

A neural model of 3D shape-from-texture: Multiple-scale filtering, boundary grouping, and surface filling-in

Stephen Grossberg ^{*,1}, Levin Kuhlmann ¹, Ennio Mingolla ¹

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

Received 3 April 2006; received in revised form 15 October 2006

Abstract

A neural model is presented of how cortical areas V1, V2, and V4 interact to convert a textured 2D image into a representation of curved 3D shape. Two basic problems are solved to achieve this: (1) Patterns of spatially discrete 2D texture elements are transformed into a spatially smooth surface representation of 3D shape. (2) Changes in the statistical properties of texture elements across space induce the perceived 3D shape of this surface representation. This is achieved in the model through multiple-scale filtering of a 2D image, followed by a cooperative-competitive grouping network that coherently binds texture elements into boundary webs at the appropriate depths using a scale-to-depth map and a subsequent depth competition stage. These boundary webs then gate filling-in of surface lightness signals in order to form a smooth 3D surface percept. The model quantitatively simulates challenging psychophysical data about perception of prolate ellipsoids [Todd, J., & Akerstrom, R. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2), 242–255]. In particular, the model represents a high degree of 3D curvature for a certain class of images, all of whose texture elements have the same degree of optical compression, in accordance with percepts of human observers. Simulations of 3D percepts of an elliptical cylinder, a slanted plane, and a photo of a golf ball are also presented.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Shape; Texture; Neural modeling; 3D vision; Visual cortex; FACADE model; BCS; FCS; Multiple scales; Perceptual grouping; Size-disparity correlation; Filling-in; Shape-from-texture

1. Introduction

Shape-from-texture (SFT) refers to 3D shape perception resulting from the projection of a surface texture onto the viewer's retina. Such projections lead to *texture gradients* in the retinal image that can be used as a cue for shape and depth (Gibson, 1950). Understanding SFT is impor-

tant because we are constantly confronted with a great many textures lying on countless object surfaces with which we interact with on a daily basis. Moreover, such an understanding would help one to create machines that could “see” the 3D world and interact with it, such as a machine that could observe the shape of a terrain (e.g., the Martian surface) and navigate it depending on whether or not the ground ahead is too steep to traverse, or a machine that could observe the shape of an object, enabling the machine to reach out and grasp the object.

This article presents a neural model of SFT called the LIGHTSHAFT (LIGHTness-and-SHApe-From-Texture) model. This model utilizes monocular visual texture information to produce a 3D percept of surface lightness. It is built upon and extends the Form-And-Color-And-DEpth (FACADE) model (Grossberg, 1987a, 1987b, 1994, 1997;

* Corresponding author. Fax: +1 617 353 7755.

E-mail addresses: steve@bu.edu (S. Grossberg), levink@cns.bu.edu (L. Kuhlmann), ennio@cns.bu.edu (E. Mingolla).

¹ Authorship is in alphabetical order. All authors were supported in part by the National Science Foundation (NSF SBE-0354378) and the Office of Naval Research (ONR N00014-01-1-0624). L.K. was also supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0657 and AFOSR F49620-01-1-0005). E.M. was also supported in part by the National Science Foundation (NSF BCS-0235398).

Grossberg & Mingolla, 1987; Kelly & Grossberg, 2000) and the 3D LAMINART model (Cao & Grossberg, 2005; Grossberg & Howe, 2003; Grossberg & Swaminathan, 2004; Grossberg & Yazdanbakhsh, 2005), which explain and simulate a large amount of psychophysical and neurobiological data about 3D vision. FACADE models 3D perceptual grouping, surface perception, and figure-ground separation. 3D LAMINART models the laminar visual cortical circuits that carry out stereopsis and solve the correspondence problem, leading to 3D boundaries and surfaces capable of representing slanted and curved 3D surfaces, and properties of transparency, neon-color spreading, and figure-ground separation. Although these models deal well with stereopsis and a variety of occlusion and transparency situations, they need to be further developed in order for them to account for other depth cues, including SFT. The model presented herein is a step in this direction. The model's main innovation is to further develop FACADE concepts about how multiple scales work together to generate percepts of 3D shape.

Two basic problems must be solved to achieve this end: (1) Patterns of spatially discrete 2D texture elements need to be transformed into a spatially smooth surface representation of 3D shape. (2) Changes in the statistical properties of texture elements across space, such as element density, length and width, need to induce corresponding changes in the perceived 3D shape of this surface representation. In the model, multiple filters of different spatial scale process the 2D image. Several filters can respond to the same texture features, but in different ways. The model clarifies how this ambiguous multiple-scale representation of shape is disambiguated using cooperative and competitive boundary interactions (Grossberg & Mingolla, 1987) that, in concert with scale-to-depth and depth-to-scale maps (Grossberg, 1994), carry out coherent perceptual groupings within depths. Across-depth competition helps to refine the 3D boundary representations. These processes take place within multiple depth-selective boundary webs (Grossberg, 1987a; Grossberg & Mingolla, 1987) before the boundary representations regulate the filling-in of a smooth 3D surface representation. The model's competence is illustrated by simulating the large set of psychophysical data (involving 25 images) from Todd and Akerstrom (1987) about the perception of prolate ellipsoids, as well as 3D shape percepts of 2D images of an elliptical cylinder, a slanted plane, and a golf ball. This model is a first step in the process of developing a neural model of shape-from-texture that can deal with real-world images, where the depth inferred from texture gradients depends on issues of segmentation, reflectance and illumination.

In Section 2, the Todd and Akerstrom experiments are described and LIGHTSHAFT model simulations of their data are summarized. In Section 3, the structure of the model is described. In Section 4, the model is used to explain the Todd and Akerstrom results and to simulate the elliptical cylinder, slanted plane and the golf ball. Section 5 compares the LIGHTSHAFT model with

previous SFT models and supportive neurophysiological data.

2. Todd and Akerstrom (1987) experiments

The Todd and Akerstrom (1987) experiments investigated the perception of curved surfaces and were able to demonstrate that models of surface perception based on local estimates of depth and/or surface orientation (e.g., Stevens, 1981) are, at best, incomplete. We use the phrase “surface-texture” to refer to the distribution of texture on a 3D surface model to be rendered, and “image-texture” to refer to optical characteristics of the projection of surface-texture onto an image plane. In particular, Todd and Akerstrom demonstrated that curvature is perceived in cases where surface-texture element size is varied (i.e., the texture is inhomogeneous on the depicted surface), or when the surface-texture elements are not isotropic. Both of these manipulators would “break” a purely local model. Moreover, they demonstrated results contrary to the findings of Cutting and Millard (1984), who argued that the perception of curvature was largely dependent on changes in image-texture element compression. This was done by creating a stimulus where depth perception of a curved surface can still be achieved for image-texture elements with constant compression. Todd and Akerstrom gave a partial explanation of their results in terms of a precursor to FACADE theory (Grossberg & Mingolla, 1985a, 1985b). This article extends work by Grossberg and Mingolla (1987) that gave a partial qualitative explanation of the Todd and Akerstrom (1987) data by using multiple-scale boundary webs, and provides an explicit FACADE theory implementation that qualitatively explains and simulates Todd and Akerstrom's results.

Stimuli such as those used in the Todd and Akerstrom (1987) experiments are shown in Fig. 1 along with depth maps computed from the LIGHTSHAFT model response. These stimuli were produced following the descriptions of Todd and Akerstrom (1987) and using techniques detailed in Mingolla and Todd (1984) and Todd and Mingolla (1984). In their experiment, Todd and Akerstrom used five image conditions for which the image-texture structure for prolate ellipsoids was varied. The first two conditions correspond to different geometrical projections: *high perspective* (HP) and *low perspective* (LP, which approximates orthographic viewing; Forsyth & Ponce, 2003). The top row of Fig. 1 demonstrates black and white images of HP prolate ellipsoids, where moving from left to right along the row corresponds to an increase in eccentricity of the ellipsoids, and thus an increase in the range of depth perceived in the surface. Just below each ellipsoid image one can see the corresponding colored depth maps computed by the model. Black corresponds to regions of the image where depth is not represented. Going from dark blue, through to green, yellow, and light yellow correlates with a shift from far to near depths. Thus, looking at the HP row of image/depth map pairs, one can see that the model

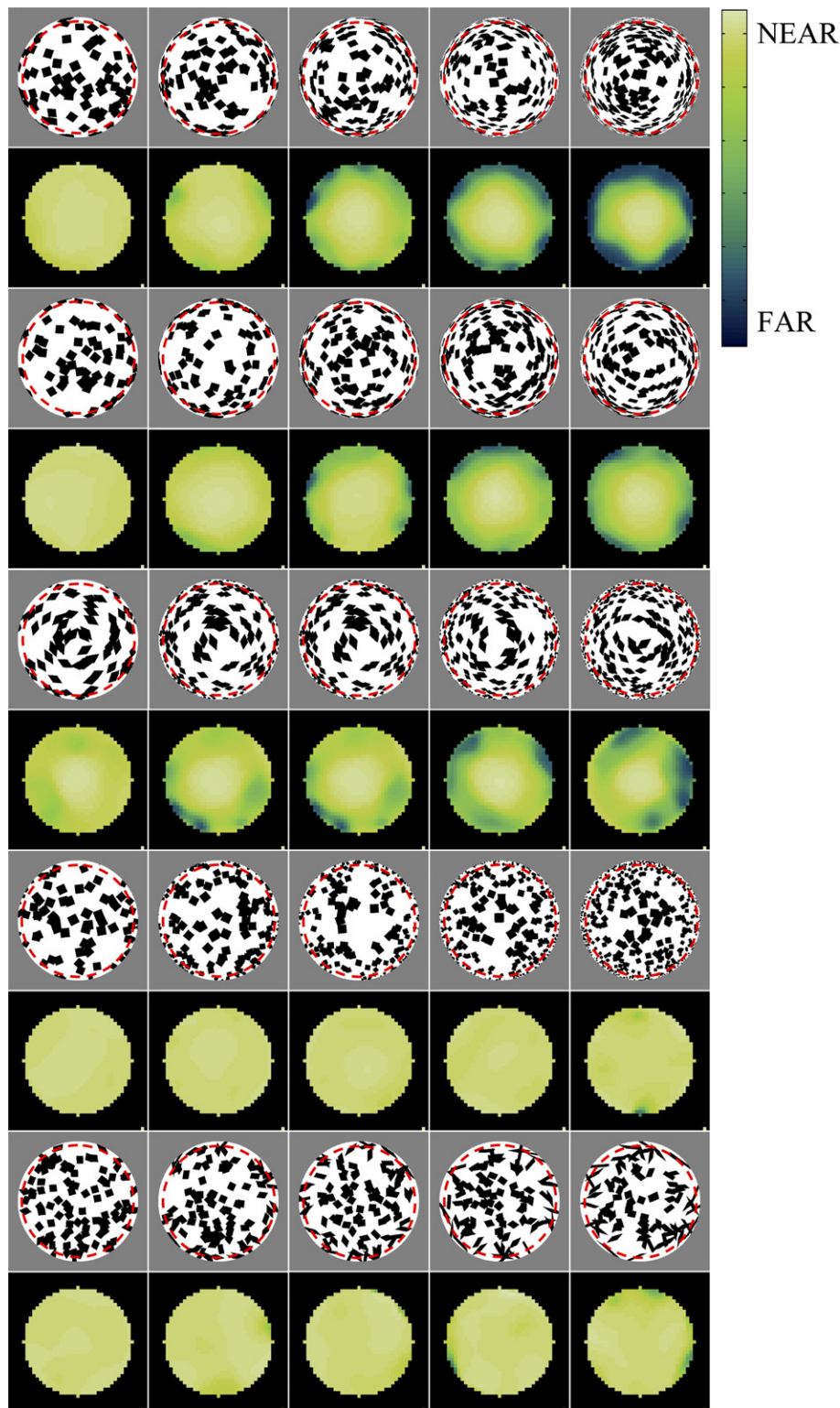


Fig. 1. Todd and Akerstrom (1987) images and the corresponding depth maps computed from the depth competition stage of the model (for the triangular map matrix case as explained in Section 3). Top row: image and depth map pairs for HP case, 2nd row: LP, 3rd row: CCE, 4th row: CCS and bottom row: RO. See text for explanation of image construction and abbreviations. Moving left to right across the rows corresponds to an increase in the simulated depth of the surface in the image. Note that the human perception of depth for this figure is reduced owing to the small size of each image (Todd et al., 2005). In each ellipsoid image, the red dashed circle indicates the boundary of the region within which the depth map was computed in order to avoid spurious effects near the occluding contour, separating the ellipsoid from the large gray background (see Section 3.5). Note that the addition of the dashed circles further diminishes the reader's impression of three-dimensional shape in these figures, which is already attenuated by the small visual angle of each depicted surface, as compared to what Todd and Akerstrom's subjects viewed (see Todd et al., 2005).

accounts for the increase of perceived depth of the ellipsoids when moving from the left to the right.

The second row of image/depth map pairs in Fig. 1 corresponds to the LP condition. One perceives the LP images as showing less relative depth than the HP images, and the depth maps show that the model accounts for this. The three remaining conditions investigated by Todd and Akerstrom are derived from the HP images. In particular, image-texture elements, computed under HP projection, were either forced to have constant compression (i.e., fixed aspect ratios) or random orientations. For image-texture elements with constant compression, they considered two cases. In the *constant compression square* (CCS) condition, image-texture area varied in correspondence with the HP projection, but each of the elements had an identical square shape and random orientation. In the *constant compression elongated* (CCE) condition, patterns were generated in a similar way except that the image-texture element was elon-

gated perpendicular to the tilt direction (i.e., the direction in the image plane along which the distance to the viewed surface increases most rapidly) with a 3:1 compression ratio. The third row of image/depth map pairs in Fig. 1 corresponds to the CCE condition. Even though the texture elements in the CCE images have constant compression, one still sees surfaces in depth and this is accounted for by the depth maps computed from the model. The fourth row of image/depth map pairs in Fig. 1 corresponds to the CCS condition. The CCS images are perceived as flat and it is argued in this paper, through simulation results, that this perceived flatness results from a lack of coherent grouping across concentric bands of texture elements (Grossberg & Mingolla, 1985a, 1985b, 1987; Todd & Akerstrom, 1987). Analysis of the CCE condition is important because it shows, contrary to the findings of Cutting and Millard (1984), that depth perception of a curved surface can still be achieved for a stimulus with constant compression. The

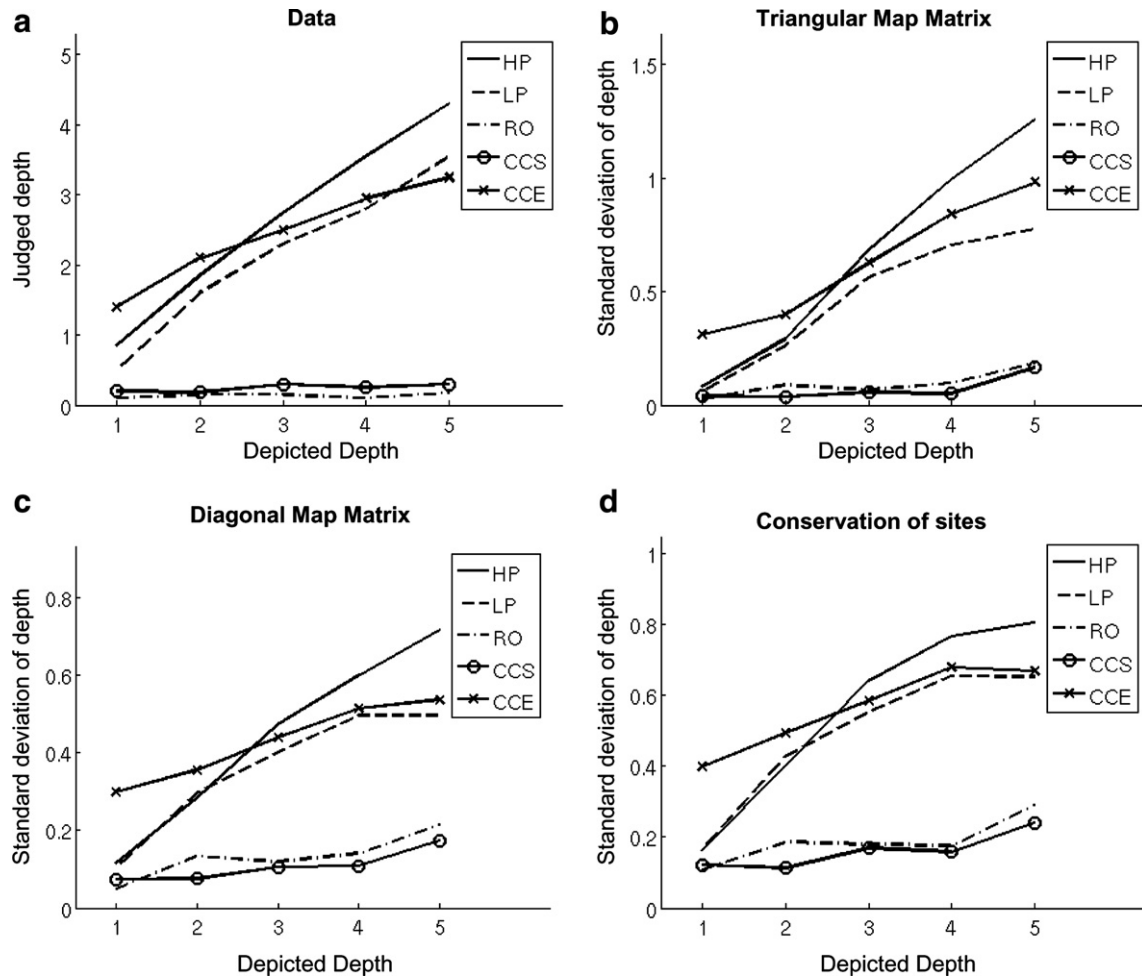


Fig. 2. Comparison of Todd and Akerstrom (1987) judged depth data with the standard deviations of depth computed from the depth maps of the three different model scale-to-depth maps. (a) Todd and Akerstrom (1987) judged depth data where the plotted curves correspond to the five different combinations of projection and texture seen in Fig. 1: HP (solid), LP (dashed), CCE (solid-with-crosses), CCS (solid-with-circles) and RO (dash-dot) conditions. (b), (c) and (d) plot the standard deviations of depth computed from the depth maps of the model with the *triangular map matrix*, the *diagonal map matrix* and *conservation of synaptic sites*, respectively. Section 3 describes the meaning of these 3 model cases. In (a) the y-axis indicates the judged depth of the ellipsoid and in (b), (c) and (d) the y-axis represents the standard deviation of each depth map. The x-axis indicates the simulated depth of the ellipsoid. (Data replotted from Todd and Akerstrom (1987).)

CCS condition further illustrates that this result requires that the image-texture elements be anisotropic (i.e., elongated) and approximately co-linearly aligned.

In the *random orientation* (RO) condition, image-texture shape varied in correspondence with the HP projection, but each of the elements had constant area, was randomly oriented, and the image-texture density was uniform. The final row of image/depth map pairs in Fig. 1 corresponds to the RO condition. As with the CCS images, the RO images are perceived as flat as a result of a lack of coherent grouping.

The main idea behind how the model works is the following: The primary source of depth information in the Todd and Akerstrom stimuli is the variation of texture element widths, *but only those that are elongated and sufficiently aligned with one another so as to form coherent groupings of texture elements*. Coherent groupings at different scales then convey the change in depth of the ellipsoid, assuming that scale and depth are largely correlated, such that large is usually near and small is usually far when a single surface slanted in depth is viewed (Gibson, 1950; Grossberg, 1987b, 1994, 1997; Grossberg & Mingolla, 1987). The HP, LP and CCE conditions all contain sufficiently elongated and aligned texture elements in order for the different scale texture elements to be grouped into the appropriate depths. The CCS and RO conditions have insufficiently

aligned texture elements, and the CCS case has square texture elements; thus any texture element groupings are weak and incoherent, and flat percepts result.

Todd and Akerstrom asked observers to judge the depths of stimuli like those presented in Fig. 1. Observers had to assign a depth value between 1 and 5, inclusive, to each ellipsoid image. Fig. 2a plots the quantitative judged depth data against the simulated depth of the ellipsoids. Considering first the judged depths corresponding to the HP condition (solid curve), one can see that observers underestimate the simulated depth. For the LP condition (dashed) the images are judged to show slightly less depth than for the HP condition. The CCE case (solid-with-crosses) is judged to show depth, but over a narrower range of depths than the high and LP conditions. The CCS (solid-with-circles) and RO (dash-dot) cases give flat judged depth curves. The following sections explain how the model is able to achieve a good match with the Todd and Akerstrom Data.

3. Model

This section describes the LIGHTSHAFT neural model for 3D SFT (see Fig. 3). Each subsection presents a stage of the model. Because of the correspondence that exists between model stages and their analogs *in vivo*, we will

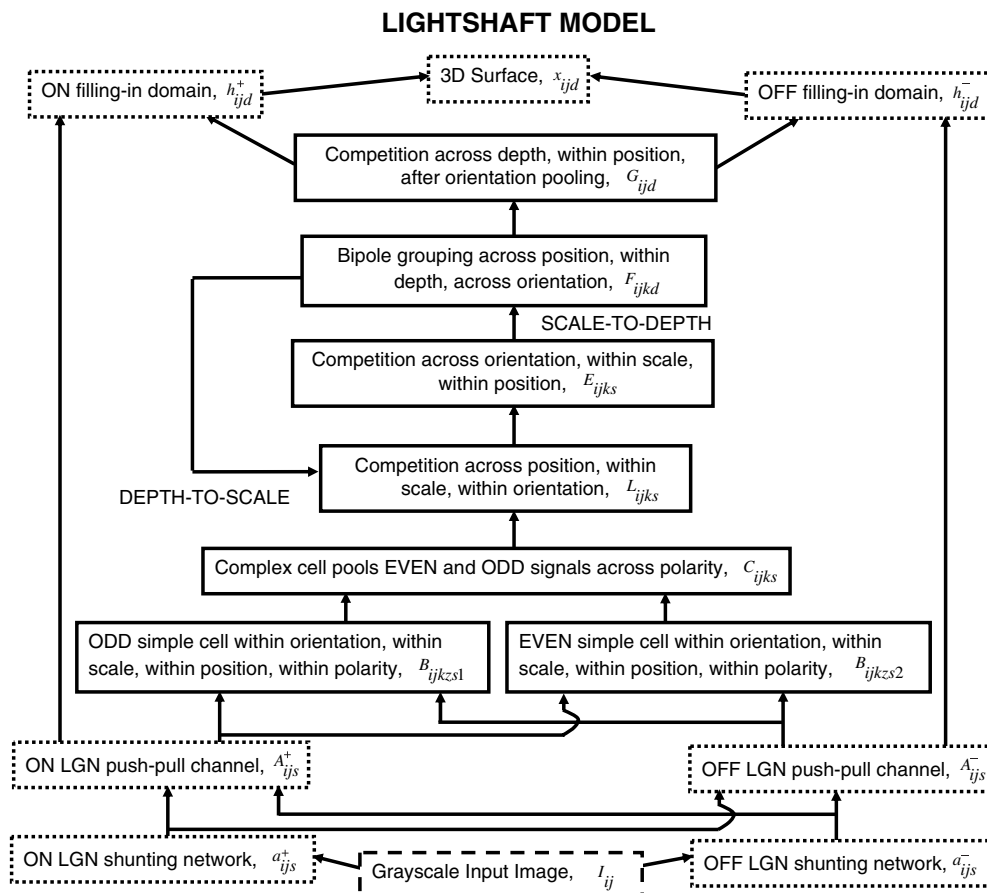


Fig. 3. LIGHTSHAFT Model Diagram. Each stage is denoted by its symbol defined in Appendix A. Boundary Contour System (BCS) stages are surrounded by solid lines, Feature Contour System (FCS) are surrounded by dashed lines. See Section 3 for an explanation of model stages.

henceforth refer to certain model stages by their corresponding physiological terms. Section 5.2.1 of the discussion reviews physiological evidence for each model stage. The stages of the model include: Lateral Geniculate Nucleus (LGN), oriented simple cells, complex cells, spatial competition, orientation competition, bipole grouping cells, filling-in cells, and 3D surface representation cells.

These stages of the model can be divided into two main systems of the FACADE model (Grossberg, 1983, 1994; Grossberg & Todorović, 1988; Kelly & Grossberg, 2000): the *Feature Contour System* (FCS) and the *Boundary Contour System* (BCS). The FCS represents surface lightness signals in depth, while the BCS provides boundaries-in-depth that can trap these surface lightness signals into perceptually appropriate configurations. The LGN provides inputs to both the FCS filling-in stages and the BCS boundary grouping stages. In particular, the LGN stage provides a multiple-scale contrast-enhanced ON and OFF channel representation of the image contrast that can be used both as a foundation for a multiple-depth boundary representation of the image and to provide contrast signals to the filling-in stages. The multiple-scale boundary representation is achieved using the simple cells, and further processing by the remaining BCS stages produces a multiple-depth boundary web (Grossberg & Mingolla, 1987), or form- and scale-sensitive plexus of boundaries, that can be used by the filling-in stages to trap the LGN's multiple-scale contrast signals and form a multiple-depth representation of surface lightness.

This section briefly describes the structure and function of the model, whose processing stages are shown in Fig. 3. Subsequent subsections “unpack” the function of each model component that is mentioned in the present overview. Moreover, we explain how each stage of the model contributes to converting a 2D textured input image into a neural representation of a 3D surface. As already noted, an input image is processed by the multiple-scale ON and OFF channels of the LGN. Multiple-scale ODD and EVEN simple cells of the Primary Visual Cortex (V1) detect local oriented contrasts through the combined processing of the ON and OFF LGN channels. In order to form complex cells in V1, the simple cell activations of opposite contrast polarity are pooled. Multiple-scale complex cells combine odd and even simple cell responses to produce “spatial-phase tolerant” responses. A cooperative-competitive feedback loop in V1, composed of multiple-scale spatial competition cells, multiple-scale orientation competition cells, and multiple-depth bipole grouping cells, realizes the scale-to-depth mapping which aids in the conversion of different sized image-texture elements into a representation of surface depth. The spatial competition cells process the complex cell responses as well as depth-to-scale feedback from the multiple-depth bipole grouping cells in order to spatially sharpen bottom-up complex cell inputs, as well as coherently grouped feedback signals at different scales. Orientation competition cells select the strongest orientations from the spatial competi-

tion cell responses to ensure that the bipole cell groupings are as coherent as possible. Then multiple-depth bipole cells try to form coherent groupings from the multiple-scale orientation-competition cell inputs. The depth competition cells in V2 select the strongest orientation-pooled bipole signals across depth, providing a multiple-depth boundary representation that can be used to describe the depth of the surface. The depth competition cells then gate the filling-in of ON and OFF LGN feature contour signals in the ON and OFF filling-in stages in V4. The final multiple-depth representation of the surface viewed in the image is given by subtracting the OFF filling-in response from the ON filling-in response, thereby providing a lightness representation of the 3D surface in depth.

The LIGHTSHAFT model does not make explicit assumptions about the homogeneity or isotropy of textured surfaces, unlike various other approaches to analyzing SFT (Aloimonos, 1986; Blake, Bülthoff, & Sheinberg, 1993; Blake & Marinos, 1990; Gårding, 1993; Li & Zaidi, 2004; Marinos & Blake, 1990; Rosenholtz & Malik, 1997; Todd & Akerstrom, 1987; Todd, Oomes, Koenderink, & Kappeers, 2004; Witkin, 1981; Zaidi & Li, 2002). While such assumptions are useful for algorithmic purposes and are consistent with certain psychophysical results, there is no neurophysiological evidence that the brain uses such assumptions to estimate shape. Instead, the LIGHTSHAFT model embodies statistical properties of the visual environment in its perceptual filtering and grouping kernels, which enable its scale-to-depth and depth-to-scale mappings to convert a multiple-scale representation of a 2D image into a multiple-depth representation of the shape perceived from that image. These statistical properties of the filtering and grouping kernels are asserted in a heuristic manner (Grossberg, 1994), are supported by neurobiological evidence (DeAngelis, Ohzawa, & Freeman, 1995; Kapadia, Westheimer, & Gilbert, 2000; Raizada & Grossberg, 2001; von der Heydt, Peterhans, & Baumgartner, 1984), and are assumed to be derived through a process of learning and development (Grossberg & Williamson, 2001). The model hereby makes use of the idea that the scale distribution/spatial frequency spectrum in an image is correlated with changes in 3D surface depth (Grossberg, 1987b, 1994; Sakai & Finkel, 1995, 1997). For the LIGHTSHAFT model, 6 scales, 16 orientations, and 6 depths were used. The choice of 16 orientations was made in order to ensure proper tracking of oriented contrast distributions for curved surfaces. The 6 scales and 6 depths were selected in order to provide a sufficiently dense discretization of scale and depth for the purposes of this study. The paragraphs that follow describe the biological plausibility, structure and purpose of each stage of the model.

3.1. ON and OFF channels of the LGN

Retinal processing contributes to the determination of an absolute lightness scale through two processes: light adaptation (Grossberg, 1983; Grossberg & Hong, 2006;

Martin, 1983; Werblin, 1971) and contrast adaptation (Baccus & Meister, 2002; Demb, 2002; Grossberg & Hong, 2006). The LGN then generates contrast signals using multiple-scales of antagonistic ON-center OFF-surrounds and OFF-center ON-surrounds (Grossberg & Hong, 2006; Hubel & Wiesel, 1961).

This article deals only with static images with similar lightness ranges. A retinal stage is not included in the model because it does not have to account for adaptation effects. Instead the LGN stage processes the image directly, producing a multiple-scale contrast-enhanced representation of the image. The LGN stage consists of six spatial scales in order to sufficiently track the changes in the size of texture in the image. The different scale representations are extracted using both ON-center OFF-surround (ON units) and OFF-center ON-surround (OFF units) kernels. The ON units are excited when light signals fall in the center of their receptive field and inhibited when light falls in their surround. OFF units react in the opposite way (Schiller, 1992). For each scale, the center-surround kernels have fixed narrow centers so as to preserve fine image lightness information for the purposes of filling-in (Grossberg & Hong, 2006; Grossberg, Mingolla, & Williamson, 1995; Mingolla, Ross, & Grossberg, 1999), while surround sizes differ in order to provide the different scale sensitivity.

The LGN stage is simulated using feedforward shunting equations (Grossberg, 1973, 1983). These equations, defined in Appendix A.1, provide a formulation for the ON and OFF channels and their push-pull interactions which are common to the LGN and V1 layer 4 (Livingstone & Hubel, 1984, 1988). The relative sizes of the LGN center-surround filters are presented in Appendix B. As mentioned above, the purpose of the LGN stage is to provide a multiple-scale contrast-enhanced representation of image contrast that can be used to produce both a multiple-depth boundary representation of the image and signals to drive the surface filling-in stage. Fig. 4 sche-

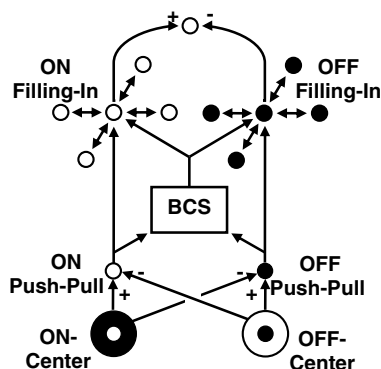


Fig. 4. Schematic of model circuit. LGN ON and OFF channels first process the image using center-surrounds and then push-pull interactions (see Appendix A.1). Their activations project to both the BCS and the ON and OFF filling-in channels. The BCS gates filling-in of the LGN lightness signals at the appropriate depths. The OFF filling-in response is subtracted from the ON filling-in response to obtain the 3D surface representation.

matizes how the ON and OFF LGN cells of the FCS provide input to both the BCS and the ON and OFF filling-in channels of the FCS. The LGN ON and OFF channels first process the image using center-surrounds, and then push-pull interactions spatially sharpen the contrast signals produced by the center-surrounds. In response to the LGN input, the BCS computes a multiple-depth boundary web that gates filling-in within ON and OFF filling-in networks whose responses are used to create a 3D surface lightness representation. Fig. 5 demonstrates how the ON and OFF LGN cells provide input to the model BCS by projecting to the odd and even simple cells.

3.2. Simple cells

One of the first stages of V1 cortical processing is performed by oriented simple cells (Hubel & Wiesel, 1959, 1962). These cells have elongated excitatory and inhibitory zones that form an oriented receptive field. This allows simple cells to produce a multiple-scale boundary representation of the image by processing the multiple-scale un-oriented signals of the LGN (Grossberg, 1983; Grossberg & Todorović, 1988; Hubel & Wiesel, 1962). In the LIGHTSHAFT model, simple cells have multiple-scale odd-symmetric and even-symmetric receptive fields which are *self-similar*, that is, for a given receptive field symmetry

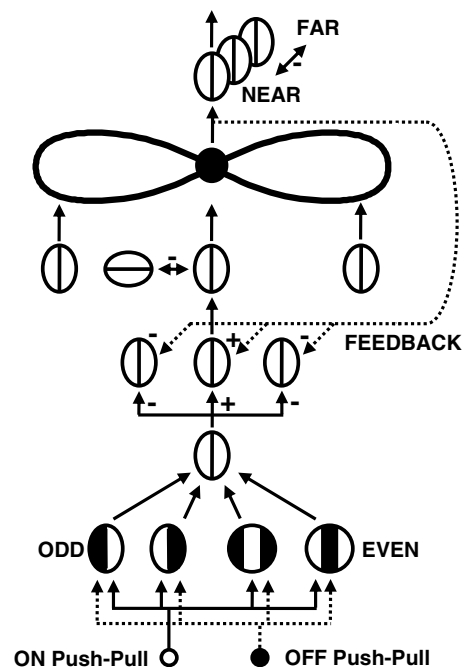


Fig. 5. Schematic of the BCS circuit. ODD and EVEN simple cells process the ON and OFF LGN input. Complex cells pool the simple cell responses. Spatial competition cells spatially sharpen the complex cell response. Orientation competition cells then orientationally sharpen the spatial competition cell responses. Bipole cells group the orientation competition signals, then feed back to the spatial competition cells to selectively enhance the groupings. Finally depth competition cells refine the bipole cell depth representation, before projecting to the ON and OFF filling-in channels.

(odd or even), different sized fields have the same structure, up to a uniform scaling of the two-dimensions of the field (Grossberg, 1983). Each receptive field consists of ON- and OFF-subregions. The ON-subregions receive excitatory ON LGN signals and inhibitory OFF LGN signals, while the OFF-subregions have the converse relation to the LGN channels (Hirsch, Alonso, Reid, & Martinez, 1998; Raizada & Grossberg, 2001; Reid & Alonso, 1995). This structural organization is embodied by the convolution of the ON- and OFF-subregions with their corresponding LGN inputs. Sixteen discrete orientations are used in order to be able to track the changes in orientation in the Todd and Akerstrom stimuli and to approximate the continuum of orientation selectivity observed for simple cells (Ringach, Shapley, & Hawken, 2002). Each receptive field is defined by either a Difference-Of-Gaussians (DOG, even-symmetric) or a Difference-Of-Offset-Gaussians (DOOG, odd-symmetric) filters (Grossberg & McLoughlin, 1997; Malik & Perona, 1990; Parker & Hawken, 1988). Moreover, these filters are sensitive to polarity of contrast. The responses of the ON- and OFF-subregions are combined to provide the multiple-scale oriented contrast-polarity-sensitive simple cell outputs (Raizada & Grossberg, 2001).

The simple cells are the main means of tracking changes in scale of image texture, because their receptive fields are self-similar; that is, they look the same up to a uniform scaling. As a result, larger simple cells need more input to fire as vigorously as smaller simple cells. The simple cells, however, provide only a local representation of scale in the image (Sakai & Finkel, 1995, 1997). Without further processing, it would be hard to generate a depth representation that is not sensitive to local scale changes or image noise since individual simple cells have high or low activity depending on how they align with individual texture elements. As a result, subsequent grouping and competition stages are necessary for estimating depth from the image. The simple cell equations are presented in Appendix A.2. The relative sizes of the simple cell DOG and DOOG filters are presented in Appendix B. Fig. 5 schematizes how the odd and even simple cell responses are derived from the ON and OFF LGN cell responses.

3.3. Complex cells

The first and simplest model of complex cells in V1 was presented by Hubel and Wiesel (1959, 1962). They demonstrated that complex cells were oriented, contrast-polarity insensitive and also phase-insensitive. They argued that these properties arose from the spatial pooling of oriented simple cells with opposite contrast-polarity sensitivities. This model has been incorporated as one piece of more elaborate filtering and boundary grouping models published by Grossberg and colleagues (Grossberg, 1983, 1987a, 1987b; Grossberg & Howe, 2003; Grossberg & Mingolla, 1987; Grossberg & Todorović, 1988; Kelly & Grossberg, 2000). However, these models typically combine

spatial-phase insensitivity with a spatial competition stage which performs competition both *across-position* and *within-orientation*. The spatial competition stage usually occurs after the pooling of opposite contrast-polarity simple cell signals into complex cell receptive fields at each point in the image; cf., Fig. 3.

In the LIGHTSHAFT model, complex cells pool both odd-symmetric and even-symmetric simple cells of opposite contrast-polarities at each point in the image (Grossberg & McLoughlin, 1997; Raizada & Grossberg, 2001). This pooling is defined in Appendix A.3. Combining both odd and even simple cells gives a more complete estimate of the local changes of scale in the image. Fig. 5 demonstrates how the odd and even simple cell responses are pooled to generate complex cell responses.

3.4. Cooperative-competitive feedback loop

The complex cells provide a multiple-scale-and-orientation boundary representation of the image. How can this boundary representation be converted into a depth representation of the image? Complex cells are sensitive to binocular disparity, but there are no binocular disparity cues in response to a monocularly-viewed textured object. Grossberg (1994) analyzed how the brain exploits the size-disparity correlation (Julesz & Schumer, 1981; Kuffler, 1978; Prince, Cumming, & Parker, 2002; Prince & Eagle, 1999; Richards & Kaye, 1974; Schor & Tyler, 1981; Schor & Wood, 1983; Schor, Wood, & Ogawa, 1984; Smallman & MacLeod, 1994; Tyler, 1975, 1983) to generate a 3D representation of a 2D scene by pooling selective responses of multiple scales of complex cells to generate a 3D boundary representation that is capable of supporting a 3D representation of surface lightness. This is accomplished by a (multiple scale)-to-(multiple depth) mapping that combines output from multiple scales to compute the total evidence for an object at a prescribed depth. This scale-to-depth mapping is embodied within a cooperative-competitive feedback loop (Grossberg, 1994; Grossberg & Mingolla, 1985a, 1985b, 1987) that coherently groups texture-element signals into the appropriate depths, thus helping to solve the problem of how to convert oriented responses to discrete texture elements in the image into a smooth three-dimensional surface representation of that image. Spatial and orientation competition stages select the input signals which have more statistical support, while the bipole cell grouping stage creates coherent groupings of the selected signals. The feedback loop takes the complex cell responses as input and involves the following sequence: competition across position → competition across orientation → bipole grouping → competition across position (see Fig. 3). The two competition stages are multiple-scale, while the bipole grouping stage is multiple-depth. Thus when the orientation competition cells project to the bipole grouping cells there is a *scale-to-depth map*, and when the bipole cells feedback to the spatial competition cells there is a *depth-to-scale map* (see Fig. 3). This framework results in bipole

cell responses that provide a multiple-depth-and-orientation boundary representation of the surface and helps to determine the final 3D surface representation.

While interactions between scale and depth have been implemented within the FACADE framework (Kelly & Grossberg, 2000), this article investigates the robustness of the LIGHTSHAFT model with novel implementations of three different scale-to-depth maps: *triangular map matrix*, *diagonal map matrix* and *conservation of synaptic sites*. Moreover, consideration of the performance resulting from a variety of scale-to-depth maps could shed some light on the, poorly understood, physical structure of the scale-to-depth map used in the brain. Since the model involves feedback, the scale-to-depth map requires both a scale-to-depth map matrix for the 6 orientation competition scales to project to the 6 bipole depths, and a depth-to-scale map matrix for the 6 bipole depths to feed back to the 6 spatial competition scales.

The triangular map matrix case involves scale-to-depth and depth-to-scale map matrices in forms that are consistent with the size-depth correlation (Brown & Weisstein, 1988; Grossberg, 1994; Sakai & Finkel, 1995) and the size-disparity correlation (Julesz & Schumer, 1981; Kuffler, 1978; Prince et al., 2002; Prince & Eagle, 1999; Richards & Kaye, 1974; Schor & Tyler, 1981; Schor & Wood, 1983; Schor et al., 1984; Smallman & MacLeod, 1994; Tyler, 1975, 1983). The size-depth correlation is important for SFT since variations in spatial frequency (i.e., size) are correlated with variations in perceived depth for SFT stimuli. The basic idea behind the size-depth correlation is that, for continuous surfaces that can be segmented from a visual scene, large surface features (i.e., low spatial frequencies) are typically perceived to be nearer than small surface features (i.e., high spatial frequencies). The size-disparity correlation, defined in psychophysical terms, is the relationship between the disparity range for binocular stereopsis and spatial frequency. Essentially, low spatial frequencies can be fused over a larger range of disparities than high spatial frequencies.

In the triangular map matrix case, the furthest depth receives inputs from all scales with the strongest connections coming from the smaller scales. Nearer depths progressively receive less input from smaller scales than further depths receive. This occurs to the point where the nearest depth only receives input from the largest scale. The exact values of the scale-to-depth and depth-to-scale map matrices, W_{ds} and W_{sd} , respectively, used by the model are presented in Appendix C. The diagonal values of the scale-to-depth and depth-to-scale map matrices are stronger than the other weights of the matrix. This indicates that the scale-to-depth map is biased by the size-depth correlation since small scales are predominately mapped to far depths and large scales are predominantly mapped to near depths.

Although LIGHTSHAFT does not model binocular vision, aspects of SFT are related to mechanisms for perceiving binocular depth. Indeed, FACADE theory laid a foundation for understanding how 3D percepts of 2D

images may be generated by using the brain's mechanisms for perceiving the binocularly viewed world in depth (Grossberg, 1994, 1997). In particular, both binocular vision and SFT depend on multiple-scale processing. The size-disparity correlation posits that large scales can fuse over both large and small disparities, while small scales can only fuse over small disparities. This can be taken to mean that large scales provide input to a larger range of depths than small scales. The off-diagonal weights of the scale-to-depth and depth-to-scale map matrices take this idea into account by allowing the larger scales to provide input to more depths than do the smaller scales.

The conservation of synaptic sites case is a variation of the triangular matrix map case where the sum of the input weights projecting to each depth in the scale-to-depth map, or each scale in the depth-to-scale map, is constant across all depths and scales, thus conserving the number of synaptic sites attaching to a bipole cell of a given depth, or a spatial competition cell of a given scale. This case is consistent with the size-depth and size-disparity correlations in a similar way as the triangular map matrix case. For this case, the values of the scale-to-depth and depth-to-scale map matrices, W_{ds} and W_{sd} , respectively, used by the model are presented in Appendix D.

Fig. 6a illustrates a schematic of the scale-to-depth map for the triangular map matrix and conservation of sites cases where the multiple-scale cells project to the multiple-depth bipole grouping cells, which in turn provide feedback to the multiple-scale cells, thus completing the cooperative-competitive feedback loop. The left bipole cell (labeled as an infinity-type symbol) corresponds to the farthest depth, and in accordance with the scale-to-depth map of these cases, receives input from, and provides feedback to, cells of all the scales (indicated by ellipses). The middle bipole cell corresponds to a mid-range depth and receives input from, and provides feedback to, only the larger scales. The right bipole cell corresponds to the nearest depth and receives input from, and provides feedback to, only the largest scale.

In the diagonal map matrix case, scale maps to depth in a one-to-one manner that is consistent with the size-depth correlation, but does not try to consider the size-disparity correlation. For this case, the values of the scale-to-depth and depth-to-scale map matrices, W_{ds} and W_{sd} , respectively, used by the model are presented in Appendix E. Fig. 6b illustrates these maps for the diagonal map case. Note that each bipole cell receives input from, and feeds back to, only one scale. The far bipole connects to the small scale, the middle depth to the middle scale, and the near bipole to the large scale.

Each stage of the cooperative-competitive feedback loop is computed using the equilibrium form of the model equations. The feedback loop is repeated until equilibrium is reached, through a process of relaxation. This speeds up simulation times compared to when numerical integration is used. Moreover, in order to speed up simulations the complex cell response was sub-sampled, by taking every

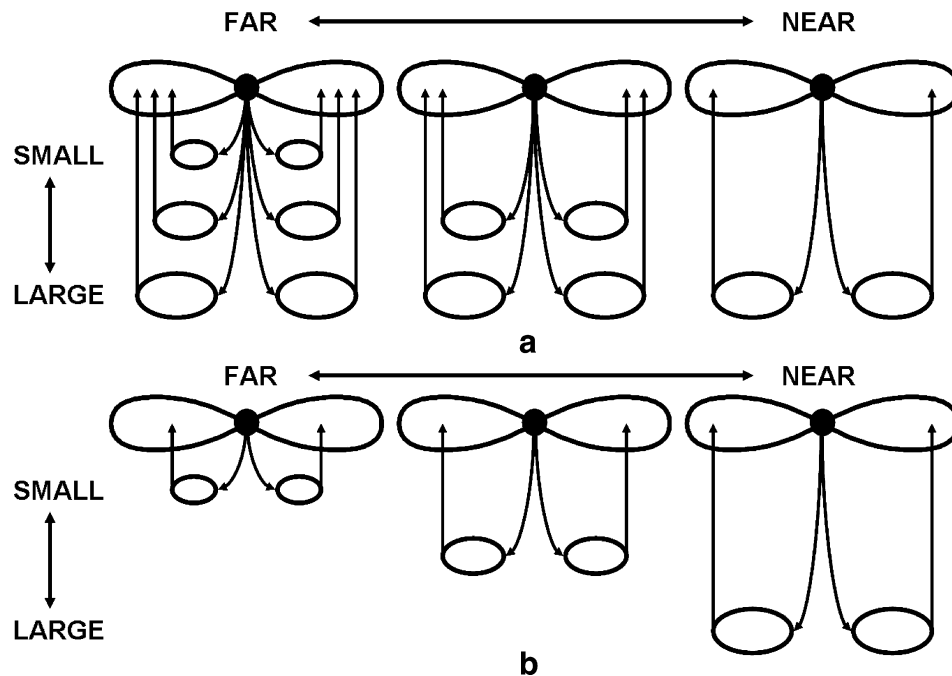


Fig. 6. (a) Schematic of scale-to-depth map for the triangular map matrix and conservation of sites cases where the multiple-scale cells (indicated by ellipses) project to the multiple-depth bipole grouping cells (indicated by infinity-type symbol), which in turn provide feedback to the multiple-scale cells, thus completing the cooperative-competitive feedback loop. The left bipole cell corresponds to the farthest depth, and in accordance with the scale-to-depth map of these cases, receives input from, and provides feedback to, all of the scales. The middle bipole cell corresponds to a mid-range depth and receives input from, and provides feedback to, only the larger scales. The right bipole cell corresponds to the nearest depth and receives input from, and provides feedback to, only the largest scale. (b) For the diagonal map matrix case the far bipole connects to the small scale, the middle depth to the middle scale, and the near bipole to the large scale.

12th pixel along both of the image dimensions. This means that the complex cell array (see Appendix A.3) which had original dimensions $481 \times 481 \times 16 \times 6$ was reduced to the size $41 \times 41 \times 16 \times 6$. These dimensions are also used by the spatial competition, orientation competition, bipole grouping and depth competition cells. When the depth competition cell responses are projected to the filling-in stage they are resized back to the original dimensions using nearest neighbors interpolation in order to preserve image detail in the final 3D surface representation.

3.4.1. Spatial competition cells

The spatial competition cells (Gove, Grossberg, & Mingolla, 1995; Grossberg & Mingolla, 1985b, 1987; Grossberg et al., 1995) are modeled as hypercomplex cells in V1. See Appendix A.4. They are represented by a multiple-scale shunting neural network which performs competition both *across position* and *within orientation*. Moreover, the spatial competition stage uses radially-symmetric kernels of different scale that are self-similar. See Appendix B for these kernels. Each scale of the spatial competition stage spatially sharpens the combination of bottom-up inputs from multiple-scale complex cells and feedback from multiple-depth bipole cells. Fig. 5 illustrates how the spatial competition cells receive lateral inhibition from the bottom-up complex cell input and top-down bipole cell feedback. This diagram depicts feedback from one bipole cell. However, in the model, a spatial competition cell of a given scale can

receive feedback from bipole cells of up to 6 depths depending on the scale-to-depth map (schematized in Fig. 6 and detailed in Appendices C–E).

The spatial competition cell responses are passed onto the multiple-scale orientation competition cells (Fig. 3). The scale-to-depth map occurs between the multiple-scale orientation competition cells and the multiple-depth bipole cells. Feedback from the bipole cells to the spatial competition cells carries out the depth-to-scale map (Fig. 3).

3.4.2. Orientation competition cells

The orientation competition cells (Gove et al., 1995; Grossberg & Mingolla, 1985a, 1985b, 1987; Grossberg et al., 1995) are also modeled as hypercomplex cells in V1. They are represented by a multiple-scale shunting neural network where competition occurs both *across orientation* and *within position*. This competitive interaction is the same for all scales. The multiple-scale orientation competition stage preserves strong-orientation signals from the spatial competition stage while inhibiting weaker orientation signals, thus helping the bipole cells to form more coherent groupings. The outputs of the orientation competition cells project to the multiple-depth bipole cells through the scale-to-depth map (see Appendices C–E). The equations for the orientation competition cells are presented in Appendix A.5. Fig. 5 shows how the orientation competition cells apply cross-orientation inhibition to the spatial competition cell outputs.

3.4.3. Bipole grouping cells

Bipole cells realize the grouping of texture elements necessary for a coherent depth representation (Grossberg & Mingolla, 1985a, 1985b, 1987; Grossberg et al., 1995; Grossberg & Swaminathan, 2004; Mingolla et al., 1999; Raizada & Grossberg, 2001; see also Field, Hayes, & Hess, 1993; Heitger, von der Heydt, Peterhans, Rosenthaler, & Kubler, 1998; Kellman & Shipley, 1991). The bipole cells are modeled as V2 neurons in layer 2/3 (Bosking, Zhang, Schofield, & Fitzpatrick, 1997; McGuire, Gilbert, Rivlin, & Wiesel, 1991; Schmidt, Goebel, Löwel, & Singer, 1997; von der Heydt et al., 1984). In particular, bipole cells can form a boundary web (Grossberg & Mingolla, 1987) through cooperative grouping and competitive selection of boundaries. Boundary webs help to trap lightness signals within appropriate surfaces in depth.

In the model, the multiple-scale orientation competition cells project to the multiple-depth bipole cells (Fig. 5). Boundary grouping then occurs in each depth using the bipole filter defined in Appendix A.6.1 and presented in Appendix B. The bipole cell formulation described in Appendix A.6 is based on the formulations used in Gove et al. (1995), Grossberg and Mingolla (1987), and Raizada and Grossberg (2001). For each bipole cell, there are two inhibitory interneurons, or interneuronal populations (Grossberg & Howe, 2003; Grossberg & Swaminathan, 2004; Grossberg & Yazdanbakhsh, 2005), rather than just one (Raizada & Grossberg, 2001). This hypothesis makes the bipole cell groupings more robust to parameter changes. Although the bipole cells provide a multiple-depth-and-orientation boundary representation of the image, this representation is enhanced by the depth competition stage which receives the orientation-pooled bipole cell activations as input. Fig. 5 depicts input from orientation competition cells of a single scale to a bipole cell of a single depth. In the model, an orientation competition cell of a given scale can connect to bipole cells of up to 6 depths depending on the scale-to-depth map (schematized in Fig. 6 and detailed in Appendices C–E).

3.5. Depth competition cells

The depth competition cells are also modeled as V2 neurons. They are represented by a multiple-depth shunting network with uniform competition *across depth* and *within position*. The depth competition cells process the orientation-pooled bipole cell responses to give a multiple-depth boundary web which represents the BCS output. These cells project to the ON and OFF filling-in domains, or FIDOs, to gate the filling-in of surface feature signals projected from the LGN. The equations describing the depth competition cells are defined in Appendix A.7. Fig. 5 illustrates how the depth competition cells apply across-depth inhibition to their bipole cell inputs. This diagram depicts only input from one bipole cell.

In the model, a depth competition cell of a given depth receives excitation and inhibition from bipole cells of all 6

depths. The depth maps, as shown in Fig. 1, are computed from the depth competition response by determining the activation-based weighted average of depth at each pixel. For the pixel at position (i, j) in the image plane, the activation-based weighted average of depth, \bar{d}_{ij} , is given by:

$$\bar{d}_{ij} = \sum_{d=1}^6 d \left(\frac{G_{ijd}}{\sum_{y=1}^6 G_{ijy}} \right), \quad (1)$$

where G_{ijd} is the activation of the depth competition cell at position (i, j) and depth d , and y is a dummy index. See Appendix Eq. (A.57). Moreover, the depth maps are computed only in the region of the image where the continuous surface is present. This assumes that there is a figure-ground segmentation process going on that separates the textured surface in the image from the background; cf., Grossberg (1994, 1997).

3.6. ON and OFF filling-in domains and the 3D surface representation

The ON and OFF FIDOs are modeled as V4 neurons. Filling-in has typically been modeled by a diffusion network (Cohen & Grossberg, 1984; Grossberg & Hong, 2006; Grossberg & Todorović, 1988; Hong & Grossberg, 2004; Pessoa, Mingolla, & Neumann, 1995). However, alternatives to a physical diffusion process in the form of long-range interactions for filling-in have been proposed and shown to run 1000 times faster (Grossberg & Hong, 2006; Hong & Grossberg, 2004). The ON and OFF FIDOs in the present model are based on the diffusion network used by Grossberg and Todorović (1988). The multiple-depth boundary webs computed by the BCS trap the diffusion of surface lightness signals generated from the LGN output. This boundary-gated surface filling-in concept has been used to describe many psychophysical data about brightness, color, and 3D figure-ground perception (Grossberg, Hwang, & Mingolla, 2002; Grossberg & Mingolla, 1985a; Grossberg & Yazdanbakhsh, 2005; Kelly & Grossberg, 2000).

When projecting the multiple-scale ON and OFF LGN signals to the multiple-depth ON and OFF filling-in channels, the scale-to-depth map described for the BCS is not used. Instead, scale-pooled LGN signals are projected to each depth. Fig. 4 schematizes how the ON and OFF FIDOs receive input from the LGN cells and the BCS. In Fig. 4, there are LGN cells of only one scale and filling-in cells of only one depth. In the model, 6 LGN scales project to each filling-in depth and each filling-in depth receives gating signals from its corresponding BCS depth.

Before inputting to the filling-in stages, the BCS multiple-depth representation computed on a sub-sampled image grid is resized back to the original image grid in order to preserve image detail. Multiple-depth filling-in of LGN signals then occurs within these depth-selected boundaries. Within each depth and position, the OFF filled-in response is subtracted from the ON filled-in

response to obtain the 3D surface representation, that is interpreted to lie in V4 (Schiller, 1994, 1995; Schiller & Lee, 1991; Zeki, 1983a, 1983b). The 3D surface representation is a multiple-depth representation of the surface perceived in the input image. The equations describing the ON and OFF FIDOs are defined in Appendix A.8, while the equation describing the 3D surface representation is defined in Appendix A.9.

4. Computer simulations

This section discusses how the LIGHTSHAFT model fits the Todd and Akerstrom (1987) data seen in Figs. 1 and 2. Simulations of an elliptical cylinder, slanted plane and golf ball image are also summarized.

4.1. LIGHTSHAFT model simulations of the Todd and Akerstrom (1987) data

Generating a 3D surface percept from a 2D input image relies on scale-to-depth and depth-to-scale maps (described in Section 3), which help to transform a multiple-scale representation of the 2D image into a multiple-depth representation of the image. To demonstrate the robustness of the LIGHTSHAFT model framework, three cases of the scale-to-depth and depth-to-scale maps were simulated: a triangular map matrix, a diagonal map matrix and conservation of synaptic sites (see Section 3 and Appendices C–E). Fig. 1 shows the depth maps computed using the triangular map matrix case. Fig. 2b–d display the standard deviation of the model's depth maps for each of the 25 ellipsoid images presented in Fig. 1 computed using the triangular map matrix, diagonal map matrix and conservation of sites cases, respectively. Comparing Fig. 2a with Fig. 2b–d shows that there is a good qualitative match of the curves for all of the Todd and Akerstrom stimulus cases and scale-to-depth map cases. For the HP, LP and CCE conditions, the model-data matches arise because the model is able to form groupings of the oriented energy of sufficiently aligned and elongated texture elements. These groupings form in different depths depending on the different scales of the texture elements, thus giving rise to model percepts with strong variations in depth (see Fig. 1; 1st row of image/depth map pairs: HP; 2nd row: LP; 3rd row: CCE). For the CCS and RO conditions, the model-data matches arise because the texture elements are either not aligned or isotropic. Thus there is little ability to form strong groupings. This allows the near depth to dominate at the depth competition stage of the model, leading to flat model percepts (see Fig. 1; 4th row: CCS; 5th row: RO).

4.1.1. Simulations of the high perspective condition

In order to better understand how the model works, consider first the number 5 HP condition (top right image in Fig. 1) simulated using the triangular map matrix case. Fig. 7a and b show the LGN ON and OFF channel responses, respectively. Each panel in each figure corre-

sponds to a scale. The top left panel corresponds to the smallest scale, then moving right along the rows and down along the columns corresponds to an increase in scale. For each panel, the brighter the image at a given point, the stronger the activation at that point. Moreover, each panel is 481×481 pixels in size. These LGN stages provide a multiple-scale, contrast-enhanced representation of the image, where each scale response has been computed using center-surround kernels that have a fixed center size, but a different surround size depending on the scale.

Fig. 7c shows the sub-sampled multiple-scale-and-oriented complex cell responses that pool odd and even simple cell responses to the ON and OFF LGN channels (Section 3.4 explains the need for sub-sampling). The complex cell responses are plotted using needle plots, wherein a longer needle of a given orientation indicates greater activation for that orientation. For each panel, there are 41×41 needle sites corresponding to the cell locations of that scale. This low spatial resolution version of the complex cell responses helps one to better visualize the prevalence of oriented energy at different image locations. The bottom panel helps one to see this more clearly by showing a zoomed in version of the largest scale response panel (3rd down on the right). From the complex cell response one can see that there is a significant amount of oriented energy in the direction perpendicular to the direction of tilt, especially in the smaller scales. The goal of the cooperative-competitive feedback loop is to take advantage of this appropriately aligned energy to form circular bands of grouping within the appropriate depths.

Fig. 8a displays the multiple-scale-and-oriented spatial competition cell equilibrium responses. These cells spatially sharpen both the bottom-up complex cell responses as well as the depth-to-scale mapped bipole grouping cell responses using center-surrounds that vary with scale. Here, small scales are sharpened on a small spatial scale while large scales are sharpened on a large spatial scale. This can be seen mainly for scales 3–5 in Fig. 8a where the thickness of the bands increases with an increase in scale.

Fig. 8b illustrates the equilibrium responses of orientation competition cells, notably their multiple-scale-and-oriented properties. These cells orientationally sharpen the spatial competition cell responses using center-surrounds across orientation that are the same for all scales. This sharpens the orientation representation of the model by eliminating weak orientation signals at each point in the image. This can be observed through the reduced variation of orientations seen when comparing Fig. 8b of the orientation competition cells with Fig. 8a of the spatial competition cells.

This orientation sharpening helps to make more coherent circular groupings within depth, as can be seen in Fig. 8c where the multiple-depth-and-oriented bipole grouping cell responses are displayed. As depth gets nearer, the circular bands become thicker and the peak activation of the bands shifts towards the center of the image; cf., Grossberg and Mingolla (1987). These strong bands arise

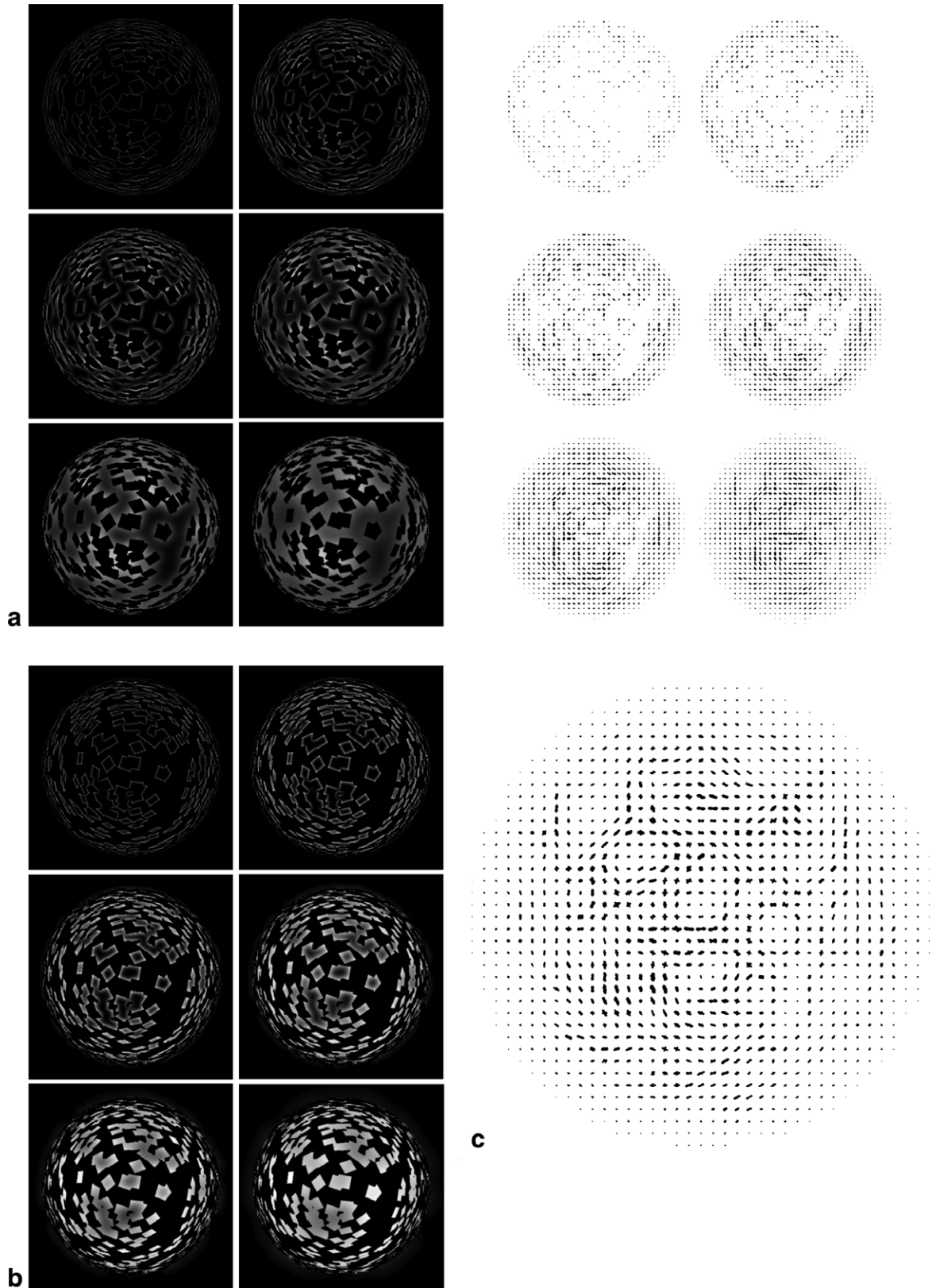


Fig. 7. Number 5 HP case: (a) *LGN ON response*, (b) *LGN OFF response*, and (c) *complex cell response*. In (a), (b) and (c), the top left panel corresponds to the smallest scale, then moving right along the rows and down along the columns corresponds to an increase in scale. The bottom panel in (c) shows a zoomed in version of the largest scale response (3rd panel in (c) down on the right). For each panel with gray scale images, here and in Figs. 8–14, the brighter the image at a given point the stronger the activation at that point. For each panel in (a) and (b), and in the filling-in stage figures, the dimensions are 481×481 pixels. For each panel in (c) and in the needle plots in Figs. 8–14, a longer needle of a given orientation indicates greater activation for that orientation. Moreover, for each panel here and the other figure needle plots, there are 41×41 needle sites.

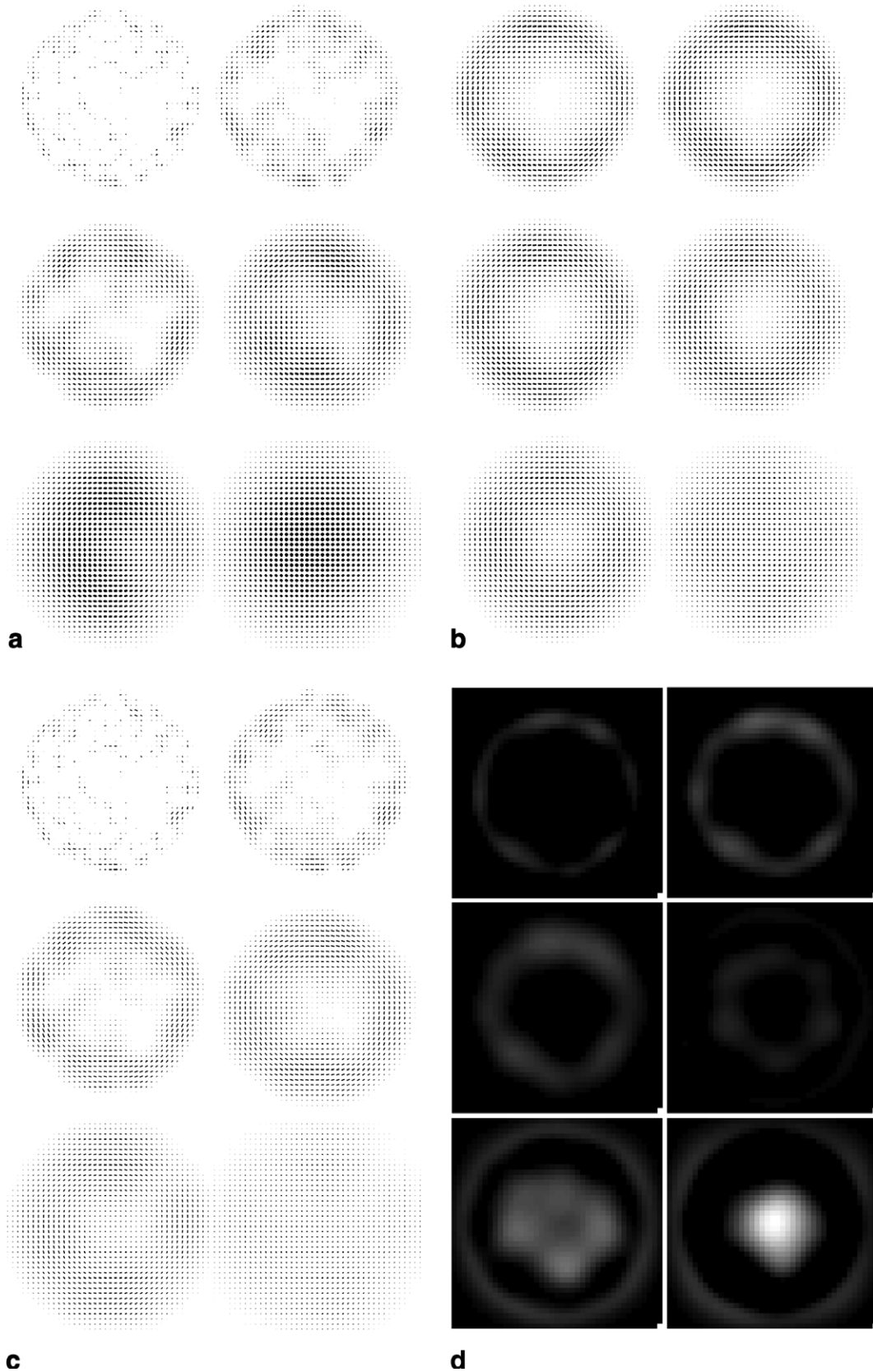


Fig. 8. Number 5 HP case: (a) *spatial competition cell response*, (b) *orientation competition cell response*, (c) *bipole grouping cell response*, and (d) *depth competition cell response*. In (a) and (b), the top left panel corresponds to the smallest scale, then moving right along the rows and down along the columns corresponds to an increase in scale. In (c) and (d), the top left panel corresponds to the furthest depth, then moving right along the rows and down along the columns corresponds to moving nearer in depth. These simulations involved the triangular map matrix case (see Section 3).

from a combination of the scale-to-depth map (when the orientation competition cells project to the bipole cells) and the bipole grouping kernel. The depth-to-scale mapped feedback from the bipole cells to the spatial competition cells also helps to reinforce the formation of these strong bands of grouping.

Fig. 8d demonstrates the multiple-depth depth competition cell responses to their bipole grouping cell input. For each panel, there are 41×41 pixels corresponding to the cell locations of that depth. This figure shows that depth competition allows one to compute a multiple-depth boundary web that can be used to trap surface lightness signals into the appropriate depths. This is seen by noting that, for the number 5 HP surface, the periphery is represented at further depths, while there is a shift towards a central representation for nearer depths. Since the boundary webs shown in Fig. 8d are thick bands of activation lying on a sub-sampled image grid, this representation is resized back to the original image dimensions using nearest neighbors interpolation, so that the model can properly gate the filling-in of the image-texture elements.

Fig. 9a and b show the ON and OFF filling-in responses, respectively. For each FIDO (ON or OFF), the filled-in responses are obtained through the diffusion of their LGN feature inputs summed across all scales (Fig. 7a and b show the LGN ON and OFF responses before summation over scale) trapped into the appropriate depths by the multiple-depth boundary representation (Fig. 8d). A final representation of the number 5 HP surface is obtained simply by subtracting the OFF filling-in response from the ON filling-in response. For each panel, there are 481×481 pixels corresponding to the cell locations of that depth. This final result (Fig. 9c) is a multiple-depth representation of surface brightness.

The simulation results for the number 1–4 HP conditions arise in a similar way to the number 5 case. However, as the depth of the surface decreases, the texture elements become more isotropic and, although perspective projection leads to colinear texture elements in the periphery, grouping signals in the further depths become progressively weaker. This increasingly allows for the nearer depths to dominate, thus giving rise to flatter percepts when decreasing the surface depth (i.e., decreasing from surface number 5 down to 1). This result is observed clearly in the top row of image/depth map pairs in Fig. 1, where moving from right to left corresponds to a decrease in perceived surface depth which is matched well by the model. To understand more directly how weak grouping signals allow depth competition to produce flatter percepts, consider the responses of the bipole grouping and depth competition stages to the number 1 HP condition in Fig. 10a and b, respectively. While the bipole grouping cells appear to show some grouping at the further depths, the actual activations are not strong enough to completely survive the depth competition stage. This is illustrated by the strong near-depth responses in Fig. 10b.

4.1.2. Simulations of the low perspective condition

The LP condition simulations provide similar, but less deep, results when compared to the HP condition. This is demonstrated by the lower slope of the LP curve compared to the HP curve in Fig. 2 for both the psychophysical data (Fig. 2a) and the model (Fig. 2b–d). Thus the model is consistent with the general result that orthographic projections (LP) give less deep stimuli than (high) perspective projections. The model is able to account for this result by the simple fact that the LP images (2nd row of input images in Fig. 1) have a narrower range of texture-element size variation when moving from the center of the image to its periphery, than do the HP images (1st row of input images in Fig. 1). Going by the size-depth correlation (see Section 3.4), the narrower range of texture-element sizes implies a narrower range of depth.

The LIGHTSHAFT model thus processes the narrower range of texture-element sizes and creates less deep percepts for the LP case than when compared to the HP case. We can see this more directly by comparing simulations of the number 5 LP and number 5 HP conditions. The end panels in the 1st and 2nd rows of input images in Fig. 1 show the images of the number 5 HP and LP cases, respectively. It can be clearly seen that the LP case has a narrower range of texture-element sizes (biased more towards large scales) compared to the HP case. This narrower range of scales is exemplified by the responses of the complex cells. Figs. 7c and 11a display complex cell responses for the HP and LP cases, respectively. For the LP case, one observes more activation in the larger scale complex cells than in the smaller scales, while for the HP case activation is more distributed across the scales.

This greater bias towards the larger scales for the LP case (resulting from the bias of larger scales in the input image as opposed to some built-in model bias) leads to a greater bias towards nearer depths. Figs. 8d and 11b show the depth competition cell responses for the number 5 HP and LP cases, respectively. If one considers just the nearest depth in these figures, the central activation is slightly broader and stronger (i.e., brighter in terms of the figure) for the LP case than for the HP case. This bias towards nearer depths for the LP case gives rise to less deep model depth maps when compared with the HP case. This can be observed by considering Fig. 1 where the 5th column displays number 5 image/depth map pairs for the HP (1st row of image/depth map pairs) and LP (2nd row of image/depth map pairs) conditions. Although the difference is only slight, one can see that the central region for the LP case has a larger distribution of white and yellow (nearest depths) than does the HP case. This qualitative observation is quantified in Fig. 2b–d where the standard deviation of the depth maps is plotted for the HP and LP cases, and for the three scale-to-depth map cases, respectively. Focusing on the HP and LP curves at the points corresponding to the number 5 surface condition (i.e., simulated depth of 5), one can see that the standard deviation of the depth maps is less for

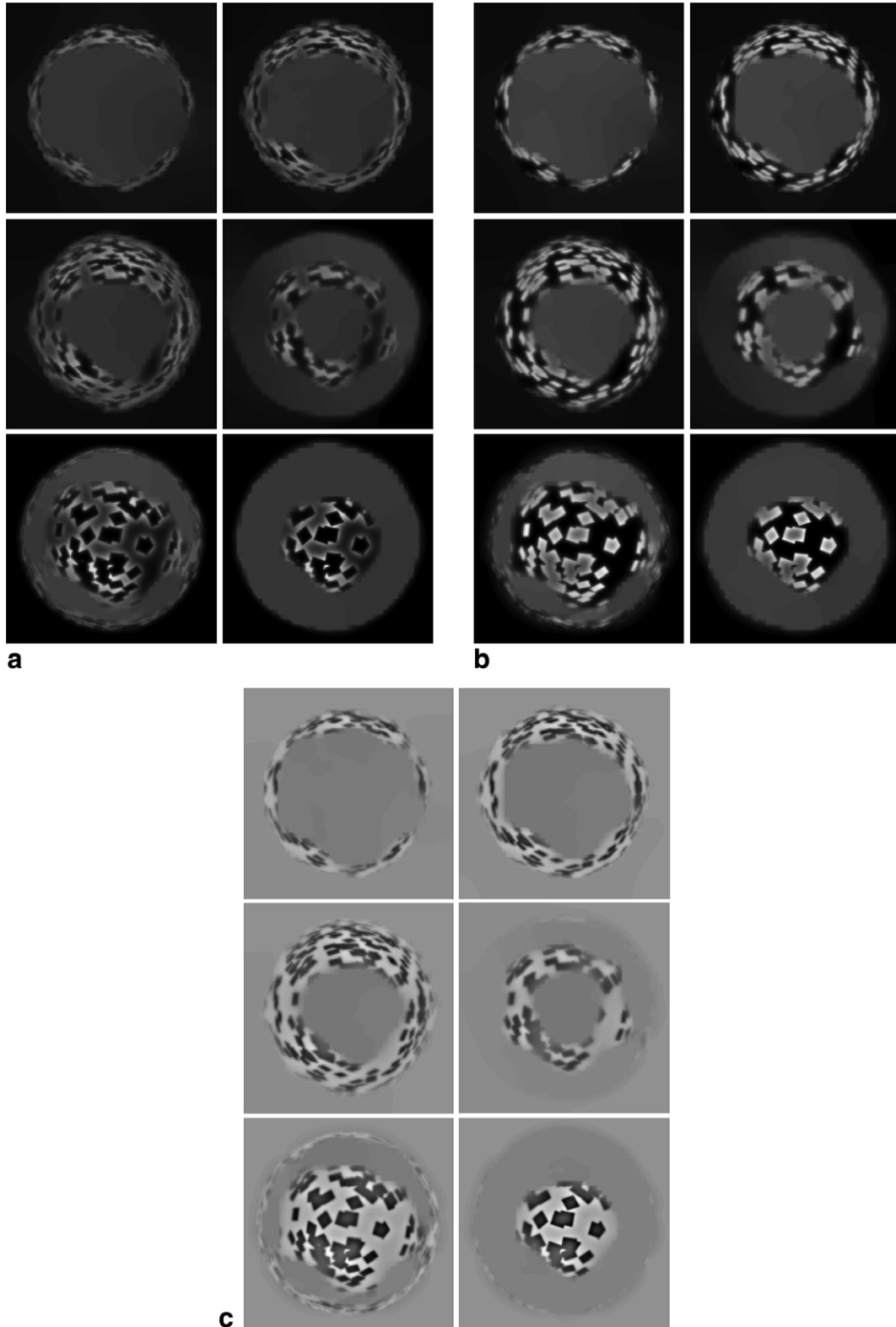


Fig. 9. Number 5 HP case: (a) *ON filling-in*, (b) *OFF filling-in*, and (c) *3D surface representation*. In (a), (b) and (c), the top left panel corresponds to the furthest depth, then moving right along the rows and down along the columns corresponds to moving nearer in depth. These simulations involved the triangular map matrix case (see Section 3).

the LP case. Thus the LIGHTSHAFT model is capable of explaining how LP projection leads to less deep percepts than HP projection.

4.1.3. Simulations of the CCE condition

As discussed in Section 2, the CCE condition (the 3rd row of image/depth map pairs in Fig. 1) was introduced

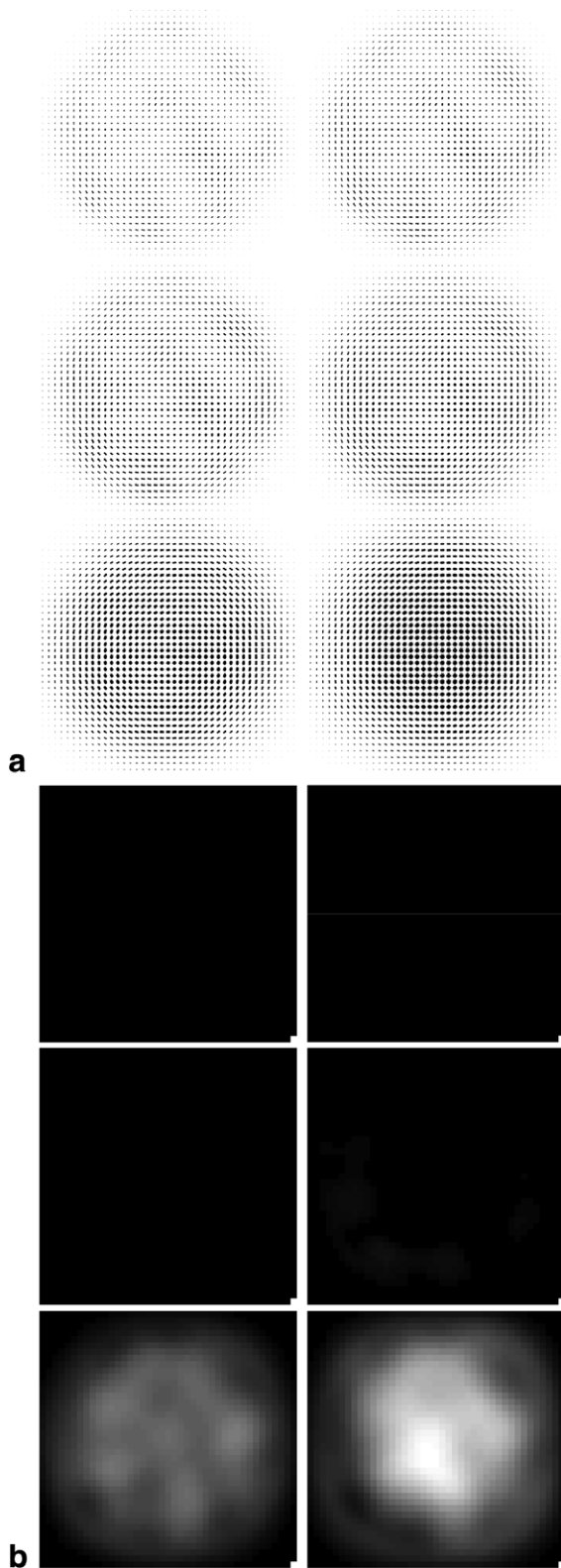


Fig. 10. Number 1 HP case: (a) *bipole grouping cell response* and (b) *depth competition cell response*. In (a) and (b), the top left panel corresponds to the furthest depth, then moving right along the rows and down along the columns corresponds to moving nearer in depth. These simulations involved the triangular map matrix case (see Section 3).

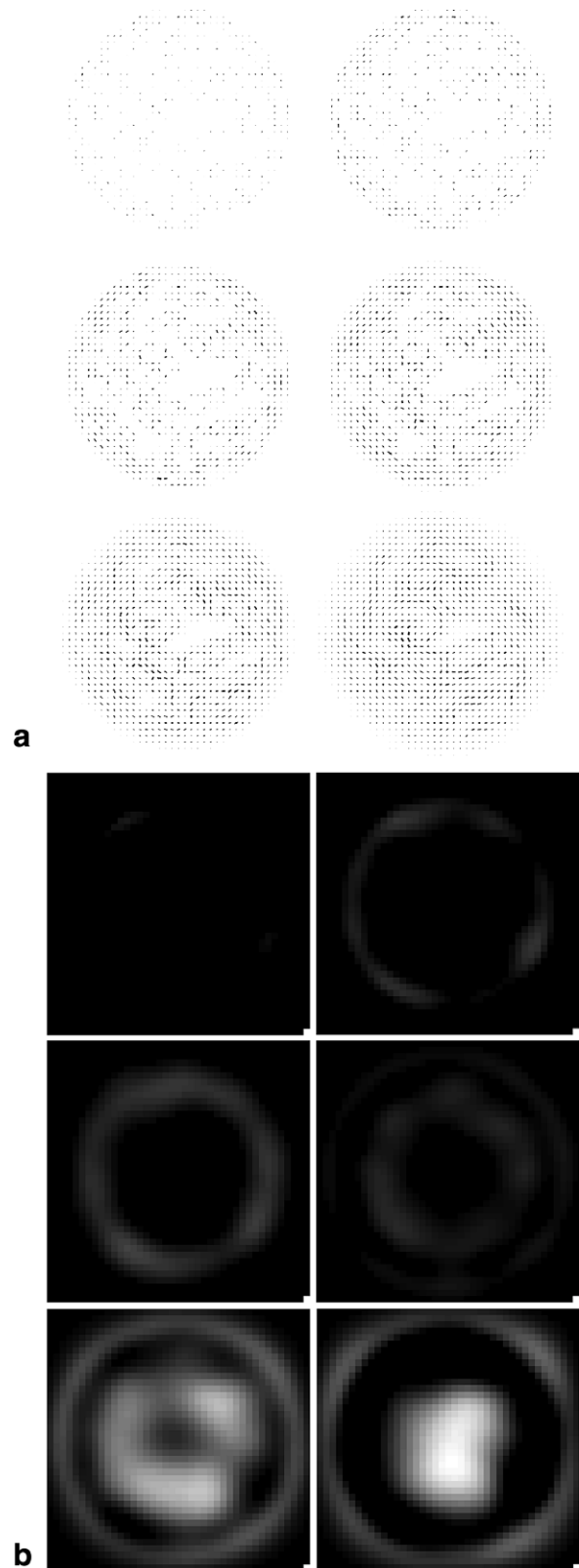


Fig. 11. Number 5 LP case: (a) *complex cell response* and (b) *depth competition cell response*. In (a), the top left panel corresponds to the smallest scale, then moving right along the rows and down along the columns corresponds to an increase in scale. In (b), the top left panel corresponds to the furthest depth, then moving right along the rows and down along the columns corresponds to moving nearer in depth. These simulations involved the triangular map matrix case (see Section 3).

to test the idea that perception of depth and curved surfaces depends largely on changes in texture-element compression (Cutting & Millard, 1984; Todd & Akerstrom, 1987). The CCE case demonstrates the Todd and Akerstrom (1987) finding that depth percepts can still be obtained when texture-element compression is constant and the texture-elements are elongated perpendicular to the tilt direction. Figs. 1 and 2 show that the model can account for the judged depths of the CCE stimuli for different simulated surface depths. The model also accounts for the relative relationships of the judged depth curves for the CCE, HP and LP conditions (i.e., for low simulated depths the CCE curve is above the HP curve, while for high simulated depths the CCE curve is below the HP curve).

The model is able to create percepts with strong variation in depth for the CCE stimuli because, even though image-texture element compression is constant (i.e., they have the same axis-ratios), *the texture elements still decrease in size when shifting from the center of the image to the periphery and they are elongated in the direction perpendicular to the tilt direction.* The colinear alignment of the elongated texture elements allows for bipole grouping, and the variations in scale enable these groupings to be formed in particular depths. This can be understood in more detail by considering the simulation of the number 5 CCE image (end of the 3rd row of image/depth map pairs in Fig. 1) using the triangular map matrix case. Fig. 12a shows that the complex cells are able to detect the oriented energy of the elongated texture elements of different scales, and Fig. 12b illustrates that the bipole cells can form groupings of these oriented signals in the appropriate depth planes. Depth competition (see Fig. 12c) then helps to refine the multiple-depth boundary representation that leads to the final 3D surface representation shown in Fig. 12d.

The model accounts for the relative relationships of the judged depth curves for the CCE, HP and LP conditions (see Fig. 2) mainly because the scale-to-depth map depends heavily on the size-depth correlation. This can be further understood by focusing on the number 1 simulated depth. For the number 1 simulated depth, the range of scale variation is largest for the CCE condition, 2nd largest for the HP condition and smallest for the LP condition (left-most images in 3rd, 1st, and 2nd row of input images in Fig. 1, respectively). The perceived depth of these images also decrease in this order (see Fig. 2a). The model complex cells detect the different scale ranges of the different images. Since the scale-to-depth map depends on the size-depth correlation, the variation of depths activated at the bipole cell stage is largely proportional to the variation of scales activated at the complex cell stage. This leads to final surface representations that match well with the perceived depth ordering of the CCE, HP and LP conditions for the number 1 simulated depths. Moreover, the model percepts also capture the perceived depth orders of the same conditions for simulated depths 2–5.

4.1.4. Simulations of the CCS condition

The CCS condition (4th row of image/depth map pairs in Fig. 1) illustrates that the size-depth correlation cannot be used as the only cue for judging depth from textured images. This is true because, even though the greater simulated depth images of the CCS case contain texture elements that vary in size moving from the center of the image to the periphery, all of the CCS images are perceived as flat (see Fig. 2a). The model is able to prevent these scale variations from being interpreted as depth variations through the bipole grouping mechanisms and the subsequent depth competition stage. For the CCS images, the image texture elements are square and randomly oriented, thus making it difficult for the model to form strong boundary groupings from the multiple-scale complex cell responses. Compared to the HP, LP and CCE cases, the groupings formed at the bipole cell stage are much weaker in response to CCS stimuli, especially for the farther depths. This allows the depth competition stage to favor the near depths, leading the model to produce flat percepts.

This result can be understood by considering the simulation of the number 5 CCS condition image (end of the 4th row of image/depth map pairs in Fig. 1) using the triangular map matrix case. Fig. 13a shows that the complex cells are able to detect the oriented energy of the texture elements of different scales, although these activations show much less colinear alignment and are weaker than those detected for the number 5 CCE condition (Fig. 12a). As a result, the groupings formed by the bipole cells (Fig. 13b) are weaker, especially for far depths when compared to the CCE condition. The depth competition stage weakens the far depth representations, thereby allowing the nearest depth to win the competition (Fig. 13c). The output of the depth competition stage is then used to gate the surface filling-in that leads to the final surface representation (Fig. 13d). Since the nearer depths are the only ones to survive the competition, filling-in within texture-element boundaries only occurs in the nearer depths. Thus the image is perceived as a flat surface.

The same mechanisms that produced a flat model percept for the number 5 CCS condition image are also the cause of the flat percepts produced for simulated depths numbers 1–4.

4.1.5. Simulations of the RO condition

The model explanation for the flat percept (Fig. 2a) of the RO condition images (5th row of image/depth map pairs in Fig. 1) is similar to that given for the CCS condition. Although the simulated depth images of the RO case contain texture elements that vary in size in a different manner to the CCS case (because of their constant area), the texture elements have very little colinear alignment. As was true for the CCS case, the model is able to prevent these scale variations from being interpreted as depth variations through the bipole grouping mechanisms and the subsequent depth competition stage. Since the RO condition image contains randomly oriented texture-elements,

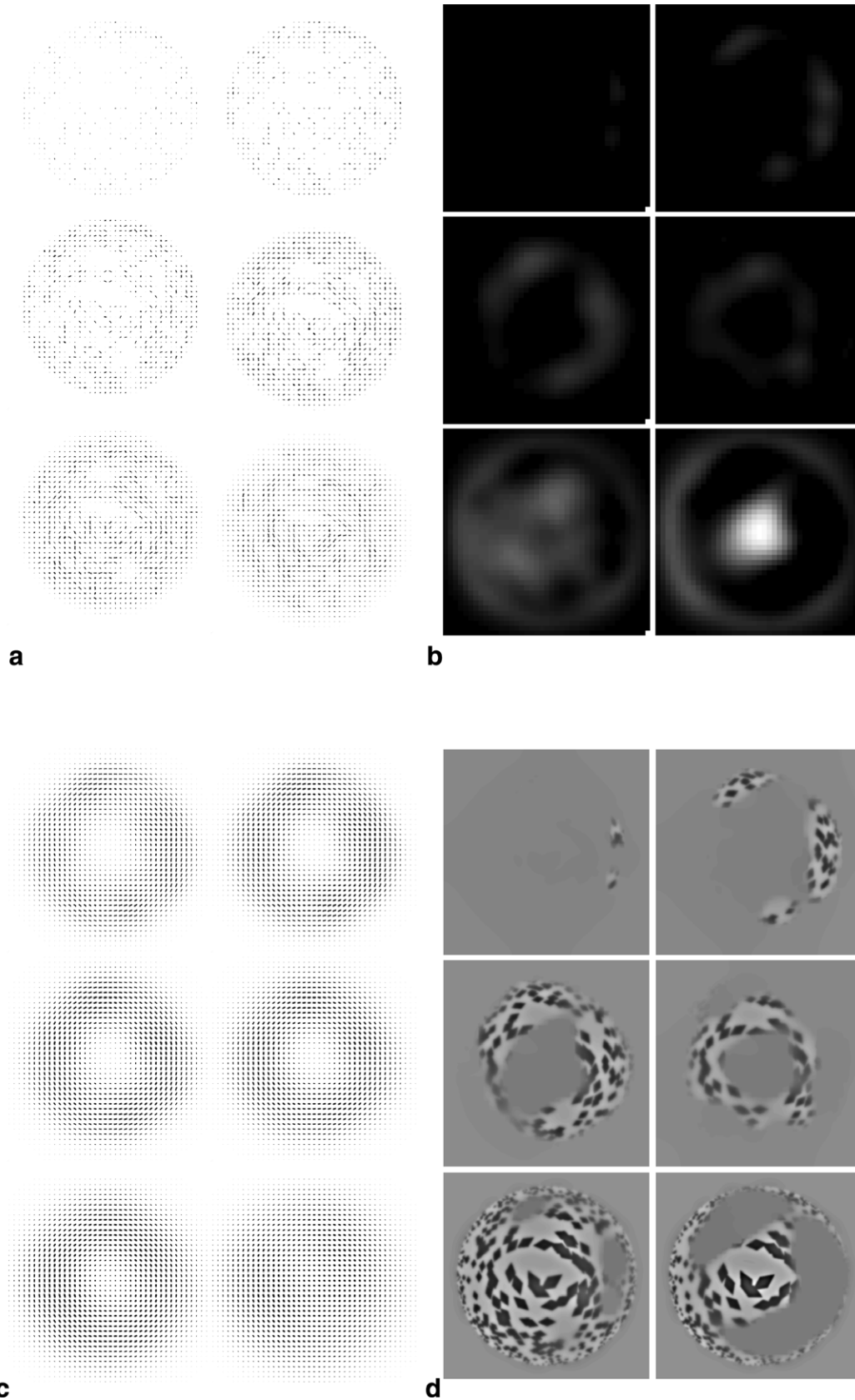


Fig. 12. Number 5 CCE case: (a) *complex cell response*, (b) *bipole grouping cell response*, (c) *depth competition cell response*, and (d) *3D surface representation*. In (a), the top left panel corresponds to the smallest scale, then moving right along the rows and down along the columns corresponds to an increase in scale. In (b), (c) and (d), the top left panel corresponds to the furthest depth, then moving right along the rows and down along the columns corresponds to moving nearer in depth. These simulations involved the triangular map matrix case (see Section 3).

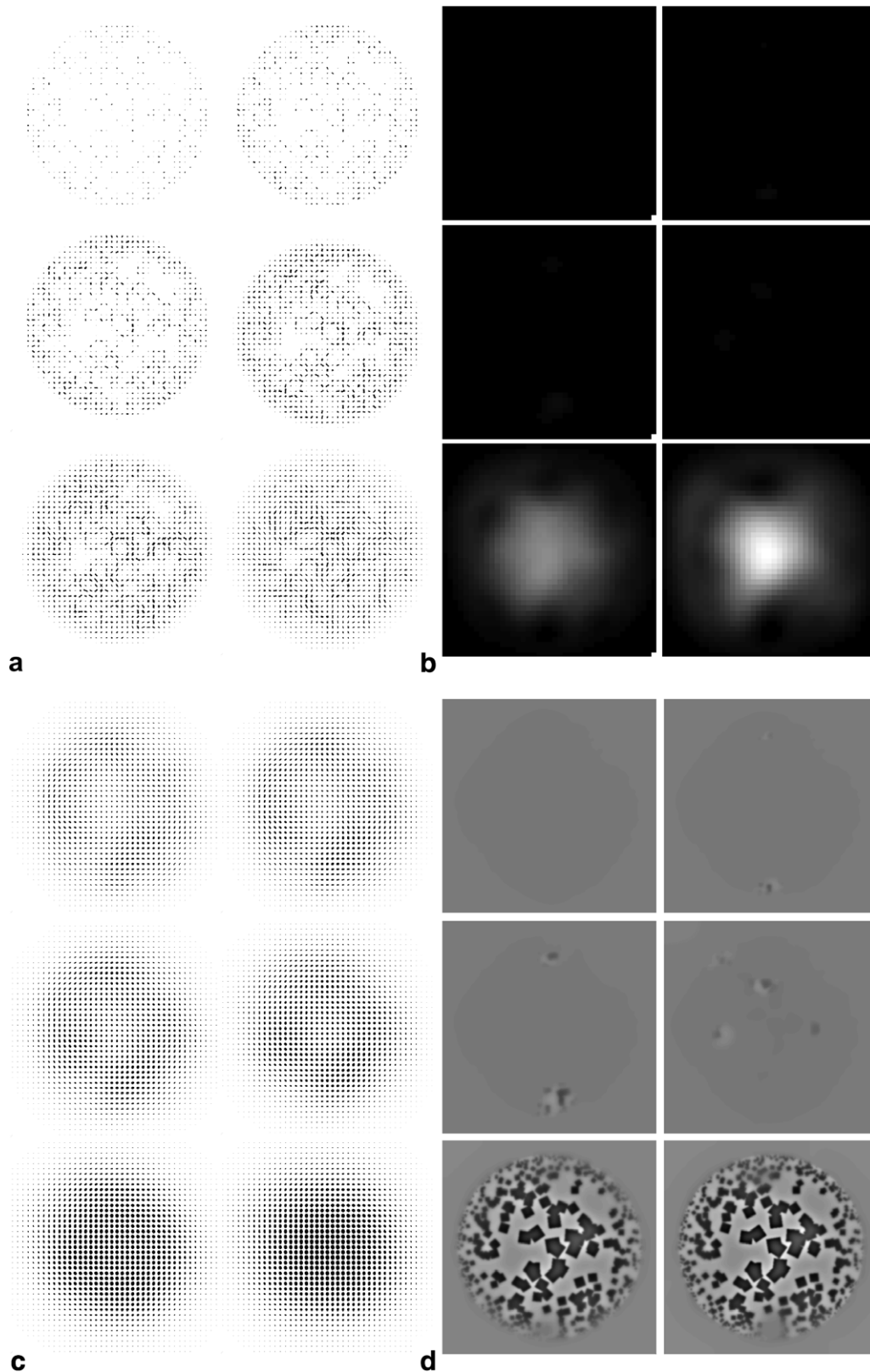


Fig. 13. Number 5 CCS case: (a) *complex cell response*, (b) *bipole grouping cell response*, (c) *depth competition cell response*, and (d) *3D surface representation*. In (a), the top left panel corresponds to the smallest scale, then moving right along the rows and down along the columns corresponds to an increase in scale. In (b), (c) and (d), the top left panel corresponds to the furthest depth, then moving right along the rows and down along the columns corresponds to moving nearer in depth. These simulations involved the triangular map matrix case (see Section 3).

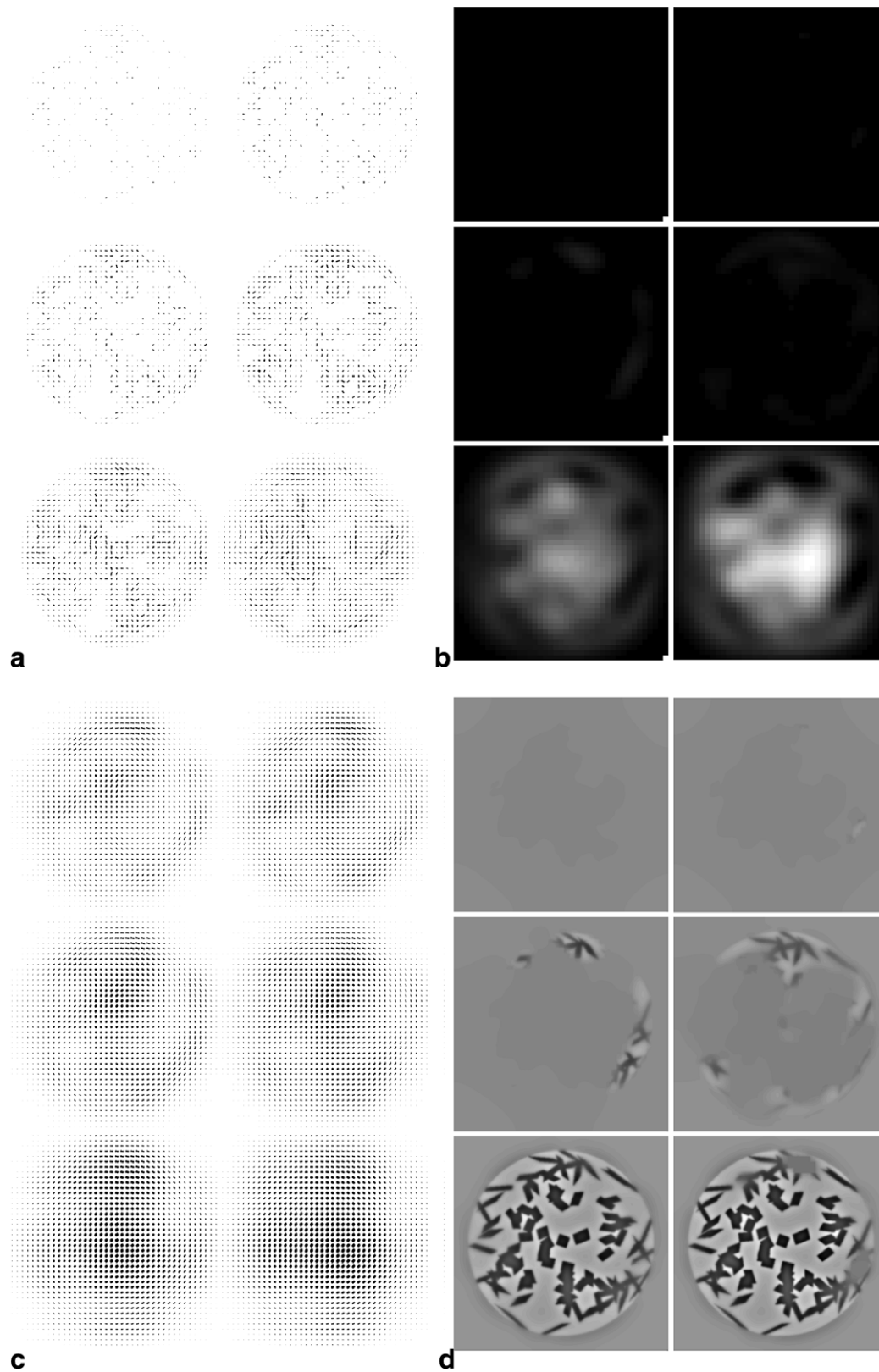


Fig. 14. Number 5 RO case: (a) *complex cell response*, (b) *bipole grouping cell response*, (c) *depth competition cell response*, and (d) *3D surface representation*. In (a), the top left panel corresponds to the smallest scale, then moving right along the rows and down along the columns corresponds to an increase in scale. In (b), (c) and (d), the top left panel corresponds to the furthest depth, then moving right along the rows and down along the columns corresponds to moving nearer in depth. These simulations involved the triangular map matrix case (see Section 3).

it is difficult for the model to form strong boundary groupings from the multiple-scale complex cell responses. Compared to the HP, LP and CCE cases, the groupings formed at the bipole cell stage are much weaker in response to randomly oriented texture-element stimuli, especially for the farther depths. Subsequent processing by the depth competition stage leads to flat percepts.

This result can be further understood by considering the simulation of the number 5 RO condition image (end of 5th row of image/depth map pairs in Fig. 1) using the triangular map matrix case. Fig. 14a shows that the complex cells are able to detect the oriented energy of the texture elements of different scales. However, these activations show much less colinear alignment than those detected for the number 5 HP condition (Fig. 7c), from which the RO condition was created. As a result, the groupings formed by the bipole cells (Fig. 14b) are weaker, especially for far depths when compared to the HP condition. The depth competition stage weakens the far depth representations, thereby allowing the nearest depth to win the competition (Fig. 14c). The output of the depth competition stage is then used to gate the surface filling-in that leads to the final surface representation (Fig. 14d). While the nearest depth dominates most of the surface representation, there is some middle depth representation of the surface, indicating that it is slightly harder for the model to create flat percept representations for the RO case than it is for the CCS case.

The same mechanisms that produced an approximately flat model percept for the number 5 RO condition image are also the cause of the flat percepts produced for simulated depths numbers 1–4.

4.2. Simulation of the elliptical cylinder, slanted plane and golf ball

The model deals with the elliptical cylinder image (top left of Fig. 15) in the same ways that it dealt with the HP, LP and CCE conditions. Namely, the variation in scale gives rise to the variation in depth seen in the depth maps for the three scale-to-depth map cases (left column of Fig. 15). Similar statements hold for the slanted plane and golf ball simulation. The slanted plane (middle column of Fig. 15) has a slant of 55° at the center of the image, assuming line of sight is perpendicular to the image plane. (Slant is defined as the angle between the surface normal at a point and the line of sight.) This slant results in a large range in the sizes of image-texture elements, which is captured in the depth maps computed by the model. The texture elements in the golf ball image (right column of Fig. 15) have a fair amount of large scale bias since the golf ball is spherical, as opposed to having a long depth axis like the prolate ellipsoids of Todd and Akerstrom. This large scale bias leads to depth maps that favor the near depths as a result of the size-depth correlation that is sensed by the scale-to-depth map. Both the elliptical cylinder and slanted plane images contain texture lying next to the image edge. Since the spatial competition stage of the

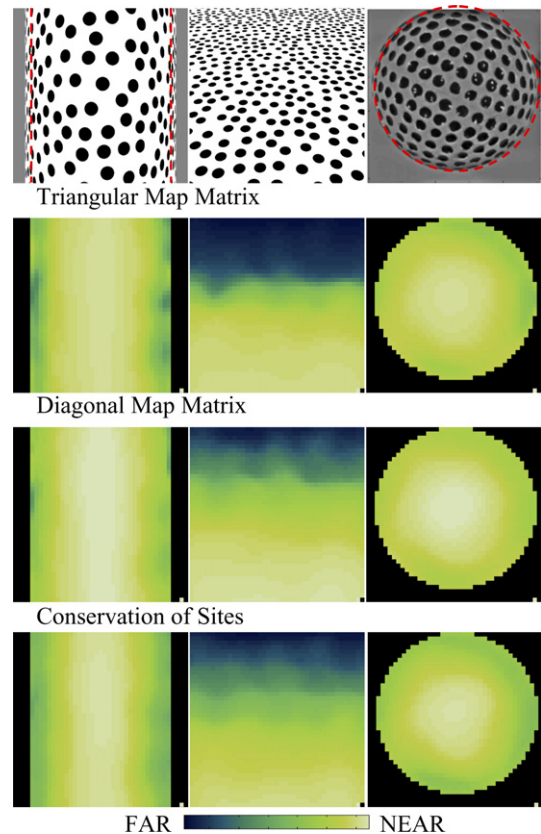


Fig. 15. Elliptical cylinder (left column), slanted plane (middle column) and golf ball (right column) depth maps computed using the triangular map matrix, the diagonal map matrix and conservation of sites (2nd, 3rd and 4th rows, respectively). In order to remove image edge-effects, the elliptical cylinder and slanted plane, shown here, are cropped versions of larger images that were used as input to the model. The golf ball image, on the other hand, was not cropped since the surface of the ball is reasonably distant from the edges of the image, except for the top edge, and is thus less subject to image edge-effects. In addition, for the elliptical cylinder and the golf ball, the red dashed lines indicate the boundary of the region within which the depth map was computed in order to avoid spurious effects near the occluding contour, separating the surface from the background (see Section 3.5). The golf ball photo was adapted from Clerc and Mallat (2002).

model uses large filters (see Appendix B), the model is subject to large edge effects in these two cases. To counteract this, the elliptical cylinder and slanted plane were simulated with initial image dimensions of 781×781 pixels and subsampling down to 71×71 pixels. The results displayed in Fig. 15 for the cylinder and plane correspond to the central 481×481 pixels of the image and the central 41×41 pixels of the computed depth map.

5. Discussion

5.1. Comparison of LIGHTSHAFT with previous SFT models and neurophysiological data

Since Gibson began the scientific study of SFT, a major research focus has been to perceptually isolate different kinds of texture gradients (i.e., scaling/perspective,

foreshortening/compression, density, or area) and study their effects on 3D shape perception (Blake et al., 1993; Braunstein, 1976; Braunstein & Payne, 1969; Cumming, Johnston, & Parker, 1993; Cutting & Millard, 1984; Stevens, 1981, 1984; Todd & Akerstrom, 1987). These perceptual studies have been accompanied with computational studies that have tried to explain how these different gradients can induce shape perception (Aloimonos, 1988; Bajcsy & Lieberman, 1976; Purdy, 1958; Witkin, 1981). Derivation of the mathematical relationship between these gradients and projective scene geometry for both planar (Stevens, 1981) and curved (Gårding, 1992) surfaces has led to the development of algorithms that use local affine texture distortion measurements to estimate local surface orientation and curvature parameters (Clerc & Mallat, 2002; Malik & Rosenholtz, 1997) from which shape can be calculated. A variety of other algorithms have also been developed that either use linear/non-linear filtering (Freeman & Torralba, 2002; Torralba & Freeman, 2002; Torralba & Oliva, 2001, 2002) or statistical approaches (Aloimonos, 1988; Blake & Marinos, 1990; Brown & Shvayster, 1990; Davis, Janos, & Dunn, 1983; Forsyth, 2001; Gårding, 1993; Ikeuchi, 1984; Kanatani, 1984; Kanatani & Chou, 1989; Marinos & Blake, 1990; Super & Bovik, 1995; Witkin, 1981) to estimate SFT. The focus of this line of research has been computational modeling of how the problem of SFT might be solved algorithmically, rather than on modeling the neural circuitry of the primate visual system. Despite the limited biological focus of these models, many of them take into account a variety of psychophysical phenomena.

Two neural models of SFT are the Sakai and Finkel (1995, 1997) Average Peak Frequency (APF) model and the Zaidi and Li (2000) Matched-Filters for Shape Identification model. The latter is a model of shape identification (i.e., it determines if a shape is concave or convex) rather than a model which provides a neural representation of perceived shape. Fleming, Torralba, and Adelson (2004) have provided a similar framework for shape based on specular reflections. The APF model begins with neurons depicting the multiple-scale-and-orientation structure of the early visual stages, but then the output of these neurons is transformed into an APF map of the image. A processed version of the APF map corresponds to the shape estimated from the textured image. Although the APF model can account for a large set of stimuli, this final representation requires that at each point in the image only one neuron is needed to encode any depth based on its firing rate. Such cells have not been found to date in the visual system, neither for SFT (Connor, 2002; Howard, 2003; Liu, Vogels, & Orban, 2004; Tsutsui, Sakata, Naganuma, & Taira, 2002; Tsutsui, Taira, & Sakata, 2005) nor for the more studied topic of stereo-disparity defined depth (Cumming & Parker, 1999, 2000; Hinkle & Connor, 2002; Janssen, Vogels, & Orban, 1999, 2000a, Janssen, Vogels, & Orban, 2000b; Nguyenkim & DeAngelis, 2001, 2003; Poggio, 1995; Prince et al., 2002; Sakata et al., 1998; Shikata, Tanaka, Nakamura, Taira, & Sakata, 1996; Taira, Tsutsui,

Jiang, Yara, & Sakata, 2000; Tsao, 2003). On the other hand, the LIGHTSHAFT model estimates SFT using a biologically plausible combination of spatial and rate coding through interactions between boundary and surface representations. In addition, the LIGHTSHAFT model provides a more complete explanation of the Todd and Akerstrom (1987) data than that given by Sakai and Finkel (1995), who simulated the perceived depths of prolate ellipsoids with polka-dot surface texture in only the LP case, and variations upon the CCS, RO and constant area cases.

Neurophysiological research on SFT, although limited, has demonstrated that cells in caudal intraparietal cortex (CIP) (Tsutsui et al., 2002) and inferior temporal cortex (IT) (Liu et al., 2004) are tuned to specific 3D surface orientations for imaged surfaces whose depth is defined by texture gradients. In particular, cells in CIP are selective to the tilt direction of a slanted plane defined by a texture gradient. Neurons in IT are selective for the tilt of texture-defined surfaces, largely independent of the surface slant. In several cases in both CIP and IT, the tilt preferences are similar regardless of whether the surface is defined by different textures or disparity gradients. Since CIP and IT are high in the visual hierarchy, it is not clear whether or not these surface orientation signals are used to represent 3D shape or depth percepts. An alternative may be that these signals are computed from a perceptual representation of shape/depth in order for the brain to perform other shape-based tasks; for example, guiding hand movements about an object, because CIP directly projects to the ‘hand’ area, in the anterior intraparietal (AIP) cortex (Sakata et al., 1998).

The structure of the LIGHTSHAFT model creates a 3D surface representation without the use of cells that explicitly encode 3D surface orientation. The present article hereby shows that many challenging SFT data can be explained by using multiple-scale boundary web perceptual groupings whose bipole grouping cells preferentially group visual features at a single depth. This bipole grouping theory has a natural extension to grouping by cells that are sensitive to 3D surface orientation, which becomes an increasingly important cue when processing surfaces that are tilted or curved in 3D. Indeed, disparity gradient cells (DeAngelis, 2000; Hinkle & Connor, 2002; Janssen et al., 2000b; Lee, 1999; Nguyenkim & DeAngelis, 2003; Ryan & Gillam, 1993; Sakata et al., 1999; Seyama, Takeuchi, & Sato, 2000; Thomas, Cumming, & Parker, 2002) can form coherent groupings across depths. These types of cells have been used by Grossberg and Swaminathan (2004) to explain developmental, attentional and bistability data about 3D perception of slanted and curved surfaces and of 2D images. In this work, these disparity gradient cells are predicted to be specialized types of bipole grouping cells. It remains to develop a theory of multiple-scale boundary webs whose bipole cells can group both within and across depth.

Such an extension would allow for strong 3D boundary groupings along any contour on a surface. For the case of prolate ellipsoids, the grouping mechanisms, which occur

within depth, of the LIGHTSHAFT model form the strongest boundary groupings along contours of minimum curvature; i.e., concentric rings on the ellipsoid surface which are oriented parallel to the image plane for the images presented here. The inclusion of bipole cells that can group across depths would also likely lead to strong 3D boundary groupings along contours of maximum curvature in a manner that is dependent upon the surface texture of the ellipsoid. Li and Zaidi (2000, 2001; Zaidi and Li, 2002) have proposed that oriented energy along contours of maximum and minimum curvature help determine our perception of SFT, in particular, the perception of the direction of surface slant, or the sign of surface curvature. Todd and colleagues argue that this is true only for a specific set of surfaces, textures, and viewing conditions (Todd & Oomes, 2002; Todd et al., 2004). An extension of the LIGHTSHAFT model, to incorporate mechanisms which can group boundaries across depth, may shed additional light on this problem because it would have the flexibility to investigate more completely how the relationship between surface texture and surface contours influences our perception of SFT.

The neural code used in the model to create the 3D surface representation of an image is a combination of spatial and rate coding within interacting boundaries and surface representations: the spatial and rate code within the boundary representations locates the neurons that code 6 distinct depths at different positions of the visual field. These boundary signals are predicted to be amodal, or invisible, within the (V1 interblob)-to-(V2 pale stripes)-to-V4 cortical processing stream (Grossberg, 1994). By gating the filling-in of visible surface brightness, lightness, and color, these boundary signals enable a second spatial and rate code to be computed within the (V1 blob)-to-(V2 thin stripe)-to-V4 cortical processing stream. Here the consciously seen lightness or brightness at a given spatial location increases with the response rate of cells at that location.

5.2. Model propositions, predictions and simplifications

The model described in this article is based upon conceptual ideas that have been developed as part of an emerging unified theory of how the visual cortex sees the world in depth, including the FACADE and 3D LAMINART models (Cao & Grossberg, 2005; Grossberg, 1987a, 1987b, 1994; Grossberg & Howe, 2003; Grossberg & Swaminathan, 2004; Grossberg & Yazdanbakhsh, 2005; Kelly & Grossberg, 2000), which have explained and predicted a large amount of neurobiological and psychophysical data. The key concepts of this theory that have been integrated and further developed within the LIGHTSHAFT model are: the size-disparity and size-depth correlations (Grossberg, 1994), the use of bipole cells within multiple-scale, cooperative-competitive feedback loops to create coherent groupings in the form of boundary webs (Grossberg, 1987b; Grossberg & Mingolla, 1987), and surface capture via the 3D gating of surface filling-in by these boundary

webs (Grossberg, 1987b, 1994). The size-disparity and size-depth correlations have been used to define the scale-to-depth and depth-to-scale maps that are incorporated into the cooperative-competitive feedback loop, which takes a multiple-scale representation of the image and converts it into a coherent multiple-depth boundary web. The model also uses a depth competition stage to refine the multiple-depth representation.

Two of the most important parameters of the model are the number of scales and the number of depths. Scale and depth, two continuous variables in the world, are represented in a quantized manner in the model because neurons are discrete elements. The choice of 6 scales and 6 depths tries to approximate the continuum of scale and depth, while at the same time keeping the model's implementation at a tractable size. In the visual cortex, cells tuned to different scales and depths sufficiently cover the continuum of these two variables (De Valois, Albrecht, & Thorell, 1982; Prince et al., 2002). Grossberg and Swaminathan (2004) have simulated how the gradually changing distribution of activity across depths can generate a continuous representation of depth.

5.2.1. Neurobiological support for the model stages

The different stages of the model embody explanations and predictions about the structure and function of the visual system. The LGN, simple cell and complex cell stages have the strongest neurobiological support. In the LGN, the different scale representations are extracted using both ON-center OFF-surround and OFF-center ON-surround kernels. This simulates the cell types having concentric center-surround receptive fields in the retina (Barlow, 1953; Cook & McReynolds, 1998; Kuffler, 1953; Werblin & Dowling, 1969) and the LGN (Dubin & Cleland, 1977; Hubel & Wiesel, 1961; Jones, Andolina, Oakely, Murphy, & Sillito, 2000). Moreover, the changes in scale of the center-surround receptive fields is consistent with the luminance/lightness coding properties of the LGN (Barlow, Snodderly, & Swadlow, 1978; Kayama, Riso, Bartlett, & Doty, 1979; Marrocco, 1972; Papaioannou & White, 1972; Rossi & Paradiso, 1999; Rossi, Rittenhouse, & Paradiso, 1996) and V1 (Bartlett & Doty, 1974; Friedman, Zhou, & Heydt, 2003; Kayama et al., 1979; Kinoshita & Komatsu, 2001; Komatsu, Murakami, & Kinoshita, 1996; MacEvoy, Kim, & Paradiso, 1998). In particular, the LIGHTSHAFT model uses a fixed narrow center kernel with the different surround scales (Grossberg & Hong, 2006; Grossberg et al., 1995; Mingolla et al., 1999) and thereby also simulates the output of a sharp center at the ganglion cells due to interactions in the retinal network (Cook & McReynolds, 1998; Roska, Nemeth, Orzo, & Werblin, 2000).

As in the brain, model simple cells possess an oriented receptive field that is formed by elongated excitatory and inhibitory zones (DeAngelis et al., 1995; De Valois & De Valois, 1980; Hawken & Parker, 1991; Heeger, 1991; Hubel & Wiesel, 1962; Jones & Palmer, 1987; Parker & Hawken,

1988; Ringach et al., 2002; Szulborski & Palmer, 1990; Tsao, Conway, & Livingstone, 2003). The model incorporates two common receptive field configurations: the odd-symmetric and even-symmetric cases (Jones & Palmer, 1987; Malik & Perona, 1990; Parker & Hawken, 1988; Szulborski & Palmer, 1990). Although these cases appear to be part of a continuum of receptive field configurations (Parker & Hawken, 1988), they have been commonly used as archetypal receptive fields in the areas of neuroscience and neural modeling (De Valois & De Valois, 1980; Grossberg, 1983; Grossberg & McLoughlin, 1997; Grossberg & Mingolla, 1987; Grossberg & Todorović, 1988; Heeger, 1991; Hubel & Wiesel, 1962; Kandel, Schwartz, & Jessell, 1991), in part because of the appeal of symmetry, and also because they can be used as basis functions for computing simple cell receptive fields that have neither odd nor even symmetry.

The properties of the spatial competition cells, orientation competition cells, and bipole grouping cells receive support from both psychophysical and neuroscience data (Callaway, 1998; Field et al., 1993; Gove et al., 1995; Grossberg & Mingolla, 1985a, 1985b, 1987; Grossberg et al., 1995; Grossberg & Swaminathan, 2004; Heitger et al., 1998; Kellman & Shipley, 1991; Mingolla et al., 1999; Peterhans & von der Heydt, 1989; Raizada & Grossberg, 2001; von der Heydt et al., 1984). Since the model tries to quantitatively simulate complex psychophysical data about SFT, simplifications are made at the level of neural circuitry. 3D LAMINART models (Cao & Grossberg, 2005; Grossberg & Howe, 2003; Grossberg & Swaminathan, 2004; Grossberg & Yazdanbakhsh, 2005; Raizada & Grossberg, 2001) are more articulated than LIGHT-SHAFT with respect to incorporating known facts of detailed laminar local circuit connectivity. Raizada and Grossberg (2001) have demonstrated the biological plausibility of bipole cells by simulation and citing evidence of excitatory horizontal connections between V1 and V2 layer 2/3 pyramidal cells with colinear, co-axial receptive fields (Bosking et al., 1997; Schmidt et al., 1997). The excitation of these horizontal connections is balanced by inputs from inhibitory interneurons in layer 2/3 that are driven by layer 2/3 pyramidal inputs (McGuire et al., 1991). This cooperative-competitive balance leads to the bipole property. This laminar bipole design is incorporated within LIGHTSHAFT.

5.2.2. Texture segmentation assumption

A major assumption applied to computing the depth maps of the model (Figs. 1 and 15) is that a texture segmentation process separates the textured surface in the image from the background. The texture segmentation process is not implemented explicitly in this model. Instead, the depth maps are computed only in the region of the image where the surface is present. If depth maps were computed over the entire image, there would also be signals coding depth in the background regions, and these signals would contribute to the computed standard deviation of the depth

map that is used to describe the Todd and Akerstrom (1987) judged depth data. The occluding boundary in each input image between the textured surface and the background contains both large and small scale information and thus it is difficult to assign it an appropriate depth. In order to avoid this problem, all depth maps are computed in the region of the textured surface that is slightly inside of the occluding contour. Grossberg (1994, 1997), Grossberg and Yazdanbakhsh (2005) and Kelly and Grossberg (2000) describe FACADE theory mechanisms that are capable of separating 3D figures from their backgrounds.

5.2.3. Sub-sampling

A simplification in implementing the model is sub-sampling of the complex cell responses from 481×481 pixels down to 41×41 . This simplification ensures that the cooperative-competitive feedback loop (spatial competition, orientation competition, and bipole grouping cells) can be simulated in a reasonable amount of time. For a single input image, at a resolution of 481×481 pixels, the LGN, simple cell and complex cell stages together take on the order of 20 min to simulate in MATLAB[®] (The Mathworks, Inc., Natick, MA, USA) on a 3 GHz Pentium IV processor and with 3 GB RAM. This duration results from the large number of convolutions required. At a resolution of 41×41 pixels, the cooperative-competitive feedback loop takes on the order of 5 min, even though equilibration requires 20 relaxation steps. At a resolution of 41×41 pixels, the depth competition stage runs on the order of 10 s. The bulk of the simulation time lies with the FIDOs, which at a resolution of 481×481 pixels require 15 min to simulate each depth plane.

Given that the sub-sampling used is nearly 12 fold, it is impressive that these simulations work as well as they do. The sub-sampling has the strongest effects for the smallest scales where the activation of the complex cells has transitions over regions only a few pixels in size. With nearly 12 fold sub-sampling, many of these transitions are removed from the overall small scale signal representation. Since large scales have slower transitions of activation over space, less information is lost from sub-sampling. Despite the reduced sampling, all scales still contain sufficient orientation specificity to form the necessary groupings required for the final 3D surface representation. In addition, the representation of the farther depths, which is driven primarily by the “less spatially informative” small scales, receives extra support from larger scales through the scale-to-depth map when either the triangular map matrix or conservation of sites cases are used. This extra support helps to generate more coherent far-depth representations.

5.2.4. Sub-sampling vs pyramid architecture

An alternative to the sub-sampling used in the current model is to use a pyramid architecture (Adelson, Anderson, Bergen, Burt, & Ogden, 1984; Burt & Adelson, 1983; Freeman & Adelson, 2002; Simoncelli & Freeman, 1995). Such

an architecture would preserve the high resolution in the small scales and reduce the resolution of the larger scales where high resolution is not really necessary. For the purposes of this project, a pyramid was not implemented for the following reasons: First, if the scales had different pixel dimensions, the weighted interactions between scales and depths would be more complicated and harder to control. Second, pyramids typically require blurring of the large scales before reducing their pixel dimensions in order to prevent aliasing (Forsyth & Ponce, 2003). Blurring would lead to unwanted delocalization of the large scale signals.

The sub-sampling technique used for the model presented here is equivalent to nearest neighbors image reduction. Since sub-sampling works well, it would be possible to create the model with a pyramid architecture that ignored potential aliasing effects, by not blurring the image before using nearest neighbors image reduction. However, one would still have to deal with the first issue of having interactions between scales and depths of different pixel dimensions. In addition, the current sub-sampling framework takes less time to simulate than a pyramid architecture, since the current sub-sampling framework samples image positions less densely than a pyramid architecture would at smaller scales.

5.2.5. Standard deviation as a measure of depth variation

The standard deviation of depth, used for comparison with the psychophysical measure of judged depth, was employed instead of any other metric because it is probably the simplest metric of depth changes that is robust to noise, or local variations in the depth map. For example, in Fig. 1 the depth maps corresponding to the number 5 RO and CCS prolate ellipsoid cases contain local variations in depth (green blips) near the occluding contour of the surface, whereas the depth across the rest of the surface is relatively constant (yellowish). For such depth maps, the standard deviation of depth over the entire depth map will be relatively insensitive to fluctuation in depth measures over small regions. A simple alternative to using standard deviation would be to measure the difference between the maximum and minimum depth values within the depth map. Such a measure would be sensitive to the local variation seen in the Fig. 1 depth maps corresponding to the number 5 RO and CCS prolate ellipsoid cases, and produce larger estimates of depth variation, even though the majority of the depth map appears relatively flat. Another alternative to standard deviation would be variance. However, variance is essentially measured in units of depth, squared, while standard deviation is measured in units of depth.

5.2.6. Real-world images, isotropy, anisotropy and oblique views

The LIGHTSHAFT model first estimates the dimensions of the local surface tokens using multiple-scale filters. Then, using secondary mechanisms involving thresholding, competition and grouping, a coherent boundary web is

formed across depth in order to trap surface lightness into the appropriate depths. In particular, the grouping mechanism, which depends on the bipole filters, combines the local surface token information over larger scales, forming a global representation of changes of scale in the image. This general form of hierarchical processing is expected to work on textures involving other patterns found in real-world images, besides dots or diamonds. However, the explicit implementation of these ideas presented in the LIGHTSHAFT model is sensitive to image illumination, thresholding, competition, spatial filter structure, and the use of a discrete representation of continuous depth. LIGHTSHAFT is able to simulate the depth percept of the real-world example of the golf ball partly because the image has reasonably constant illumination throughout, and the contrast of the dots on the golf ball is sufficiently high, such that filter responses survive thresholding.

When dealing with real-world images, while thresholding and competition can eliminate noise, it also has the potential to eliminate signals with relevant depth information depending on the sensitivity and the dynamic nature of the thresholding and competition, and the variations of illumination in the visual scene (Grossberg & Hong, 2006; Land, 1986; Rahman, Jobson, & Woodell, 2004). The LIGHTSHAFT model achieves desired functional properties in a tractable amount of simulation time by simplifying its thresholding and competitive dynamics and using a discrete representation of continuous depth with six scales and six depths. Grossberg and Swaminathan (2004) have shown how a small number of scales can generate a continuous percept of surface depth. Additional research will be needed to test if these hypotheses are sufficient to process natural scenes.

Given that the spatial filters in the LIGHTSHAFT model are isotropic, one might think that the model best detects isotropic image changes. Thus, the model would seem less capable of dealing with foreshortened or anisotropic surface textures (e.g., grass). This idea is, however, confounded by the fact that the model involves stages of competition and grouping, which also involve the scale-to-depth map. In particular, the orientation competition stage helps detect local anisotropic information in the image by biasing the strongest orientations at different scales. In addition, grouping detects anisotropic information at large scales in the image by using anisotropic bipole filters. Isotropic and anisotropic information are further integrated across depth in the case of the triangular matrix scale-to-depth map, where multiple scales can map to a single depth.

Treatment of oblique views is also important when dealing with real-world images. The LIGHTSHAFT model was developed primarily to account for fronto-parallel views, as in Todd and Akerstrom (1987) data. In general, geometric correlations between visible obliquities and relative depths influences our depth estimation, and while the size-depth correlation is one heuristic used by

LIGHTSHAFT to do this, it is actually the interaction of the size-depth correlation with orientation information that influences the depth percepts created by the model. In the model, the size-depth correlation is embodied by the scale-to-depth map, which maps multiple-scale-and-orientation information into multiple-depth-and-orientation information, dependent on competitive and grouping interactions which detect anisotropies in the image. Thus the model is sensitive to changes in both scale and anisotropic information, and this implies that various cues for depth estimation, including size, foreshortening (i.e., anisotropic compression) and density (Gibson, 1950; Rosenholtz & Malik, 1997), play a role within the LIGHTSHAFT model. While there is evidence to support the idea that foreshortening is a dominant cue for 3D shape perception (Cutting & Millard, 1984), including under different views (Koenderink, van Doorn, Kappers, & Todd, 2004), it is not always the dominant cue. This is illustrated by the constant compression elongated prolate ellipsoid case developed by Todd and Akerstrom (1987), and simulated by the LIGHTSHAFT model, which shows that a stimulus without changes in foreshortening of the texture elements (i.e., anisotropic changes) can create a percept which varies in depth. Rather, the depth percept arises from size changes in texture elements (i.e., isotropic changes) provided that the elements are sufficiently elongated and aligned such that they can be grouped and mapped into the appropriate depths. Further work is needed to determine whether the current structure of the LIGHTSHAFT model is capable of estimating the shape of surfaces with anisotropic textures, or under different views. In particular, it would be important to determine if anisotropic variation in the scale of spatial filters is necessary to account for the psychophysics relating to both isotropic and anisotropic surface textures, and the angle of view.

6. Conclusion

This article develops the LIGHTSHAFT model of how visual cortical areas interact to convert a textured 2D image into a surface representation of 3D shape. This was achieved by incorporating and extending ideas from the FACADE and 3D LAMINART models, with an emphasis on *multiple-scale* filtering, boundary grouping, and surface filling-in. The successful simulation of the Todd and Akerstrom (1987) data, the elliptical cylinder, the slanted plane and the golf ball depended on four main facets of the model: the scale-to-depth and depth-to-scale maps, cooperative-competitive bipole grouping, depth competition, and boundary-gated surface capture and filling-in. These components ensure that coherently grouped texture elements can be assigned to the appropriate depths, while incoherently grouped texture elements are typically mapped to a single depth. It remains to integrate the model into a larger FACADE theory that is also capable of 3D figure-ground separation.

Appendix A. Model equations

Here we define the equations of the model described in Section 3. The parameter values used in the simulations are presented in Table 1 of Appendix F.

Each model neuron is typically modeled as a single voltage compartment in which the membrane potential, $v(t)$, is given by

$$C_m \frac{d}{dt} v(t) = - (v(t) - E_{\text{leak}}) g_{\text{leak}} - (v(t) - E_{\text{excit}}) g_{\text{excit}}(t) - (v(t) - E_{\text{inhib}}) g_{\text{inhib}}(t), \quad (\text{A.1})$$

where E represents reversal potentials, g_{leak} is a constant leakage conductance, and the time-varying conductances $g_{\text{excit}}(t)$ and $g_{\text{inhib}}(t)$ represent the total inputs to the cell (Grossberg, 1973; Hodgkin, 1964). Most of the following network equations are instances of this general membrane equation, where for simplicity, the capacitance term C_m was set equal to 1, the leakage conductance is relabeled as $g_{\text{leak}} = A$, the excitatory and inhibitory reversal potentials are relabeled as: $E_{\text{excit}} = B$ and $E_{\text{inhib}} = C$, and the leakage reversal potential is set to: $E_{\text{leak}} = 0$. Then Eq. (A.1) can be rewritten as a membrane, or shunting equation:

$$\frac{d}{dt} v = -Av + (B - v)g_{\text{excit}} - (v + C)g_{\text{inhib}}, \quad (\text{A.2})$$

where A is a constant decay rate, B is a excitatory saturation potential, C is a hyperpolarization parameter, g_{excit} is the total excitatory input, and g_{inhib} is the total inhibitory input.

A.1. Stage 1: LGN shunting network

The first stage of the model is the ON and OFF shunting network that processes the input image directly. These equations represent multiple-scale single-opponent networks where scale is indexed by $s = 1, \dots, 6$ ($s = 1$ for smallest, $s = 6$ for largest). The activity a_{ijs}^+ of the ON cell at image position (i, j) and scale s obeys an on-center off-surround shunting equation:

$$\frac{d}{dt} a_{ijs}^+ = -Aa_{ijs}^+ + (B - a_{ijs}^+) \sum_{pq} \Phi_{pqijs} I_{pq} - (a_{ijs}^+ + C) \sum_{pq} \Gamma_{pqijs} I_{pq}. \quad (\text{A.3})$$

In Eq. (A.3), parameter A represents the decay rate, B represents the excitatory saturation potential, Φ_{pqijs} represents the isotropic normalized Gaussian on-center kernel as defined in Eq. (A.7), I_{pq} is the input image at position (p, q) , C is the hyperpolarization parameter, and Γ_{pqijs} represents the isotropic normalized Gaussian off-surround kernel as defined in Eq. (A.8). At equilibrium, Eq. (A.3) yields:

$$a_{ijs}^+ = \frac{B \sum_{pq} \Phi_{pqijs} I_{pq} - C \sum_{pq} \Gamma_{pqijs} I_{pq}}{A + \sum_{pq} \Phi_{pqijs} I_{pq} + \sum_{pq} \Gamma_{pqijs} I_{pq}}. \quad (\text{A.4})$$

The activity a_{ijs}^- is the OFF cell activation at image position (i, j) and scale s obey the equation:

$$\frac{d}{dt} a_{ijs}^- = -A a_{ijs}^- + (B - a_{ijs}^-) \sum_{pq} \Gamma_{pqijs} I_{pq} - (a_{ijs}^- + C) \sum_{pq} \Phi_{pqijs} I_{pq}. \quad (\text{A.5})$$

In Eq. (A.5) the excitatory and inhibitory kernels of the ON cell in Eq. (A.3) are exchanged. At equilibrium, Eq. (A.5) obeys:

$$a_{ijs}^- = \frac{B \sum_{pq} \Gamma_{pqijs} I_{pq} - C \sum_{pq} \Phi_{pqijs} I_{pq}}{A + \sum_{pq} \Phi_{pqijs} I_{pq} + \sum_{pq} \Gamma_{pqijs} I_{pq}}. \quad (\text{A.6})$$

The two isotropic Gaussian kernels, Φ_{pqijs} and Γ_{pqijs} , obey the following equations:

$$\Phi_{pqijs} = \frac{1}{\sum_{uv} \exp\left(-\frac{1}{2\phi_s^2}((u-i)^2 + (v-j)^2)\right)} \times \exp\left(-\frac{1}{2\phi_s^2}((p-i)^2 + (q-j)^2)\right), \quad (\text{A.7})$$

and

$$\Gamma_{pqijs} = \frac{1}{\sum_{uv} \exp\left(-\frac{1}{2\eta_s^2}((u-i)^2 + (v-j)^2)\right)} \times \exp\left(-\frac{1}{2\eta_s^2}((p-i)^2 + (q-j)^2)\right), \quad (\text{A.8})$$

where for the center kernel, Φ_{pqijs} , the standard deviation, ϕ_s , varies with the scale indexed by s . Likewise, for the surround kernel, Γ_{pqijs} , the standard deviation, η_s , also varies with the scale indexed by s . In the normalizing sums in the denominators of Eqs. (A.7) and (A.8), u and v represent dummy positional variables.

The activity A_{ijs}^+ of the ON push-pull cell at image position (i, j) and scale s , spatially sharpens the contrast-enhanced LGN ON cell activity:

$$A_{ijs}^+ = [a_{ijs}^+ - a_{ijs}^-]^+, \quad (\text{A.9})$$

where a_{ijs}^+ is the ON cell activation, and a_{ijs}^- is the OFF cell activation. The activity A_{ijs}^- of the OFF push-pull cell at image position (i, j) and scale s , spatially sharpens the contrast-enhanced LGN OFF cell activity:

$$A_{ijs}^- = [a_{ijs}^- - a_{ijs}^+]^+. \quad (\text{A.10})$$

These ON and OFF push-pull cells provide a multiple-scale contrast-enhanced representation of the input image. This formulation for the model's simplified LGN stage has been adapted from Grossberg et al. (1995).

A.2. Stage 2: simple cells

The simple cell response depends on the LGN response convolved with odd or even symmetric kernels. The activity

of the ON, R_{ijkzsy} , and OFF, L_{ijkzsy} , sub-regions of the simple cells at position (i, j) , orientation k , contrast polarity z , scale s and filter symmetry y (i.e., odd or even) obeys:

$$R_{ijkzsy} = \sum_{pq} [X_{pqijkzsy}]^+ A_{ijs}^+, \quad (\text{A.11})$$

and

$$L_{ijkzsy} = \sum_{pq} [-X_{pqijkzsy}]^+ A_{ijs}^-, \quad (\text{A.12})$$

where $X_{pqijkzsy}$ represents the odd ($y = 1$) and even ($y = 2$) symmetric simple cell filters (defined in Appendix A.2.1) of orientation k , contrast polarity z , scale s , and symmetry y . A_{ijs}^+ and A_{ijs}^- represent the ON and OFF LGN responses, respectively. In the model implementation, the orientations take the range $k = 1, \dots, 16$, while contrast polarities are positive or negative, $z = -1, 1$. The simple cell activity, B_{ijkzsy} , at position (i, j) , orientation k , contrast polarity z , scale s , and symmetry y , is then determined by the following equation, in which activity is the rectified sum of the activities of each sub-region, minus their absolute difference:

$$B_{ijkzsy} = [R_{ijkzsy} + L_{ijkzsy} - |R_{ijkzsy} - L_{ijkzsy}|]^+, \quad (\text{A.13})$$

where R_{ijkzsy} and L_{ijkzsy} are the ON and OFF subregions of the simple cells, respectively. This formulation for the simple cells is adapted from Gove et al. (1995) and Raizada and Grossberg (2001).

A.2.1. The odd and even filters

The *odd-symmetric* simple cell filter, $X_{pqijkzst}$, of orientation k , contrast polarity z , scale s and symmetry $y = 1$ is defined by a Difference Of Offset Gaussians (DOOG):

$$X_{pqijkzst} = K_{ijkzs} \Omega_{pqijkzs}, \quad (\text{A.14})$$

where the normalization factor, K_{ijkzs} , is given by

$$K_{ijkzs} = \frac{1}{\sum_{pq} |\Omega_{pqijkzs}|}, \quad (\text{A.15})$$

and the odd-symmetric DOOG filter, $\Omega_{pqijkzs}$, with orientation k , contrast polarity z and scale s is given by

$$\Omega_{pqijkzs} = z \left(G_{pq} \left(i - \delta_s \cos \left(- \left(\theta_k - \frac{\pi}{2} \right) \right), \right. \right. \\ \left. \left. j - \delta_s \sin \left(- \left(\theta_k - \frac{\pi}{2} \right) \right), \varphi_s, \iota_s, \theta_k \right) \right. \\ \left. - G_{pq} \left(i + \delta_s \cos \left(- \left(\theta_k - \frac{\pi}{2} \right) \right), \right. \right. \\ \left. \left. j + \delta_s \sin \left(- \left(\theta_k - \frac{\pi}{2} \right) \right), \varphi_s, \iota_s, \theta_k \right) \right). \quad (\text{A.16})$$

In Eq. (A.16), G_{pq} represents a generalized 2D Gaussian, defined by Eq. (A.17), that can be elongated and oriented. The parameter φ_s represents the standard deviation of the Gaussian in the direction perpendicular to the direction of orientation (width), while ι_s corresponds to the standard deviation of the Gaussian in the direction of orientation (length). Parameter δ_s controls the separation of the Gaussians along the direction perpendicular to the direction of

orientation. The parameters φ_s , l_s and δ_s are all indexed by scale s , and θ_k defines the orientations of the DOOG filter.

The generalized 2D Gaussian, $G_{pq}(i, j, \alpha, \beta, \theta)$, used in Eq. (A.16) is given by:

$$G_{pq}(i, j, \alpha, \beta, \theta) = \frac{1}{\sum_{uv} \exp(f_{uv}(i, j, \alpha, \beta, \theta))} \times \exp(f_{pq}(i, j, \alpha, \beta, \theta)), \quad (\text{A.17})$$

where positions (i, j) vary over the filter support (centered at position (p, q)). In the sum, u and v are used as dummy variables. Parameters α and β are constants determining the standard deviations of the major and minor axes of the Gaussian, respectively. They determine how elongated the Gaussian is. On the other hand the parameter θ determines the orientation of the Gaussian. The function $f_{pq}(i, j, \alpha, \beta, \theta)$ is given by:

$$f_{pq}(i, j, \alpha, \beta, \theta) = \frac{1}{2\alpha^2\beta^2} \left(a(\alpha, \beta, \theta)(p-i)^2 + 2b(\alpha, \beta, \theta)(p-i)(q-j) + c(\alpha, \beta, \theta)(q-j)^2 \right), \quad (\text{A.18})$$

where

$$a(\alpha, \beta, \theta) = \beta^2 \cos^2 \left(\theta - \frac{\pi}{2} \right) + \alpha^2 \sin^2 \left(\theta - \frac{\pi}{2} \right), \quad (\text{A.19})$$

$$b(\alpha, \beta, \theta) = (\alpha^2 + \beta^2) \cos^2 \left(\theta - \frac{\pi}{2} \right) + \sin^2 \left(\theta - \frac{\pi}{2} \right), \quad (\text{A.20})$$

and

$$c(\alpha, \beta, \theta) = \alpha^2 \cos^2 \left(\theta - \frac{\pi}{2} \right) + \beta^2 \sin^2 \left(\theta - \frac{\pi}{2} \right). \quad (\text{A.21})$$

The *even-symmetric* simple cell filter, $X_{pqijkzs2}$, of orientation k , contrast polarity z , scale s , and symmetry $y = 2$ is defined by a Difference Of Gaussians (DOG):

$$X_{pqijkzs2} = M_{ijkzs} \Theta_{pqijkzs}, \quad (\text{A.22})$$

where the normalization factor, M_{ijkzs} , is given by

$$M_{ijkzs} = \frac{1}{\sum_{pq} |\Theta_{pqijkzs}|}, \quad (\text{A.23})$$

and the even-symmetric DOG filter, $\Theta_{pqijkzs}$, with orientation k , contrast polarity z and scale s is given by

$$\Theta_{pqijkzs} = z(G_{pq}(i, j, \varphi_s, l_s, \theta_k) - G_{pq}(i, j, \kappa_s, l_s, \theta_k)). \quad (\text{A.24})$$

In Eq. (A.24), G_{pq} represents the generalized 2D Gaussian defined by Eq. (A.17). The parameter φ_s represents the standard deviation of the narrower Gaussian in the direction perpendicular to the direction of orientation (inner width). As was the case with the DOOG, the parameter l_s corresponds to the standard deviation of the Gaussian in the direction of orientation (length), while κ_s corresponds to the standard deviation of the outer Gaussian in the same direction (outer width). The parameters φ_s , l_s and κ_s are all indexed by scale s , and θ_k defines the orientations of the DOG filter.

A.3. Stage 3: complex cells

The activity c_{ijks} of the complex cell at position (i, j) , orientation k , and scale s , depends on the pooled odd and even simple cell responses of opposite-polarity, in a manner adapted from Grossberg and McLoughlin (1997):

$$c_{ijks} = \sum_{z \in \{-1, 1\}, y \in \{1, 2\}} [B_{ijk(z)sy} - B_{ijk(-z)sy}]^+, \quad (\text{A.25})$$

where

$$w(x, y) = [x - \lambda]^+ = \max(x - \lambda, 0), \quad (\text{A.26})$$

for which x is a real variable and λ is a real parameter (in the case of Eq. (A.25), $\lambda = 0$). In Eq. (A.25), the sum is over the simple cell index z which has the values of either 1 or -1 depending on the contrast polarity of the filter, and index y which has the values $y = 1$ for odd and $y = 2$ for even filter symmetry. For the corresponding position, orientation and scale, $B_{ijk(z)sy}$ and $B_{ijk(-z)sy}$ represent the odd and even simple cell activations of opposite contrast polarity. The output signal of the complex cells, C_{ijks} , results from thresholding the activity, c_{ijks} :

$$C_{ijks} = w(c_{ijks}, \gamma), \quad (\text{A.27})$$

where the signal function $w(\dots)$ is given by Eq. (A.26) and γ is the threshold parameter.

A.4. Stage 4: spatial competition cells

The spatial competition cells are represented by a shunting network with concentric DOG interaction kernels. The bottom-up input comes from the multiple-scale complex cells, C_{ijks} , and the top-down feedback comes from the multiple-depth bipole cells, F_{ijkd} . Both these inputs come from position (i, j) and orientation k . However, the C_{ijks} input comes from scale s , while the F_{ijkd} input comes from depth d . These inputs combine to give a total input T_{ijks} to the spatial competition cell at position (i, j) , orientation k , and scale s :

$$T_{ijks} = C_{ijks} \left(1 + D \sum_d W_{ds} F_{ijkd} \right). \quad (\text{A.28})$$

Here depth is indexed by $d = 1, \dots, 6$ ($d = 1$ for furthest, $d = 6$ for nearest) and D is the top-down gain parameter. Parameters W_{ds} correspond to the weights that realize the scale-to-depth mapping. The activity l_{ijks} of spatial competition cells at position (i, j) , orientation k , and scale s obeys the following on-center off-surround shunting equation:

$$\frac{d}{dt} l_{ijks} = -E l_{ijks} + (F - l_{ijks}) \sum_{pq} P_{pqijs} T_{pqks} - (l_{ijks} + G) \sum_{pq} Y_{pqijs} T_{pqks}, \quad (\text{A.29})$$

where E represents the decay rate, F represents the excitatory saturation potential, P_{pqijs} represents the on-center kernel, T_{ijks} is the total input to the spatial competition

cells, G is the hyperpolarization parameter, and Y_{pqijs} represents the off-surround kernel. Moreover, the two isotropic Gaussian kernels, P_{pqijs} and Y_{pqijs} , obey the following equations:

$$P_{pqijs} = \frac{1}{\sum_{uv} \exp\left(-\frac{1}{2\vartheta_s^2}((u-i)^2 + (v-j)^2)\right)} \times \exp\left(-\frac{1}{2\vartheta_s^2}((p-i)^2 + (q-j)^2)\right), \quad (\text{A.30})$$

and

$$Y_{pqijs} = \frac{1}{\sum_{uv} \exp\left(-\frac{1}{2\zeta_s^2}((u-i)^2 + (v-j)^2)\right)} \times \exp\left(-\frac{1}{2\zeta_s^2}((p-i)^2 + (q-j)^2)\right), \quad (\text{A.31})$$

respectively, where for the on-center kernel, P_{pqijs} , the standard deviation of ϑ_s is used for each of the scales indexed by s , and for the off-surround kernel, Y_{pqijs} , the standard deviation of ζ_s is used for each of the scales indexed by s . In the sums of Eqs. (A.30) and (A.31), u and v represent dummy variables. At equilibrium, Eq. (A.29) obeys:

$$l_{ijks} = \frac{F \sum_{pq} P_{pqijs} T_{pqks} - G \sum_{pq} Y_{pqijs} T_{pqks}}{E + \sum_{pq} P_{pqijs} T_{pqks} + \sum_{pq} Y_{pqijs} T_{pqks}}. \quad (\text{A.32})$$

The output signal of the spatial competition cells is the half-wave rectification of the activity l_{ijks} :

$$L_{ijks} = [l_{ijks}]^+. \quad (\text{A.33})$$

These equations for the spatial competition stage have been adapted from Grossberg et al. (1995).

A.5. Stage 5: orientation competition cells

The orientation competition cells receive input from the spatial competition cells, L_{ijks} . The activity e_{ijks} of the orientation competition cell at position (i, j) , orientation k , and scale s obeys the following on-center off-surround shunting equation:

$$\frac{d}{dt} e_{ijks} = -H e_{ijks} + (I - e_{ijks}) \sum_r N_{kr} L_{ijrs} - (e_{ijks} + J) \sum_r Z_{kr} L_{ijrs}, \quad (\text{A.34})$$

where H represents the decay rate, I represents the excitatory saturation potential, $r = 1, \dots, 16$ indexes over the orientations, N_{kr} represents the orientation on-center kernel, L_{ijks} is the input from the spatial competition

cells, J is the hyperpolarization parameter, and Z_{kr} represents the orientation off-surround kernel. The two Gaussian kernels, N_{kr} and Z_{kr} , both obey the following equations:

$$N_{kr} = K \exp(-\mu(r-k)^2) \quad (\text{A.35})$$

and

$$Z_{kr} = L \exp(-v(r-k)^2), \quad (\text{A.36})$$

for which k and r index orientation, K and L are weighting parameters and μ and v control the widths of the kernels. At equilibrium, Eq. (A.34) obeys:

$$e_{ijks} = \frac{I \sum_r N_{kr} L_{ijrs} - J \sum_r Z_{kr} L_{ijrs}}{H + \sum_r N_{kr} L_{ijrs} + \sum_r Z_{kr} L_{ijrs}}. \quad (\text{A.37})$$

The output signal of the orientation competition cells is the half-wave rectification of the activity e_{ijks} :

$$E_{ijks} = [e_{ijks}]^+. \quad (\text{A.38})$$

These equations for the orientation competition stage have been adapted from Grossberg and Mingolla (1985b, 1987).

A.6. Stage 6: bipole grouping cells

The multiple-depth bipole cells group signals from the multiple-scale orientation competition cells. The input U_{ijkd} to the bipole cells at position (i, j) , orientation k , and depth d is defined as a weighted and thresholded sum of the orientation competition cell responses, E_{ijks} :

$$U_{ijkd} = w \left(\sum_s W_{ds} E_{ijks}, \sigma \right), \quad (\text{A.39})$$

where E_{ijks} and U_{ijkd} lie at position (i, j) and orientation k . However, E_{ijks} lies at scale s , while U_{ijkd} lies at depth d . Signal function $w(\cdot, \cdot)$ is defined by Eq. (A.26). Parameters W_{ds} are the weights that realize the depth-to-scale mapping. The activity f_{ijkd} of the bipole cells at position (i, j) , orientation k , and depth d obey the following shunting equation:

$$\frac{d}{dt} f_{ijkd} = -M f_{ijkd} + (N - f_{ijkd}) \times \left[O U_{ijkd} + \sum_{pqr} H_{pqrijkd}^L U_{pqrd} + \sum_{pqr} H_{pqrirkd}^R U_{pqrd} \right] - (f_{ijkd} + p) [Z_{ijkdL} + Z_{ijkdR}], \quad (\text{A.40})$$

where M represents the decay rate, N represents the excitatory saturation potential, O represents the bottom-up input weight, U_{ijkd} is the total input to the bipole cells, $H_{pqrijkd}^L$ and $H_{pqrirkd}^R$ represent the left and right lobes of the bipole grouping filter, respectively, and are defined in the immediately following Appendix A.6.1, P is the hyperpolarization parameter, and Z_{ijkdL} and Z_{ijkdR} correspond to the left and right inhibitory interneurons (see Eqs. (A.42) and (A.43)) that act to assure the selectivity of each bipole cell. At equilibrium, Eq. (A.40) yields

$$f_{ijkl} = \frac{N[OU_{ijkl} + \sum_{pqr} H_{pqrijkd}^L U_{pqrd} + \sum_{pqr} H_{pqrijkd}^R U_{pqrd}] - P[Z_{ijkdL} + Z_{ijkdR}]}{M + OU_{ijkl} + \sum_{pqr} H_{pqrijkd}^L U_{pqrd} + \sum_{pqr} H_{pqrijkd}^R U_{pqrd} + Z_{ijkdL} + Z_{ijkdR}}. \quad (\text{A.41})$$

The activations of the left, Z_{ijkdL} , and right, Z_{ijkdR} , inhibitory interneurons at position (i, j) , orientation k , and depth d are given at equilibrium by:

$$Z_{ijkdL} = \frac{\sum_{pqr} H_{pqrijkd}^L U_{pqrd}}{1 + QZ_{ijkdR}}, \quad (\text{A.42})$$

and

$$Z_{ijkdR} = \frac{\sum_{pqr} H_{pqrijkd}^R U_{pqrd}}{1 + QZ_{ijkdL}}, \quad (\text{A.43})$$

where for the left inhibitory interneuron response, Z_{ijkdL} , the left lobe of the bipole filter, $H_{pqrijkd}^L$, is convolved with the total bipole input, U_{ijkd} , and the right inhibitory neuron activation, Z_{ijkdR} , inhibits the left inhibitory interneuron through the denominator; and vice versa for the right inhibitory interneuron response, Z_{ijkdR} . Parameter Q is the interneuron cross-inhibitory weight.

The output of the bipole cells F_{ijkd} results from thresholding the activity f_{ijkd} :

$$F_{ijkd} = w(f_{ijkd}, \tau), \quad (\text{A.44})$$

where τ is the output threshold parameter and the signal function $w(\dots)$ is defined by Eq. (A.26). These equations defining the bipole cells have been adapted from Raizada and Grossberg (2001).

A.6.1. Bipole kernel

The kernels $H_{pqrijkd}^L$ and $H_{pqrijkd}^R$ are obtained from the set of bipole kernels $H_{pqrijkd}$ with orientation k , and depth d . For each orientation k , the kernel $H_{pqrijkd}^L$ is the left (L) lobe of the bipole that is defined in Eq. (A.53), while $H_{pqrijkd}^R$ is the right (R) lobe of the bipole that is defined in Eq. (A.54). The bipole kernel $H_{pqrijkd}$ for each orientation k is formed by rotating the coordinate frame of the kernel below, which corresponds to the vertical bipole filter $H_{pqrij1d}$ (i.e., $k=1$) and has been adapted from Gove et al. (1995):

$$H_{pqrij1d} = RK_{ijd} \Psi_{pqrijd}, \quad (\text{A.45})$$

where R is a weighting factor, the normalization factor K_{ijd} for depth d is

$$K_{ijd} = \frac{1}{\sum_{pqr} \Psi_{pqrijd}}, \quad (\text{A.46})$$

and the vertical bipole kernel Ψ_{pqrijd} for depth d is

$$\Psi_{pqrijd} = (p-i) \exp \left(-\frac{(A_{pqij} - \rho_d)^2}{2(\zeta_d)^2} - \frac{(B_{pqij})^2}{2(\psi_d)^2} - \frac{(\theta_r - B_{pqij})^2}{2(\zeta_d)^2} \right). \quad (\text{A.47})$$

In Eq. (A.47), the first term in the exponential modulates filter values based on their distance A_{pqij} from the bipole's center:

$$A_{pqij} = \sqrt{(p-i)^2 + (q-j)^2}. \quad (\text{A.48})$$

The parameter ρ_d determines the position of the peak weights from the center of the bipole filter for each depth d , and parameter ζ_d determines the overall length of the bipole.

The second term inside the exponential of Eq. (A.47) determines the spread of bipole sensitivity about the axis of orientation. This term depends on the slope of the tangent at position (p, q) of the circle with radius T_{pqij} , centered at $(0, T_{pqij})$, which passes through $(0, 0)$ (the bipole's origin) and (p, q) . The circle and its tangent are defined by Eqs. (A.51) and (A.52), respectively. The term B_{pqij} penalizes orientations in the filter that have large tangent values:

$$B_{pqij} = \begin{cases} \tan^{-1} \left(\frac{(p-i)}{T_{pqij} - (q-j)} \right) & \text{if } q < p, q > -p \text{ and } q \neq 0, \\ \text{OR if } q > p, q < -p \text{ and } q \neq 0, \\ \infty & \text{otherwise.} \end{cases} \quad (\text{A.49})$$

The parameter ψ_d constrains the angular spread of the bipole weight strengths away from the axis of orientation. In Eq. (A.49), the circle radius T_{pqij} is given by:

$$T_{pqij} = \begin{cases} \frac{(p-i)^2 + (q-j)^2}{2(q-j)} & \text{if } q-j \neq 0, \\ \infty & \text{otherwise.} \end{cases} \quad (\text{A.50})$$

In particular T_{pqij} is the radius of the following circle equation:

$$p^2 + (q - T_{pqij})^2 = (T_{pqij})^2, \quad (\text{A.51})$$

the tangent of which is

$$\frac{dq}{dp} = \frac{p}{T_{pqij} - q}. \quad (\text{A.52})$$

The third term inside the exponential of Eq. (A.47) determines the spread of the bipole's cross-orientational sensitivity. The parameter θ_r represents the angle of each cross-orientation r , and the parameter ζ_d constrains the angular spread with which an arbitrary cross-orientation can influence the bipole weights. When computing $\theta_r - B_{pqij}$ in Eq. (A.47) it is important to make sure that angular values between $[0, 2\pi]$ are used for θ_r and B_{pqij} .

The left $H_{pqrij\lambda d}^L$ and right $H_{pqrij\lambda d}^R$ lobes of the vertical bipole filter ($k = 1$) for depth d are given as follows:

$$H_{pqrij\lambda d}^L = [H_{pqrij\lambda d}]^+, \tag{A.53}$$

and

$$H_{pqrij\lambda d}^R = [-H_{pqrij\lambda d}]^+. \tag{A.54}$$

A.7. Stage 7: depth competition cells

The depth competition cells pool the bipole cell responses from all orientations and then compete across depth with uniform strength. The activity g_{ijd} of the depth competition cell at position at position (i, j) and depth d thus obeys the following shunting equation:

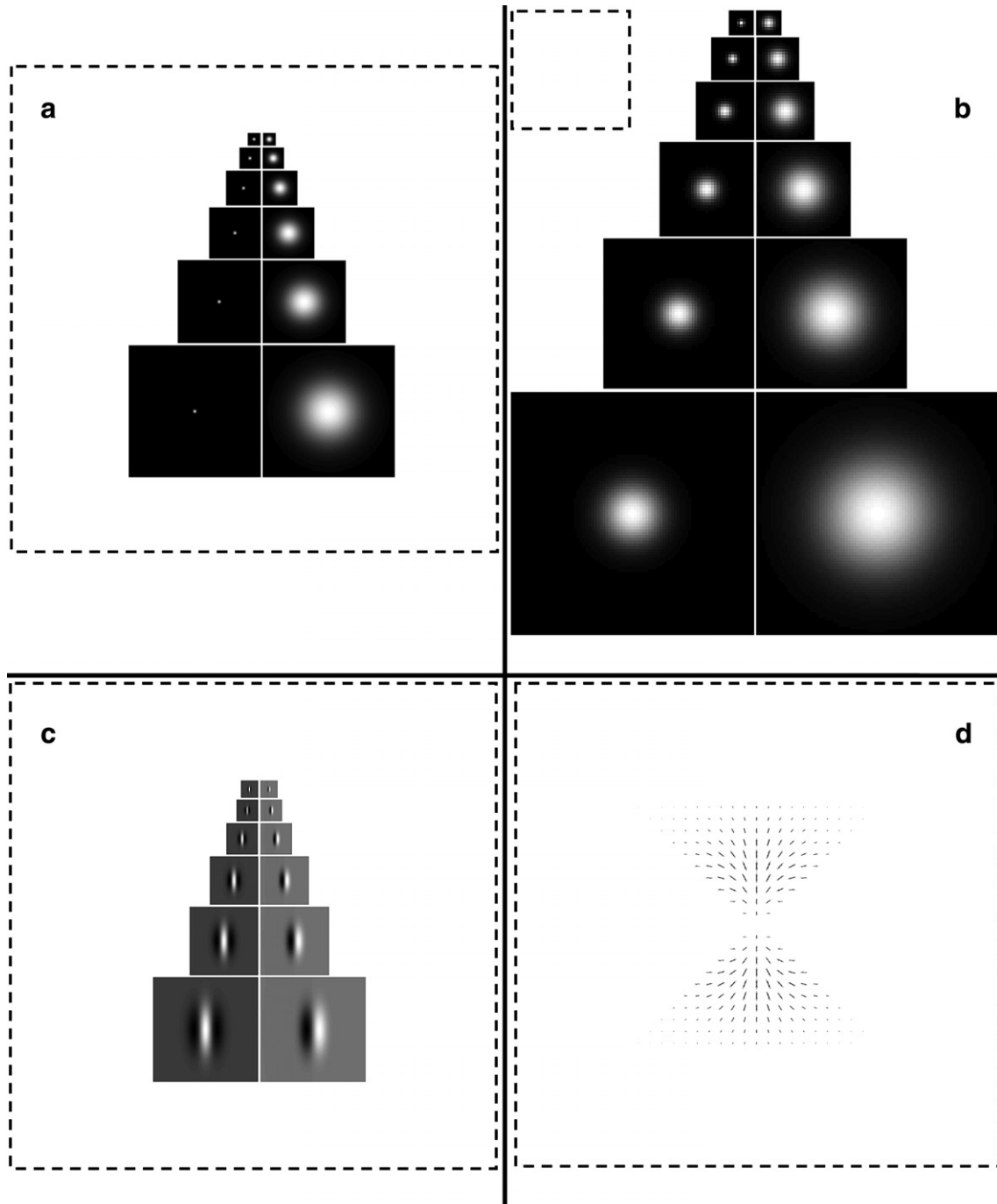


Fig. 16. Spatial filters of the model: (a) The relative sizes of the LGN center-surrounds for all six scales. The centers are positioned on the left and the surrounds on the right. (b) The relative sizes of the vertical simple cell DOGs (left) and DOOGs (right). (c) The relative sizes of the spatial competition stage center-surrounds for all six scales. The centers are positioned on the left and the surrounds on the right. (d) The vertical bipole filter. The longer a needle of a given orientation the stronger the filter weight for that orientation. In (a)–(d), The dashed box indicates the size of the input image.

$$\begin{aligned} \frac{d}{dt}g_{ijd} = & -Sg_{ijd} + (T - g_{ijd}) \sum_k w(F_{ijkd}, v) \\ & - (g_{ijd} + U)V \sum_{k,v \neq d} w(F_{ijkv}, v), \end{aligned} \quad (\text{A.55})$$

where S represents the decay rate, T represents the excitatory saturation potential, F_{ijkd} represents the input from the bipole cell at position (i, j) , orientation k , and depth d , v represents the input threshold, U is the hyperpolarization parameter, V is the across-depth inhibitory weight parameter, and v is the dummy variable for the sum over depth. At equilibrium, Eq. (A.55) obeys:

$$g_{ijd} = \frac{T \sum_k w(F_{ijkd}, v) - UV \sum_{k,v \neq d} w(F_{ijkv}, v)}{S + \sum_k w(F_{ijkd}, v) + V \sum_{k,v \neq d} w(F_{ijkv}, v)}. \quad (\text{A.56})$$

The output of the depth competition cells G_{ijd} results from thresholding the equilibrium activity g_{ijd} :

$$G_{ijd} = w(g_{ijd}, \omega), \quad (\text{A.57})$$

where ω is the output threshold parameter and $w(\dots)$ is defined by Eq. (A.26).

A.8. Stage 8: surface filling-in domain cells

The surface Filling-In Domain, or FIDO, cells within the FCS, prevent inputs from the ON and OFF LGN cells (pooled over scale) from spreading across the multiple-depth boundary inputs that they receive from the BCS. The total input V_{ij}^{\pm} from the ON (+) or OFF (−) LGN cells, respectively, at position (i, j) and summed over scale s , takes the form:

$$V_{ij}^{\pm} = \sum_s A_{ijs}^{\pm}, \quad (\text{A.58})$$

where A_{ijs}^{\pm} , as in Eq. (A.9), represents the activity of the ON (+) or OFF (−) LGN cell at position (i, j) and scale s . The input from the BCS is the output G_{ijd} of the depth competition cells, indexed by position (i, j) and depth d , as in Eq. (A.57). These boundary signals gate the spread of filling-in. The activity h_{ijd}^{\pm} of the ON (+) and OFF (−) FIDO cell at position (i, j) and depth d thus obeys the following diffusion equation:

$$\frac{d}{dt}h_{ijd}^{\pm} = -Wh_{ijd}^{\pm} + \sum_{(p,q) \in \Delta_{ij}} (h_{pqd}^{\pm} - h_{ijd}^{\pm})P_{pqijd} + V_{ij}^{\pm}, \quad (\text{A.59})$$

where W is the decay rate, the boundary gating term equals

$$P_{pqijd} = \frac{\delta}{1 + \varepsilon(G_{pqd} + G_{ijd})}, \quad (\text{A.60})$$

and

$$\Delta_{ij} = \{(i, j - 1), (i - 1, j), (i + 1, j), (i, j + 1)\} \quad (\text{A.61})$$

are the nearest-neighbor cells with which the diffusion occurs around cell (i, j) . The parameters δ and ε are non-negative constants. Increasing δ increases diffusion of the

feature signals, while increasing ε reduces diffusion where the boundaries, G_{ijd} , are present.

A.9. Stage 9: 3D surface representation

The activity x_{ijd} of the 3D surface representation at position (i, j) and depth d is determined by subtracting the OFF FIDO cell responses h_{ijd}^- from the ON FIDO cell responses h_{ijd}^+ :

$$x_{ijd} = h_{ijd}^+ - h_{ijd}^-. \quad (\text{A.62})$$

Appendix B. Model filters

Four sets of spatial filters were used in this model: LGN center-surrounds, simple cell DOGs and DOOGs, spatial competition center-surrounds, and bipole grouping filters. The sizes of the filter supports are given in Table 1 of Appendix F. Fig. 16a diagrams the relative sizes of the LGN center-surrounds for all six scales. The centers are positioned on the left and the surrounds on the right. Fig. 16b diagrams the relative sizes of the vertical simple cell DOGs (left) and DOOGs (right). Fig. 16c diagrams the relative sizes of the spatial competition stage center-surrounds for all six scales. The centers are positioned on the left and the surrounds on the right. Fig. 16d shows the vertical bipole filter which is the same size for all depths.

Appendix C. Scale-to-depth map, case 1: triangular map matrix

Scale-to-depth map, W_{ds} , occurs when the multiple-scale orientation competition cells project to the multiple-depth bipole cells (see Appendix A.6), while the depth-to-scale map, W_{sd} , occurs when the multiple-depth bipole cells project to the multiple-scale spatial competition cells (see Appendix A.4). Both of these mappings are defined by a matrix representing ideas based on the size-depth and

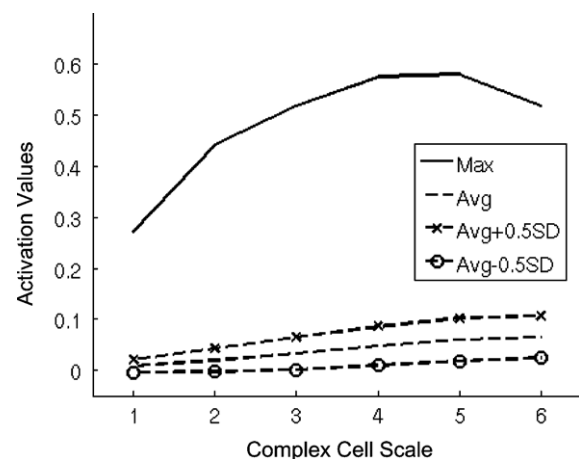


Fig. 17. Activation values across complex cell scale for the number 5 HP case. Scale size increases from 1 to 6.

Table 1

Name	Symbol	Value
<i>Stage 1: LGN shunting network</i>		
Decay rate	A	1
Excitatory saturation potential	B	1
Hyperpolarization parameter	C	1.01
Center kernel standard deviation for each of the 6 scales	ϕ_s	1
Surround kernel standard deviation for each of the 6 scales	η_s	2.0000, 3.2000, 5.1200, 8.1920, 13.1072, 20.9715
Center and surround kernel support size (pixels) for each of the 6 scales		13, 21, 35, 51, 83, 131
<i>General oriented cell parameters</i>		
Total number of orientations		16
Orientation angles (radians)	θ_k	From $\frac{\pi}{2}$ stepping by $\frac{\pi}{16}$ to $\frac{7\pi}{16}$
<i>Stage 2: simple cells</i>		
Odd filter Gaussian separation parameter for each of the 6 scales	δ_s	0.5000, 0.8000, 1.2800, 2.0480, 3.2768, 5.2429
Odd and even filter Gaussian width parameter for each of the 6 scales	φ_s	0.5000, 0.8000, 1.2800, 2.0480, 3.2768, 5.2429
Odd and even filter Gaussian length parameter for each of the 6 scales	t_s	1.5000, 2.4000, 3.8400, 6.1440, 9.8304, 15.7286
Outer width parameter of the even filter Gaussian for each of the 6 scales	κ_s	0.8000, 1.2800, 2.0480, 3.2768, 5.2429, 8.3886
Odd and even filter support size (pixels) for each of the 6 scales		17, 21, 31, 47, 67, 103
<i>Stage 3: complex cells</i>		
Threshold parameter	γ	0.01
<i>Stage 4: spatial competition cells</i>		
Depth-to-scale map matrix	W_{ds}	See Appendices C–E
Bipole-feedback modulation parameter	D	17
Decay rate	E	1
Excitatory saturation potential	F	1
Hyperpolarization parameter	G	0.5
Center kernel standard deviation for each of the 6 scales	ϑ_s	0.6819, 1.0911, 1.7457, 2.7931, 4.4690, 7.1504
Surround kernel standard deviation for each of the 6 scales	ζ_s	1.3638, 2.1821, 3.4914, 5.5862, 8.9380, 14.3007
Kernel support size (pixels) for each of the 6 scales		9, 15, 21, 33, 53, 85
<i>Stage 5: orientation competition cells</i>		
Decay rate	H	1
Excitatory saturation potential	I	1
Hyperpolarization parameter	J	0.7
Orientation-center weight parameter	K	1
Orientation-surround weight parameter	L	0.25
Orientation-center width parameter	μ	0.3
Orientation-surround width parameter	v	0.00006
<i>Stage 6: bipole cells</i>		
Scale-to-depth map matrix	W_{sd}	See Appendices C–E
Decay rate	M	4
Excitatory saturation potential	N	1
Bottom up input weight factor	O	0.01
Hyperpolarization parameter	P	1
Interneuron cross-inhibitory weight	Q	50
Bipole filter weight factor	R	2
Input threshold	σ	0.00001
Output threshold	τ	0.00001
Bipole filter term 1 peak-weight shifting parameter for each of the 6 depths	ρ_d	4.8880
Bipole filter term 1 standard deviation for each of the 6 depths	ξ_d	2.7931
Bipole filter term 2 standard deviation for each of the 6 depths	ψ_d	1.6
Bipole filter term 3 standard deviation for each of the 6 depths	ζ_d	0.2
Bipole filter support size (pixels) for each of the 6 depths		21
<i>Stage 7: Depth Competition Cells</i>		
Decay rate	S	1
Excitatory saturation potential	T	1
Hyperpolarization parameter	U	1
Inhibitory weighting parameter	V	0.2
Input threshold for each of the 6 depths	v	0 for map matrix cases 1 and 2, 0.001 for map matrix case 3
Output threshold	ω	0
<i>Stage 8: filling-in cells</i>		
Decay rate	W	10
Diffusion strength parameter	δ	100,000
Boundary strength parameter	ε	100,000

size-disparity correlations. The scale-to-depth map matrix is given by

$$W_{ds} = \begin{cases} \omega_d \exp(-\alpha(d-s)^2) & \text{if } d \leq s, \\ 0 & \text{otherwise} \end{cases} \quad (\text{C.1})$$

where $\omega_d = 0.47, 0.41, 0.40, 0.43, 0.65, 1.25$ for depths $d = 1, \dots, 6$, respectively, ($d = 1$ for furthest, $d = 6$ for nearest) and $\alpha = 0.08$. The depth-to-scale map matrix is given by $W_{sd} = W_{ds}$. The bowing of the depth weights, ω_d , is consistent with the idea that the middle scales are activated more optimally than the outlying, large and small, scales since the optimal sensitivity of the middle scales lies near the center of the expected statistical distribution of texture element sizes of the images investigated. In particular, the depth weights, ω_d , are smaller for the middle scales than they are for the outlying scales. Fig. 17 visualizes this property by showing that the maximum activation (solid line) of the complex cells in response to the number 5 high perspective (HP) ellipsoid is greatest for the middle scales. This maximum activation curve best reflects the optimal activation of the complex cells across scale. The average activation (dashed line) of the complex cells across scale does not show the same trend. Instead, the average activation increases with scale size probably because the large scales are sensitive to more distant regions of the image, allowing them to respond at more points in the image, whether the responses be optimal or sub-optimal. The dash-cross and dash-circle lines represent the average activation plus or minus half of one standard deviation, respectively.

Appendix D. Scale-to-depth map, case 2: conservation of synaptic sites

This case is a variation upon the triangular map matrix which also represents ideas based on the size-depth and size-disparity correlations. This case ensures that the sum of the input weights projecting to each depth, or each scale, is constant across all depths and scales, thereby conserving the number of synaptic sites attaching to a bipole cell of a given depth, or a spatial competition cell of a given scale. The scale-to-depth map matrix is given by

$$W_{ds} = \frac{\chi}{\sum_s M_{ds}} M_{ds}, \quad (\text{D.1})$$

where $\chi = 1.3$,

$$M_{ds} = \begin{cases} \exp(-\varpi(d-s)^2) & \text{if } d \leq s, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{D.2})$$

and $\varpi = 0.08$. The depth-to-scale map matrix is given by

$$W_{sd} = \frac{\chi}{\sum_d M_{sd}} M_{sd}, \quad (\text{D.3})$$

where

$$M_{sd} = \begin{cases} \exp(-\varpi(d-s)^2) & \text{if } s \leq d, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{D.4})$$

In this case the depth weights, $\chi/\sum_s M_{ds}$, and scale weights, $\chi/\sum_d M_{sd}$, increase with scale size because the near depths and large scales receive input from fewer scales and depths, respectively. Thus they need larger weights to ensure that the sum of the input weights to each depth or each scale is the same.

Appendix E. Scale-to-depth map, case 3: diagonal map matrix

For this case, a purely diagonal map matrix was used such that the scale-to-depth and depth-to-scale maps were one-to-one. Thus this matrix only represents the size-depth correlation. The scale-to-depth map matrix is given by

$$[W_{ds}] = \begin{bmatrix} 0.66 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.35 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.33 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.34 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 \end{bmatrix}. \quad (\text{E.1})$$

The depth-to-scale map matrix is given by $W_{sd} = W_{ds}$. In this case, the diagonal depth weights also bow, however there is greater weight at the far depths, compared with the triangular map matrix case, because they now only receive small scale input in a one-to-one manner.

Appendix F. Table of model parameters

Table 1.

References

- Adelson, E., Anderson, C., Bergen, J., Burt, P., & Ogden, J. (1984). Pyramid methods in image processing. *RCA Engineer*, 29(6), 33–41.
- Aloimonos, Y., (1986). Detection of surface orientation from texture. i. the case of planes. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 584–593.
- Aloimonos, Y. (1988). Shape from texture. *Biological Cybernetics*, 58, 345–360.
- Baccus, S., & Meister, M. (2002). Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36, 909–919.
- Bajcsy, R., & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing*, 5, 52–67.
- Barlow, H. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology (London)*, 119, 69–88.
- Barlow, H., Snodderly, D., & Swadlow, H. (1978). Intensity coding in primate visual system. *Experimental Brain Research*, 31, 163–177.
- Bartlett, J., & Doty, R. (1974). Response of units in striate cortex of squirrel monkeys to visual and electrical stimuli. *Journal of Neurophysiology*, 37, 621–641.
- Blake, A., Bülthoff, H., & Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*(12), 1723–1737.
- Blake, A., & Marinos, C. (1990). Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence*, 45, 323–380.
- Bosking, W., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, 17(6), 2112–2127.
- Brown, J., & Weisstein, N. (1988). A spatial frequency effect on perceived depth. *Perception and Psychophysics*, 44, 157–166.

- Braunstein, M. (Ed.). (1976). *Depth perception through motion*. New York: Academic Press.
- Braunstein, M., & Payne, J. (1969). Perspective and form ratio as determinants of relative slant judgements. *Journal of Experimental Psychology*, 81, 584–590.
- Brown, L., & Shvayster, H. (1990). Surface orientation from projective foreshortening of isotropic texture autocorrelation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6), 584–588.
- Burt, P., & Adelson, E. (1983). The laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, 31, 532–540.
- Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21, 47–74.
- Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18, 515–578.
- Clerc, M., & Mallat, S. (2002). The texture gradient equation for recovering shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), 536–549.
- Cohen, M., & Grossberg, S. (1984). Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. *Perception and Psychophysics*, 36, 428–456.
- Connor, C. (2002). Reconstructing a 3D world. *Science*, 298, 376–377.
- Cook, P., & McReynolds, J. (1998). Lateral inhibition in the inner retina is important for spatial tuning of ganglion cells. *Nature Neuroscience*, 1, 714–719.
- Cumming, B., Johnston, E., & Parker, A. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, 33(5/6), 827–838.
- Cumming, B., & Parker, A. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *Journal of Neuroscience*, 19(13), 5602–5618.
- Cumming, B., & Parker, A. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *Journal of Neuroscience*, 20(12), 4758–4767.
- Cutting, J., & Millard, R. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, 113(2), 198–216.
- Davis, L., Janos, L., & Dunn, S. (1983). Efficient recovery of shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(5).
- De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545–559.
- De Valois, R., & De Valois, K. (1980). Spatial vision. *Annual Review of Psychology*, 31, 309–341.
- DeAngelis, G. (2000). Seeing in three dimensions: the neurophysiology of stereopsis. *Trends in Cognitive Sciences*, 4(3), 80–90.
- DeAngelis, G., Ohzawa, I., & Freeman, R. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neuroscience*, 18, 451–458.
- Demb, J. (2002). Multiple mechanisms for contrast adaptation in the retina. *Neuron*, 36, 781–783.
- Dubin, M., & Cleland, B. (1977). Organization of visual inputs to interneurons of lateral geniculate nucleus of the cat. *Journal of Neurophysiology*, 40, 410–427.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local “association field”. *Vision Research*, 33, 173–193.
- Forsyth, D. (2001). Shape from texture and integrability. *Proceedings of the IEEE International Conference on Computer Vision*, 447–453.
- Forsyth, D., & Ponce, J. (Eds.). (2003). *Computer vision: A modern approach*. Upper Saddle River, New Jersey: Prentice Hall.
- Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular Reflections and the Perception of Shape. *Journal of Vision*, 4(9), 798–820.
- Freeman, W., & Adelson, E. (2002). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 891–906.
- Freeman, W., & Torralba, A. (2002). Shape recipes: Scene representations that refer to the image. *MIT AI Lab. Memo 016*, 1–13.
- Friedman, H., Zhou, H., & Heydt, R. V. D. (2003). The coding of uniform color figures in monkey visual cortex. *Journal of Physiology*, 584(2), 593–613.
- Gårding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, 2(4), 327–350.
- Gårding, J. (1993). Shape from texture and contour by weak isotropy. *Journal of Artificial Intelligence*, 64, 243–297.
- Gibson, J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Gove, A., Grossberg, S., & Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience*, 12(6), 1027–1052.
- Grossberg, S. (1973). Contour enhancement, short-term memory and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 217–257.
- Grossberg, S. (1983). The quantized geometry of visual space: The coherent computation of depth, form and lightness. *Behavioural and Brain Sciences*, 6, 625–692.
- Grossberg, S. (1987a). Cortical dynamics of three-dimensional form, color and brightness perception, I: Monocular theory. *Perception and Psychophysics*, 41, 87–116.
- Grossberg, S. (1987b). Cortical dynamics of three-dimensional form, color and brightness perception, II: Binocular theory. *Perception and Psychophysics*, 41, 117–158.
- Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, 55(1), 48–120.
- Grossberg, S. (1997). Cortical Dynamics of three-dimensional figure-ground perception of two-dimensional figures. *Psychological Review*, 104, 618–658.
- Grossberg, S., & Hong, S. (2006). A neural model of surface perception: Lightness, anchoring and filling-in. *Spatial Vision*, 19, 263–321.
- Grossberg, S., & Howe, P. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research*, 43, 801–829.
- Grossberg, S., Hwang, S., & Mingolla, E. (2002). Thalamocortical dynamics of the mcolough effect: boundary surface alignment through perceptual learning. *Vision Research*, 42, 1259–1286.
- Grossberg, S., & McLoughlin, N. (1997). Cortical dynamics and three-dimensional surface perception: Binocular and half-occluded scenic images. *Neural Networks*, 10(9), 1583–1605.
- Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures and neon color spreading. *Psychological Review*, 92, 173–211.
- Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries and emergent segmentations. *Perception and Psychophysics*, 38, 141–171.
- Grossberg, S., & Mingolla, E. (1987). Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading. *Computer Vision, Graphics and Image Processing*, 37, 116–165.
- Grossberg, S., Mingolla, E., & Williamson, J. (1995). Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. *Neural Networks*, 8(7/8), 1005–1028.
- Grossberg, S., & Swaminathan, G. (2004). A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention and bistability. *Vision Research*, 44, 1147–1187.
- Grossberg, S., & Todorović, D. (1988). Neural dynamics of 1D and 2D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics*, 43, 241–277.
- Grossberg, G., & Williamson, J. R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual groupings and learning. *Cerebral Cortex*, 11, 37–58.
- Grossberg, S., & Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research*, 45, 1725–1743.

- Hawken, M., & Parker, A. (1991). Spatial receptive field organization in monkey V1 and its relationship to the cone mosaic. In M. Landy & J. Movshon (Eds.), *Computational models of visual processing* (pp. 119–133). Cambridge, Massachusetts: The MIT Press.
- Heeger, D. (1991). Nonlinear model of neural responses in cat visual cortex. In M. Landy & J. Movshon (Eds.), *Computational models of visual processing* (pp. 119–133). Cambridge, Massachusetts: The MIT Press.
- Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., & Kubler, O. (1998). Simulation of neural contour mechanisms: representing anomalous contours. *Image and Vision Computing*, 16, 407–421.
- Hinkle, D., & Connor, C. (2002). Three-dimensional orientation tuning in macaque area V4. *Nature Neuroscience*, 5(7), 665–670.
- Hirsch, J., Alonso, J., Reid, R., & Martinez, L. (1998). Synaptic integration in striate cortical simple cells. *Journal of Neuroscience*, 18(22), 9517–9528.
- Hodgkin, A. L. (1964). *The conduction of the nervous impulse*. Springfield, IL: Charles C. Thomas.
- Hong, S., & Grossberg, S. (2004). A neuromorphic model for achromatic and chromatic surface representation of natural images. *Neural Networks*, 17, 787–808.
- Howard, I. (2003). Neurons that respond to more than one depth cue. *Trends in Neuroscience*, 26(10), 515–517.
- Hubel, D., & Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574–591.
- Hubel, D., & Wiesel, T. (1961). Integrative interaction in the cat's lateral geniculate body. *Journal of Physiology*, 155, 385–398.
- Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's striate cortex. *Journal of Physiology*, 160, 106–154.
- Ikeuchi, K. (1984). Shape from regular patterns. *Journal of Artificial Intelligence*, 22, 49–75.
- Janssen, P., Vogels, R., & Orban, G. (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proceedings of the National Academy of Sciences USA*, 96, 8217–8222.
- Janssen, P., Vogels, R., & Orban, G. (2000a). Three-dimensional shape coding in inferior temporal cortex. *Science*, 288, 2054–2056.
- Janssen, P., Vogels, R., & Orban, G. (2000b). Three-dimensional shape coding in inferior temporal cortex. *Neuron*, 27, 385–397.
- Jones, H., Andolina, I., Oakely, N., Murphy, P., & Sillito, A. (2000). Spatial summation in lateral geniculate nucleus and visual cortex. *Experimental Brain Research*, 135, 279–284.
- Jones, J., & Palmer, L. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58, 1212–1232.
- Julesz, B., & Schumer, R. (1981). Early visual perception. *Annual Review of Psychology*, 32, 572–627.
- Kanatani, K. (1984). Detection of surface orientation and motion from texture by a stereological technique. *Artificial Intelligence*, 23, 213–237.
- Kanatani, K., & Chou, T. (1989). Shape from texture: General principle. *Journal of Artificial Intelligence*, 38, 1–48.
- Kandel, E., Schwartz, J., & Jessell, T. (Eds.). (1991). *Principles of neural science* (3rd ed.). East Norwalk, Connecticut: Appleton and Lange.
- Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial contribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, 84, 2048–2062.
- Kayama, Y., Riso, R., Bartlett, J., & Doty, R. (1979). Luxotonic responses of units in macaque striate cortex. *Journal of Neurophysiology*, 42, 1495–1517.
- Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23(2), 141–221.
- Kelly, F., & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception and Psychophysics*, 62, 1596–1619.
- Kinoshita, M., & Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *Journal of Neurophysiology*, 86, 2559–2570.
- Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2004). Pointing out of the picture. *Perception*, 33, 513–530.
- Komatsu, H., Murakami, I., & Kinoshita, M. (1996). Surface representation in the visual system. *Cognitive Brain Research*, 5, 97–104.
- Kuffler, S. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37–68.
- Kuffler, S. (1978). Limit of single vision in stereopsis depends on contour sharpness. *Nature*, 275, 126–127.
- Land, E. H. (1986). Recent advances in Retinex theory. *Vision Research*, 26(1), 7–21.
- Lee, B. (1999). Aftereffects and the representation of stereoscopic surfaces. *Perception*, 28, 1155–1169.
- Li, A., & Zaidi, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Research*, 40, 217–242.
- Li, A., & Zaidi, Q. (2001). Information limitations in perception of shape from texture. *Vision Research*, 41, 2927–2942.
- Li, A., & Zaidi, Q. (2004). Three-dimensional shape from non-homogeneous textures: Carved and stretched surfaces. *Journal of Vision*, 4(10), 860–878.
- Liu, Y., Vogels, R., & Orban, G. (2004). Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *Journal of Neuroscience*, 24(15), 3795–3800.
- Livingstone, M., & Hubel, D. (1984). Anatomy and physiology of the color system in the primate visual cortex. *Journal of Neuroscience*, 4, 309–356.
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement and depth: Anatomy, physiology and perception. *Science*, 240, 740–749.
- MacEvoy, S., Kim, W., & Paradiso, M. (1998). Integration of surface information in primary visual cortex. *Nature Neuroscience*, 1, 616–620.
- Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7(5), 923–932.
- Malik, J., & Rosenholtz, R. (1997). Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 23(2), 149–168.
- Marinos, C., & Blake, A. (1990). Shape from texture: The homogeneity hypothesis. *Proceedings of the International Conference on Computer Vision*, 350–353.
- Marrocco, R. (1972). Maintained activity of monkey optic tract fibers and lateral geniculate nucleus cells. *Vision Research*, 12, 1175–1181.
- Martin, G. (1983). Schematic eye models in vertebrates. *Progress in Sensory Physiology*, 4, 44–81.
- McGuire, B., Gilbert, C., Rivlin, P., & Wiesel, T. (1991). Targets of horizontal connections in macaque primary visual cortex. *Journal of Comparative Neurology*, 305(3), 370–392.
- Mingolla, E., Ross, W., & Grossberg, S. (1999). A neural network for enhancing boundaries and surfaces in synthetic aperture radar images. *Neural Networks*, 12, 499–511.
- Mingolla, E., & Todd, J. (1984). Computational techniques for graphic simulation of quadric surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 740–745.
- Nguyenkim, J., & DeAngelis, G. (2001). MT neurons are selective for 3D surface orientation defined by disparity gradients. *Society of Neuroscience Abstracts*, 165.8.
- Nguyenkim, J., & DeAngelis, G. (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *Journal of Neuroscience*, 23, 7117–7128.
- Papiaionnou, J., & White, A. (1972). Maintained activity of lateral geniculate nucleus neurons as a function of background luminance. *Experimental Neurology*, 34, 558–566.
- Parker, A., & Hawken, M. (1988). Two-dimensional spatial structure of receptive fields in monkey striate cortex. *Journal of the Optical Society of America A*, 5, 598–605.
- Pessoa, L., Mingolla, E., & Neumann, H. (1995). A contrast- and luminance-driven multiscale network model of brightness perception. *Vision Research*, 35, 2201–2223.
- Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *Journal of Neuroscience*, 9, 1749–1763.

- Poggio, G. (1995). Mechanisms of stereopsis in monkey visual cortex. *Cerebral Cortex*, 3, 193–204.
- Prince, S., Cumming, B., & Parker, A. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology*, 87, 209–221.
- Prince, S., & Eagle, R. (1999). Size-disparity correlation in human binocular depth perception. *Proceedings of the Royal Society of London - series B*, 266, 1361–1365.
- Purdy, W., (1958). The hypothesis of psychophysical correspondence in space perception. University Microfilms No. 58–5594.
- Rahman, Z., Jobson, D. J., & Woodell, G. A. (2004). Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13(1), 100–110.
- Raizada, R., & Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention and orientation contrast. *Visual Cognition*, 8(3/4/5), 431–466.
- Reid, R., & Alonso, J. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378, 281–284.
- Richards, W., & Kaye, M. (1974). Local versus global stereopsis: Two mechanisms. *Vision Research*, 14, 1345–1347.
- Ringach, D., Shapley, R., & Hawken, M. (2002). Orientation selectivity in macaque V1: Diversity and laminar dependence. *Journal of Neuroscience*, 22(13), 5639–5651.
- Rosenholtz, R., & Malik, J. (1997). Surface orientation from texture: Isotropy or homogeneity (or both)? *Vision Research*, 37(16), 2283–2293.
- Roska, B., Nemeth, E., Orzo, L., & Werblin, F. S. (2000). Three levels of lateral inhibition: a space-time study of the retina of the tiger salamander. *Journal of Neuroscience*, 20, 1941–1951.
- Rossi, A., & Paradiso, M. (1999). Neural correlates of perceived brightness in the retina, lateral geniculate nucleus and striate cortex. *Journal of Neuroscience*, 19, 6145–6156.
- Rossi, A., Rittenhouse, C., & Paradiso, M. (1996). The representation of brightness in primary visual cortex. *Science*, 273, 1104–1107.
- Ryan, C., & Gillam, B. (1993). A proximity-contingent stereopsis depth aftereffect: evidence for adaptation to disparity gradients. *Perception*, 22, 403–418.
- Sakai, K., & Finkel, L. (1995). Characterization of the spatial-frequency spectrum in the perception of shape from texture. *Journal of the Optical Society of America A*, 12(6), 1208–1224.
- Sakai, K., & Finkel, L. (1997). Spatial-frequency analysis in the perception of perspective in depth. *Network: Computation in Neural Systems*, 8, 335–352.
- Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tanaka, Y., & Tsutsui, K. (1998). Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, 353, 1363–1373.
- Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tsutsui, K., Tanaka, Y., et al. (1999). Neural representation of three-dimensional features of manipulation objects with stereopsis. *Experimental Brain Research*, 128, 160–169.
- Schiller, P. H. (1992). The on and off channels of the visual system. *Trends in Neuroscience*, 15, 86–92.
- Schiller, P. H. (1994). Area V4 of the primate visual cortex. *Current Directions in Psychological Science*, 3, 89–92.
- Schiller, P. H. (1995). Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature*, 376, 342–344.
- Schiller, P. H., & Lee, K. (1991). The role of primate extrastriate area V4 in vision. *Science*, 251, 1251–1253.
- Schmidt, K., Goebel, R., Löwel, S., & Singer, W. (1997). The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in primary visual cortex. *European Journal of Neuroscience*, 9, 1083–1089.
- Schor, C., & Tyler, C. (1981). Spatio-temporal properties of Panum's fusional area. *Vision Research*, 21, 683–692.
- Schor, C., & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research*, 23, 1649–1654.
- Schor, C., Wood, I., & Ogawa, J. (1984). Binocular sensory fusion is limited by spatial resolution. *Vision Research*, 24, 661–665.
- Seyama, J., Takeuchi, T., & Sato, T. (2000). Tilt dependency of slant aftereffect. *Vision Research*, 40, 349–357.
- Shikata, E., Tanaka, Y., Nakamura, H., Taira, M., & Sakata, H. (1996). Selectivity of the parietal visual neurones in 3D orientation of surface of stereoscopic stimuli. *NeuroReport*, 7, 2389–2394.
- Simoncelli, E., & Freeman, W., (1995). The steerable pyramid: a flexible architecture for multiscale derivative computation. In: 2nd Annual International Conference. IEEE, pp. 584–593.
- Smallman, H., & MacLeod, D. (1994). Size-disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A*, 11(8), 2169–2183.
- Stevens, K. (1981). The information content of texture gradients. *Biological Cybernetics*, 42, 95–105.
- Stevens, K. (1984). On gradients and texture “gradients”. *Journal of Experimental Psychology: General*, 113, 217–220.
- Super, B., & Bovik, A. (1995). Shape from texture using local spectral moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, 333–343.
- Szulforski, R., & Palmer, L. (1990). The two-dimensional spatial structure of nonlinear subunits of the receptive fields of complex cells. *Vision Research*, 30(2), 249–254.
- Taira, M., Tsutsui, K., Jiang, M., Yara, K., & Sakata, H. (2000). Parietal neurons represent surface orientation from the gradient of binocular disparity. *Journal of Neurophysiology*, 83, 3140–3146.
- Thomas, O. M., Cumming, B. G., & Parker, A. J. (2002). A specialization for relative disparity in V2. *Nature Neuroscience*, 5(5), 472–478.
- Todd, J., & Akerstrom, R. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human perception and performance*, 13(2), 242–255.
- Todd, J., & Mingolla, E. (1984). Simulation of curved surfaces from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 734–739.
- Todd, J., & Oomes, A. (2002). Generic and non-generic conditions for the perception of surface shape from texture. *Vision Research*, 42, 837–850.
- Todd, J., Oomes, A., Koenderink, J., & Kappers, A. (2004). The perception of doubly curved surfaces from anisotropic textures. *Psychological Science*, 15(1), 40–46.
- Todd, J. T., Thaler, L., & Dijkstra, T. M. (2005). The effects of field of view on the perception of 3D slant and texture. *Vision Research*, 45(12), 1501–1517.
- Torralba, A., & Freeman, W. (2002). Properties and applications of shape recipes. *MIT AI Lab. Memo 019*, 1–10.
- Torralba, A., & Oliva, A. (2001). Global depth perception from familiar scene structure. *MIT AI Lab. Memo 036*, 1–22.
- Torralba, A., & Oliva, A. (2002). Depth estimation from image structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9), 1226–1238.
- Tsao, D., (2003). Stereopsis. Ph.D. thesis, HMS Department of Neurobiology, Harvard University.
- Tsao, D., Conway, B., & Livingstone, M. S. (2003). Receptive fields of disparity-tuned simple cells in macaque V1. *Neuron*, 38, 103–114.
- Tsutsui, K., Sakata, H., Naganuma, T., & Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science*, 298, 409–412.
- Tsutsui, K., Taira, M., & Sakata, H. (2005). Neural mechanisms of three-dimensional vision. *Neuroscience Research*, 51(3), 221–229.
- Tyler, C. (1975). Spatial organization of binocular disparity sensitivity. *Vision Research*, 15, 583–590.
- Tyler, C. (1983). Sensory processing of binocular disparity. In C. Schor & K. CuiFredda (Eds.), *Vergence eye movements* (pp. 199–295). Boston: Butterworth.
- von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224(4654), 1260–1262.
- Werblin, F. (1971). Adaptation in a vertebrate retina: intracellular recording in necturus. *Journal of Neurophysiology*, 34, 228–241.

- Werblin, F., & Dowling, J. (1969). Organization of the retina of the mud puppy, *necturnus maculosus*. ii. intracellular recording. *Journal of Neurophysiology*, 32, 339–355.
- Witkin, A. (1981). Recovering surface shape and orientation from texture. *Journal of Artificial Intelligence*, 17, 17–45.
- Zaidi, Q., & Li, A. (2000). Neural model of shape from texture: developable surfaces. *Investigative Ophthalmology and Visual Science (Supplement)*, 41, S219.
- Zaidi, Q., & Li, A. (2002). Limitations on shape information provided by texture cues. *Vision Research*, 42, 815–835.
- Zeki, S. (1983a). Color coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colors. *Neuroscience*, 9, 741–766.
- Zeki, S. (1983b). Color coding in the cerebral cortex: The responses of wavelength-selective and colour coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience*, 9, 767–791.