Endowing AI with vision: A biological and computational perspective

Steven W. Zucker

McGill University, Research Center for Intelligent Machines, 3480 University St., Montreal, Canada H3A 2A7

Evolutionary biology indicates that vision is a key sense for obtaining information about our physical environment. Since Artificial Intelligence is rapidly moving to include the function of automata that are physically situated in the world — robots — the importance of a computational understanding of vision becomes even more central. However, visually-mediated automata have been less than totally successful. In this article we review selected research in computational vision, and highlight certain of the difficulties experienced in classical approaches. Certain new approaches to vision are then described, which open intriguing biological and mathematical connections.

Introduction

One of the most striking messages from evolutionary biology is that vision is a key sense for obtaining information about the physical environment. As new species evolved, increasingly larger fractions of the available neural architecture have become specialized for vision. In contrast, as research in Artificial Intelligence evolved, vision has become more encapsulated and removed from the central focus. Many began to view it as a peripheral activity, appropriate for applications, but employing techniques different from the rest of AI. This difference was exacerbated, because vision seems so effortless and immediate to us, while cognitive and linguistic activities often seem so difficult and consciously demanding. Now the situation is changing, however. It is becoming clear that the encapsulation was actually around classical AI, which had cut itself off from the physical world. AI is now moving rapidly to change this, by considering the function of situated automata: robots. Hence the importance of a computational understanding of vision is becoming much more central. In this article we review aspects of computational vision, with an aim being to diagnose several of its strengths and its weaknesses, and to indicate a few current research trends. We focus primarily on early vision, with a bias toward biological influences. Other reviews, which focus on the different stages through which computational vision has evolved can be found in ref. 1.

Early vision and high-level vision

Images are created by the projection of photometric properties of objects in the world onto an array of sensors. The task of vision can be summarized as inverting this; i.e. as that of inferring the three-dimensional structure of the physical environment from the two-dimensional intensity structure in images. This is clearly an immensely complex task, whose apparently effortless solution in primates involves more than half of our brains[2]! Thus the presupposition that vision is easy and immediate must be supplanted by a more measured evaluation[3]. One way to manage complexity is decomposition, and the most prominent one in vision separates early processing, or the extraction of object outlines, from high-level processing, or the recognition of objects. For the next few sections in this article we shall focus on these two components, but will indicate others toward the end.

Edge detection in early vision

Imagine a dark cube against a white background. The task of early vision is to abstract a description of this cube sufficiently rich to enable its recognition, while segmenting it (as a figure) from the background. Such a description must certainly involve the bounding contour around this cube, and it is the task of boundary detection to recover this contour. Complexity issues arise immediately, however, because the cube may subtend a large visual angle covering an enormous number of pixels. If processing could be done locally and in parallel, then these pixels could be rapidly processed. The potential of parallelism is strengthed, moreover, because the image contains a distinguishing signature in the neighborhood of a boundary of the cube: a dark image region separated from a light image region by a line. Thus the classical approach to boundary detection is parallel, local edge detection followed by a grouping process to join the edge elements together. Differentiation, which accentuates such dark/light adjacent differences, provides the mathematical basis for edge detection, and it is implemented in parallel over local image windows.

Various image filtering techniques have been used to design local operators, e.g. ref. 4.

This parallelism is prevalent in neurobiology as well, which has provided important conceptual support for computer vision. In particular, patterns of light projected onto the retina influence the activity of cells in the visual system, either in an excitatory manner, leading to an increase in that cell's firing rate, or an inhibitory one, leading to a decrease (possibly below a 'resting' or spontaneous level). The resultant map of activity as a function of light distribution is called a receptive field, and in visual cortex several varieties are found[5]. *Simple cells* (see Figure 1) are those cells whose receptive fields most resemble line and edge detectors in computer vision; they are orientationally selective, as well as selective for a number of other properties including stimulus contrast, direction of motion, and stereo disparity. *Complex cells* resemble simple cells, but their receptive fields lack the distinctive sub-field structure, and are typically larger. Finally, *endstopped*, or hypercomplex cells, appear in both simple and complex varieties, and have additional endzones which inhibit their response when lines or edges extend into them. All types appear at a range of different sizes, or are optimally selective to different spatial frequencies, or to bars of different widths.

Viewed in the large, the architecture of visual cortex appears ideally suited as a boundary detection machine. In computer vision, local edge detection is typically accomplished in two steps: (i) the convolution of an operator against the image and (ii) some process for interpretation of the operator's responses. Or stated in more general terms, the steps consist of (i) a measurement process followed by (ii) a detection process which locates those positions at which the first derivative (spatial gradient) is high; or where the second derivative crosses zero. As with edge operators, simple cells have been modeled as linear operators followed by a squaring non-linearity, and endstopped cells are thought to signal boundary endings. Furthermore, this neural wetware is arranged into *orientation hypercolumns*, so that each edge orientation can be checked for at each position.

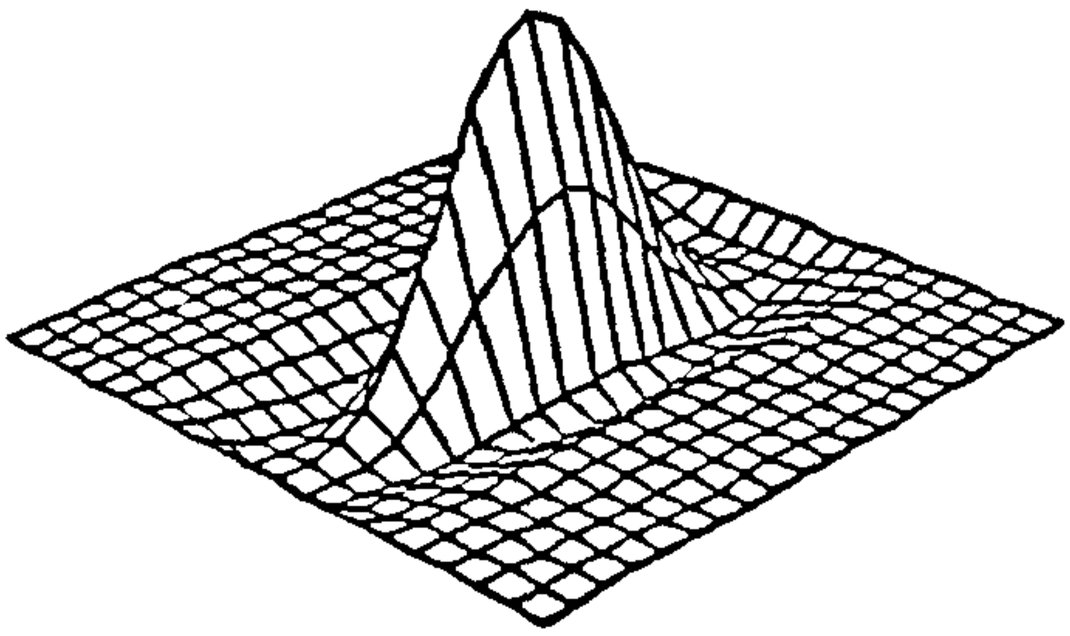


Figure 1. A model of a simple cell receptive field of the sort that could be found in primate visual cortex. Note the elongated central 'excitatory' region, surrounded by inhibitory side-bands. Viewed as an operator for computer vision, such structures are known as 'line detectors'.

The commitment to a local approach to edge detection has been powerful, given this confluence of ideas from computation and biology. Indeed, edge detectors would work perfectly at the sides of the ideal cube discussed above; however, when the responses to realistic images are examined, the quality of the results is disappointing; (see Figure 2). Such a poor front-end response has been one of the most frustrating aspects of trying to build computer vision systems, and has greatly limited their utility.

To deal with this frustration, researchers have turned to both mathematics and to neurobiology. Canny[6], for example, developed an edge operator based on optimality principles, while Marr and Hildreth[7] were motivated by the neurobiological observation that receptive fields span a range of sizes. They postulated a coincidence principle for responses across size[8]. However, neither of these approaches identified the real problems. Computationally, although the edge of the cube above can be modeled as a step change in intensity plus additive noise, realistic edge profiles do not match this[9]. Nevertheless, Canny's mathematics still follows this assumption. Neurobiologically, receptive fields modeled as derivates of Gaussians must smooth around corners. They thus destroy a feature essential for identifying the cube, and for separating objects standing in occlusion relationships to one another (observe that the point of intersection of bounding contours from separate objects must exhibit a 'corner', or a singularity in orientation). This is a fundamental shortcoming of the Marr–Hildreth approach, and of regularization-based approaches as well[10]. Furthermore, the biology is much more complicated. For example, there are different percentages of endstopping, ranging from fully endstopped cells to those exhibiting no endstopping at all. Curiously, most cells are endstopped to an intermediate extent; what function could this partial endstopping subserve? It seems unlikely that partial end-of-line detectors exist. Finally, there is an aspect of boundary inference that exceeds the physical stimulus, and is completely abstract[11] (see Figure 3).

To summarize, we can identify the following problems with the classical approach to edge detection:

- *Assumption of linearity.* Edge operators are restricted to a local measurement; this local measurement is taken to be linear for either ease of implementation, analysis, or because the biology appears to perform that way. However, such local linear measurements 'blur' nearby structure to-

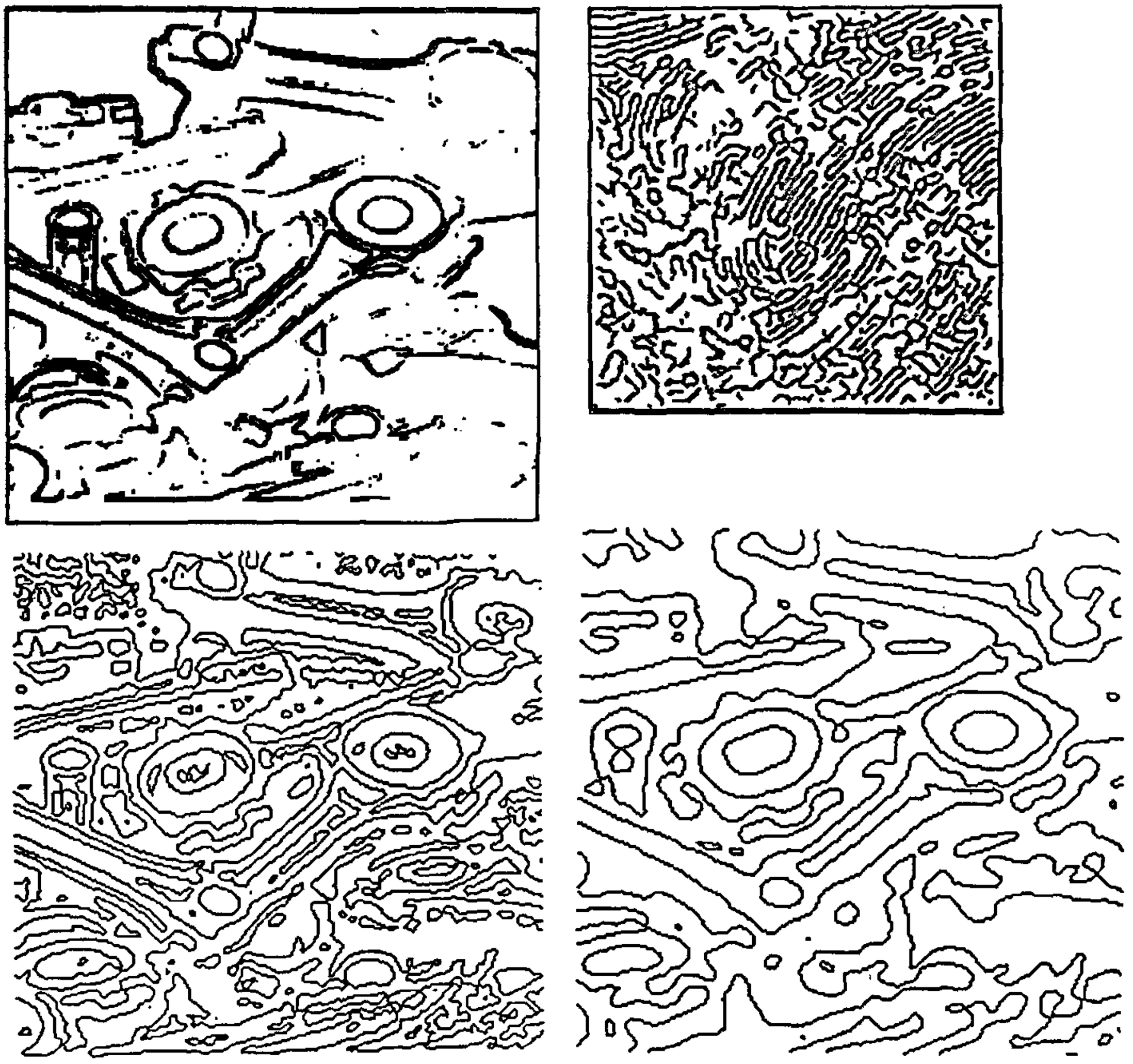


Figure 2. Illustration of classical edge detection. The upper left figure shows maximal responses from a Sobel edge detector[38] to an image of automobile engine components. Observe how the outline is broken where it should indicate a solid, bounding contour, but connected between physically different objects. The upper right figure is the response of the Canny operator to an image of a fingerprint. Note how the dense swirls are randomly merged together. The lower figures illustrate the Marr–Hildreth operator evaluated over the auto parts image at two scales. Their response combines the shortcomings of both the Sobel and the Canny operators.

gether. This blurring not only merges distinct objects, but smooths around corners as well. Such corners are key to segmentation.

- *Assumption of a single value at each position.* The interpretation of local operators as signaling the edges of objects is heuristic; mathematically, local edge detectors would be interpreted as signaling the tangent to the boundary curve. As such, the tangent is well defined, and has a single value, precisely at those positions for which the curve is regular, or smooth. The tangent is undefined at corners! Thus any operation designed to return a single edge (tangent) orientation at a position must fail at singular ones.

The local to global transition

Local edge detector responses must somehow be glued together into global contours. In his early system,

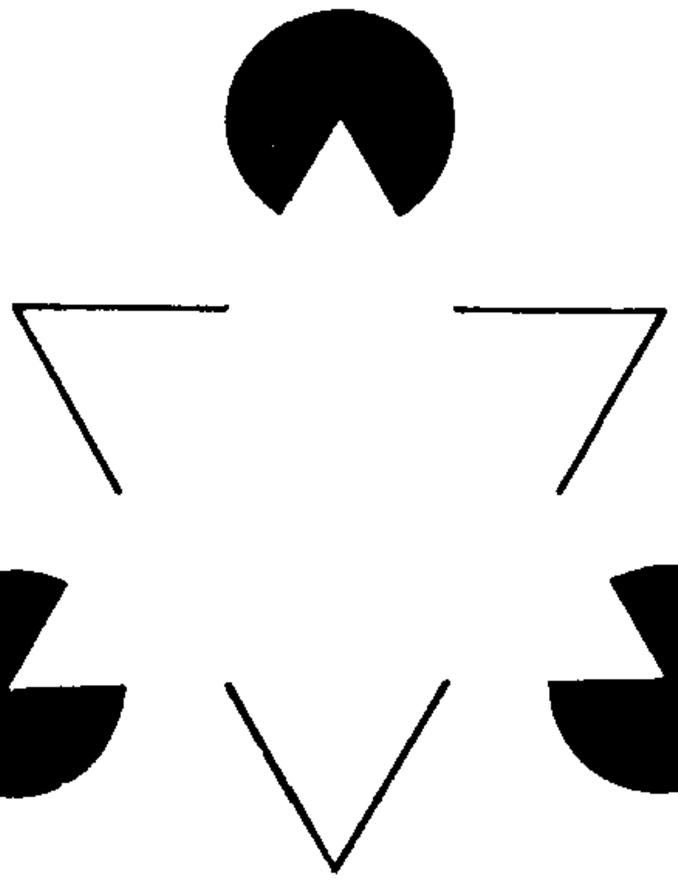


Figure 3. The Kanisza subjective edge illustrates that boundary detection is an inferential process that builds upon the information in the image, rather than being limited to it.

Roberts[12] fit long straight lines to edge detector outputs, because he was working in a blocks world of objects such as cubes, and more modern investigators are still fitting lines[13]. However, the general problem of grouping is much more subtle. The Gestalt psychologists (e.g. ref. 14) postulated a series of grouping principles, such as the principle of good continuation for curves, which suggests why we perceive a figure '8' as a single, non-simple curve that crosses itself, rather than two 'circles', one on top of the other. The Kanisa subjective edge further illustrates the constructive aspect of the process.

One view of grouping is that it is a noise problem. Since there are bogus responses from the local detectors, a global estimation procedure is necessary to eliminate them[15,16]. Another is that it is simply an image-domain phenomenon, linked to scale[17]. Since larger operators have more image support, they should be less susceptible to local variations. However, they are also more likely to average across images of different objects. Thus, we question both of these assumptions.

- *Assumption that grouping is a noise reduction operation.* The refinement and grouping of local operator responses is a geometric problem, not solely a noise estimation problem. Therefore, geometric techniques must be used in refining them. Grouping problems cannot be solved by simply looking across scale in the image (cf. Figure 2, bottom).

- *Assumption that scale is an image-domain phenomenon.* The scale of events in the world is an object-domain issue; the amount of smoothing for noise elimination is an image-domain operation.

We will show how this latter notion of scale arises in the shape portion of the paper, and how image-domain delicacies are best handled by the introduction of certain non-linearities.

A recent approach to boundary detection

We begin with the observation that a wide gulf exists between the initial, broadly-tuned measurements of edge operators and global curves, and we propose an intermediate structure — the discrete tangent field — to fill this gulf. The result is a computational solution to curve detection using it, which is biologically plausible. The solution suggests that there are two different styles of computation involved: in the first stage, hypotheses are represented explicitly and coarsely in a fixed, preconfigured architecture, while in the second hypotheses they are represented implicitly and more finely in a dynamically-constructed architecture. As we proceed, it should become clear how these different representations mediate the complexity issues raised in the Introduction.

The representation chosen for the first stage is intimately connected to the biological notion of hypercolumns, and provides an alternative solution to the corner detection problem. The standard approach is to design specialized 'corner detectors', but then an agent must be postulated to decide whether a corner or a point of high curvature is present. Another approach is to assume that large values of a variational parameter (e.g. bending energy) signal discontinuities, except this approach assumes (i) there is no difference between high curvatures and discontinuities; and (ii) orientation can be measured accurately enough to locate large values in its derivative. This second point indicates the chicken-and-egg nature of the problem: since it is necessary to know where the discontinuities are before orientation can be estimated accurately, how can estimates of orientation be used to locate discontinuities! We have been exploring an alternate approach to discontinuities, in which they are represented by multiple values of orientation at a given point[18]. Along a curve, for example, this amounts to taking the limit in both directions into the discontinuity. The connection to hypercolumns is now clear: they provide the substrate for representing multiple values of orientation at a single point. Mathematically such techniques are related to the Zariski tangent space in algebraic geometry[19].

The first stage of our model incorporates endstopped neurons, but, as we show below, they are used to infer curvatures, not line endings. This explains the intermediate values of endstopping that have been observed

by physiologists, and also suggests novel numerical techniques for measuring curvature. We also make use of an iterative network to enforce differential geometric constraints, what we call curvature consistency, via co-circularity[20]. The co-circularity relationships mediate interactions between orientation hypercolumns. The final result is a reliable but coarse description of local differential properties of curves (i.e. their discrete trace, tangent, and curvatures).

The second stage of the algorithm synthesizes the global curves through the tangent field. The idea behind our approach is to recover the global curve by computing a *covering* of it; i.e. a set of objects whose union is equivalent to the original curve, rather than attempting to compute the global curve directly. The elements of the covering are unit-length dynamic splines, and global curves are recovered to sub-pixel accuracy. The recovery of this covering is mediated through the construction of a potential distribution, and it is in the construction of this potential that the local-to-global transition is effected. An overview of the two stages of curve detection is provided next, beginning with two preliminary notions of non-linearities in local operators and coarse but stable measurements of curvature. See also Figure 4.

Logical/linear operators. The examples of edge detection shown above indicate that nearby image structure can interact to obscure the proper outline. This inappropriate interaction is mediated by the linear summation within the operator's support. For curve detection, we have developed a set of non-linearities that significantly improve the sensitivity of initial operators over (optimal) linear ones[21]. These non-linearities implement a test on continuity of support along the preferred direction of the operator, and a test on variation across it; see Figure 5. The non-linearities are formulated within a logic that accumulates positively consistent evidence linearly, but in which incompatible evidence enters nonlinearly. Thus the operators appear linear for one class of stimuli but markedly nonlinear for others—we call these logical–linear operators. They exhibit dual advantages: they are considerably more stimulus-specific than purely linear operators, while more robust to incidental stimulus variation than logical operators. Their improved performance as edge detectors is illustrated in Figure 6. As models of visual cortical neurons (e.g. simple cells) they are consistent with the well-known 'linear' properties (e.g. sensitivity to spatial frequency gratings) while exhibiting the nonlinear behaviour associated with high vernier sensitivities, strong suppressive effects for opposite contrast segments, and for cross-orientation inhibition.

Endstopping and curvature estimation. Curvature can

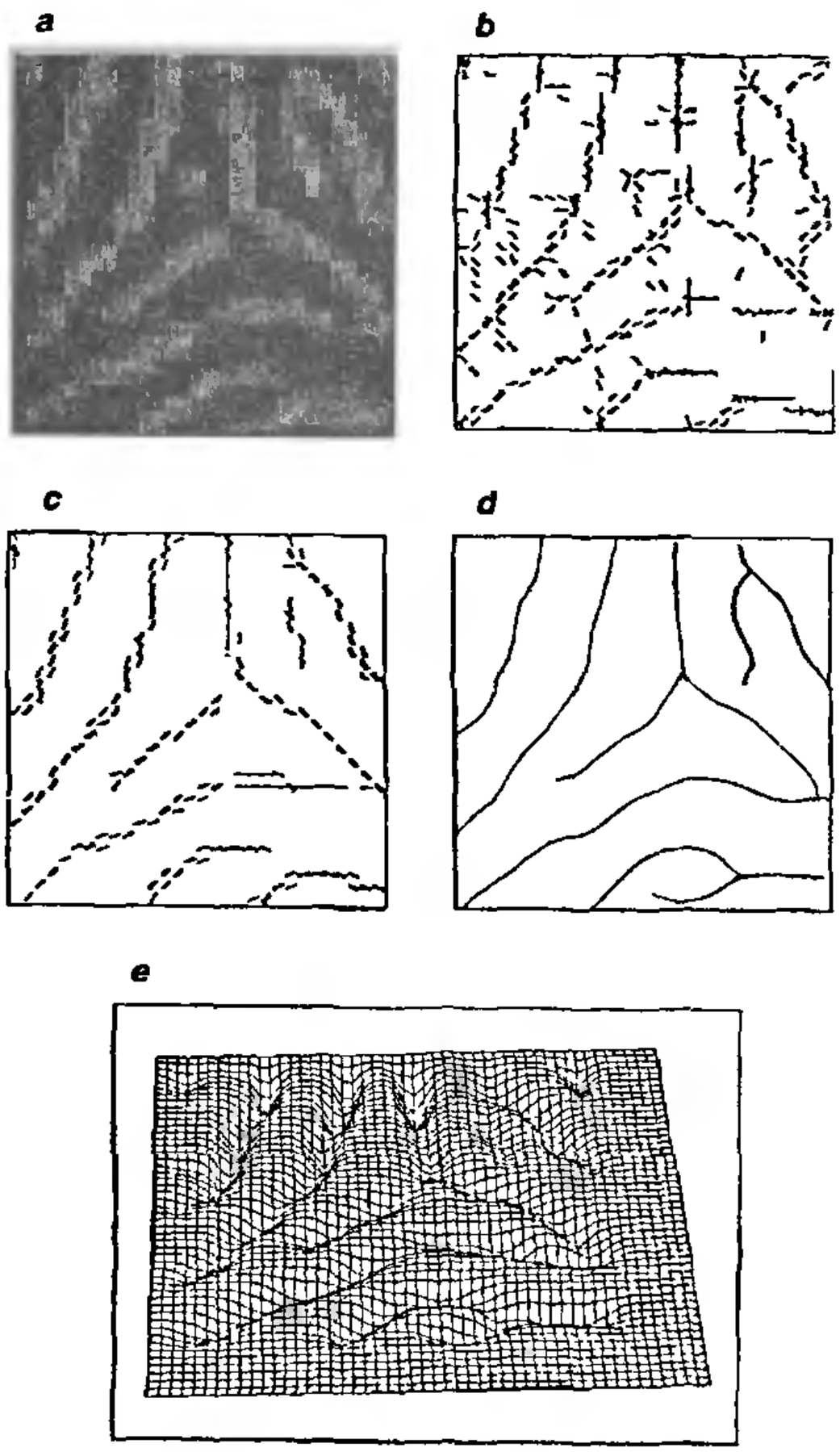


Figure 4. An illustration of the different stages of curve detection. *a*, A small fingerprint image. The first stage is broken into two steps. In the first step, initial measurements are performed to estimate the local curvature and orientation. *b*, In the second step these initial measurements are refined by a relaxation labeling process and *c*, shows the final tangent field (2 iterations). There are also two steps in the second stage, first *e*, the construction of a potential distribution from the entries in the tangent field, and second, *d*, the covering of the global curve by a family of short curves, or snakes.

be thought of as a deviation from straightness. In ref. 22 we develop a computational model for endstopping, and show how it amounts to a 'non-linear difference' between simple cell responses. We now illustrate how this can provide the basis for curvature measurements.

The model is based on the observation that there are simple cells whose receptive field size differs as a function of cortical layer. In particular, Layer VI of cat primary visual cortex contains cells with receptive fields notably longer than those in the laminae above it[23]. Now, given the response of a short simple cell, and the response of a long simple cell, with receptive fields

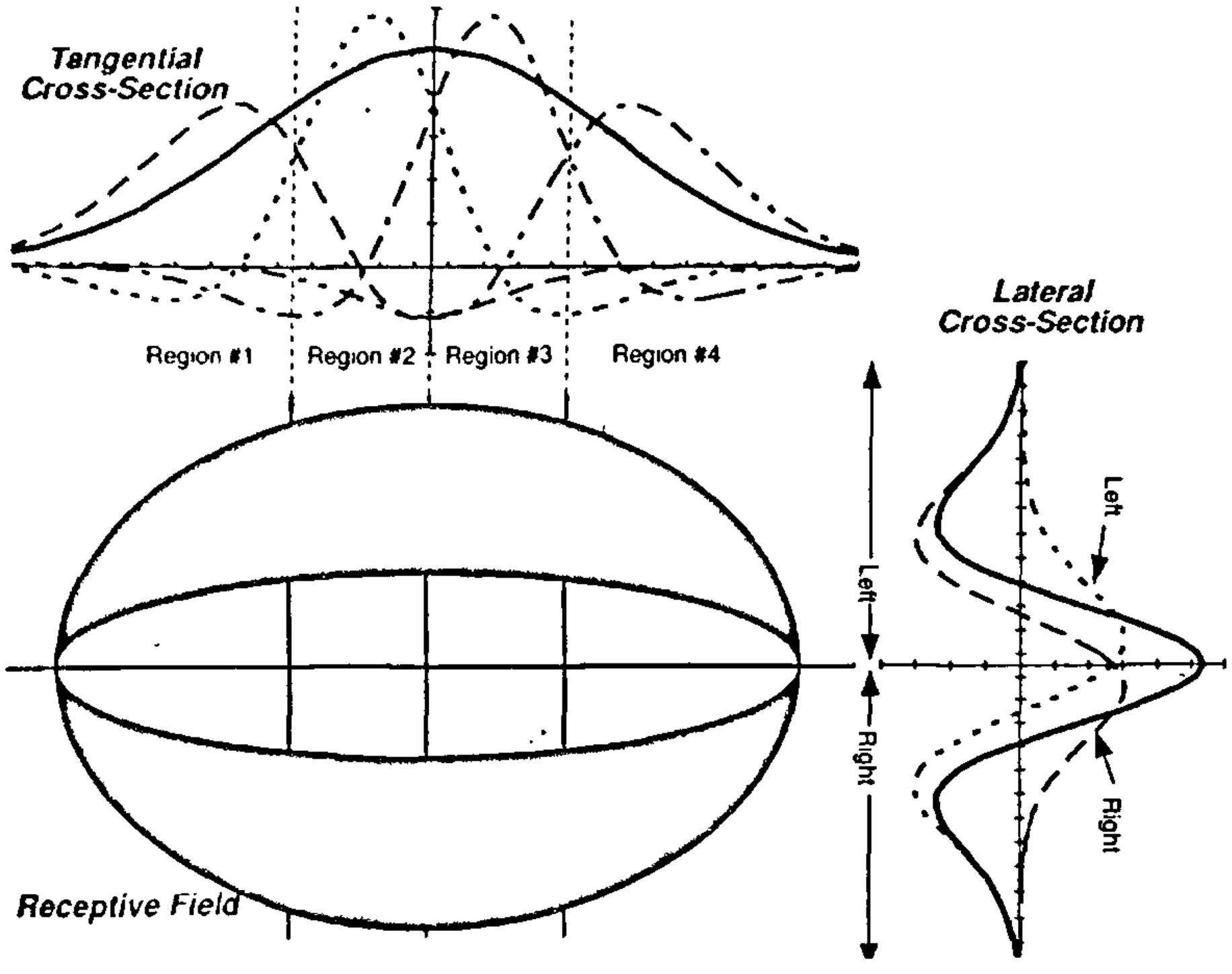


Figure 5. The subunit structure of logical/linear operators. The subunits in the normal direction evaluate contrast differences, enforcing the requirement that a bright line, for example, must have a positive contrast relative to the flank on either side. Continuity of contrast is enforced in the tangential direction. If all of the above tests are passed, then the operator appears to have a linear structure; hence we refer to this structure as hidden non-linearities.

Figure 6. The response of the logical/linear operator to the fingerprint image. Note how the curves are more appropriately represented, and compare it with the Canny response in Figure 2.

sharing orientation preference and centered at the same retinal location, then their 'difference' models the endstop property (the response stops for stimuli exceeding some length). But more importantly, the response of such an operator varies systematically with curvature. Biologically, the model has provided quantitative predictions (now verified) about the response of endstopped simple neurons to curved stimuli.

Stage 1: Inferring the tangent field

With this background, we can now sketch the first stage of our system for curve detection. The goal is to infer the trace of the curve, or the set of points (in the image) through which the curve passes, its (approximate) tangent and curvature at those points, and their discontinuities[24]. This is the *tangent field*, and note that, since the initial measurements are discrete, this will impose constraints on the (inferred) tangents, curvatures, and discontinuities[20].

This first stage of orientation selection is in turn modeled as a two-step process:

- Step 1. *Initial measurement* of the local fit at each point to estimate orientation and curvature. These

estimates derive from a model of simple cell receptive fields instantiated at multiple scales and orientations at each image position. Non-linear differences between orientation estimated over different extents provide the curvature estimate. We thus propose that endstopped neurons in the visual cortex represent joint hypotheses about orientation and curvature, and that their firing rate along an (endstopped) orientation hypercolumn represents how well these hypotheses match the local image structure. This is our representation for the first stage and, in our system at each of eight orientations a small number of simple cell instances are combined to define five discrete curvature classes—two on either side of the zero curvature class. The four curved classes are obtained from endstopped instances and the zero curvature estimate from a nonendstopped simple instance. Typical results above noise are shown in Figure 4*b*; although they convey a rough idea of what the curve structure is, there are both responses where there is no curve, and ambiguous (multi-valued) responses where there is a single curve. We contend that no local operator can solve these problems in general, and further that a spatially-interactive process can. Thus we require

- Step 2. *Interpretation* into an explicit distributed representation of tangent and curvature by establishing consistency between the local measurements. Consider an arc of a curve, and observe that tangents to this arc must conform to certain position and orientation constraints for a given amount of curvature; we refer to such constraints geometrically as *co-circularity* (Figure 7). Discretizing all continuous curves in the world that project into the columnar space of coarse (orientation, curvature) hypotheses partitions these curves into equivalence classes[20]. Interpreting the (orientation, curvature) hypotheses as endstopped neurons, such co-circularly-consistent relaltionships are expected between endstopped neurons in nearby orientation hypercolumns given such a curve as stimulus.

Such inter-columnar interactions can be viewed physiologically as excitatory and inhibitory projections between endstopped cells at nearby positions (adjacent hypercolumns), and can be used as follows. Since curvature is a relationship between tangents at nearby positions, two tangents should support one another if and only if they agree under a curvature hypothesis, and co-circularity provides the measure of such support. In addition, two tangents that disagree with the curvature estimate should detract support from one another. Relaxation labeling provides a formal mechanism for defining such support, and for specifying how to use it[25]. Mathematically it amounts to gradient descent; computationally it is a generalization of Hopfield-like neural networks[26], and physiologically it can be viewed as the computation implemented by pyramidal neurons as they combine information from adjacent (endstopped) orientation hypercolumns. Since only 2–3 iterations are required for convergence (empirically), it is natural to propose that these are accomplished by the forward- and back-projecting pyramidal neurons within visual area V1 and connecting areas V1 and V2[18].

Stage 2: Inferring a covering of the curve

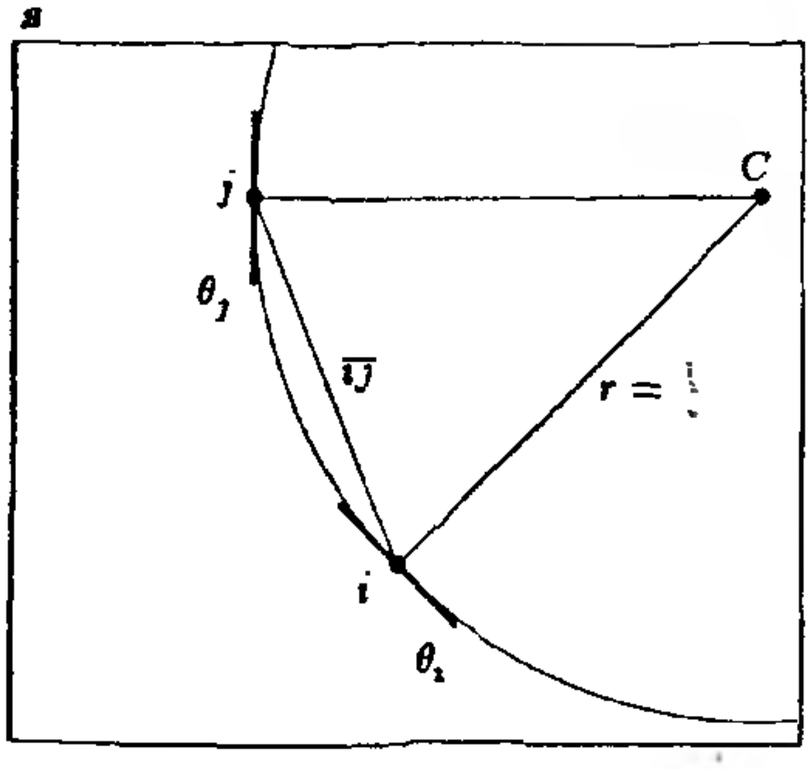
Since the tangent is the first derivative of a curve (with respect to arc length), the global curve can be recovered as an integral through the tangent field. Such a view typically leads to sequential recovery algorithms, as in Newton's method in numerical analysis. But these algorithms require global parameters, starting points, and some amount of topological structure (i.e. which tangent point follows which); in short, they are biologically implausible. In contrast, we propose an approach in which a collection of short, dynamically modifiable curves move in parallel ('snakes' in computer vision); see ref. 27.

Recovering the global curve by computing a *covering* of it; i.e. a set of objects whose union is equivalent to the original curve, avoids the prerequisite global problems. Let the elements of the covering be unit-length dynamic splines, initially equivalent to the elements of the tangent field, but which then evolve according to a potential distribution constructed from the tangent field. The evolution takes two forms: (i) a migration in position to achieve smooth coverings; and (ii) a 'growth' to triple their initial length.

Again, there are two conceptually distinct steps to Stage 2 of the algorithm:

- Step 1. *Constructing the potential distribution* from the discrete tangent field. Each entry in the tangent field actually represents a discretization of the many possible curves in the world that could project onto that particular (tangent, curvature) hypothesis. Now these pieces must be put together. Assuming the curves are continuous but not necessarily differentiable everywhere, each contribution to the potential can be modeled as a Gaussian (the Wiener measure) oriented in the direction of the tangent field entry. The full potential distribution is their pointwise sum; see Figure 4.

- Step 2. *Spline dynamics.* The discrete entities in the tangent field are converted into unit splines



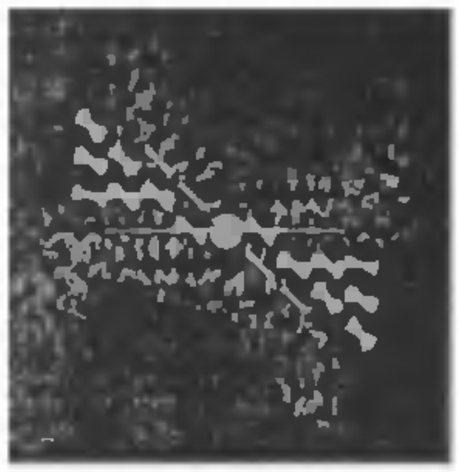


Figure 7. *a*, The geometric relationships necessary for defining the compatibilities between two label pairs at nearby points $i=(x_i, y_i)$ and $j=(x_j, y_j)$. *b*, Compatibilities between coarse (orientation, curvature) hypotheses at nearby positions. Eight distinct orientations and seven curvatures were represented, and three examples are shown. (*top*) The labels which give positive (white) and negative (black) support for a diagonal orientation with no curvature; (*middle*) positive and negative support for a small curvature class; (*bottom*) positive and negative support for the maximum curvature class. The magnitude of the interactions varies as well, roughly as a Gaussian superimposed on these diagrams. Physiologically these projective fields could represent inter-columnar interactions implemented by pyramidal neurons.

initialized in the valleys of the potential distribution. They evolve according to a variational scheme that depends on spline properties (tension and rigidity) as well as the global potential (Figure 8).

The potential distribution is created by adding together contributions from each element in the tangent field. Changing the representation from the tangent field to the potential distribution changes what is explicit and what is implicit in the representation, and local information is combined into global information. In Stage 1 there were discrete coarse entities; now there are smooth valleys that surround each of the global curves, with a separation between them. The 'jaggies' imposed by the initial image sampling have been eliminated, and interpolation to sub-pixel resolution is viable.

To recover the curves through the valleys, imagine creating, at each tangent field entry, a small spline of unit length oriented according to the tangent and curvature estimates (Figure 4). Since each spline is born in a valley of the tangent field potential distribution, they are then permitted to migrate to both smooth out the curve and to find the true local minima. The union of these local splines is the global cover. But the splines must overlap, so that each point on every curve is covered by at least one spline. We therefore let the splines extend in length while they migrate in position, until they reach a prescribed length. The covering is thus composed of these extensible splines which have grown in the valleys of the tangent field potential. Their specific dynamics and properties are described more fully in refs. 18 and 28.

Shape description

Given the descriptions of global curves, we next switch

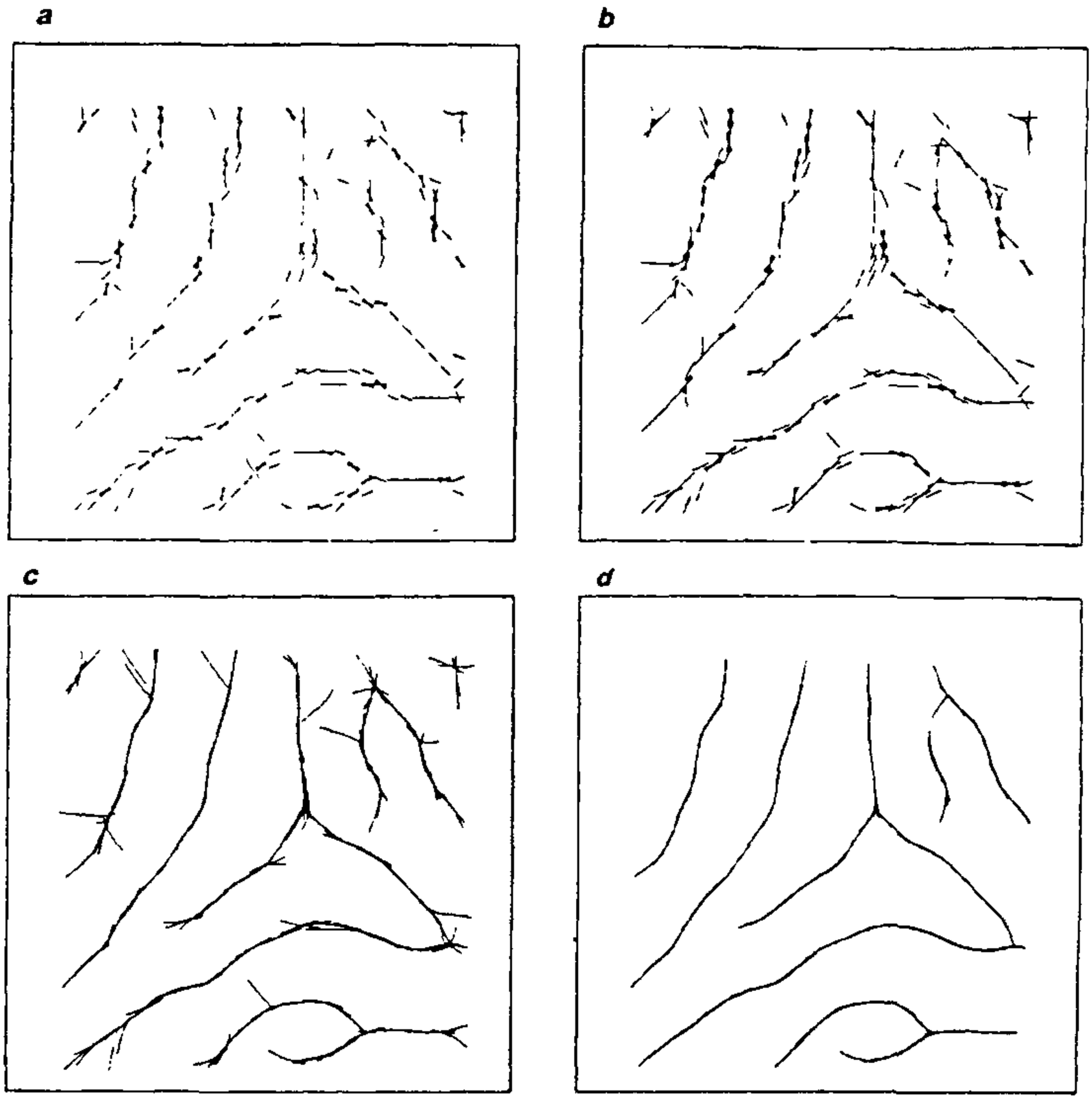


Figure 8. Illustration of the splines in motion. Initially, each spline is born at a tangent field location, with unit length. Then they migrate in position to minimal valleys in the potential distribution, and in length, so that they overlap and fill in short gaps. The length of each spline triples by convergence.

to the problem of shape description. While there is a sense in which the meaning of shape is effortlessly and intuitively understood, a formal definition of it has been elusive: there is currently no generally accepted definition of shape in either computational vision or psychology; but see refs. 29–31. This gap in our understanding is important, because shape may be considered as the bottleneck between early visual processes operating on edges, texture, color, shading, etc., and higher level processes acting on representations of objects. We therefore seek a theory of shape sufficiently powerful to provide a language for describing shapes. It follows that such a theory must be robust to variations within scenes, e.g. those variations due to small changes in viewpoint, to the changing appearance of objects due to local motion and emergent occlusions, as well as to variations within objects, e.g. due to flexibility, growth, and inflation.

To meet these needs, our approach to shape is organized around two basic intuitions: first, if a boundary were changed only slightly, then, in general, its shape would change only slightly. This leads us to propose an operational theory of shape based on incremental contour deformations. It differs from other approaches to shape based on dynamics (e.g. ref. 32). The second intuition is that not all contours are shapes, but rather only those that can enclose 'physical' material.

In a formal theory of contour deformation derived from these principles, we are able to prove that arbitrary local deformations of a curve in an arbitrary direction are qualitatively captured by a linear combination of two basis deformations along the normal: (*i*) a constant deformation and (ii) a deformation that

varies with the curvature[33]. This is a reaction/diffusion evolution equation which is defined only for smooth curves, since the normal is explicitly required. As there is no tangent at a corner, there is no normal, either. However, we have been able to abstract the above mathematical framework considerably, by showing that the deformations are equivalent to a hyperbolic conservation law with viscosity[34]. This is significant because such nonlinear conservation laws lead to the formation of shocks and to a notion of entropy. We are now finally able to close the loop back to shape, by showing how different classes of shocks in the solution to this conservation law correspond to the computational elements of shape. In particular,

1. *First-order shocks* correspond to *protrusions*;
2. *Second-order shocks* correspond to *parts*;
3. *Third-order shocks* correspond to *bends*;
4. *Fourth-order shocks* correspond to *seed points* for placing the material of shapes.

We close with two examples. First, we illustrate the notion of deformation and how it leads to robust descriptions of parts. Figure 9 contains four images of

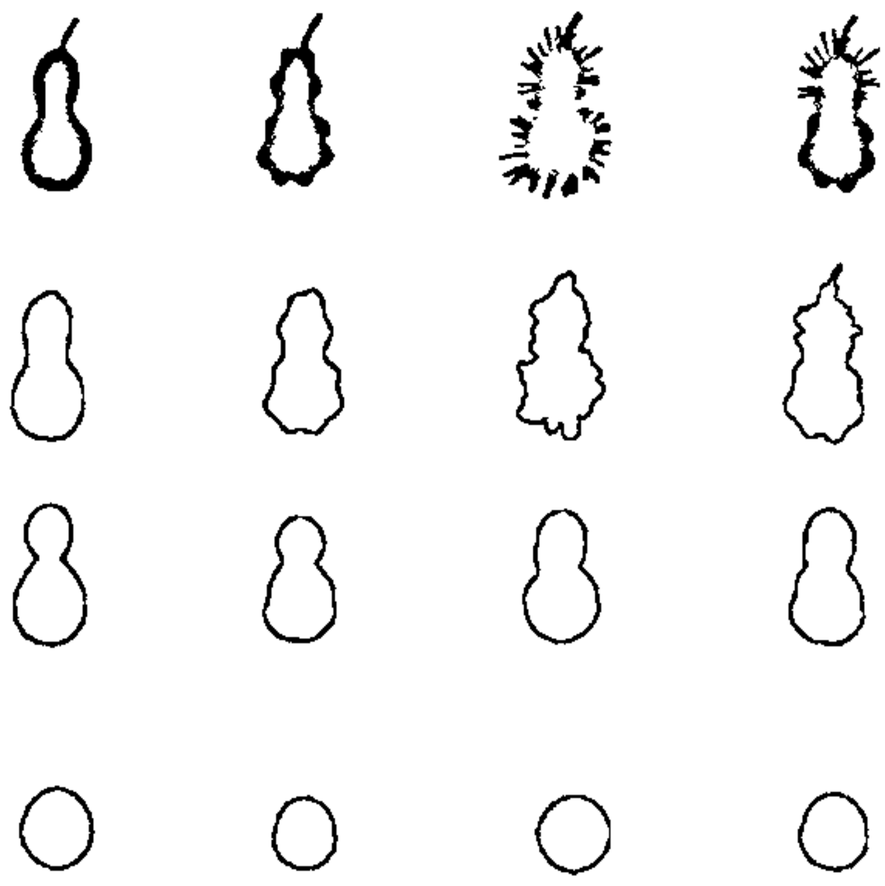


Figure 9. An illustration of how our deformation approach to shape leads to natural descriptions despite large quantities of noise and texture. These four pears were proposed by Hoffman and Richards as gross modifications of a single object category (pear). The original shapes are across the top, in black. Each column contains samples from a continuous sequence in which the bounding contour has evolved according to our deformation rules. The samples were chosen to illustrate how the deformation process eliminates the noise (first row) to reveal the fundamental part structure for the pear (second row). This structure is a pair of lobes, with the most significant one on the bottom. The part structure is signaled by the shocks (discontinuities) that develop on the contour in opposing pairs. Note how the lobe structure, and the dominant lobe (bottom row) are comparable for each of these different pear images, even though the noise and texture were so prominent. The description for each of these pears in our framework is a variant of 'a large bottom, a small middle and a very small top.'

pears, presented by Richards *et al.*[35], and which were intended as gross modifications of an object category (pear). The original shapes are across the top, and each column contains samples from a continuous sequence in which the bounding contour has evolved according to our deformation rules. The continuous space of shapes which supports such descriptions is called the *Entropy Scale-Space*[34]. It satisfies the need for object domain scale spaces referred to earlier.

A second example, Figure 10, illustrates the notion of hierarchy in more detail. An image of a doll was chosen to show how the different 'parts' emerge according to our natural intuitions about significance. Note how hands and feet are less significant than limbs, which are in turn less significant than the torso. This example also illustrates that several different types of shocks arise within our system, with first-order shocks signaling deformations, second-order shocks signaling part connections, third-order shocks signaling bends, and fourth order shocks signaling part centers. Note that occlusion will not affect decomposition into parts, a desirable feature for recognition.

Different organizational decompositions

To conclude this article, we observe that the processing described thus far focuses entirely on bounding contours. We began with edge detection, then grouped the local edge elements into global contours, and finally examined how shape analysis can proceed from deformations of these contours. The deformation analysis was particularly interesting from an AI perspective, because the different discrete components of shape were derived from continuous mathematics, not from the more symbolic perspective of models.

However, intra-surface events project into images as well, and vision is used for more than object recognition. Thus we close with a discussion of two alternative pathways for vision, which decompose the problem in different manners than those commonly addressed.

Color, contrast, and texture

In 1978 Margaret Wong-Riley stained sections of squirrel monkey striate cortex for the activity of the mitochondrial enzyme, cytochrome oxidase, and noticed a periodic distribution of 'puffs' of increased enzyme activity in layers 2 and 3. This discovery revealed an entire sub-organization within the visual cortex that supports information processing of a completely different variety than the border system already discussed[36]. When the cortex is viewed from above, the 'puffs' form a periodic array intercalated within a lattice

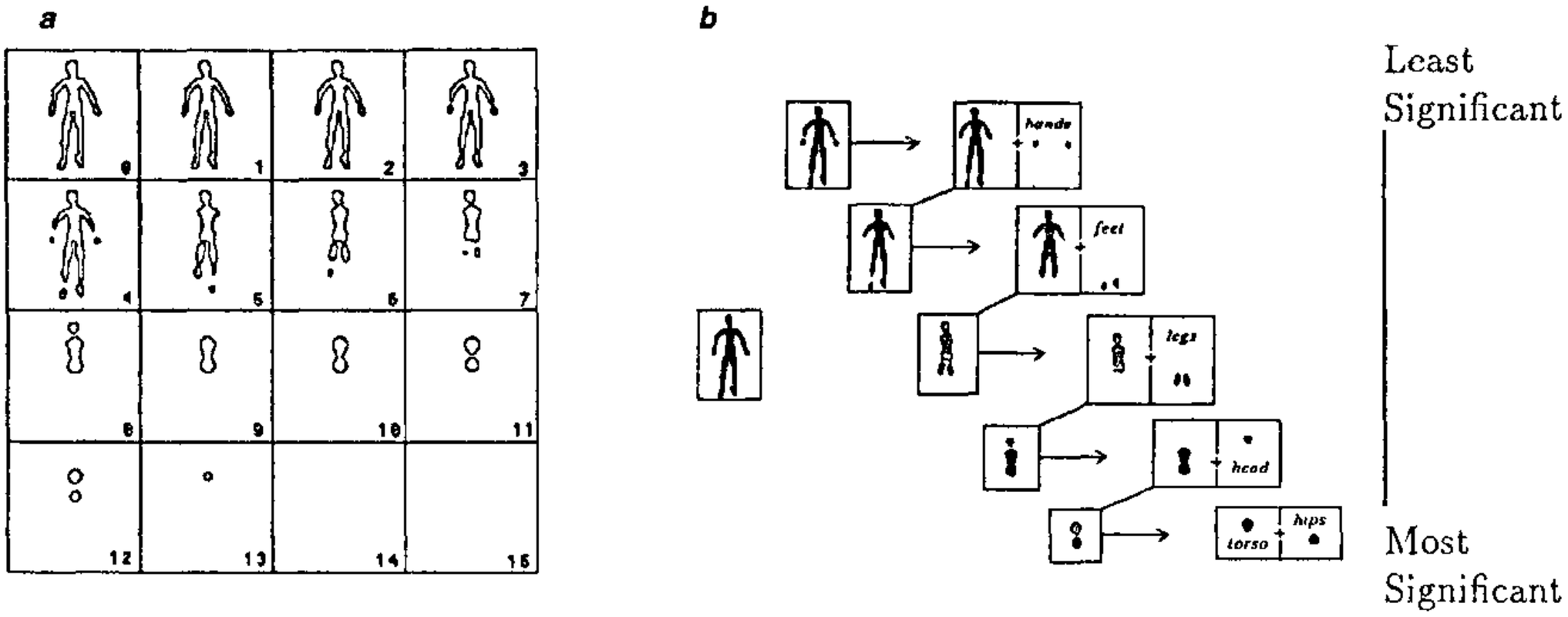


Figure 10. *a*, The evolution of shocks leads to parts, protrusions, and bends. This figure shows the development of an image of a doll (National Research Council of Canada Laser Range Image Library CNRC9077 Cat No 422; 128 × 128). The contour shown in box *N* corresponds to increasing boundary evolution (time) steps. Observe that the 'feet' partition from the 'legs' (via second-order shocks) between frames 3 and 4, and the 'hands' from the 'arms' between frames 2 and 3. Following these second-order shocks, first-order shocks develop as the 'arms' and are 'absorbed' into the chest. Running this process in the other direction would illustrate how the arms 'protrude' from the chest. *b*, The hierarchical decomposition of a doll into parts. Selected frames were organized into a hierarchy according to the principle that the significance of a part is directly proportional to its survival duration.

of lower cytochrome oxidase activity. Allman and Zucker[37] proposed that the distinction between the 'puffs' and the lattice is related to two different modes for representing stimulus variables.

We submit that *scalar* variables related to *intensity* of the stimulus are represented in the 'puffs'. The scalar variables include color, contrast, and texture density. Such intensity information is encoded explicitly over a very broad dynamic range, in which firing rate is proportional to the intensity variable (for example, contrast). Such an encoding strategy requires that neurons have the energetic capacity to sustain a broad range of activity levels, which in turn is related to the high concentration of cytochrome oxidase.

In contrast, the surrounding lattice of lower cytochrome oxidase activity supports *geometrical* variables with cells that are orientation selective. This is the system of simple, complex, and endstopped cells already discussed. In particular, we showed how each orientation is possible at any position, and each is represented explicitly within an orientation hypercolumn. Firing rate for these neurons now varies largely with how well each individual orientation *matches* image structure at that location. But there is rarely more than one orientation at any retinotopic location, so on average most oriented cells in each hypercolumn are quiet. The average level of neural activity over time is thus much less in the lattice than in the puffs, which is consistent with the lower levels of cytochrome oxidase in the lattice.

Object recognition vs. navigation

This final decomposition is perhaps even more far reaching, since it involves not intercalated substructures within the same area, but rather different regions of the brain. Conceptually, there is a fundamental theoretical difference in sensing for recognition as compared to sensing for navigation. Recognition problems have been at the center of activities in computational vision, and have been the main subject of this review. Central to recognition has been the bounding surface around an object, because it is this surface that provides the foundation for characterizing shape. The paradigm of *active vision* evolved specifically so that descriptive uncertainties could be eliminated to the point that objects could be reliaby identified. To summarize this activity, the types of questions asked could range from (most coarsely) 'is there a small object in the room' to 'is there a cup in the room' to (most specifically) 'is my favorite cup in the room'. Object recognition in primates likely involves the temporal lobe of the brain.

Vision for navigation poses rather different questions. Most coarsely, one could ask: 'can I approach coordinate area B'; more specifically, 'can I return to where I was 5 minutes ago', and most specifically, 'can I get from A to B'. Answering these questions raises dramatically different requirements for vision (and sensing) systems, a set of requirements that we submit are *complementary* to those for object recognition.

Instead of the focus on bounding surfaces, the focus now is on free-space. That is, instead of working from *without* the object, and attempting to describe the surface bounding it, now we are working from *within* free-space, and attempting to characterize the horizon. Notions of 'part', so central to shape, get replaced by topological requirements of 'reachability'. Descriptive uncertainty is now to be expected, since moving 'north' may not require elaborating a description of what is 'west' of current position. The global requirements that motived active visual paradigms are replaced by the local requirements of getting from here to there without falling into a hole. Vision for navigation likely involves the parietal cortex in primates.

Conclusions

While vision is a basic source of informational feedback from the world, systems that use it have been less than universally successful. We argued that a significant reason for this has been the poor quality of early vision, in particular, the difficulty of dealing with incorrect boundary information. This has made a complex vision problem nearly impossible, except in constrained circumstances. But recent efforts are producing much more veridical results, and a perspective from modern neurobiology and differential geometry is helping substantially. While much remains to be done, the foundations are certainly being out into place.

1. Zucker, S. W., The emerging paradigm of computational vision, *Annu. Rev. Comput. Sci.*, 1987, **2**, 69–89.
2. Van 'Essen, D. and Maunsell, J., Hierarchical organization and functional streams in the visual cortex, *Trends Neurosci.*, 1983, **6**, 370–375.
3. Tsotsos, J., A 'Complexity-level' analysis of intermediate vision, *Int. J. Comput. Vision*, 1988, **1**, 303–320.
4. Fleet, D. and Jepson, A., Hierarchical construction of orientation and velocity filters, *IEEE Trans. PAMI*, 1989, **11**, 315–324.
5. Hubel, D. and Wiesel, T., Functional architecture of macaque monkey visual cortex, *Proc. R. Soc. London*, 1977, **B198**, 1–59.
6. Canny, J., Finding edges and lines in images, AI TR 720, MIT, 1984.
7. Marr, D. and Hildreth, E., Theory of edge detection, *Proc. R. Soc. London*, 1980, **B207**, 187–217.
8. Marr, D., *Vision*, Freeman, 1982.
9. Horn, B., Understanding image intensities, *Artif. Intell.*, 1977, **8**, 201–231.
10. Poggio, T., Torre, V. and Koch, C., Computational vision and regularization theory, *Nature*, 1985, **317**, 314–319.
11. Kanisza, G., *Organization in Vision*, Praeger, New York, 1979.
12. Roberts, L., Machine perception of 3-dimensional solids, *Optical and Electro-Optical Information Processing* (ed. Tippett, J) MIT Press, Cambridge, 1965.
13. Faugeras, O., Steps toward a flexible 3-D vision system for robotics, in *Robotics Research: The Second International Symposium* (eds. Hanufusa, H. and Inoue, H.), MIT Press, Cambridge, 1985.
14. Koffka, K., *Gestalt Psychology*, Harcourt, Brace and World, New York, 1935.
15. Matthies, L., Kanade, T. and Szeliski, R., Kalman filter-based algorithms for estimating depth from image sequences, *Int. J. Comput. Vision*, 1989, **3**, 181–208.
16. Szeliski, R., *Bayesian Modeling of Uncertainty in Low-Level Vision*, Kluwer, Boston, 1989.
17. Lowe, D., Organization of smooth curves at multiple scales, *Int. J. Comput. Vision*, 1989, **3**, 119–130.
18. Zucker, S. W., Dobbins, A. and Iverson, L., Two stages of curve detection suggest two styles of visual computation, *Neural Comput.*, 1989, **1**, 68–81.
19. Hartshorne, R., *Algebraic Geometry*, Springer-Verlag, New York, 1977.
20. Parent, P. and Zucker, S. W., Trace Inference, Curvature Consistency, and Curve Detection, *IEEE Trans. PAMI*, 1989, **11**, 823–839.
21. Iverson, L. and Zucker, S. W., Logical/linear operators for image curves, *IEEE Trans. PAMI*, submitted.
22a. Dobbins, A., Zucker, S. W. and Cynader, M. S., Endstopping in the visual cortex as a substrate for calculating curvature, *Nature*, 1987, **329**, 438–441.
22b. Dobbins, A., Zucker, S. W. and Cynader, M., Endstopping and curvature, *Vision Res.*, 1989, **29**, 1371–1387.
23. Gilbert, C. D., *J. Physiol.*, 1977, **268**, 391–421.
24. Zucker, S. W., The computational connection in vision: Early orientation selection, *Behav. Res. Methods Instrum., Comp.*, 1986, **18**, 608–617.
25. Hummel, R. and Zucker, S. W., On the foundations of relaxation labelling processes, *IEEE Trans. PAMI*, 1983, **5**, 267–287.
26. Miller, D. and Zucker, S. W., Efficient simplex-like methods for equilibria of nonsymmetric analog networks, *Neural Comput.*, 1992, **4**(2), 167–190.
27. Kass, M., Witkin, A. and Terzopoulos, D., SNAKES: active contour models, *Int. J. Comput. Vision*, 1988, **1**, 321–332.
28. David, C. and Zucker, S. W., Potentials valleys, and dynamic global coverings, *Int. J. Comput Vision*, 1990, **5**, 219–238.
29. Biederman, I., Human image understanding: Recent research and theory, *Comput. Vision, Graphics, and Image Proc.*, 1985, **32**, 29–73.
30. Koenderink, J., *Solid Shape*, MIT Press, Cambridge, MA, 1989.
31. Richards, W. A. (ed.), *Natural Computation*, MIT Press, Cambridge, Ma, 1988.
32. Koenderink, J. and van Doorn, A., Dynamic shape, *Biol. Cybern.*, 1986, **53**, 383–396.
33. Kimia, B., Tannenbaum, A. and Zucker, S. W., On the evolution of curves via a function of curvature, I: The classical case, *J. Math. Anal. Appl.*, 1992, **163**(2), 438–458.
34. Kimia, B., Tannenbaum, A. and Zucker, S. W., Shapes, shocks, and deformations, I., *Int. J. Comput. Vision*, submitted; Technical Report LEMS-105, Division of Engineering, Brown University, Providence, RI, June, 1992.
35. Richards, W., Dawson, B. and Whittington, D., Encoding contour shape by curvature extrema, *J. Opt. Soc. Am.*, 1986, 1483–1489.
36. Livingstone, M. and Hubel, D., Anatomy and Physiology of a color system in primate visual cortex, *J. Neurosci.*, 1984, **4**, 309–356.
37. Allman, J. and Zucker, S. W., Cytochrome oxidase and functional coding in primate striate cortex: An hypothesis, *Cold Spring Harbor Symp. Quant. Biol.*, 1990, **55**, 979–982.
38. Duda, R. and Hart, P., *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

ACKNOWLEDGEMENTS. Research supported by NSERC, MRC, and AFOSR. I thank my many collaborators on the different parts of this project, including John Allman, Allan Dobbins, Robert Hummel, Lee Iverson, Ben Kimia, Pierre Parent, and Allen Tannenbaum.