Contour and surface perception

Heiko Neumann[†] *and Ennio Mingolla* [‡]

[†] *Fakultät für Informatik, Abteilung Neuroinformatik, Universität Ulm, Germany*

Email: `hneumann@neuro.informatik.uni-ulm.de`

[‡] Corresponding author

Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA

Tel: 1-617-353-9485

Fax: 1-617-353-7755

Email: `ennio@cns.bu.edu`

INTRODUCTION

Accumulating evidence from psychophysics and neurophysiology indicates that the computation of visual object representations is organized into parallel interacting sub-systems, or streams, which consist of mechanisms following complementary processing strategies. *Boundary formation* proceeds by spatially linking oriented contrast measures along smooth contour patterns, whereas *perceptual surface* attributes, such as lightness or texture, are derived from local ratio measures of image contrast of regions taken along contours. Mechanisms of both sub-systems mutually interact to resolve initial ambiguities and to generate coherent representations of surface layout.

Even when viewing single grey level images, people can discern important characteristics of visual scenes, including brightness, reflectance, surface orientation, texture, transparency, or relative depth. It has been proposed that these inferences are based on neural mechanisms which compute distinct and spatially registered representations of such characteristics or features, so-called intrinsic images. This approach has been formalized on the basis of statistical inference theory in which the a posteriori likelihood of estimating several scene characteristics given the available image is optimized. Common to these approaches is the requirement that representations of intrinsic scene characteristics are constrained in order to guarantee the consistency of the set of solutions, which often involve smoothness assumptions for correlated feature estimates. These consistency constraints are typically based on the laws of physical image generation, such as Lambertian surface reflectance properties, and thus the overall process comprises an ideal observer seeking for an optimal solution to the problem of inferring physical scene properties from images.

After briefly summarizing fundamental approaches to computation of intrinsic scene characteristics, we consider an alternate literature of neural models of boundary and surface computation.

INTRINSIC IMAGES AND THE NEURAL PROCESSING OF SURFACE LAYOUT

Barrow and Tenenbaum (1978) proposed a framework for machine vision based on computation of view-centered families of representations of local estimates of intrinsic scene characteristics, such as illumination, surface reflectance, and orientation. The resulting images are spatially registered with the single achromatic luminance image in which all the attributes are encoded pointwise. In order to regularize the inverse problem of computing several attributes from image data, several constraints, or assumptions, need to be incorporated into the recovery mechanisms. For example, in order to achieve the continuity of homogeneous surface attributes each intrinsic image incorporates lateral smoothing of values in a discrete grid unless an edge breaks the local continuity. Discontinuities need to be in registration with intensity edges and the recovered values for attributes need to fulfill the image irradiance equation at each grid point. This approach has been formalized in terms of coupled Markov random field (MRF) models in which the a posteriori probability for scene interpretation given an image, $p(scene|image) \propto \exp(-E/T)$ (E is an energy function that sums the image and prior constraints, T is a temperature term), can be maximized utilizing a stochastic sampling scheme (simulated annealing; Gamble *et al.*, 1989). The MRF formulation allows us to incorporate a line process (representing discontinuities) in the form of additional associated energy terms that can break the smoothness within an intrinsic attribute representation at locations where the tension of the recovered surface exceeds certain limits. These interactions correspond to the bidirectional coupling of boundary and surface computation depicted in Fig. 1. More recent developments of the MRF approach take into account the computation of perceptual surface attributes, such as transparency (Kersten, 1991).

These approaches share the same basic computational principles: Representations of smooth surface characteristics are generated by interpolating sparse data from initial estimates, and an edge map is extracted to depict the discontinuities in one or several attributes. In order to generate mutually consistent maps of intrinsic image characteristics, or attributes, the energy function used in the MRF model must utilize the prior probabilities of image characteristics which, in turn, are derived from physical imaging models, e.g., for reflectance or transparency. In that sense the estimation process defines an ideal observer that utilizes knowledge about image formation by trying to invert the optical image generation process. Perceptual findings, however, indicate that, for example, an attribute such as lightness (the perceived surface reflectance) depends on surface size and that the judgment of surface orientation is unreliable over variations in surface curvature or lighting conditions. These observations cannot uniquely be accounted for by the intrinsic image approach, since size effects impose constraints not considered by such local processes as are employed in implementing an inverse optics solution.

In addition, line processes representing physical discontinuities in a surface property are modeled as separate MRFs. In order to be computationally tractable their prior probability structure is formulated over a small local pixel neighborhood only. Empirical observations again suggest that contour processes act over long spatial distances in order to reliably integrate contour fragments to form surface boundaries under variable imaging conditions. In all, these observations motivate the investigation of the neural mechanisms underlying the computation of perceptual boundaries and the subsequent assignment of surface attributes.

ELEMENTS OF SPATIAL LONG-RANGE INTEGRATION IN BOUNDARY FINDING

Formal approaches to the Gestalt concept of *good continuation* have garnered increasing attention since Field *et al.* (1993) popularized the notion of an "association field," an elongated spatial zone aligned with oriented contour segments denoting facilitatory perceptual interactions with other segments. This geometry of spatial integration is summarized in the "bipole" icon of Fig. 2, a figure-eight shaped zone that was introduced as a "cooperative cell" unit in a neural network model for grouping by Grossberg and Mingolla (1985). The strength of spatial integration in the presently considered models is always some function of the distances and relative alignments of oriented units, such as contrast edges. Imagine that a contrast pattern occurs along the orientations denoted by the long axes of the dark and the light ellipses at locations $\mathbf{x}$ and $\mathbf{x}'$. The influence of the edge at the light ellipse on the representation of the contour strength in the region of the dark ellipse is calculated using the following fundamental quantities: (1) the distance, $\mathbf{r}$, between the centers of the two ellipses; (2) the angle, ϑ, between the ray passing through the centers of the two ellipses and the principle axis of the dark ellipse; and (3) the difference in orientation, φ, of the principle axes of the two ellipses.

The bipole shape expresses the region of relatively high coupling strength for the influence of contour segments remote from the central ellipse on a unit whose positional and orientational preferences are denoted by the ellipse at the center. Its shape and connectivity pattern is justified by recent investigations in psychophysics, physiology, and anatomy. For example, Geisler *et al.* (2001) measured psychophysically the co-occurrence of oriented contrasts in natural scenes, and Bosking *et al.* (1997) directly demonstrated that cortical cells of similar orientation selectivity in tree shrew are preferentially connected along the axis of their visuotopic alignment to cells of compatible orientational preference. Thus the bipole icon for grouping of contour segments – as originally suggested on intuitive grounds – is validated by a growing body of empirical data.

The bipole concept can be embedded in a framework in which the pattern of connectivity among relatively tightly-coupled neural units is described formally by a *spatial weighting*, or kernel, function, coding the strength of connections between units in a spatial array. Items that are closely spaced are more likely to be grouped than candidates that are located far apart. The underlying neighborhood function often selectively facilitates a sector of a spatial surround to define an anisotropic coupling that is compatible with the feature domain. Elementary features along dimensions such as, e.g., (tangential) orientation, motion direction, and disparity provide the dimensions of the visual representation space. The feature *relatability* (or *compatibility*) between stimulus items is defined along these dimensions. Herein, the most likely appearance of meaningful structure is encoded to represent "what feature goes with what". In the following treatment, we will focus on grouping mechanisms for static form processing and therefore consider only *orientation* as the relevant feature dimension.

The result of integration of items defines the *support* of a localized feature measurement at a given spatial location based on the configuration of other items. Based on the bipole structure of left and right sub-fields (Fig. 2), the support of a target item can be defined as a function

$$\text{support}_{xy,feature} = \left\{\text{left-input}_{xy,feature}\right\} \circ \left\{\text{right-input}_{xy,feature}\right\} \tag{1}$$

where '$\circ$' denotes some operation to define the *combination of sub-fields.* Subscripts xy denote spatial locations in a 2-D retinotopic map and $feature$ identifies the feature dimension involved in the grouping process, *orientation* in our case. The *activation* of the corresponding target cell results as a function $f(\cdot)$ of the support. The formal description of the mechanisms underlying the computation of activations 'left-input' and 'right-input', respectively, necessitates the detailed specification of the interaction of activities in grouping.

In most models the connectivity pattern of the relatable features is pre-specified, or *programmed*, referring to some measure of geometric entities. These are designed to encode static efficacies between n-tuples, e.g. pairs, of feature measurements at given locations optimizing a given functionality. To date, few approaches have investigated the possible *self-organization* of such lateral interactions in a neural architecture. Note that the spatial weighting function and the feature relatability define the components of the net *coupling strength.* This function specifies a metric for the similarity measure in the $< \ xy, feature >$-space and thus defines the distance function of *feature cooperation* for the clustering of a visual pattern in accordance to a relatability measure that underlies the visual interpolation in spatial grouping for object recognition (Kellman and Shipley, 1991). *Input activations* at the different sites are necessary to gather any support. Here, two different types of interactions can be distinguished: (1) a convergent feedforward mechanism is defined when the *bottom-up* input is integrated at the target location, whereas (2) a mechanism of (non-linear) *lateral* interaction is defined when activity is horizontally integrated within a neural layer.

Taken together, the activation of, say, 'left-input' derived from the feature integration process using the bipole mechanism is computed by

$$\text{left-input}_{\mathbf{x}\theta} = \sum_{\mathbf{x}'\phi} \left\{ \text{act}_{\mathbf{x}'\phi} \cdot \text{relate}_{\mathbf{x}\mathbf{x}'\theta\phi} \cdot \text{weight}^{L}_{\mathbf{x}\mathbf{x}'\theta} \right\} \tag{2}$$

where 'weight' denotes the spatial weighting kernel, 'relate' the feature relatability, and 'act' the input activations. The 'right-input' activity is computed similarly. Here, $\mathbf{x} = (x, y)$ and θ correspond to the location and orientation of the target feature, respectively. Other parameters refer to the specific location $< \mathbf{x}', \phi >$ in the space-orientation neighborhood (see Fig. 2).

GROUPING MODELS AND THEIR COMPONENTS

The initial processing stages of these models consist of some form of filtering the input luminance image. A (possibly non-linear) center-surround mechanism (resembling segregated ON and OFF contrast channels at the retina and LGN) computes contrast ratios throughout the image. These outputs drive oriented simple and complex cells, which are often modeled as localized spatial frequency filters, such as Gabor or wavelet kernels. Their output of different orientation fields defines the interface representation for subsequent grouping processes. We focus on those models which (1) have their roots in the explanation of empirical data and (2) were most influential for subsequent scientific developments. A comprehensive treatment can be found in Neumann and Mingolla (2001).

The Boundary Contour System (BCS) has been developed as part of a unified modeling framework called FACADE ("Form-And-Color-And-DEpth"). As described by Ross *et al.* (2000) the BCS consists of a series of boundary detection, competition, and cooperation stages. The general layout of basic processing stages is presented in Fig. 3 (left). Long-range boundary cooperation of this scheme (stage 3) accomplishes the grouping of consistent boundaries and the completion of interrupted boundaries. Spatial weighting and relatability, as elements of the support function, in this model are defined as $\text{weight}^{L/R}_{\mathbf{xx}'\theta} = \left[\pm\Gamma^{rad}_{\mathbf{xx}'} \cdot \Gamma^{ang}_{\mathbf{xx}'\theta}\right]^+$ (for a scheme separable in polar coordinates; $[\cdot]^+$ denoting half-wave rectification) and $\text{relate}_{\mathbf{xx}'\theta\phi} = \cos^q\left(2\tan^{-1}\left((y'-y)/(x'-x)\right) - (\theta+\phi)\right)$ (for a co-circularity constraint; q controlling the angular width of the lobes). The support and activation function is computed employing bipole cells which only fire if both lobes of their integration fields are sufficiently activated. The left/right subfield combination thus realizes a multiplicative, or AND gate, combination of input terms. The cooperative-competitive (CC) feedback loop between stages 1 and 3 (Fig. 3, left) acts to complete and enhance spatially and orientationally consistent boundary groupings while inhibiting inconsistent ones, thereby also suppressing noise. This mechanism of long-range completion is also capable to signal boundaries over gaps in the image, that is, over regions void of any contrast signals, and can thus generate illusory contours for images in which humans perceive them (Lesher, 1995).

Relaxation labeling schemes have been developed as optimization procedures to find consistent interpretations for measurement problems with uncertainty. The computational goal seeks to achieve a consistent labeling assigning graded activations, or probabilities, to a limited set of labels for nodes in a graph representation. For contour integration, the labeling problem can be formulated as one of finding the most likely set of orientations for discrete grid locations corresponding to the graph nodes (a “no-line” label is taken into account to represent the non-responsiveness of cells at locations that are not boundary elements). Parent and Zucker (1989) determine unambiguous orientation estimates along contours by evaluating consistency constraints based on a measure of compatibility between pairs of orientations. Initial responses are generated by oriented filters (Fig. 3, left, stage 1) that were normalized in order to allow filter activations treated as probabilities for assigning orientation labels (stage 2). The individual strengths for orientation measures at a given image location are iteratively updated through the support that is gathered from relatable activities in a spatial neighborhood. Spatial weighting and relatability are defined as $\text{weight}_{\mathbf{xx}'\theta}^{L/R} = P_{\mathbf{xx}'\theta}^{length,L/R} \cdot d(\mathbf{x}, \mathbf{x}')$, where P denotes a predicate to compensate for path length differences on discrete grids, and $d(\cdot)$ compensates for differences in inter-pixel distances, and $\text{relate}_{\mathbf{xx}'\theta\phi} = c_{\mathbf{xx}'\theta\phi} \cdot K_{\mathbf{xx}'\theta\phi}^{k} \cdot C_{\mathbf{xx}'\theta\phi}^{kk'}$. Here the co-circularity measure c is augmented by two binary predicates to exclude any candidates of incompatible local contour curvature. Input activations in the orientation field are thinned by non-maximum suppression in order to generate localized representations of contours. The support and activation function is computed by integrating responses from the sub-fields of the spatial weighting functions (stage 3). Activities are iteratively updated by a nonlinear recurrent competitive/cooperative mechanism.

Contrary to the previous approaches, Heitger *et al.* (1998) proposed a feed-forward scheme of successive filtering for contour grouping that selectively integrates activities from oriented single end-stopped (ES) filters, which respond to, e.g., line ends and corners. The result of such grouping is combined with the representation of oriented contrast responses to generate a final contour map. The core mechanism of grouping again utilizes spatial weighting functions (bipoles) virtually equivalent to those of the BCS (Fig. 3, left, stage 3). Two grouping rules are distinguished for corners (para grouping) and line ends (ortho grouping). The responses of ortho and para grouping are linearly interpolated. Different curvature classes are distinguished similar to the relaxation scheme (see above) by partitioning the bipoles into subfields such that only some of them in the left lobe can cooperate with their counterparts in the right lobe. Left/right lobes are combined in a multiplicative fashion via an AND-gating mechanism, similar to the BCS approach. This makes the grouping scheme selective to complete activations between localized ES features. The elongated bipole lobes are also capable to signal boundaries over gaps and, thus, generate illusory contours. This model does not incorporate any feedback mechanism and must, therefore, employ some stages of postprocessing in order to sharpen the final boundary response.

Based on the previous "core models" other approaches have been developed that elaborate aspects of the general framework. Williams and Jacobs (1997) investigated random walks of particles in a discrete lattice of the sampled space-orientation domain to determine the paths of spatially relatable items at a given location (x, y, θ) and another point (i, j, ϕ). Particles were initiated at sparse keypoint locations corresponding to localized responses of ES cells. The probability densities in the stochastic completion field represent the strengths – and therefore likelihood – of smooth paths connecting pairs of key points. Li (1998) investigated how V1 horizontal long-range integration could functionally account for the enhancement of texture region boundaries and pop-out effects in visual search. Here, the spatial integration field is subdivided into spatially non-overlapping parts of excitatory and inhibitory contribution. An elongated bipole integrates activities of cells oriented such that they form smooth interpolations with the target cell at the bipole center. Activities of like orientation from a sector orthogonal to the target cell orientation generate the inhibitory signal. In a similar spirit Yen and Finkel (1998) proposed a scheme of excitatory long-range integration that consists of two spatial regions: One coaxial bipole with cocircular relatability and one sector that extends orthogonally from the target cell's orientation axis and integrates units oriented parallel to that of the cell. The latter component contributes to a facilitation of simple symmetric shape axes.

Each cell inhibits itself based on a threshold of the average input from its immediate neighbors so that only salient arrangements produce a net output.

Contrast integration for boundary grouping addresses the particularly important question about what the core principles are that underlie the establishing of spatial integration and grouping. Yet there still is an ongoing debate about the role of feedforward and feedback mechanisms involved in spatial grouping and the dominance of their individual contributions in visual processing. Aspects of *temporal coding* principles based on oscillator mechanisms or spiking neurons may play another important role in grouping tasks. Several neurophysiological studies indicate that distributed representations of related scene fragments are linked by temporally correlated, or synchronized, neural activation. The temporal coding hypothesis studied in isolation appears, however, to be incomplete. The temporal establishment of grouping addresses the signaling of binding, but not the "how" or "what" of its computation. The mechanisms of oriented long-range interactions for integration provide the underlying basis for grouping to establish perceptual items related to surface boundaries.

COMPLEMENTARY MECHANISMS OF SURFACE PERCEPTION

The computation of perceptual surface qualities complements the formation of their boundaries. Image regions that are sufficiently homogeneous in luminance or statistical distribution of contrasts can give rise to the impression of color, texture, brightness, or lightness, also known as achromatic color. How is the generation of smooth representations of surface qualities accomplished? Paradiso and Nakayama (1991) provided compelling psychophysical evidence for a long-hypothesized neural process that propagates local estimates of lightness from boundaries into region interiors. Further psychophysical as well as physiological investigations of temporal properties of brightness and texture filling-in revealed further details of the neural machinery underlying the integration of perceptual surface properties (Pessoa *et al.,* 1998.)

Computational models for the generation of perceptual surface quantities generally pursue one of three basic strategies: (1) filtering and rule-based symbolic interpretation, (2) spatial integration via inverse filtering and labeling, or (3) filling-in. In the first scheme, (non-linear) combinations of filter responses and subsequent rule-based decision operations lead to the final prediction of lightness. The latter two approaches both begin with local luminance ratios estimated along boundaries. Computation of surface qualities proceeds by propagating local estimates into region interiors in order to generate a spatially contiguous representation of homogeneous properties (Fig. 3, right). In the spatial integration approach, the lateral propagation is the consequence of an iterative process to invert previous spatial derivative operations labeling region interiors with estimates of quantities derived at region boundaries. Filling-in approaches spread activity in a neural map such that at locations coding region interiors activation is generated by a spatial diffusion process integrating estimates of local contrasts from remote boundaries.

Spatial filtering approaches utilize initial stages of either isotropic or oriented filters, or both, over multiple bandpass channels of spatial frequency. The filter outputs are scaled non-linearly by a gain control mechanism and thresholded, individually interpreted by a set of rules and finally combined over several scales. In two dimensions this strategy leads to results that depend on the direction of the sequential application of interpretation rules. Authors pursue a simple averaging of results from forward and backward scanning over different orientations (McArthur and Moulden, 1999). Approaches of this category follow a tradition that the input luminance distribution is processed through a sequence of filtering steps during the early stages of the visual system. It is assumed that specific features in the responses directly contribute to observable brightness effects by implicitly propagating local qualities into region interiors as a consequence of applying some interpretation rules.

Spatial integration models attempt to recover object lightness of a surface by utilizing the sequence of processing stages Filtering/Differentiation → Boundaries/Thresholding → Integration. Differentiation and subsequent thresholding operations are intended to detect salient changes in the luminance signal which, in turn, trigger the integration from local luminance ratios. Luminance ratios provide the basis to infer surface-related properties that are invariant against gradual illumination changes, thus discounting the illuminant. Furthermore, lightness can be influenced by luminances at regions remote from each other. The Retinex algorithm by Land and McCann (1971) accounts for these observations by integrating contrasts along several pathways of different lengths. The logarithms of luminance ratios at thresholded contrast locations are summed such that, as a net effect, ratios measured between the target region and distant regions are integrated and averaged. A center-surround interaction accounts for this process in an approximate form. Alternative formulations of lightness integration numerically invert the differentiated 2D luminance image. Under certain boundary conditions a unique solution exists that can be computed by numerical techniques (Hurlbert, 1986). Changes in reflectance properties occur locally, thus comprise a high spatial frequency pattern, while inhomogeneous illumination constitutes a phenomenon at low spatial frequencies. Suppression of influences in low spatial frequencies could be achieved by homomorphic filtering in which low frequency components are reduced or even suppressed and higher frequencies are amplified. Following this idea approximate approaches for lightness computation utilize multiple spatial frequency channels of center-surround mechanisms based on divisive, or shunting, center-surround interactions of different scales (Jobson *et al.,* 1997).

As a second computational strategy of spatial integration models, *filling-in* approaches proceed by taking local ratios along boundaries and subsequently propagate these measures into the void spaces of bounded regions. Generating a representation of surface quality utilizes the processing stages Filtering $\rightarrow$ Boundaries $\rightarrow$ Filling-in. The algorithm proposed by Grossberg and Todorović (1988) was the first implementation in 2D of the complementary operations of the combined BCS and Feature Contour System (FCS) model of brightness perception that was able to explain a wide variety of phenomena, such as simultaneous contrast and the Craik-O'Brien-Cornsweet illusion, among others. The model was later demonstrated to also account for the temporal properties of brightness filling-in (Paradiso and Nakayama, 1991). Filling-in of local contrast ratios taken along extended figural boundaries is laterally propagated in a diffusion process. Such a diffusion is controlled by a gradual permeability function Υ utilizing a boundary signal such as the one generated by mechanisms of long-range integration (compare Fig. 2 and Fig. 3, right). This renders filling-in a spatially inhomogeneous diffusion process which generates a maximum likelihood (regularized) solution of a brightness surface given the sparse contrast estimates at the boundaries (Neumann *et al.,* 1998), where Υ is monotonically decreasing to split apart regions of homogeneous surface properties. In the case that the function Υ is solely controlled by an auxiliary boundary signal the function solves a simple gradient descent of finding a dense activity distribution representing surface brightness. In general, Υ could also depend on gradients of the filling-in signal itself such that boundary *and* surface computation becomes mutually interdependent (dotted arrow in Fig. 1) and the overall system becomes non-linear.

Ross and Pessoa (2000) describe an important extension of previous models by suggesting a context sensitive weighting of contrast measures prior to a final integration stage. Emphasis is put on a computational mechanism that tags boundaries from contrast detection and grouping to (partially) segment an image into different context domains. The segmentation is triggered by the presence of T-junctions which are used as seed points to propagate the tagging signal along the roof of the T's and the adjoining smooth boundary segments, thus generating a map of context boundaries. These boundaries are subsequently used to suppress initial contrast measures along the corresponding contours via a gating mechanism. As a result, the contribution of contrast measures that could lead to erroneous lightness estimates over segmentation boundaries is reduced or even suppressed. This computational mechanism is capable of producing region lightness estimates that account for several effects that have been shown previously to be unexplainable by simple local mechanisms of contrast integration, e.g. White's effect, the Benary cross, and Adelson's folded-card stimuli.

Although not described here, the FACADE model, to which the Ross and Pessoa model has certain parallels, proposes related explanations of the just-mentioned and related lightness effects (Grossberg, 1994). FACADE has also been developed to account for the perception of such phenomena as amodal surface completion, whereby occluded portions of objects are sensed as being present in specific locations behind foreground objects, and, more generally, the separation of surface regions seen as "figures" from backgrounds. The development of this and other models (Heitger *et al.*, 1998; Yen and Finkel, 1998) that attempt to account for more complex aspects of surface perception than are captured in planar arrays indicates that an exciting new phase of inquiry into surface perception is underway.

STOCHASTIC FORMULATION OF BOUNDARY/SURFACE COMPUTATIONS

In the tradition of intrinsic image architectures some approaches formulate the above outlined boundary and surface computations in a Bayesian framework. For example, Lee (1995) utilized a variant of a non-linear diffusion mechanism that incorporated piecewise smoothness of filling-in domains which were separated by contour segments denoting breaks in the smooth surface signal. Statistical signals from initial responses of oriented Gabor filters of different frequency selectivity served as filling-in signals. These inputs were subsequently diffused in the space-frequency domain incorporating a spatial modulation by the boundary signal. Mutual interaction between boundary and surface processes were modeled as MRF processes in which Bayesian priors propagate bidirectionally and interact through local connections in each area. Since this approach is focusing on texture region segregation the explicit modeling of the optical image generation process, as in classical intrinsic image approaches, is not necessary here. This renders the model to be a Bayesian interpretation of above mentioned neural processes for surface/boundary extraction.

DISCUSSION

We have seen that the key stages of a common framework of computational models for generating perceptual surface representations utilize separate sub-systems for boundary and surface computation. Depending on the modeling framework different processes of mutual interaction are defined to achieve the goal of generating a coherent representation of object surfaces and their attributes. Modeling the *neural* processes of perceptual surface layout has focused on processes of contour completion based on long-range integration based on some variation of the structure of the bipole kernel and a number of additional computational principles in evaluating related activities in feature space. The bipole kernel itself graphically visualizes the oriented fan-like connectivity between sites in a space-orientation feature space.

In comparison to processes of boundary finding, the situation with respect to surface quality perception is considerably less developed, most especially with respect to attributes such as texture perception, which we have not reviewed. All models presented are based on some mechanism to laterally propagate activities to interpolate and smooth sparse estimates. Improved empirical techniques juxtaposed with hypotheses developed by computational modelers offer the hope that coming years will see a convergence in this area, as has already occured in the field of contour completion.

REFERENCES

Barrow, H.G. and Tenenbaum, J.M. (1978). Recovering intrinsic scene characteristics from images. In A.R. Hanson, E.M. Riseman (eds.), Computer vision systems. New York: Academic Press, 3-26.

Bosking, W.H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. Journal of Neuroscience, 17:2112-2127.

Field, D.J., Hayes, A., and Hess, R.F. (1993). Contour integration by the human visual system: Evidence for a local "association field". Vision Research, 33:173-193.

Gamble, E.B., Geiger, D., Poggio, T., and Weinshall, D. (1989). Integration of vision modules and labeling of surface discontinuities. IEEE Transactions on Systems, Man, and Cybernetics, 19:1576-1581.

Geisler, W.S., Perry, J.S., Super, B.J., and Gallogly, D.P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. Vision Research, 41:711-724.

Grossberg, S. (1994) 3-D vision and figure ground separation by visual cortex. Perception and Psychophysics, 55(1):48-120.

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentation. Perception and Psychophysics, 38:141-171.

Grossberg, S. and Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of Classical and recent phenomena. Perception & Psychophysics, 43:723-742.

Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., and Kübler, O. (1998). Simulation of neural contour mechanisms: Representing anomalous contours. Image and Vision Computing, 16:407-421.

(*) Hurlbert, A. (1986). Formal connections between lightness algorithms. Journal of the Optical Society of America A, 3:1684-1693.

Jobson, D.J., Rahman, Z., and Woodell, G.A. (1997). Properties and performance of a center/surround Retinex. IEEE Transactions on Image Processing, 6, 451-462.

(*) Kellman, P.J. and Shipley, T.F. (1991). A theory of visual interpolation in object perception. Cognitive Psychology, 23:141-221.

(*) Kersten, D. (1991). Transparency and the computation of scene attributes. In M.S. Landy, J.A. Movshon (eds.), Computational models of visual processing. Cambridge: MIT Press, 209-228.

Land, E.H. and McCann, J.J. (1971). Lightness and Retinex theory. Journal of the Optical Society of America, 61:1-11.

Lee, T.S. (1995). A Bayesian framework for understanding texture segmentation in the primary visual cortex. Vision Research, 35:2643-2657.

(*) Lesher, G.W. (1995). Illusory contours: Toward a neurally based perceptual theory. Psychonomic Bulletin and Review, 2:279-321.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. Neural Computation, 10:903-940.

McArthur, J.A and Moulden, B. (1999) A two-dimensional model of brightness perception based on spatial filtering consistent with retinal processing. Vision Research, 39:1199-219.

(*) Neumann, H. and Mingolla, E. (2001). Computational neural models of spatial integration in perceptual grouping. In T.F. Shipley, P.J. Kellman (eds.), From fragments to objects: Grouping and segmentation in vision. Amsterdam: Elsevier, 354-400.

(*) Neumann, H., Pessoa, L., and Mingolla, E. (1998). A neural architecture of brightness perception: Non-linear contrast detection and geometry-driven diffusion. Image and Vision Computing, 16:423-446.

Parent, P. and Zucker, S. (1989). Trace inference, curvature consistency, and curve detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11:823-839.

Pessoa, L., Thompson, E., and Noë, A. (1998). Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. Behavioral and Brian Sciences, 21(6):723-748.

Ross, W.D., Grossberg, S., and Mingolla, E. (2000). Visual cortical mechanisms of perceptual grouping: Interacting layers, networks, columns, and maps. Neural Networks, 13:571-588.

Ross, W.D. and Pessoa, L. (2000). Lightness from contrast: A selective integration model. Perception & Psychophysics, 62:1160-1181.

Williams, L.R. and Jacobs, D.W. (1997). Local parallel computations of stochastic completions fields. Neural Computation, 9:859-881.

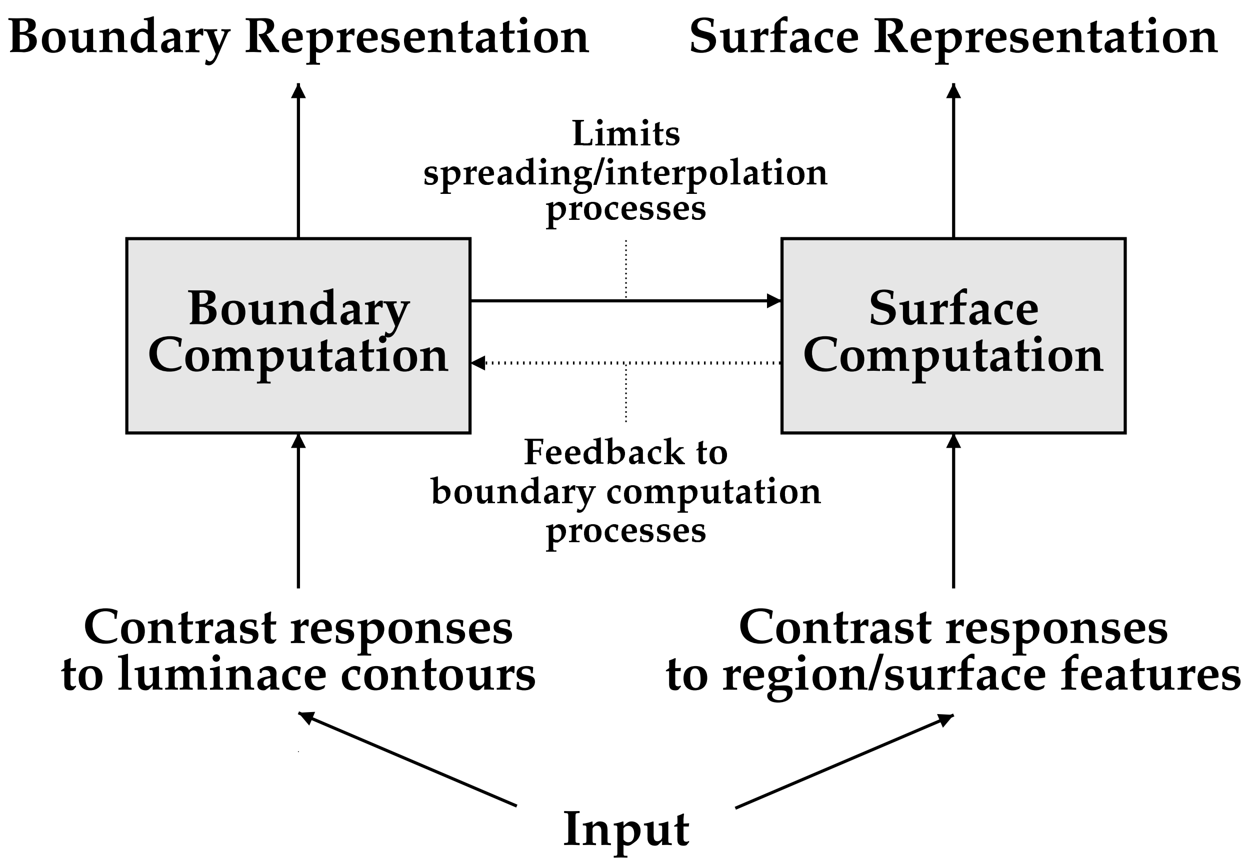
Yen, S.-C. and Finkel, L.H. (1998). Extraction of perceptually salient contours by striate cortical networks. Vision Research, 38:719-741.

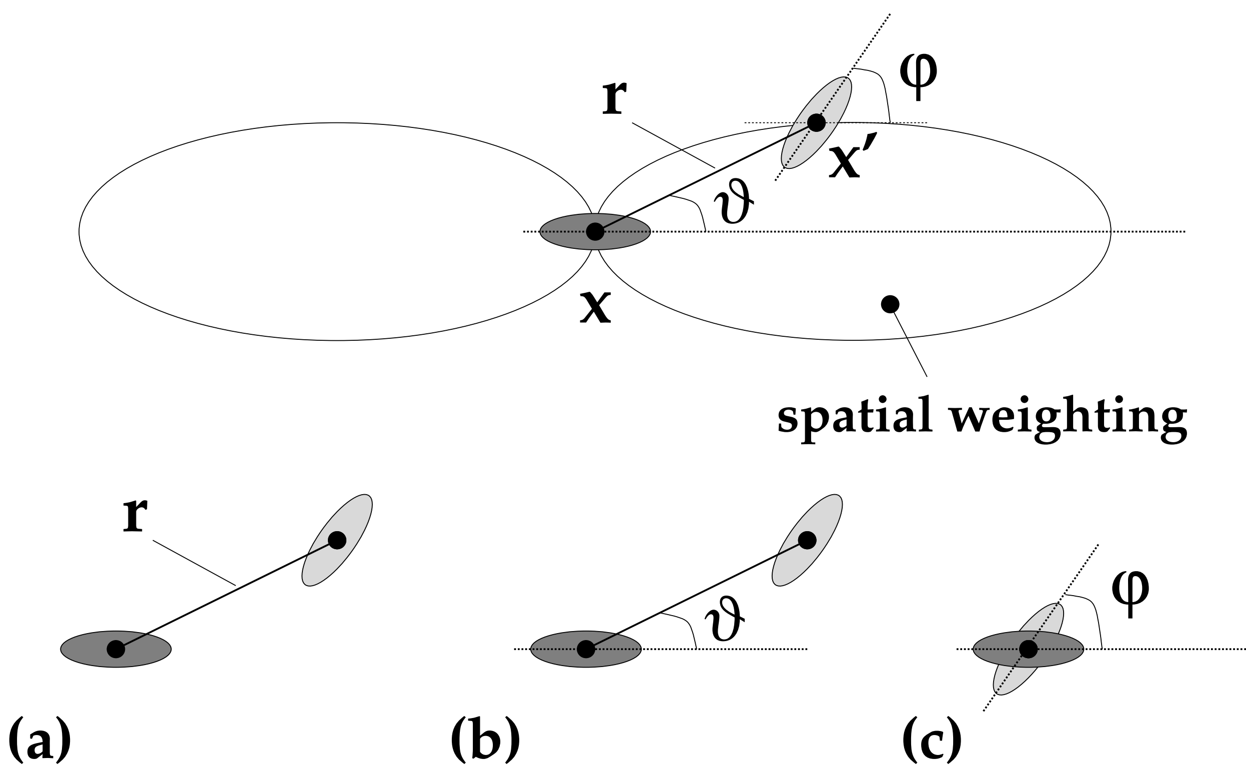
Figure Captions

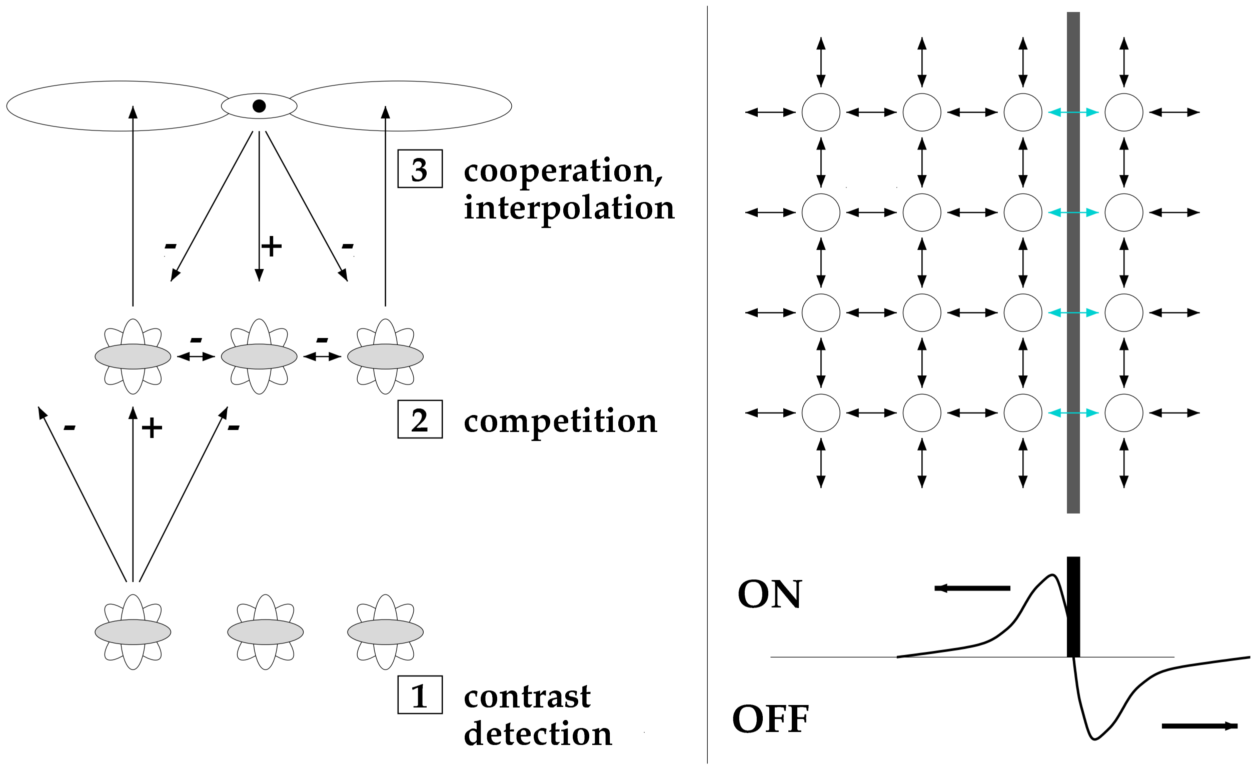
Figure 1. Schematic of the macroscopic computational elements for computing boundaries and surfaces. The flow of computation is segregated into parallel but interacting streams to determine boundaries and surface attributes. See text for details.

Figure 2. The "bipole icon" for modeling feature integration in spatial grouping (top). The quantity to be assessed is the "contribution" of activation in an oriented unit, denoted by the light ellipse, to the activation at the center unit, denoted by the dark ellipse. The figure-eight shape of a bipole expresses relations among three fundamental quantities (bottom): (a) the distance, $\mathbf{r}$, between spatial locations (b) the angle, ϑ, between the virtual line and the target orientation, and (c) the difference in orientation, φ, of the two ellipses.

Figure 3. Details of the mechanisms involved in boundary and surface computations. Left: The three major stages for *boundary computation* involve contrast detection and competition between responses in space and orientation domain. Resulting responses are integrated over longer distances utilizing oriented bipoles, which lead to on-center/off-surround feedback interaction. Right: Lateral interaction for *surface computation* between sites of a regular grid. Local contrast signals measured near boundaries are laterally propagated to fill in regions void of activities. The lateral interaction is controlled by an inhibitory mechanism to reduce or switch off lateral couplings by high boundary activation, as indicated by the light shaded arrows. See text for details.


Boundary Representation
Surface Representation
Limits spreading/interpolation processes
Boundary Computation
Surface Computation
Feedback to boundary computation processes
Contrast responses to luminace contours
Contrast responses to region/surface features
Input


r
φ
x′
ϑ
x
spatial weighting
r
ϑ
φ
(a)
(b)
(c)


3 cooperation, interpolation
2 competition
1 contrast detection
ON
OFF