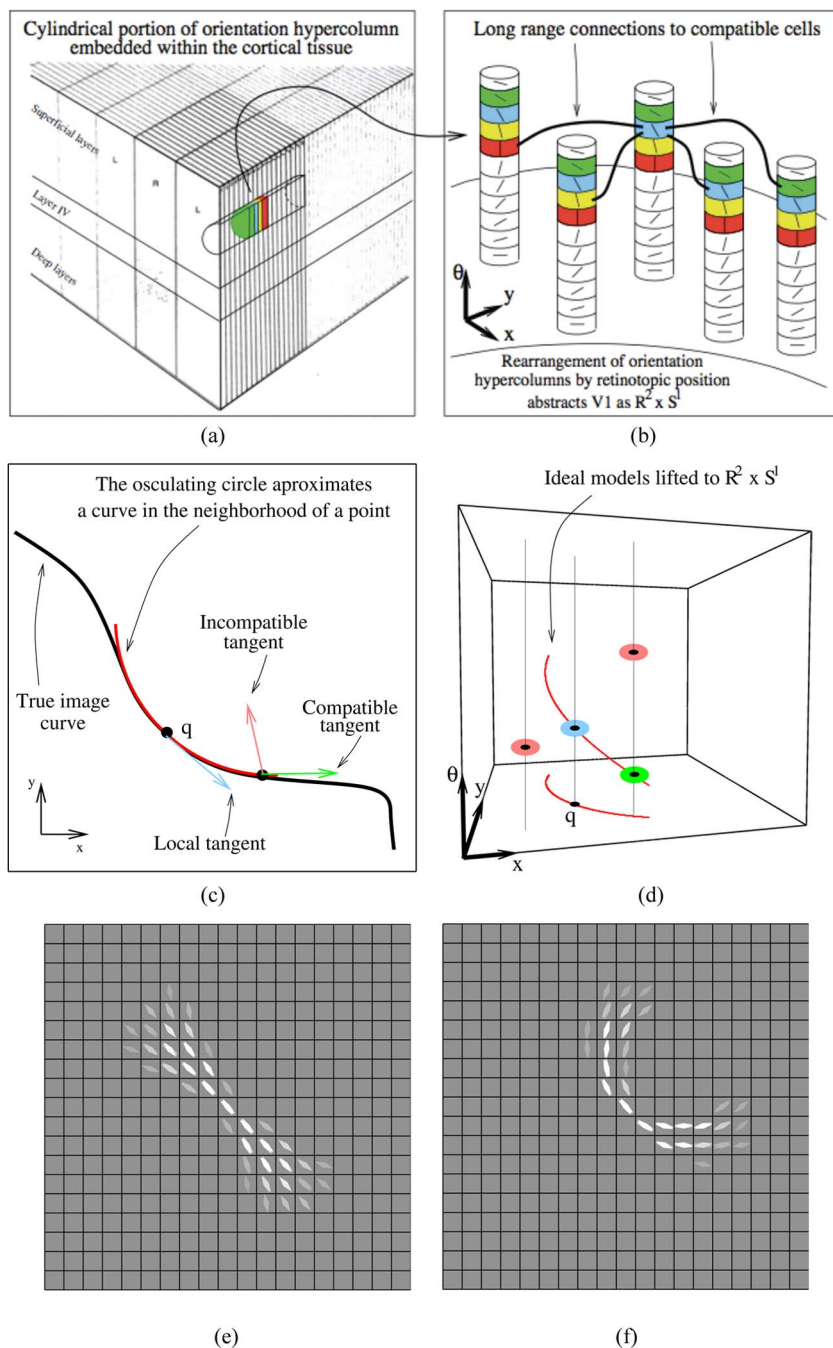


Fig. 6. Columnar organization of visual cortex. (a) A group of cells selective for different orientations at about the same location in the visual field. **(b)** This column of cells is rearranged in (position, orientation) coordinates. (Long range horizontal) connections between cells relate an orientation signal θ at position (x, y) to another orientation θ' at (x', y') . **(c)** If each cell signals a tangent to a contour, then transport along the contour can reveal consistency among nearby tangents. **(d)** Using the osculating circle as a local approximation to the curve, transport over short distances in (x, y, θ) is movement along a helix. By identification with (b), these helices are a model for the horizontal connections. They are a function of curvature, either straight (e) or curved (f). Figures after [9].

areas [3], [88], [112]. V2, for example, has an elaborate organization into subzones, including the thin, thick, and pale stripe areas [102]. It is thought these participate in stereo and color computations, to which we will turn shortly. There are also feedback projections from higher

visual areas [3], [101]. Since receptive fields are larger in higher areas, this could involve contour computations over a larger scale [128].

Differential geometry specifies how orientations align along a contour. Following [87], let $\beta : I \rightarrow \mathbb{E}^2$ with

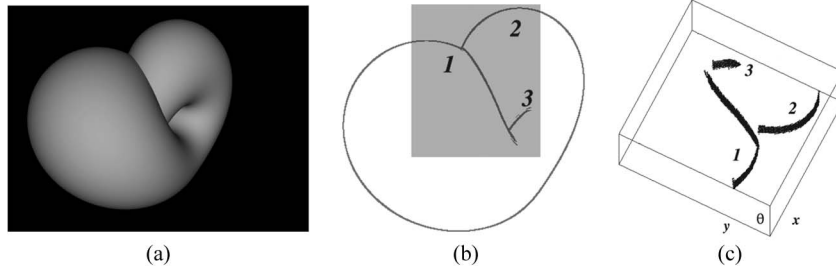


Fig. 7. Discontinuities in (x, y, θ) -space are represented by multiple orientations at the same location. (a) Image of a Klein bottle with edges (b). Such edges signal monocular occlusion events. When lifted into (x, y, θ) -space there are multiple values at a position, shown (c) by tilting (x, y, θ) -space so that the fibers are at an angle.

$\|\beta'(s)\| = 1, s \in I$ denote the unit-speed curve defined by the differentiable map from the interval I into Euclidean 2-space. The unit tangent is $T = \beta'$, from which we get $T' = \beta''$, the curvature vector field. Observe that T' is orthogonal to T (just differentiate $T \cdot T = 1$). The direction of the curvature vector is normal to β , and its length $\kappa(s) = \|T'(s)\|$ defines the curvature. The vector field $N = T'/\kappa$ defines the principal normal.

The Frenet frame field on β is the pair (T, N) such that $T \cdot T = N \cdot N = 1$, all other dot products = 0, and the above conditions hold. The elegance of cortical geometry derives from the fact that derivatives of the frame can be expressed in terms of the frame itself. For $\kappa > 0$, we have

$$\begin{pmatrix} T' \\ N' \end{pmatrix} = \begin{bmatrix} 0 & \kappa \\ -\kappa & 0 \end{bmatrix} \begin{pmatrix} T \\ N \end{pmatrix}. \quad (6)$$

The lift from the image into cortical coordinates [Fig. 6(b) and (d)] reveals a rich connection to Gestalt principles [121]. Good continuation [125] for curves—that slow changes in orientation should be preferred to sudden, abrupt ones—has a special realization in (x, y, θ) -coordinates: at the crossing point of a figure “8” are two line orientations—tangents—but these are separated along the column—the fiber—of orientations. Good continuation means that there is no big jump along a fiber; the connections to nearby tangents are “shorter” by passing through the junction. The nonsimple curve in the plane becomes a simple curve in (x, y, θ) . The contact geometry for this has been worked out [106]; see also [126].

Discontinuities are a different story, however (Fig. 7). Now multiple orientations at the same position signal what often amounts to a monocular occlusion event [13], [128]; a contour ending can signal a cusp [68].

Before moving on, we draw a lesson from the columnar organization. The column is a representational architecture that contains each possible curve tangent at every position;

the bundle of columns contains every possible curve. This architecture will be repeated for other problems.

B. Texture and DTI

Orientation-defined textures [11], [53], [98], [114] arise when oriented elements are dense in two directions rather than one, in effect weaving edges together into a tapestry. Again the orientation column/fiber bundle structure works ideally to represent such patterns, and again there is a high-order curvature dependency. The mathematics are generalized, with the Frenet curvature replaced by a Cartan connection form [87] (Fig. 8). The form at each location is denoted (E_T, E_N) and transport is generalized from tangential motion along a streamline to the entire tangent plane. This will lead to richer projective fields.

The transport equations are analogous to the curve case, except now it is possible to move the frame in any (tangent plane) direction rather than only along a contour. This requires the use of covariant derivatives rather than standard ones, and a one-form w for the curvature

$$\begin{pmatrix} \nabla_V E_T \\ \nabla_V E_N \end{pmatrix} = \begin{bmatrix} 0 & w_{12}(V) \\ -w_{12}(V) & 0 \end{bmatrix} \begin{pmatrix} E_T \\ E_N \end{pmatrix}. \quad (7)$$

The Cartan connection equations resemble the Frenet–Serret formulas but involve the connection form $w_{12}(V)$. Such forms “take” a vector as “input” and “output” a scalar. Just as surface curvature can be expressed in terms of principal curvatures, for general oriented patterns there are two basic curvatures

$$\begin{aligned} \text{tangential curvature: } \kappa_T &= w_{12}(E_T) \\ \text{normal curvature: } \kappa_N &= w_{12}(E_N). \end{aligned} \quad (8)$$

Psychophysically, we are sensitive to these curvatures [8], [12], [52], [84]. Knowledge of $E_T, E_N, \kappa_T, \kappa_N$ at a point (x_0, y_0) allows us to develop an osculating flow field

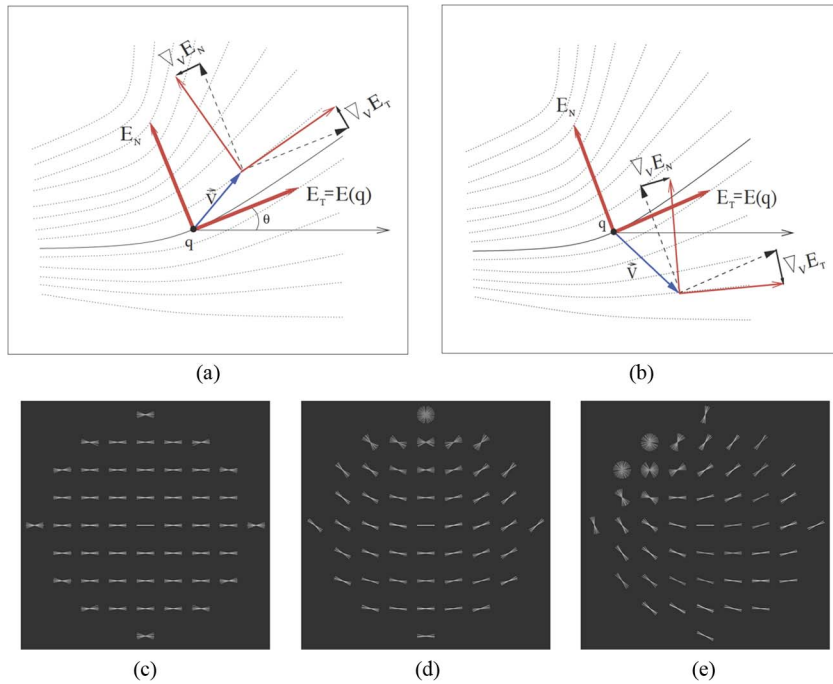


Fig. 8. Cartan connections define the curvature structure and connection patterns in orientation-defined textures. The rotation can now be in any direction in the tangent plane. (a) Displacement in the V -direction yields a rotation of the frame according to the covariant derivative ∇_V . (b) Displacement in another direction V . Note the rotation is different. (c) Excitatory connections between neurons for $\kappa_T = \kappa_N = 0$. (d) $\kappa_T = 0.2, \kappa_N = 0$. (e) $\kappa_T = 0.2, \kappa_N = 0.2$. Figure after [11].

analogous to the osculating circle in cocircularity, and the right helicoid has several natural properties. Letting $\theta(x, y)$ denote the field of orientations around (x_0, y_0)

$$\theta(x, y) = \tan^{-1} \left(\frac{\kappa_{TX} + \kappa_{NY}}{1 + \kappa_{NX} - \kappa_{TY}} \right). \quad (9)$$

The long-range horizontal connections could again implement them as projective fields [9]. [Compare Fig. 8(c)–(e) with Fig. 6(e) and (f).] The importance of the (x, y, θ) representation is further illustrated with nonsimple patterns, such as crossing textures; see Fig. 9.

The helicoid can be generalized to an orientation field in a volume from one in the plane, and has been used for applications such as modeling hair patterns [92]. This illustrates the serendipity that can be achieved with mathematical models: although we began with cortical connections in mind, generalizations have arisen to other anatomical applications. Many of these are triggered by the development of new imaging technologies such as diffusion MRI or diffusion tensor imaging (DTI). This technology is able to image the diffusion of water molecules in biological tissues, such as white matter fibers in the brain. Because many of these fiber tracks cross, regularization must be conducted “along” the fibers and

not between them [73]. The geometry illustrated in Fig. 9 illustrates precisely this.

Another geometrically related application is to the arrangement of myofibers in the heart wall [107]; see Fig. 10. Individual myofibers have the form of helices and shorten in length during contraction. The generalized helicoid model extends from fibers to distributions of fibers, in particular providing optimal volume change without tangling.

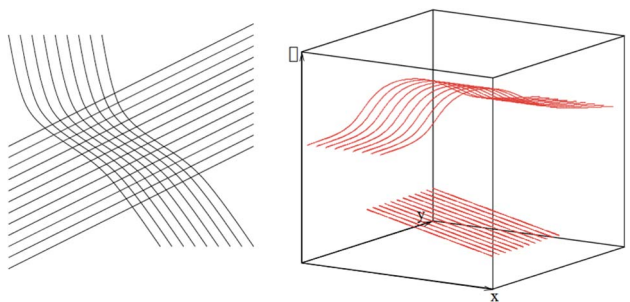


Fig. 9. Crossing textures separate in the (x, y, θ) representation. These are analogous to crossing fiber tracks in brain imaging. For motion analysis in computer vision these are called layered representations [123].

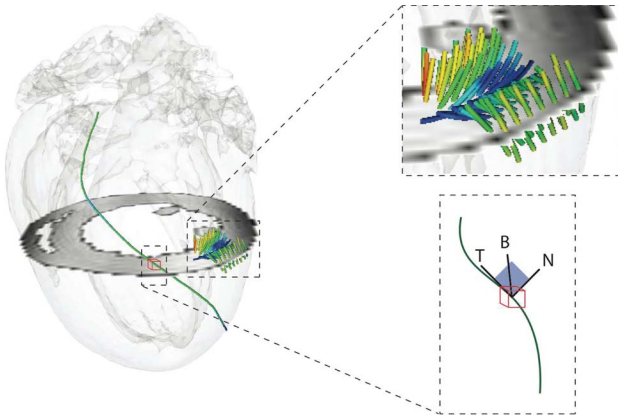


Fig. 10. *The left ventricle in the heart is surrounded by myofibers that provide contractile strength. Each individual fiber follows a helical geometry; the ensemble of fibers is arranged as a generalized helicoid. The instantaneous “angle” of each fiber rotates smoothly as it wraps around, and also varies smoothly across fibers moving from the exterior to the interior of the ventricle wall. Figure adapted from [107].*

III. SURFACES AND INTERMEDIATE-LEVEL VISION

Early (2-D) vision was based on the lift of image properties into (position, orientation column) organizations. Such organizations have natural “good continuation” properties, with curvature relating nearby orientations. We now consider surfaces and 3-D inferences. These naturally involve products of earlier representations.

A. Orientation-Based Stereo Correspondence

Stereo infers depth by integrating the different images striking our eyes. It begins in V1, where cells exist that are selective to positional or phase shifts in Gabor-like receptive fields [97]. This positional disparity is not all of the story, however: it must be integrated over larger distances to yield a consistent depth percept. Evidence of recurrent computation is now appearing [104], [105], in analogy with curve inferences. How might such recurrent computations be structured?

Almost all disparity selective neurons in V1 are also orientationally selective [93]. The second visual area, V2, is also very rich in disparity processing with orientationally selective cells [102]. Therefore, we ask how positional differences and orientation could combine in stereo correspondence.

To understand how Euclidean space and cortical coordinates relate, consider the border of an object as a space curve in 3-D. For image boundaries, we studied good continuation in 2-D; now we will study good continuation in the world. But this is not what is given; it is what is sought. We start with a pair of images, one to the left eye and one to the right which, in visual cortex (following the previous abstraction), amounts to columnar representa-

tions of boundary tangents in the left and right images (Fig. 11). The method for putting them together follows [74]. We move beyond spatial disparity to determine which tangent in the left image goes with which tangent in the right image. This is the correspondence problem.

For 2-D boundaries tangents were transported along cocircular approximations to establish consistency. Orthogonal to the tangent was the normal vector. The situation in 3-D is conceptually the same (Fig. 11), except now the tangent vector is a 3-D vector and the full geometry is captured by transporting a (tangent, normal, binormal) or (T, N, B) frame. Again “curvatures” connect frame components. Torsion, a kind of curvature out of the osculating (T, N) plane, is the second rotation [87].

We now develop tangent correspondence between the left and right images by first considering the forward problem. The (T, N, B) frame at a point along a space curve in 3-D projects to a pair of 2-D (T, N) frames [Fig. 11(d)]. In general, these 2-D frames are different. Their points of attachment in image coordinates will be displaced; this is the spatial disparity. But just as importantly, their angles will be different; this is orientation disparity. All of this structure derives only from the projection of a single frame.

Solving stereo correspondence is an inverse problem: find those pairs of (left, right) tangents, such that the resultant 3-D tangent can be inferred. This inverse problem is inherently ambiguous in the same way that the 2-D curve inference problem was ambiguous, so we solve the 3-D problem in an analogous fashion. Good continuation for 2-D curves came from transporting a tangent via cocircularity and reinforcing those that agreed. In 3-D, a single tangent projects into each of the two image planes. Moving slightly along the 3-D space curve again requires an approximation; in this case, a short piece of a helix generalizes the 2-D osculating circle. Now, considering a second (3-D) tangent slightly further along the space curve from the first one, it will project to another pair of tangents [Fig. 11(e)]. Thus, the stereo problem is solved by determining which tangent pairs, when transported along a helix, match which other pairs. This is how the results in Fig. 11(c) were obtained.

The machinery to implement this computation could be formulated as a set of neural connections, perhaps realized in the V1 → V2 projection, within V2, or in higher areas. A major constraint that derives from this model is that the accuracy at which orientation is represented needs to be sufficient to support orientation disparity estimates; perhaps this explains why the stereo task is relegated to higher visual areas. There exists evidence that such responses are available by V4 [40] and psychophysics supports (at least) colinear facilitation in depth [47]. Moreover, rivalry results when nonmatching-oriented patterns are used [51].

As with 2-D curves, the good continuation approach to solving stereo correspondence for space curves relies on curvatures. Another leap is required when stereo for surfaces is considered (Fig. 12). Now, instead of a tangent

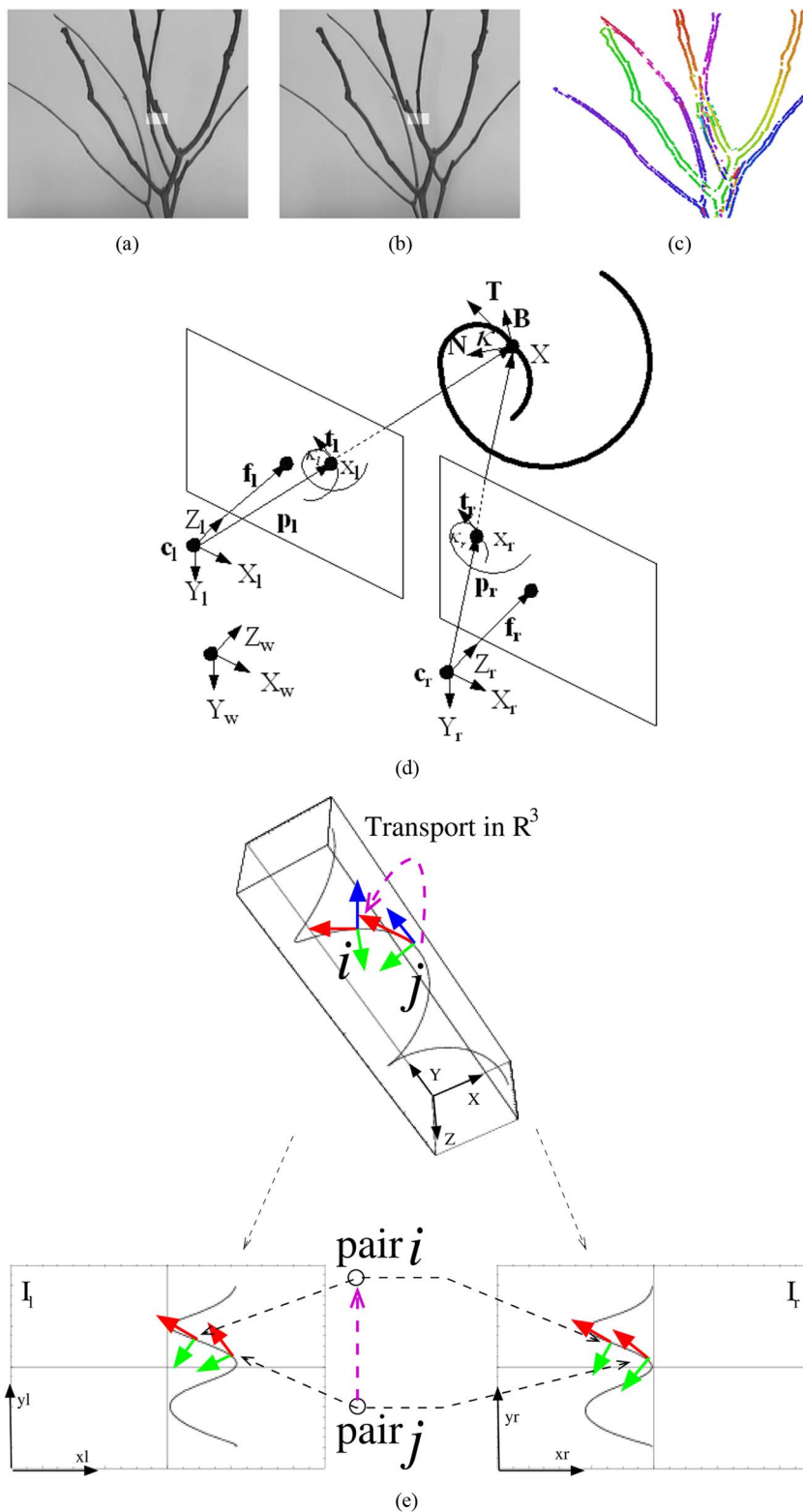


Fig. 11. The stereo correspondence problem for space curves. (a) and (b) A left-right image pair demonstrating that structure may appear in a different ordering when projected into the left and right eyes (highlighted box). (c) Color-coded depth inferred along the tree branches. Note how it varies smoothly along a branch but abruptly between branches. (d) Geometrical setup: the spiral curve in 3-D projects to two image curves. Points along the space curve have (T, N, B) associated frames, while the 2-D curves have (T, N) frames. Notice how a tangent to the space curve projects to a pair of (2-D) tangents, one in the left image and one in the right image. (e) Stereo correspondence between pairs of (left-right) pairs of tangents. Figure after [74].

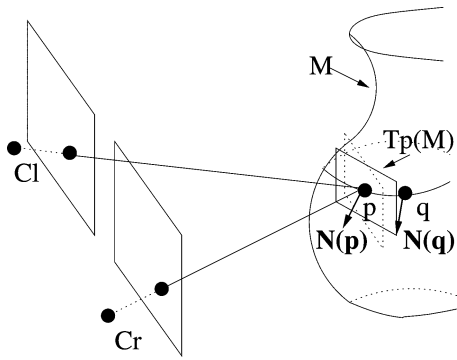


Fig. 12. Stereo for surfaces. The surface normal $N(p)$ varies smoothly and generally differs from nearby normal $N(q)$. Each is orthogonal to the tangent plane (e.g., T_p at that point). Moving an infinitesimal distance along the curve connecting p and q induces a small rotation in the normal (or, equivalently in the tangent plane); this rotation is a type of surface curvature in that direction. Taking all possible directions into account yields the shape operator, another curvature form. Figure after [75].

to a surface, there is a tangent plane, and it rotates depending on the direction in which it is transported. To build intuition, consider slicing an apple: for every direction in which the knife is pointed (the direction of transport) a different cut (surface curve) is made. Each cut defines a curvature, which specifies how the surface normal varies as it is transported in different directions (the shape operator). Details for how to solve the stereo problem for surfaces can be found in [75]. Now, we turn to another way to get surface information: shading analysis.

B. Orientation-Based Shape From Shading

Ernst Mach may have been the first to formulate a shape-from-shading inference problem as a PDE [99], a tradition taken up with enthusiasm in computer vision

[46]. Typically, one seeks a map from image intensities to some representation of the surface (usually surface normals) under a given shading model (usually Lambertian). Various ways to formulate the PDEs [23], [77], [85] or regularization conditions [96] have been proposed.

Ambiguity arises at several levels. Even with a simple Lambertian model, many different surface normals could account for a given image intensity given a light source; and in general there are many possible light sources [67]. Perhaps the most common solution is to place a global prior on the light source [33]; or an assumption on the class of surfaces [72], [91]; or to try to estimate the source, albedo, and shape simultaneously [5], [127]. At the base is a global bas-relief ambiguity. In general, there is a deep sense of frustration around this problem, exacerbated by the fact that we “seem” to be able to do it so easily (although this is in part an illusion [29], [58], [59]).

In seeking ways that our brains could infer shape from shading, we begin not with the image but with how the image would be represented in visual cortex (Fig. 13). Ideally, cells tuned to low spatial frequencies will respond maximally when, e.g., the excitatory receptive field domain is aligned with brighter pixels; the inhibitory domain of an oriented receptive field will then align with the darker regions. These maximally responding cells define the shading flow field in cortical space [17]; it is the tangent map to the image isophotes [57].

Working with the shading flow removes some ambiguity—it is invariant to arbitrary monotonic intensity transformations [56]—and it reduces image noise. But the biologically motivated algorithms with which we have been working suggest a more radical advantage: consider the shading flow as a vector field, or section through the bundle of possible shading flows, and apply the machinery of differential geometry to it. This research program is being carried out now [63]–[65], and we report current progress in it.

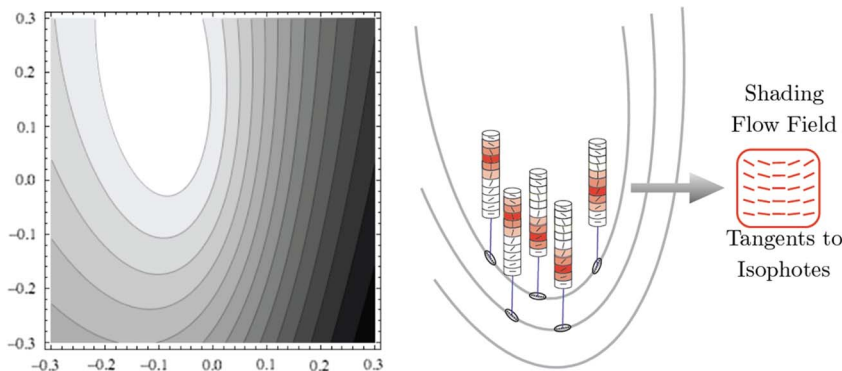


Fig. 13. Representation of shading information in visual cortex. Cells with oriented receptive fields, tuned to low spatial frequencies, will respond optimally when aligned along isophotes, or contours of constant brightness. Activity in (x, y, θ) space is thus the tangent map to these isophotes—the shading flow field. This is analogous to the lift of oriented textures. Figure after [66].

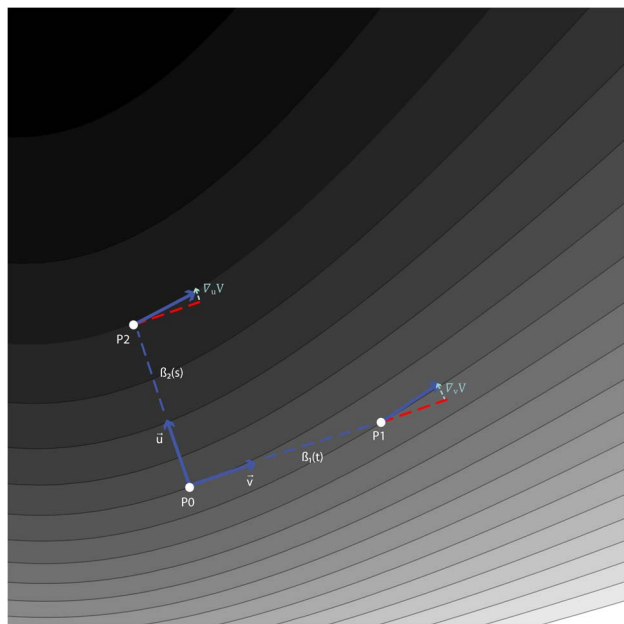


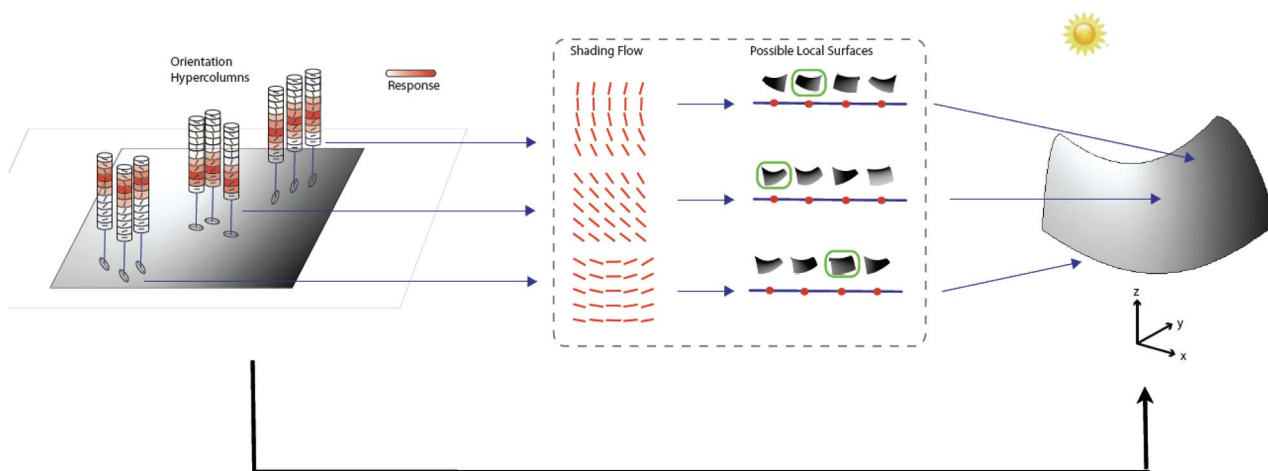
Fig. 14. Geometry of motion through a shading flow field. Moving along the curve $\beta_1(t)$, from P_0 to P_1 in the isophote direction \vec{v} , implies that the flow field $V(x, y)$ changes by $\nabla_{\vec{v}}V$. Moving in direction \vec{u} along $\beta_2(s)$, which is perpendicular (in the image) to the isophote causes the flow field to change by $\nabla_{\vec{u}}V$. These changes can be formally related to the surface curvatures and the light source direction.

Corresponding to the shading flow is an illuminated surface and, generalizing from earlier ideas about transport, the trick is to analyze what happens on the surface as you move through the shading flow field (Fig. 14). Walking

in the direction of a tangent corresponds to walking along an isophote on the surface. According to, e.g., Lambertian reflectance, the tangent plane has to rotate precisely so the brightness remains constant. Or, moving normal to the shading flow says the brightness gradient must be changing in another fashion. Together, these constraints on the flow changes correspond to changes in the surface curvatures and result in a system of differential equations that can be solved in certain circumstances. Apart from bas-relief ambiguity, they reveal a family of possible surface patch/light source combinations for each patch of shading flow. These patches include the classical bas-relief “cup” versus “bump” ambiguity, plus a number of twisted ones [64].

Putting the possible patches together suggests finding a section through a more complex bundle than previously reviewed (Fig. 15). Some boundary conditions are available to select from among these, for example, the manner in which surfaces curve as they approach a boundary [49], [55], but, in general, this is not sufficient to reduce ambiguity to bas-relief.

Having developed the differential equations that allow calculation of surfaces from shading flows permits another type of analysis: one can ask for which features is the ambiguity minimal (Fig. 16). This turns out to be not just around certain boundary conditions but also for ridges and related structures. We conjecture that this is the reason why shading analysis appears to work so well—it is rather nicely defined in certain circumstances—and may clarify why certain boundaries are important in viewing art and drawings [22]. When ambiguity is extensive, almost all reasonable prior assumptions will be questionable, so perhaps shading analysis should not even be attempted.



Standard Computer Vision Formulation

Fig. 15. Inference of shape-from-shading information as a problem in perceptual organization. For each patch of the shading flow field there is a family of possible surfaces; this family is a kind of column of possibilities analogous to the orientation column in early visual cortex. Selecting from among these families according to boundary and interior conditions reveals a surface just as selecting orientations reveals a contour. Figure after [66].

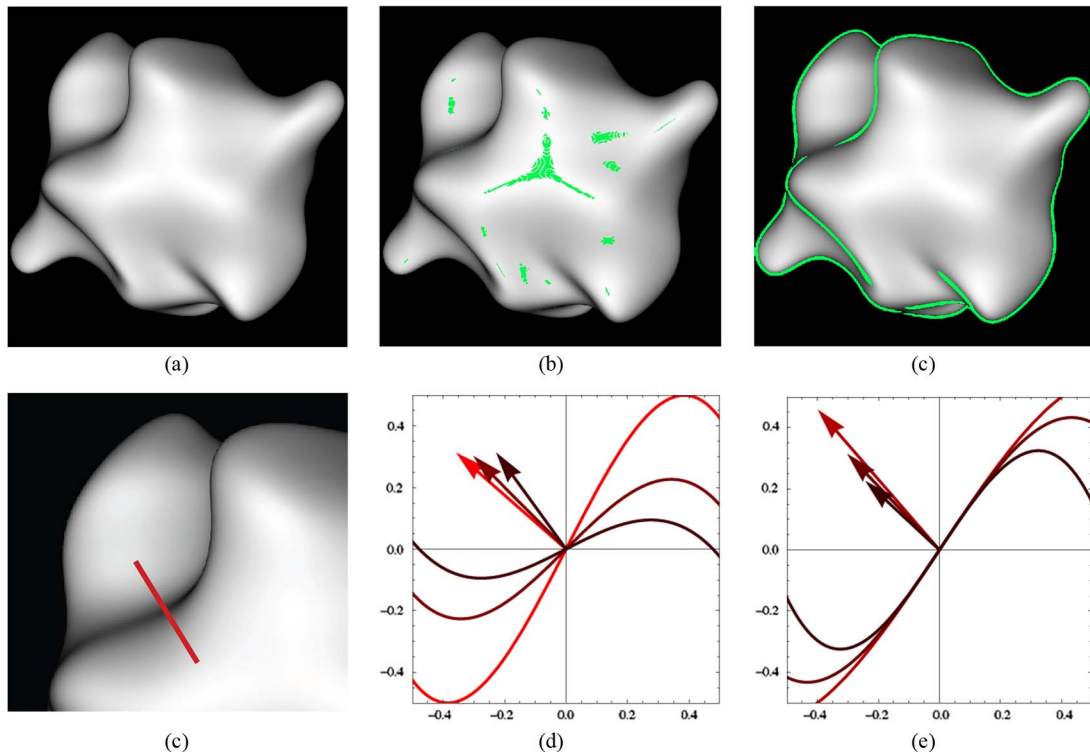


Fig. 16. Shape inferences are highly constrained in certain neighborhoods of a shape, and less at others. (a) A shaded surface is well constrained at (b) highlight points and (c) along boundaries and ridges. (d) Zooming in on a ridge, the red line defines a normal plane. Taking a cross section along this plane shows (e) how the second derivative of intensity along the line constrains possible cross sections. Note that the tangent plane changes. (f) Various cross sections and associated light sources with the tangent plane fixed; the projected light source hardly changes. These two types of transformations characterize the possible cross sections and illustrate how constrained they are. Figure after [65].

Instead, a solution could be interpolated across the ambiguous positions and anchored by minimal ones. This interpolation could be accomplished by the manner in which shape is represented in higher visual areas [90].

C. Orientation-Based Color Processing

While shading inferences were naturally expressed in differential geometric terms, color would seem to be very different. Typically, one thinks of the short–medium–long wavelength retinal cones and the single opponent processing in retinal ganglion cells [Fig. 17(a)]. Such opponency is readily characterized by efficient coding principles [111]. But something rather different emerges when nonlinear dimensionality reduction techniques are used [Fig. 17(c)]. Munsell patches can be viewed as a collection of points in wavelength space. When this is projected by diffusion maps [21] to three coordinates, the intensity–hue–saturation representation emerges [10]. Now, attaching a unit vector to each image position defines a flow in hue. Such flows have arisen in image denoising and in painting applications [117].

How might these hue flows be realized in primate visual cortex? There is a rich representation of color information in the form of oriented double-opponent cells [109], shown in Fig. 17(b). Just as the receptive fields

shown in Fig. 1(b) provided an oriented contrast measurement, one can also characterize oriented color-contrast measurements. These would be Gabor-like filters with (say) red–green subzones rather than dark–light ones. Visual cortex goes one step further, however: double-opponent-oriented receptive fields with red–green-oriented opponency contrasted with green–red-oriented opponency. There are also oriented blue–yellow double-opponent flows. These oriented double-opponent flows relate to the information processing questions that we considered in the Introduction (Fig. 18).

The variation of pigment across the surface of a fruit suggests another type of ambiguity in images even more primitive than those considered in Section III-B: which brightness variations correspond to shading variations and which to material changes. Interpreting pigment variations as shading variations would lead to huge shape errors.

Color and brightness variations are correlated on surfaces, which suggests checking for this [110]. While it can be done locally or at edge points [118], the flow structure is even richer: it exists across surfaces. Following the cue in shading analysis, we seek isohue flows. These are naturally expressed in the red–green/blue–yellow-oriented double-opponent basis [43]; see Fig. 19. Most

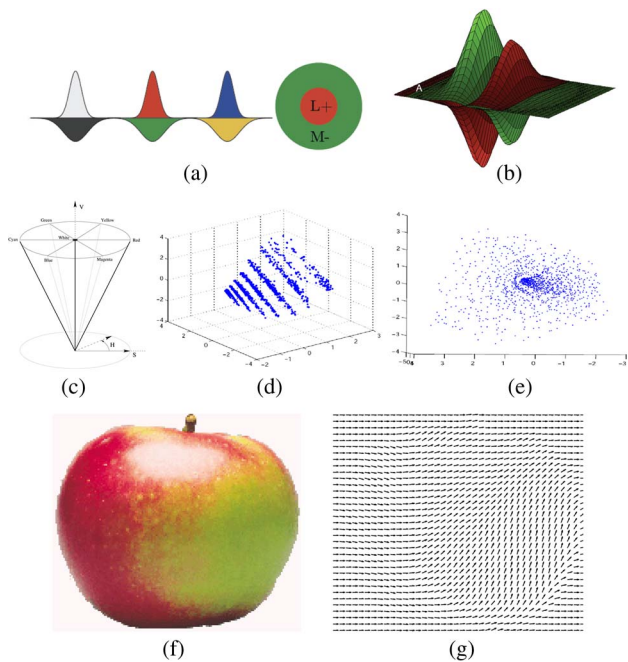


Fig. 17. Representation of color. (a) The retina and lateral geniculate exhibit circular surround receptive fields that are single opponent in brightness, red-green, and blue-yellow. (b) In visual cortex, cells exhibit oriented, double-opponent receptive fields. (c) The intensity-hue-saturation representation, in which hue lies on a circle. Nonlinear embeddings of Munsell patches reveal this representation, shown in (d) side and (e) top views. Natural objects (f) are rich in color variation, as shown in the hue flow (g). Figure (b) after [109]. Figures (c)–(f) after [10].

importantly, when the isophote and the isohue flows are parallel, it means they are covarying over a region; this is highly unlikely to occur naturally unless they have a common source such as pigment variation. On the other hand, when the flows are transverse, it implies that structure is developing differently over a region. In this latter case, the brightness information can be interpreted as shading.

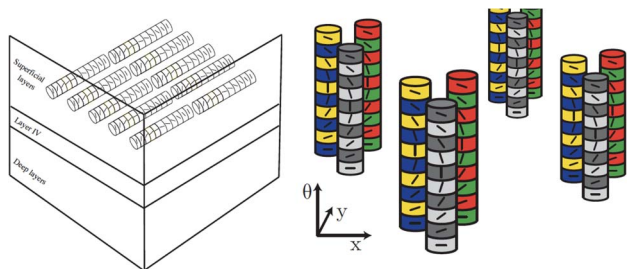


Fig. 18. Color in cortex. (a) The model in Fig. 6 can be generalized to postulate both red-green and blue-yellow double-opponent columns. These provide a natural frame basis for isohue flows.

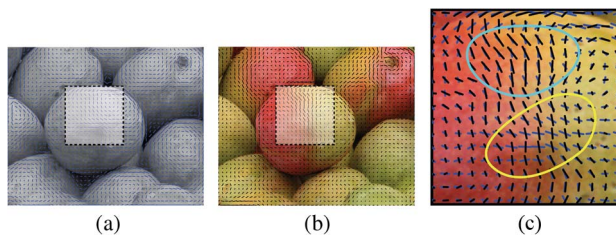


Fig. 19. Image of a mango showing the interaction of hue and brightness. (a) The shading flow and (b) the isohue flow. When compared in the highlighted region, it is clear that in some locations the flows are parallel, indicating a material event, and in others, the flows are transverse, indicating that the brightness variation can be interpreted as shading. Figure after [43].

A psychophysical demonstration illustrates how compelling flow interaction can be (Fig. 20). Two colored versions of a shaded image were created by adding isoluminant color images: one with an isohue flow parallel to the shading flow and the other transverse to the shading flow. In the aligned, parallel case, the depth relief is reduced, even though the brightness distribution remains unchanged. It is as if this specific color pattern masks that depth effect, thus providing a new role for color perception different from shadow and boundary detection.

IV. CONCLUSION

Neural circuitry has inspired generations of biologically motivated computer vision algorithms. Beginning with the identification of receptive fields with edge operators, many of the ingredients of computer vision classes are the same as the ingredients of visual perception classes. While this

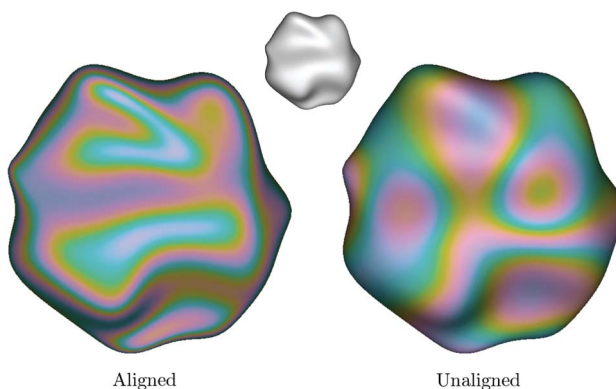


Fig. 20. Combining color and shading information. The gray-level shaded figure has two different isoluminant color images added to it. In the aligned case, the shading flow and the isohue flow are parallel and the depth relief seems to disappear; in the unaligned case, the color information appears “painted” onto the surface. Figure after [43].

shows an intuitive identification of vision algorithms with visual processes, the intuition is difficult to realize with productive systems.

We adopted a much more limited view in this review, based on the centrality of orientation fields in both neural modeling and differential geometry. The analogy was established around the bottom-up boundary detection problem, and developed into stereo, shading, and color. The advantage for stereo was parallel realization of spatial and orientation disparity in computing stereo correspondence. The advantage for shading was the pullback of transport operations on the shading flow to reveal curvature forms on the surface. Finally, the advantage for color was uncovering a role for isohue flows in a primitive discrimination between surface and material changes.

These results are concrete and can be put into practice for computer vision applications. Three general lessons emerged. First, there exists useful high-order structure in the world for which geometry can serve as a proxy. This

was illustrated with edge statistics. Second, understanding constraints between problems can help to make them better posed. This was illustrated by the color-shading interaction. Third, the shading analysis suggests that perhaps one should not seek a full, global solution to a problem, especially when it is very ill-posed. Rather, there may be islands of (almost) well-posed subproblems within them that can serve as anchors for a more general, overall solution. Nailing 3-D structure around boundaries and ridges could be a case in point. Although our percepts seem globally veridical, in fact much of what we perceive is an hallucination. Perhaps this is all that our computational vision algorithms should be asked to accomplish. ■

Acknowledgment

The author would like to thank O. Ben-Shahar, D. Holtmann-Rice, B. Kunsberg, and M. Lawlor for their contributions to this research.

REFERENCES

- [1] C. Rother, A. Blake, and P. Kohli, Eds., *Markov Random Fields for Vision and Image Processing*. Cambridge, MA, USA: MIT Press, 2011.
- [2] Y. Adini, D. Sagi, and M. Tsodyks, "Excitatory-inhibitory network in the visual cortex: Psychophysical evidence," in *Proc. Nat. Acad. Sci. USA*, 1997, vol. 94, pp. 10426–10431.
- [3] A. Angelucci, J. B. Levitt, E. J. S. Walton, J.-M. Hupe, J. Bullier, and J. S. Lund, "Circuits for local and global signal integration in primary visual cortex," *J. Neurosci.*, vol. 22, no. 19, pp. 8633–8646, 2002.
- [4] J. August and S. W. Zucker, "The curve indicator random field: Curve organization via edge correlation," in *Perceptual Organization for Artificial Vision Systems*. New York, NY, USA: Springer-Verlag, 2000, pp. 265–288.
- [5] J. Barron and J. Malik, "Shape, illumination, reflectance from shading," *Tech. Rep.*, 2013.
- [6] H. G. Barrow and J. M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Computer Vision Systems*, A. Hanson and E. Riseman, Eds. New York, NY, USA: Academic, 1978.
- [7] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vis. Res.*, vol. 37, no. 23, pp. 3327–3338, 1997.
- [8] O. Ben-Shahar, "Visual saliency and texture segregation without feature gradient," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 42, pp. 15704–15709, 2006.
- [9] O. Ben-Shahar and S. W. Zucker, "Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex," *Neural Comput.*, vol. 16, pp. 445–476, 2003.
- [10] O. Ben-Shahar and S. W. Zucker, "Hue geometry and horizontal connections," *Neural Netw.*, vol. 17, pp. 753–771, 2004.
- [11] O. Ben-Shahar and S. W. Zucker, "The perceptual organization of texture flow: A contextual inference approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 401–417, Apr. 2003.
- [12] O. Ben-Shahar and S. W. Zucker, "Sensitivity to curvatures in orientation-based texture segmentation," *Vis. Res.*, vol. 44, no. 3, pp. 257–277, 2004.
- [13] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychol. Rev.*, vol. 94, no. 2, pp. 115–147, 1987.
- [14] C. M. Bishop, "Pattern recognition and machine learning," in *Information Science and Statistics*. New York, NY, USA: Springer-Verlag, 2006.
- [15] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick, "Orientation selectivity and the arrangement of horizontal connections in the tree shrew striate cortex," *J. Neurosci.*, vol. 17, no. 6, pp. 2112–2127, Mar. 15, 1997.
- [16] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [17] P. Breton and S. W. Zucker, "Shadows and shading flow fields," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 1996, pp. 782–789.
- [18] N. Caporale and Y. Dan, "Spike timing-dependent plasticity: A Hebbian learning rule," *Annu. Rev. Neurosci.*, vol. 31, pp. 25–46, 2008.
- [19] V. A. Casagrande and J. H. Kaas, "The afferent, intrinsic, efferent connections of primary visual cortex in primates," in *Cerebral Cortex, Primary Visual Cortex in Primates*, A. Peters and K. S. Rockland, Eds. New York, NY, USA: Springer-Verlag, 1994, pp. 201–259.
- [20] V. Caselles, J.-M. Morel, and C. Sbert, "An axiomatic approach to image interpolation," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 376–386, Jul. 1998.
- [21] R. R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part I: Diffusion maps," *Proc. Nat. Acad. Sci. USA*, vol. 21, pp. 7426–7431, May 2005.
- [22] D. DeCarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella, "Suggestive contours for conveying shape," *ACM Trans. Graphics*, vol. 22, no. 3, pp. 848–855, 2003.
- [23] P. Diefel and J. Sylvester, "Some remarks on the shape-from-shading problem in computer vision," *J. Math. Anal. Appl.*, vol. 84, no. 1, pp. 235–248, 1981.
- [24] J. Dorsey, H. Rushmeier, and F. Sillion, *Digital Modeling of Material Appearance*. San Mateo, CA, USA: Morgan Kaufmann, 2008.
- [25] R. J. Douglas and K. A. C. Martin, "Neuronal circuits of the neocortex," *Annu. Rev. Neurosci.*, vol. 27, pp. 419–451, 2004.
- [26] J. H. Elder, "Bridging the dimensional gap: Perceptual organization of contour in two-dimensional shape," in *Oxford Handbook of Perceptual Organization*, J. Wagemans, Ed. Oxford, U.K.: Oxford Univ. Press, 2013.
- [27] J. H. Elder and R. M. Goldberg, "Ecological statistics of gestalt laws for the perceptual organization of contours," *J. Vis.*, vol. 2, no. 4, 2002, DOI: 10.1167/2.4.5.
- [28] J. H. Elder and R. M. Goldberg, "The statistics of natural image contours," in *Proc. IEEE Workshop Perceptual Organisation Comput. Vis.*, 1998. [Online]. Available: http://marathon.csee.usf.edu/~sarkar/pocv_program.html
- [29] R. Erens, A. Kappers, and J. J. Koenderink, "Perception of local shape from shading," *Percept. Psychophys.*, vol. 54, no. 2, pp. 145–156, 1993.
- [30] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
- [31] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cereb. Cortex*, vol. 1, pp. 1–47, 1991.
- [32] D. Field, A. Hayes, and R. Hess, "Contour integration by the human visual system: Evidence for a local association field," *Vis. Res.*, vol. 33, pp. 173–193, 1993.
- [33] W. T. Freeman, "The generic viewpoint assumption in a framework for visual perception," *Nature*, vol. 368, pp. 542–545, 1994.
- [34] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition

- unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.
- [35] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," *Vis. Res.*, vol. 41, no. 6, pp. 711–724, 2001.
- [36] D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," INRIA, Paris, France, Tech. Rep. 2757, Sep. 1996.
- [37] C. D. Gilbert and T. N. Wiesel, "Clustered intrinsic connections in cat visual cortex," *J. Neurosci.*, vol. 3, no. 5, pp. 1116–1133, May 1983.
- [38] P. Halmos, "What does the spectral theorem say?" *Amer. Math. Monthly*, vol. 70, no. 3, pp. 241–247, 1963.
- [39] R. F. Hess, K. A. May, and S. O. Dumoulin, "Contour integration: Psychophysical, neurophysiological and computational perspectives," in *Oxford Handbook of Perceptual Organization*, J. Wagemans, Ed. Oxford, U.K.: Oxford Univ. Press, 2013.
- [40] D. A. Hinkle and C. E. Connor, "Three-dimensional orientation tuning in macaque area v4," *Nature Neurosci.*, vol. 5, no. 7, pp. 665–670, 2002.
- [41] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, pp. 1527–1554, 2006.
- [42] G. E. Hinton and T. J. Sejnowski, "Optimal perceptual inference," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1983, pp. 448–453.
- [43] D. Holtmann-Rice, E. Alexander, R. Fleming, and S. W. Zucker, "When color flows with shading: Making depth disappear," *J. Vis.*, vol. 13, no. 9, 2013, DOI: 10.1167/13.9.467.
- [44] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci. USA*, vol. 81, pp. 3088–3092, 1984.
- [45] B. Horn and M. J. Brooks, *Shape From Shading*. Cambridge, MA, USA: MIT Press, 1989.
- [46] B. K. P. Horn and M. J. Brooks, Eds., *Shape from Shading*. Cambridge, MA, USA: MIT Press, 1989.
- [47] P.-C. Huang, C.-C. Chen, and C. W. Tyler, "Collinear facilitation over space and depth," *J. Vis.*, vol. 12, no. 2, pp. 1–9, 2012.
- [48] D. H. Hubel and T. N. Wiesel, "Functional architecture of macaque monkey visual cortex," *Proc. Roy. Soc. Lond. B*, vol. 198, pp. 1–59, 1977.
- [49] P. Huggins and S. W. Zucker, "Folds and cuts: How shading flows into edges," in *Proc. 8th Int. Conf. Comput. Vis.*, Washington, DC, 2001, pp. 153–158.
- [50] R. Hummel and S. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 3, pp. 267–287, Mar. 1983.
- [51] J. J. Hunt, J. B. Mattingley, and G. J. Goodhill, "Randomly oriented edge arrangements dominate naturalistic arrangements in binocular rivalry," *Vis. Res.*, vol. 64, pp. 49–55, 2012.
- [52] M. Kapadia, M. Ito, C. Gilbert, and G. Westheimer, "Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys," *Neuron*, vol. 15, pp. 843–856, 1995.
- [53] M. Kass and A. Witkin, "Analyzing oriented patterns," *Comput. Vis. Graph. Image Process.*, vol. 37, pp. 362–385, 1987.
- [54] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, pp. 321–331, 1988.
- [55] J. J. Koenderink, *Solid Shape*. Cambridge, MA, USA: MIT Press, 1990.
- [56] J. J. Koenderink and A. J. van Doorn, "Two-plus-one-dimensional differential geometry," *Pattern Recognit. Lett.*, vol. 15, pp. 439–443, 1994.
- [57] J. J. Koenderink and A. J. Van Doorn, "Photometric invariants related to solid shape," *Optica Acta*, vol. 27, no. 7, pp. 981–996, 1980.
- [58] J. J. Koenderink and A. J. Van Doorn, "Shape and Shading," in *The Visual Neurosciences*, L. M. Chappuis and J. S. Werner, Eds. Cambridge, MA, USA: MIT Press, 2004, pp. 1090–1105.
- [59] J. J. Koenderink, A. J. Van Doorn, and A. M. L. Kappers, "Ambiguity and the 'mental eye' in pictorial relief," *Optica Acta*, vol. 30, pp. 431–448, 2001.
- [60] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.
- [61] D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, and M. Mishkin, "The ventral visual pathway: An expanded neural framework for the processing of object quality," *Trends Cogn. Sci.*, vol. 17, no. 1, pp. 26–49, 2013.
- [62] N. Krüger, "Collinearity and parallelism are statistically significant second-order relations of complex cell responses," *Neural Process. Lett.*, vol. 8, no. 2, pp. 117–129, 1998.
- [63] B. Kunsberg and S. W. Zucker, "The differential geometry of shape from shading: Biology reveals curvature structure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 39–46.
- [64] B. Kunsberg and S. W. Zucker, "Characterizing ambiguity in light source invariant shape from shading," 2013. [Online]. Available: <http://arxiv.org/abs/1306.5480>
- [65] B. Kunsberg and S. W. Zucker, "Why shading matters along contours," in *Neurogeometry of Vision*, G. Citti and A. Sarti, Eds. New York, NY, USA: Springer-Verlag, 2014.
- [66] B. Kunsberg and S. W. Zucker, "How shading constraints surface patches without knowledge of light sources," *SIAM J. Imaging Sci.*, 2014.
- [67] M. Langer and S. W. Zucker, "Casting light on illumination: A computational model and dimensional analysis of sources," *Comput. Vis. Image Understand.*, vol. 65, no. 2, pp. 322–335, 1997.
- [68] M. Lawlor, D. Holtmann-Rice, P. Huggins, O. Ben-Shahar, and S. W. Zucker, "Boundaries, shading, border ownership: A cusp at their interaction," *J. Physiol.*, vol. 103, pp. 18–36, 2009.
- [69] M. Lawlor and S. W. Zucker, "Third-order edge statistics: Contour continuation, curvature, cortical connections," *Advances in Neural Information Processing Systems 26*. Cambridge, MA, USA: MIT Press, 2013, pp. 1763–1771.
- [70] Y. Lecun and Y. Bengio, "Convolutional networks for images, speech, time series," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA, USA: MIT Press, 1995.
- [71] T.-S. Lee and D. Mumford, "Hierarchical Bayesian inference in the visual cortex," *J. Opt. Soc. Amer. A*, vol. 20, no. 7, pp. 1434–1447, 2003.
- [72] S. R. Lehky and T. J. Sejnowski, "Network model of shape-from-shading: Neural function arises from both receptive and projective fields," *Nature*, vol. 333, no. 2, pp. 452–454, 1988.
- [73] C. Lenglet, J. S. W. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, D. Wassermann, A. Anwender, R. Deriche, G. B. Pike, G. Sapiro, K. Siddiqi, and P. Thompson, "Mathematical methods for diffusion MRI processing," *NeuroImage*, vol. 45, no. 1, pp. S111–S122, 2009.
- [74] G. Li and S. W. Zucker, "Contour-based binocular stereo: Inferring coherence in stereo tangent space," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 59–75, 2006.
- [75] G. Li and S. W. Zucker, "Differential geometric inference in surface stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 72–86, Jan. 2010.
- [76] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York, NY, USA: Springer-Verlag, 2009.
- [77] P.-L. Lions, E. Rouy, and A. Tourin, "Shape-from-shading, viscosity solutions and edges," *Numer. Math.*, vol. 64, pp. 323–353, 1993.
- [78] A. K. Mackworth, "Consistency in networks of relations," *Artif. Intell.*, vol. 8, pp. 99–118, 1978.
- [79] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, no. 4262, pp. 283–287, 1976.
- [80] D. A. Miller and S. W. Zucker, "Efficient simplex-like methods for equilibria of nonsymmetric analog networks," *Neural Comput.*, vol. 4, pp. 167–190, 1992.
- [81] D. A. Miller and S. W. Zucker, "Computing with self-excitatory cliques: A model and an application to hyperacuity-scale computation in visual cortex," *Neural Comput.*, vol. 11, pp. 21–66, 1999.
- [82] K. D. Miller, "Understanding layer 4 of the cortical circuit: A model based on cat V1," *Cereb. Cortex*, vol. 13, pp. 73–82, 2003.
- [83] U. Montanar, "Networks of constraints: Fundamental properties and applications to picture processing," *Inf. Sci.*, vol. 7, pp. 95–132, 1974.
- [84] H. C. Nothdurft, "Orientation sensitivity and texture segmentation in patterns with different line orientation," *Vis. Res.*, vol. 25, no. 4, pp. 551–560, 1985.
- [85] J. Oliensis, "Uniqueness in shape from shading," *Int. J. Comput. Vis.*, vol. 2, no. 6, pp. 75–104, 1991.
- [86] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
- [87] B. O'Neill, *Elementary Differential Geometry*, 2nd ed. Burlington, MA, USA: Elsevier, 2006.
- [88] P. A. Salin and J. Bullier, "Corticocortical connections in the visual system: Structure and function," *Physiol. Rev.*, vol. 75, pp. 107–154, 1995.
- [89] P. Parent and S. W. Zucker, "Trace inference, curvature consistency, curve detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 8, pp. 823–839, Aug. 1989.

- [90] A. Pasupathy and C. E. Connor, "Population coding of shape in area V4," *Nature Neurosci.*, vol. 5, no. 12, pp. 1332–1338, 2002.
- [91] A. Pentland, "Local shading analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 3, pp. 170–187, Mar. 1984.
- [92] E. Piuze, P. G. Kry, and K. Siddiqi, "Generalized helicoids for modeling hair geometry," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 247–256, 2011.
- [93] G. F. Poggio and B. Fisher, "Binocular interaction and depth sensitivity of striate and pre-striate cortical neurons of the behaving rhesus monkey," *J. Neurophysiol.*, vol. 40, no. 1, pp. 392–405, 1977.
- [94] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, Sep. 1985.
- [95] R. B. Potts, "Some generalized order-disorder transitions," *Proc. Cambridge Philosoph. Soc.*, vol. 48, pp. 106–109, 1952.
- [96] E. Prados and O. Faugeras, "Shape from Shading," in *Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. New York, NY, USA: Springer Science, 2006, pp. 375–388.
- [97] N. Qian, "Binocular disparity and the perception of depth," *Neuron*, vol. 18, pp. 359–368, 1997.
- [98] A. A. Rao and R. C. Jain, "Computerized flow field analysis: Oriented texture fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 7, pp. 693–709, Jul. 1992.
- [99] F. Ratliff, *Mach Bands: Quantitative Studies on Neural Networks in the Retina*. San Francisco, CA, USA: Holden-Day, 1965.
- [100] K. S. Rockland and J. S. Lund, "Widespread periodic intrinsic connections in the tree shrew visual cortex," *Science*, vol. 215, pp. 1532–1534, 1982.
- [101] K. S. Rockland and A. Virga, "Terminal arbors of individual 'feedback' axons projecting from area v2 to V1 in the macaque monkey: A study using immunohistochemistry of anterogradely transported phaseolus vulgaris-leucoagglutinin," *J. Compar. Neurol.*, vol. 285, pp. 54–72, 1989.
- [102] A. W. Roe and D. Y. Ts'ao, "The functional architecture of area v2 in the macaque monkey," in *Extrastriate Cortex in Primates*, vol. 12, K. S. Rockland, J. H. Kaas, and A. Peters, Eds. New York, NY, USA: Plenum, 1997, pp. 295–333.
- [103] A. Rosenfeld, R. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, no. 6, pp. 420–433, Jun. 1976.
- [104] J. Samonds, B. Potetz, and T. S. Lee, "Cooperative and competitive interactions facilitate stereo computations in macaque primary visual cortex," *J. Neurosci.*, vol. 29, no. 50, pp. 15780–15795, 2009.
- [105] J. Samonds, B. Potetz, C. Tyler, and T. S. Lee, "Recurrent connectivity can account for the dynamics of disparity processing in V1," *J. Neurosci.*, vol. 33, no. 7, pp. 2934–2946, 2013.
- [106] A. Sarti, G. Citti, and J. Petitot, "The symplectic structure of the primary visual cortex," *Biol. Cybern.*, vol. 98, no. 1, pp. 33–48, 2008.
- [107] P. Savadjiev, G. J. Strijkers, A. J. Bakermans, E. Piuze, S. W. Zucker, and K. Siddiqi, "Heart wall myofibers are arranged in minimal surfaces to optimize organ function," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 24, Jun. 2012, DOI: 10.1073/pnas.1120785109.
- [108] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, Mar. 2007.
- [109] R. Shapley and M. J. Hawken, "Color in the cortex: Single- and double-opponent cells," *Vis. Res.*, vol. 51, pp. 701–717, 2011.
- [110] S. K. Shevell and F. A. Kingdom, "Color in complex scenes," *Annu. Rev. Psychol.*, vol. 59, pp. 143–166, 2008.
- [111] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, 2001.
- [112] L. C. Sincich and J. C. Horton, "Divided by cytochrome oxidase: A map of the projections from V1 to V2 in macaques," *Science*, vol. 295, pp. 1734–1737, 2002.
- [113] H. Sompolinsky and R. Shapley, "New perspectives on the mechanisms for orientation selectivity," *Curr. Opinion Neurobiol.*, vol. 7, pp. 514–522, 1997.
- [114] K. A. Stevens, "Computation of locally parallel structure," *Biol. Cybern.*, vol. 29, pp. 19–28, 1978.
- [115] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 2010.
- [116] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.
- [117] B. Tang, G. Sapiro, and V. Caselles, "Color image enhancement via chromaticity diffusion," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 701–707, May 2001.
- [118] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1459–1472, Sep. 2005.
- [119] G. Tkačik, J. S. Prentice, J. D. Victor, and V. Balasubramanian, "Local statistics in natural scenes predict the saliency of synthetic textures," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 42, pp. 18 149–18 154, 2010.
- [120] W. E. Vinje and J. L. Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision," *Science*, vol. 287, no. 5456, pp. 1273–1276, 2000.
- [121] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt, "A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization," *Psychol. Bull.*, vol. 138, no. 6, pp. 1172–1217, 2012.
- [122] C. Wang, N. Komodakis, and N. Paragios, "Markov random field modeling, inference and learning in computer vision and image understanding: A survey," *Comput. Vis. Image Understand.*, vol. 117, pp. 1610–1627, 2013.
- [123] J. Y. A. Wang and E. H. Adelson, "Layered representation for motion analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1993, pp. 361–366.
- [124] Y. Weiss, "Deriving intrinsic images from image sequences," in *Proc. Int. Conf. Comput. Vis.*, 2001, pp. 68–75.
- [125] M. Wertheimer, "Untersuchungen zur lehre von der gestalt II," *Psychologische Forschung*, vol. 4, pp. 301–350, 1923.
- [126] B. Z. Yao, X. Yang, L. Lin, M. W. Lee, and S.-C. Zhu, "12T: Image parsing to text description," *Proc. IEEE*, vol. 98, no. 8, pp. 1485–1508, Aug. 2010.
- [127] Q. Zheng and R. Chappella, "Estimation of illuminant direction, albedo, shape from shading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 7, pp. 680–702, Jul. 1991.
- [128] S. W. Zucker, A. Dobbins, and L. Iverson, "Two stages of curve detection suggest two styles of visual computation," *Neural Comput.*, vol. 1, pp. 68–81, 1989.

ABOUT THE AUTHOR

Steven W. Zucker (Fellow, IEEE) received the B.Eng. degree in electrical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1969 and the M.Eng. and Ph.D. degrees in biomedical engineering from Drexel University, Philadelphia, PA, USA, in 1972 and 1975, respectively.

He is the David and Lucile Packard Professor at Yale University, New Haven, CT, USA. He is Professor of Computer Science, Biomedical Engineering, and a member of the Program in Applied Mathematics and the Interdisciplinary Neuroscience Program. Before moving to Yale in 1996, he was Professor of Electrical Engineering at McGill



University, Montreal, QC, Canada, and Director of the Program in Artificial Intelligence and Robotics of the Canadian Institute for Advanced Research, Toronto, ON, Canada. He was a Postdoctoral Research Fellow in Computer Science at the University of Maryland, College Park, MD, USA, and an SERC Fellow at the Isaac Newton Institute for Mathematical Sciences, University of Cambridge, Cambridge, U.K. He has authored or coauthored more than 300 papers on computational vision, biological perception, artificial intelligence, robotics, and (most recently) computational biology.

Dr. Zucker was elected a Fellow of the Canadian Institute for Advanced Research and a Fellow of the Churchill College, Cambridge, U.K. He won the Siemens Award, a number of Best Paper prizes, and most recently was named a Distinguished Investigator by the Paul Allen Foundation.