Available online at www.sciencedirect.com

ScienceDirect

Vision Research 47 (2007) 2585–2602

Vision Research

www.elsevier.com/locate/visres

Solving da Vinci stereopsis with depth-edge-selective V2 cells

Andrew Assee *, Ning Qian *

Center for Neurobiology and Behavior and Department of Physiology and Cellular Biophysics, Columbia University, 1051 Riverside Drive, Box 87, Kolb Research Annex, Room 519, New York, NY 10032, USA

Received 17 November 2006; received in revised form 22 June 2007

Abstract

We propose a new model for da Vinci stereopsis based on a coarse-to-fine disparity energy computation in V1 and disparity-boundary-selective units in V2. Unlike previous work, our model contains only binocular cells, relies on distributed representations of disparity, and has a simple V1-to-V2 feedforward structure. We demonstrate with random-dot stereograms that the V2 stage of our model is able to determine the location and the eye-of-origin of monocularly occluded regions, and improve disparity map computation. We also examine a few related issues. First, we argue that since monocular regions are binocularly defined, they cannot generally be detected by monocular cells. Second, we show that our coarse-to-fine V1 model for conventional stereopsis explains double matching in Panum's limiting case. This provides computational support to the notion that the perceived depth of a monocular bar next to a binocular rectangle may not be da Vinci stereopsis per se [Gillam, B., Cook, M., & Blackburn, S. (2003). Monocular discs in the occlusion zones of binocular surfaces do not have quantitative depth—a comparison with Panum's limiting case. *Perception 32*, 1009–1019.]. Third, we demonstrate that some stimuli previously deemed invalid have simple, valid geometric interpretations. Our work suggests that studies of da Vinci stereopsis should focus on stimuli more general than the bar-and-rectangle type and that disparity-boundary-selective V2 cells may provide a simple physiological mechanism for da Vinci stereopsis.
Published by Elsevier Ltd.

Keywords: Binocular disparity; Ocularity; Half-occlusion; Stereovision; Disparity step

1. Introduction

When we view a scene with opaque objects at various depths, the lateral separation of our eyes creates not only binocular disparity but also monocularly occluded regions. These are the regions seen by one eye but not the other, and they arise frequently under normal viewing conditions since near surfaces often occlude far surfaces to different extents in the two eyes. Fig. 1a shows a well-known example. The left-eye-only monocular region, right-eye-only monocular region, and binocular region are represented by black, white, and gray squares, respectively. Lines drawn from each eye indicate the extent of occlusion by the near surface of the far background. Assuming that the near surface is the fixation plane with zero disparity, Fig. 1b shows the alignment of the regions visible to the two eyes. Correspondence between the binocular regions is indicated by dotted lines linking the two eyes' views. In contrast, the monocular regions have no correspondence.

While it is well-known that disparity between binocular regions is responsible for conventional stereopsis, many experiments also show that monocular occlusion contributes significantly to depth perception, among other things (Gillam & Nakayama, 1999; Liu, Stevenson, & Schor, 1997; Nakayama & Shimojo, 1990; Shimojo & Nakayama, 1990, 1994). The perceptual effects of monocular regions like those in Fig. 1 have been termed da Vinci stereopsis by Nakayama and Shimojo (1990). Based on the occlusion geometry of opaque objects as in Fig. 1, Nakayama and Shimojo (1990) classified monocular regions into valid and invalid types. Valid monocular regions are (1) a left-eye-only region to the left of a near surface (Fig. 2a), or

[*] Corresponding authors.
E-mail addresses: ada2007@columbia.edu (A. Assee), nq6@columbia.edu (N. Qian).

0042-6989/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.visres.2007.07.003

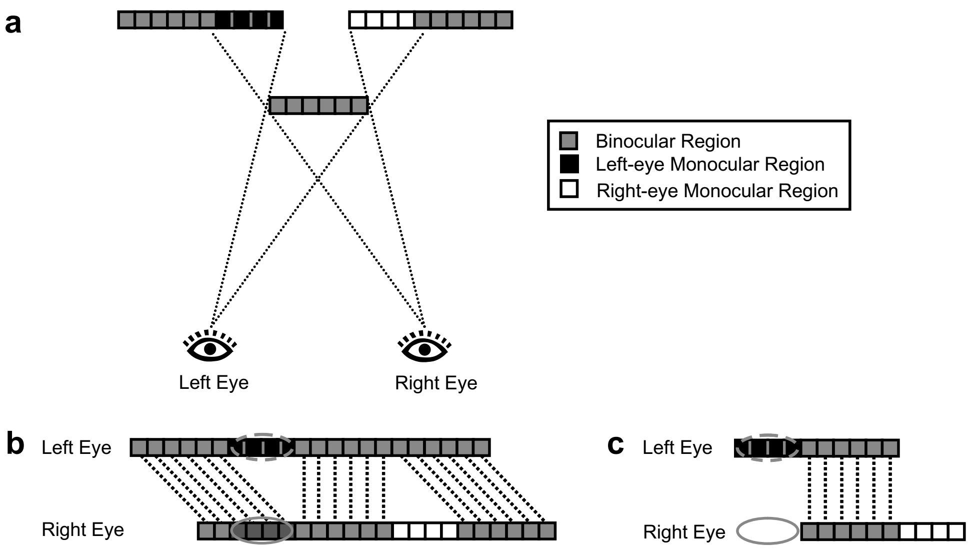

Fig. 1. Occlusion geometry and the role of monocular vs. binocular cells in solving da Vinci stereopsis. (a) Schematic diagram of a scene where a near surface occludes a background. The dotted lines indicate the extent to which the near surface occludes the far surface from each eye. (b) Images seen by the left and right eyes for the scene in (a) when fixation is at the near surface. (c) A special case of (b) when the binocular background is assumed to be featureless. For all panels, gray squares indicate binocular regions, and black and white squares represent left- and right-eye-only monocular regions, respectively. In (b and c), the dotted lines indicate correspondence between two eyes' images. Ovals indicate the RFs of monocular cells, with dashed and solid lines representing left- and right-eye-only RFs, respectively. The vertical dimension of RFs is not shown to scale. Only for the special case in (c) can the relative responses of the left- and right-eye-only monocular cells determine the monocular regions.

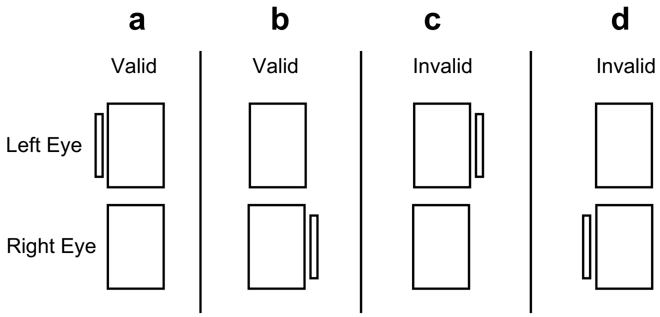

Fig. 2. Stimuli used by Nakayama and Shimojo (1990). Each stimulus is composed of a binocular rectangle and monocular bar. Nakayama and Shimojo (1990) classified the four possible configurations into two valid (a and b) and two invalid (c and d) cases. Redrawn from Nakayama and Shimojo (1990).

(2) a right-eye-only region to the right of a near surface (Fig. 2b). The two remaining cases (Fig. 2c and d) are called invalid.

Models for conventional stereopsis, including the physiologically based disparity energy model (Ohzawa, DeAngelis, & Freeman, 1990; Qian, 1994), measure disparities of binocular regions. Consequently, these models are not directly applicable to da Vinci stereopsis which relies on monocular regions whose disparity is not defined. Numerous models for da Vinci stereopsis have also been proposed (Cao & Grossberg, 2005; Hayashi, Maeda, Shimojo, & Tachi, 2004; Watanabe & Fukushima, 1999; Zitnick & Kanade, 2000). Most of them (e.g., Watanabe & Fukushima, 1999; Zitnick & Kanade, 2000) are based on the Marr and Poggio model (1976) that assigns one unit for each potential match between a specific feature (such as a dot or an edge) in the left image and a specific feature in the right image. To deal with da Vinci stereopsis, additional units responding to monocular features are then added. Finally, geometric constraints are introduced as excitatory and inhibitory connections among the units to do the main computation. Although these and related models are important for machine vision applications, they are non-physiological because they do not use anything like receptive fields (RFs) of visual cells. Instead, they rely on binary units, each responding to nothing but one specific binocular match or monocular feature. As we argued previously (Qian, 1997), such a binary representation is inconsistent with physiology. Real visual cells are tuned to a range of disparities or monocular features (distributed representation) based on the cells' RF structures. Even the most sharply tuned V1 cells have a disparity-tuning width of around 0.2°, and the majority of cells have much broader tuning (Ohzawa et al., 1990; Poggio, Gonzalez, & Krause, 1988) (see Qian, 1997, for a more detailed discussion).

One might think that these non-physiological models could be readily made physiological by adding realistic RFs at the front end to generate a distributed representation. This appears to be what Hayashi et al. (2004) attempted to do to Watanabe and Fukushima's da Vinci stereopsis model (Watanabe & Fukushima, 1999). However, although their work provides important new insights into binocular rivalry, Hayashi et al. did not really make the model physiological. After applying RFs to stimuli, they immediately converted the resulting distributed representation into a binary representation. The main computation was done with the non-physiological binary units just as in Watanabe and Fukushima's original model. Our own experience (Qian & Sejnowski, 1989) with the Marr–Poggio 1976 model suggests that the difficulty with making this class of models physiological is that once the binary representation is replaced by a realistic distributed representation of disparity, there is no known method for implementing the required constraints on the model units. For example, most of these models use either a strong or weak form of the uniqueness constraint, namely that only one depth is allowed along each line of sight. With the binary representation, this constraint can be readily implemented as inhibitory connections among the units representing different depths along each line of sight. However, with a distributed representation, a given unit will respond to many depths. This will generate widespread, unintended inhibition and jeopardize the computation. Thus, the style of computation in the brain must be fundamentally different from that in these models.

An earlier model that also starts with binocular RFs has been proposed by Grossberg et al. (Cao & Grossberg, 2005; McLoughlin & Grossberg, 1998). This model makes a large number of assumptions and explains an impressive list of phenomena, including da Vinci stereopsis. The work is particularly interesting in suggesting potential functions for different cell groups and cortical areas. However, like the studies discussed above, Grossberg et al. used a binary depth representation to do the core computation; they converted the distributed representation from the RFs into a binary representation by sampling responses at five depth planes that are far apart from each other, and ignoring responses between the planes. Consequently, each unit only responds to a single depth.

Another problem with many existing models for da Vinci stereopsis is that they use monocular cells to detect monocular regions. This may sound reasonable, superficially, as one could argue that if at a given location, the left-eye-only monocular cells respond well while the right-eye-only monocular cells do not, then the location must belong to a left-eye-only monocular region. Unfortunately, this is only true for the special case shown in Fig. 1c where the monocular region is defined by a textured area in one eye (black squares) and a corresponding featureless area in the other eye. (The dashed and solid ovals represent the RFs of the left-eye-only monocular cells and RFs of the right-eye-only monocular cells, respectively.) This special case happens only if the far surface is completely featureless except for a narrow region whose relationship to the near surface is just right to be monocularly occluded. A more general situation is shown in Fig. 1b where both the near and far surfaces are textured. The monocular zone in one image does not align with a blank region in the other image; instead, it aligns with a binocular region which is also textured (gray squares). Monocular cells cannot detect this monocular region because the relative response levels of the left-eye-only and right-eye-only monocular cells depend on the textures and contrasts of the surface patches in the RFs, and will have little to do with whether the region is monocular.

It has been suggested that information indicating through which eye the monocular region is seen, known as eye-of-origin or ocularity, is crucial for solving da Vinci stereopsis (Shimojo & Nakayama, 1990). However, this does not necessarily imply that monocular cells have to be used to preserve the eye-of-origin information. In a sense, the eye-of-origin information is also needed in conventional, disparity-based stereopsis, for otherwise we would not be able to tell positive and negative disparities apart. As many stereo algorithms have shown (see, e.g., Qian, 1994) disparity sign, and thus the eye-of-origin information, in conventional stereopsis can be readily extracted by a proper set of binocular cells. Since monocular regions are a binocularly defined property—they are the regions seen by one eye AND not by the other eye—binocular cells may provide a better and more general mechanism to detect them.

In addition to finding the location and ocularity of monocular regions, a related issue in da Vinci stereopsis is what rule determines the depth (or the equivalent disparity) of a monocular region. In the general case where a binocular background is not featureless (Fig. 1b), the answer is straightforward: psychophysical observations suggest that a monocular region simply takes the disparity of the adjacent background surface (Julesz, 1971; Shimojo & Nakayama, 1990). However, in the special case of a featureless binocular background (Fig. 1c), the disparity of the background cannot be measured, so the simple rule above does not apply. Nakayama and Shimojo (1990) used stimuli belonging to this special case (Fig. 2) and demonstrated that perceived depth of a valid monocular bar depends on the distance of the bar from the edge of a binocular rectangle. The finding appears to suggest the rules for assigning depth to monocular regions in the general case and in the special case are different. However, subsequent studies suggest that double matching, as in Panum's limiting case, can explain the distance dependence (Gillam, Cook, & Blackburn, 2003). This raises the possibility that the special case of a featureless binocular background may be treated as conventional stereopsis instead of da Vinci stereopsis, and that models for da Vinci stereopsis should focus on the general case where the depth assignment rule is simple.

In this paper, we first show that the disparity energy model for conventional stereopsis can indeed explain double matching, and thus the distance dependence in the special configuration of Nakayama and Shimojo (1990). We also demonstrate that the distinction between valid and invalid monocular regions based on the geometry of projection is harder to make than previously thought, particularly for the special case with featureless binocular background. Therefore, models for da Vinci stereopsis should focus on the general case. We then propose a physiologically plausible model for da Vinci stereopsis by extending the V1 disparity energy model to include disparity-edge-selective cells, as found in V2 (von der Heydt, Zhou, & Friedman, 2000). We constructed the V2 cells by selectively combining V1 cells via feedforward connections. No binary disparity representation or monocular cells are required in our model. We demonstrate that at the V2 stage, our model not only determines the location and ocularity of monocular regions but also improves the accuracy of disparity maps computed in V1. Preliminary results have been reported in abstract form (Assee & Qian, 2006).

2. Methods

Our model consists of a V1 stage and a V2 stage. At the V1 stage, we used the disparity energy model (Ohzawa et al., 1990; Qian, 1994) with a coarse-to-fine process incorporated (Chen & Qian, 2004; Menz & Freeman, 2003). The importance of the coarse-to-fine process in the current study will be explained in Section 3. As we have shown previously (Chen & Qian, 2004; Qian, 1994; Qian & Zhu, 1997), the V1 stage can effectively compute disparities of binocular regions of stereograms. At the V2 stage, we combined responses from V1 cells via feedforward connections to form disparity-edge-detectors. We then used V2 population responses to determine the location, ocularity, and equivalent disparity of monocular regions in a stereogram, and to refine disparity map estimation.

2.1. V1 stage

The V1 disparity energy model uses the standard quadrature pair method to simulate complex cell responses from simple cell responses; the simple cell responses, in turn, are determined by filtering the stimulus through binocular RFs (Ohzawa et al., 1990; Qian, 1994). The details of our analysis and implementation of the model, the inclusion of a coarse-to-fine process, and discussions on biological relevance, have been published elsewhere (Chen & Qian, 2004; Qian, 1994; Qian & Zhu, 1997). Here we present a brief summary and some implementation specifics. Coarse-to-fine computation refers to the procedure by which an image is first processed at a coarse scale, i.e., by relatively large RFs, and the result is then used to guide more refined processing at a finer scale. To see the rationale, note that cells with large RFs can cover a large range of stimulus disparity, but the estimated disparity and its spatial location are inaccurate. Conversely, small RFs can give accurate results but only cover a narrow range of disparity. This dilemma can be alleviated by a coarse-to-fine procedure: first use large RFs to estimate disparity crudely and then apply the estimate to offset stimulus disparity so that the residual disparity after the offset is small and can be determined precisely by small RFs (Marr & Poggio, 1979). Recent single-unit recordings in V1 by Menz and Freeman (2003) provide physiological evidence for coarse-to-fine processing in stereovision. Their data suggest that a significant portion of V1 cells reduce their spatial scales over time. For ease of implementation, we used separate sets of cells for different scales.

Our coarse-to-fine algorithm combines the two well-known RF models for V1 binocular cells: the position-shift and phase-shift models (see Qian, 1997, for a review). At each scale, disparity is always estimated by the phase-shift RF mechanism to take advantage of its higher reliability, and a position-shift component equal to the estimated disparity in the previous, coarser scale is used to offset the stimulus disparity (Chen & Qian, 2004). At the coarsest scale where computation begins, the position-shift component is set to 0. In our original implementation, we used two-dimensional gabor RFs for simple cells, and pooled across orientation and a local spatial area at each scale before disparity estimation (Chen & Qian, 2004). In the current study, we used one-dimensional gabor RFs to speed up simulations, and thus did not pool across orientation.

Mathematically, the simple cell response is given by:

$$R_s = \int_{-\infty}^{\infty} [f_L(x)I_L(x) + f_R(x)I_R(x)]\,dx \tag{1}$$

where

$$f_L(x) = \exp\left(\frac{-(x - x_0 + d/2)^2}{2\sigma^2}\right)\cos\left(\omega(x - x_0 + d/2) + \varphi_L\right) \tag{2a}$$

$$f_R(x) = \exp\left(\frac{-(x - x_0 - d/2)^2}{2\sigma^2}\right)\cos\left(\omega(x - x_0 - d/2) + \varphi_R\right) \tag{2b}$$

represent the gabor RF for the left and right eyes, respectively, and $I_L(x)$ and $I_R(x)$ are the left and right images of a stereo pair. For Eqs. (2a) and (2b), σ and ω are the guassian width and spatial frequency of the gabor RF, x_0 is the RF center location, φ_L and φ_R are the phase parameters for the left and right RFs, respectively, and d is the position-shift component of the RFs divided equally between the left and right RFs. Note that the position-shift component was never used to estimate disparity; instead, it is used to offset disparity computed by the phase-shift component (Chen & Qian, 2004). The set of scales we used followed the geometric sequence $\sigma = 8, 4\sqrt{2}, 4, 2\sqrt{2}$, and 2 pixels, and we set $\omega = \pi/\sigma$ at each scale so that the bandwidth was constant across scales. The phase difference between the left and right RFs, represented by $\Delta\varphi = \varphi_R - \varphi_L$, was sampled uniformly from $[-\pi, \pi)$ with $\pi/4$ increments. A fourth-order polynomial was used to interpolate across sampled responses.

The complex cell response equals the squared sum of a quadrature pair of simple cells:

$$R_c = R_{s1}^2 + R_{s2}^2 \tag{3}$$

where R_{s1} and R_{s2} represent each simple cell's response with σ, ω, and $\Delta\varphi$ the same for the pair. The quadrature relationship requires $(\varphi_L + \varphi_R)/2$ between the two simple cells to have a phase difference of $\pi/2$. According to our previous analysis of the disparity energy model (Qian, 1994), at each scale, we can estimate the stimulus disparity at each location by using an appropriate set of complex cells centered at that location. The cells all have the same parameters except that their left–right RF phase difference ($\Delta\varphi$) covers the full 2π range. For broad-band stimuli such as lines and dots, the stimulus disparity at a given location can be estimated according to (Qian, 1994):

$$D \approx \Delta\varphi^*/\omega \tag{4}$$

where $\Delta\varphi^*$ is the phase difference of the most responsive cell in the population.

2.2. Double matching in panum's limiting case

Before presenting our V2 model for da Vinci stereopsis, we describe how we applied the V1 disparity energy model to explain double matching in Panum's limiting case and thus the perceived depth in the bar-and-rectangle stimuli used by Nakayama and Shimojo (1990). We considered a stimulus configuration with two lines in the left eye and one line in the right eye. Mathematically, the left and right eye's images can be represented by delta functions as follows:

$$I_L(x) = \delta(x - x_1) + \delta(x - x_2) \tag{5a}$$

$$I_R(x) = \delta(x - x_1) \tag{5b}$$

where x_1 and x_2 are the positions of the two lines. Line 1 is binocular with 0 disparity and line 2 is seen by the left eye only. This stimulus is identical to the one used by Gillam et al. (2003) to demonstrate double matching in Panum's limiting case. It also captures the essence of the stimulus used by Nakayama and Shimojo (1990) (Fig. 2). The right edge of the large rectangle in the right eye of Fig. 2a is unlikely to double match with the monocular bar in the left eye because of the large lateral separation. The top and bottom edges are also unlikely to double match with the monocular bar because of the orthogonal orientations. Therefore, only the left edge in the right eye could potentially double match with the monocular bar of the left eye.

We applied both the original disparity energy algorithm with a single scale (Qian, 1994) and the multi-scale, coarse-to-fine energy algorithm (Chen & Qian, 2004) to this problem. We obtained results with the single-scale algorithm analytically. We considered binocular simple cells whose left and right RFs are described by Gabor functions as in Eqs.

(2a) and (2b) (with position-shift $d = 0$ as there is no other scale present to determine the value of d.). Using Eqs. (1) and (3), and the properties of the delta function, we can derive the simple cell response as:

$$\begin{aligned} R_s &= \exp\left(-(x_1-x_0)^2/2\sigma^2\right)\cos\left(\omega(x_1-x_0)+\varphi_L\right) \\ &\quad + \exp\left(-(x_2-x_0)^2/2\sigma^2\right)\cos\left(\omega(x_2-x_0)+\varphi_L\right) \\ &\quad + \exp\left(-(x_1-x_0)^2/2\sigma^2\right)\cos\left(\omega(x_1-x_0)+\varphi_R\right) \end{aligned} \tag{6}$$

and the complex cell response as:

$$\begin{aligned} R_c &= \exp\left(-(x_2-x_0)^2/\sigma^2\right) + 2\exp\left(-(x_1-x_0)^2\Big/\sigma^2\right) \\ &\quad + 2\exp\left(\left(-(x_1-x_0)^2-(x_2-x_0)^2\right)\Big/2\sigma^2\right)[\cos\left(\omega(x_1-x_2)\right) \\ &\quad + \cos\left(\omega(x_1-x_2)+\Delta\varphi\right)] \\ &\quad + 2\exp\left(-(x_1-x_0)^2\Big/\sigma^2\right)\cos(\Delta\varphi). \end{aligned} \tag{7}$$

Eq. (7) determines the population response for each stimulus location x_0. We then found the phase difference ($\Delta\varphi^*$) of the most responsive cell at each location by setting

$\partial R_c/\partial\Delta\varphi = 0$, and estimated the stimulus disparity according to Eq. (4). The solution is given by:

$$\Delta\varphi^* = \arctan(A/B) \tag{8}$$

where

$$A = \exp\left(\left(-(x_2-x_0)^2\right)/2\sigma^2\right)\sin\left(\omega(x_2-x_1)\right),$$

and

$$B = \exp\left(\left(-(x_2-x_0)^2\right)/2\sigma^2\right)\cos\left(\omega(x_2-x_1)\right) + \exp\left(-(x_1-x_0)^2/2\sigma^2\right).$$

In psychophysical experiments, subjects usually report the perceived depth of line 2 with respect to line 1. We therefore determined disparities for line 1 and line 2 separately by setting x_0 equal to x_1 and x_2 separately, and the difference between the two disparities are considered as the perceived disparity of line 2. By repeating this procedure across a range of lateral separations between the lines ($x_2 - x_1$), we obtained plots in Fig. 5a for comparison with psychophysics (see Section 3). We let σ equal to 2, 4, 6, and 16 min separately, and set $\omega = \pi/\sigma$ to keep the frequency bandwidth of RFs constant.

For applying our multi-scale, coarse-to-fine algorithm to this problem, we similarly computed disparities of line 1 and line 2 separately, and plotted the difference as a function of ($x_2 - x_1$) in Fig. 5b for comparison with psychophysics.

2.3. V2 stage

von der Heydt et al. (2000) found that approximately 20% of recorded cells in V2 responded preferentially to only one edge of a figure occluding a background. For example, for the scene depicted in Fig. 1a, some of these disparity-boundary-selective cells only respond to the depth discontinuity at the left edge of the front surface, while others respond only to the depth discontinuity at the right edge. These cells have been used to model stereo segmentation (Zhaoping, 2002) but not da Vinci stereopsis. Since monocular regions are always associated with depth discontinuity, we modeled the V2 disparity-boundary-selective cells via feedforward connections from V1, and examined whether V2 population responses could determine the location and ocularity of monocular regions. A V2 cell receives inputs from multiple (4 in our simulations except when noted otherwise) V1 binocular complex cells whose RFs are centered at different but nearby horizontal locations (Fig. 3a). Therefore, the RFs of V2 cells are larger than those of corresponding V1 cells (Burkhalter & Van Essen, 1986; Gattass, Gross, & Sandell, 1981; Smith, Singh, Williams, & Greenlee, 2001; Zeki, 1978). To generate disparity-boundary-selectivity observed in V2, we let the left and right halves of the V1 input cells have various combinations of preferred disparities. (Note that here "left" and "right" do not refer to eyes, but to V1 RF centers relative to the V2 RF center.)

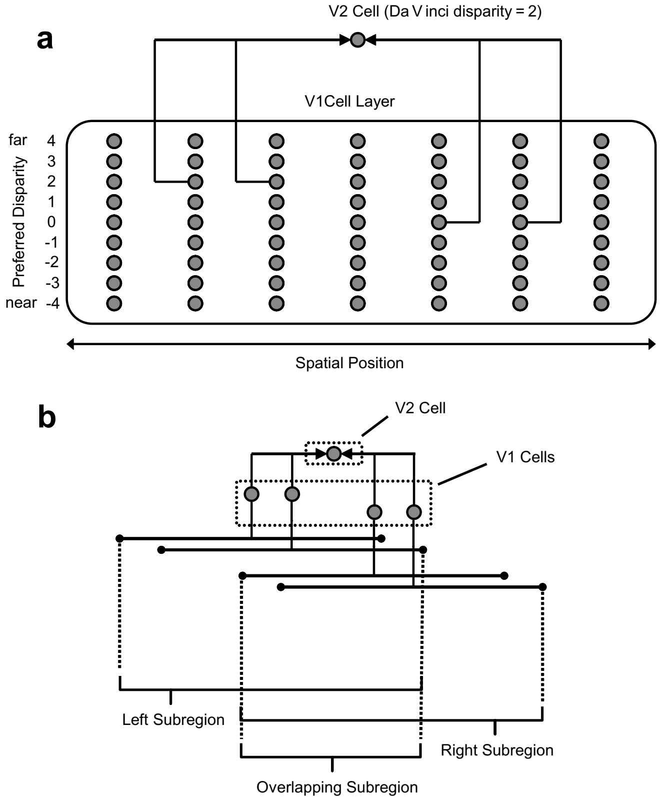

Fig. 3. Schematic representation of the V1–V2 circuitry in our model. (a) An example V2 cell that receives inputs from two V1 cells with a preferred disparity of 2 pixels to the left and two V1 cells with a preferred disparity of 0 pixels to the right. The farther of the two preferred disparities, 2 pixels in this example, is termed the cell's da Vinci disparity. (b) The tripartite receptive field of the model V2 cell in (a).

Fig. 3a shows an example of a V2 cell that receives inputs from two V1 cells tuned to a 2-pixel disparity on the left and from two other V1 cells tuned to 0 disparity on the right. The response of a V2 cell is determined by summing the normalized responses of all its V1 input cells. The normalization is done separately for each set of V1 cells with the same RF location; it ensures that contributions from different locations to a V2 cell are equally weighted. We also normalized each V2 cell's response by the number of V1 cells from which it receives inputs, which is four for our simulations except when noted otherwise. Although this V1-to-V2 connectivity could be applied to each scale used in our V1 model, we only implemented it for the finest V1 scale as the estimated disparity at the finest scale is the most accurate. Due to the location and extent of the V1 RFs, each V2 cell has a tripartite RF structure (Fig. 3b). For example, the V2 cell in Fig. 3a has a preferred disparity of 2 pixels on the left side of its RF, 0 on the right side, and a combined preference in the center.

By generating a set of V2 cells at each image location where all combinations of preferred disparities between the left side and right side V1 inputs are represented, including equal preferred disparities, we can then use the V2 population response to determine whether there is a depth boundary, and thus a monocular region, at that location. For our simulations, the preferred disparities for each side of V2 RFs ranged from −8 to 8 pixels by 1 pixel increments and were used to produce a V2 population which represents 81 disparity combinations. To examine the V2 population response, we used a two-dimensional plot with the left and right preferred disparities along the abscissa and ordinate, respectively (Fig. 4). We interpolated the responses with a two-dimensional surface using Matlab function interp2. By locating the maximally responding cell relative to the diagonal line in the plot, we can categorize the underlying image loca-

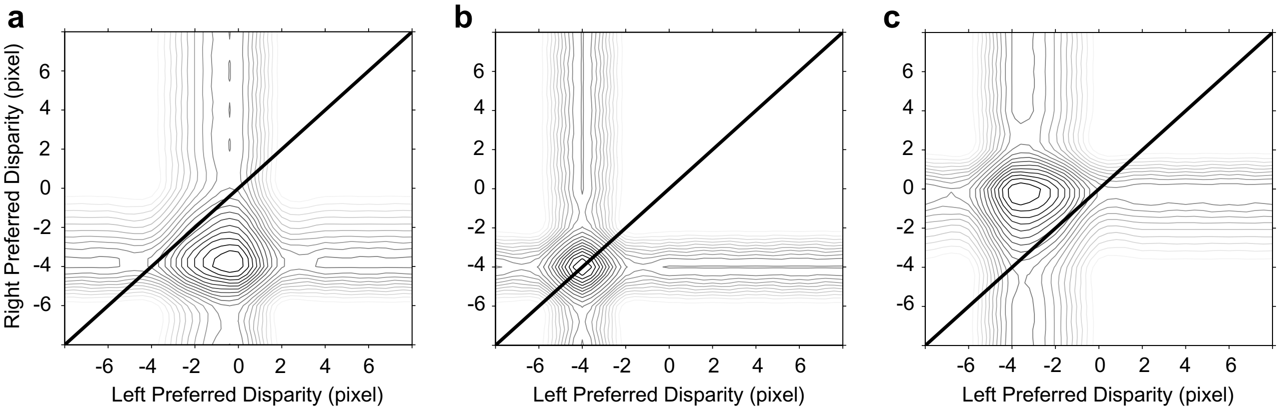

Fig. 4. Three types of population responses of our model V2 cells. Each plot shows the responses of all V2 cells for a given image location. The peak response is below the diagonal line (a), on the diagonal line (b), or above the diagonal line (c). These three types of population responses indicate a left-eye-only monocular region, a binocular region, and a right-eye-only monocular region, respectively (see text for a detailed explanation).

tion into one of three groups. (1) If the maximally responding cell appears below the diagonal (Fig. 4a), the stimulus on the right side of the V2 cells' RFs must have a nearer depth than the left side. Based on the geometry of Fig. 1a, the image location should belong to a left-eye-only monocular region. (2) If the maximally responding cell appears on the diagonal (Fig. 4b), two sides of the V2 RFs must have the same depth and there is no occlusion. The image location must belong to a binocular region. (3) If the maximally responding cell appears above the diagonal (Fig. 4c), the near occluding surface must be on the left side of the V2 cells' RFs, and the image location should belong to a right-eye-only monocular region.

The above discussion suggests a simple method for determining the ocularity of an image location by finding the most responsive cell from the V2 population centered at that location. The method is to compute the difference between the two preferred disparities (D_L and D_R) of the most responsive cell according to:

$$D_{diff} = D_R - D_L \tag{9}$$

The deviation of D_{diff} from zero indicates how likely the image location is monocular, while the sign of D_{diff} indicates whether the location is left-eye only or eight-eye only. Alternatively, one could compute the perpendicular distance from the peak of the population response to the diagonal line, with the sign of the distance determined by whether the peak is above or below the diagonal line. Since the signed distance differs from D_{diff} by a constant factor of $\sqrt{2}$, the two methods are mathematically equivalent.

After determining the location and ocularity of monocular regions, we can assign an equivalent disparity to the monocular regions according to the rule where monocular regions take the depth of the far surface (Julesz, 1971; Shimojo & Nakayama, 1990). This is particularly easy to implement in our model because we only need to assume that each V2 cell always signals the farther of the two preferred disparities within its RF. By doing so, we can generate a more accurate depth map in V2 since the depth across the monocular regions can now be computed. We call the farther of the two preferred disparities of a V2 cell its da Vinci disparity (Fig. 3a).

Unless indicated otherwise, we present representative simulation results using random-dot stereograms of the size 100 (width) × 20 (height) pixels, and dot density 50%. The middle one third of the stereograms has either a near disparity of −4 pixels, or a far disparity of 4 pixels, across the full height of the image. The remaining two thirds of the stereogram have 0 disparity.

To verify that our model's performance was not constrained to the above parameters, we varied the number of V1 input cells to a V2 cell from 2 to 12 (Fig. 10). In addition, we computed disparity from random-dot stereograms with a 33%, 50%, or 66% dot density each with disparities ranging from −6 to 6 pixels taken at 2 pixel increments (results not shown). We found that our model performs reliably across a broad range of parameters.

3. Results

Before presenting results from our V2 model for da Vinci stereopsis, we will first consider double matching in Panum's limiting case and the distance dependence of perceived depth in the special stimulus configuration used by Nakayama and Shimojo (1990) with featureless binocular background (Fig. 2), and evaluate the proposed distinction between valid and invalid monocular regions. This investigation is important because it addresses the issue of whether the results in Nakayama and Shimojo (1990) is da Vinci stereopsis or conventional disparity-based stereopsis, and allows us to focus our da Vinci stereopsis model on the general case of monocular occlusion.

3.1. Double matching in panum's limiting case explained by the V1 disparity energy model

As we mentioned in Section 1, Nakayama and Shimojo (1990) used a stimulus configuration where the far, background surface is featureless except for a bar which is monocularly occluded by a near rectangle (Fig. 2). They found that the perceived depth of the bar increases with the lateral distance between the bar and the rectangle. Although the finding was initially considered as evidence for da Vinci stereopsis, Gillam et al. (2003) later suggested that the results may be explained by double matching as in Panum's limiting case. In other words, the left edge of the right eye's rectangle in Fig. 2a could be matched to the left edge of the monocular bar as well as to the corresponding edge of the rectangle in the left eye.

Can the distance dependence be explained by double matching in a model for conventional stereopsis without evoking mechanisms of da Vinci stereopsis? We used the V1 disparity energy model to examine this issue since it is the most physiologically plausible model for conventional stereopsis. We considered a stimulus configuration with two lines in the left eye and one line in the right eye (Eq. (5) in Section 2). One line in the left eye is located at the same position as the line in the right eye and thus has 0 disparity. The other line in the left eye is monocular but if it double matches with the line in the right eye, it will have a disparity equal to the lateral separation between the lines. This stimulus is identical to the one used by Gillam et al. (2003) to demonstrate double matching in Panum's limiting case.

We applied both the original disparity energy algorithm with a single scale (Qian, 1994) and the multi-scale, coarse-to-fine energy algorithm (Chen & Qian, 2004) to this problem as detailed in Section 2. The results from the two algorithms, in the form of the relative disparity between the lines as a function of the lateral separation between the lines, are shown in panels b and c of Fig. 5, respectively. For comparison, the psychophysical data of one subject taken from Gillam et al. (2003) is shown in panel a.

For the single-scale model, we derived the result analytically (see Section 2). In Fig. 5b, we plot the analytical result for four scales with RF Gaussian width $\sigma = 2$, 4, 8, and 16 arcmin separately. For clarity, we only show a portion of each curve well within the disparity range covered by the scale; with further extension, the curve will jump in the opposite direction due to disparity wrap around in the energy model (Qian, 1994). The main finding is that the single-scale model shows qualitatively the same trend as the data, but the magnitude of the computed disparity is much smaller than the observation. This can be understood analytically by doing a Taylor expansion of Eq. (8) for small line separation. It can be shown that when relatively disparity between the lines is computed, the first-order term of expansion cancels, and since there is no second-order term, the main contribution comes from a very small third-order term.

In contrast, results from our coarse-to-fine model in Fig. 5c agree with experimental data quantitatively. The computed relative disparity between the lines roughly equals the lateral line separation as observed (Gillam et al., 2003). At the largest line separation in Fig. 5c, the computed relative disparity starts to turn toward zero. This is also consistent with experimental data (see Fig. 7 in Gil-

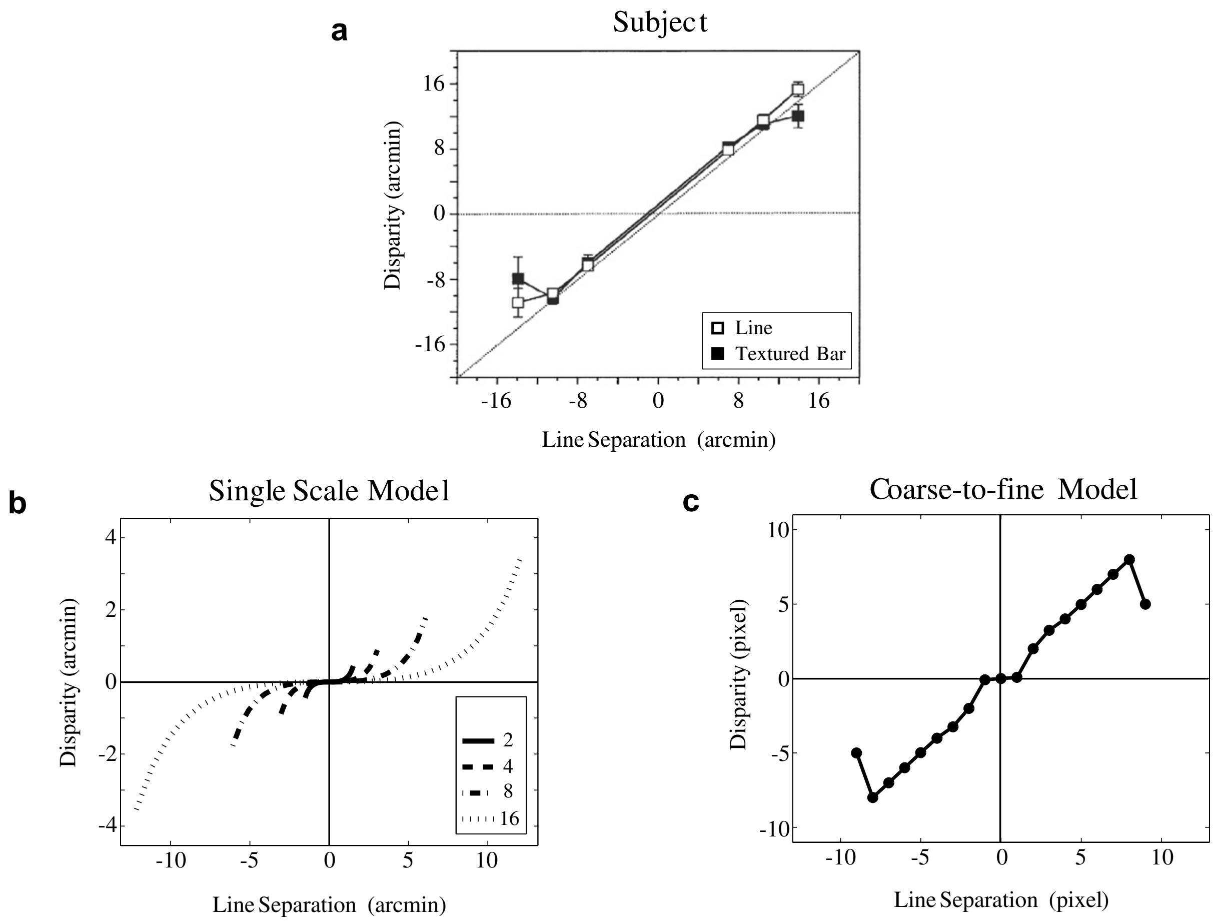

Fig. 5. Modeling double matching in Panum's limiting case. (a) Psychophysical data of one subject reprinted from Gillam et al. (2003) with permission from Pion Limited, London. The perceived disparity of the monocular line is plotted as a function of the separation between the monocular and binocular lines. (b) Analytical results for the single-scale disparity energy model plotted for four scales with $\sigma = 2$, 4, 8, and 16 min. (c) Simulation results from the multi-scale, coarse-to-fine disparity energy model.

lam et al., 2003). We therefore conclude that double matching in Panum's limiting case can be explained by our coarse-to-fine energy algorithm for conventional stereopsis.

3.2. Valid and invalid monocular regions

The results above lend support, from a modeling perspective, to the suggestion that the distance-dependent depth of the monocular bar in the experiments of Nakayama and Shimojo (1990) is not da Vinci stereopsis but instead is double matching in conventional stereopsis (Gillam et al., 2003). However, Nakayama and Shimojo (1990) distingushed between valid and invalid monocular occlusions (Fig. 2), and they reported that only valid monocular bars show the distance dependence. The double matching interpretation does not distinguish between valid and invalid cases. Interestingly, Hakkinen and Nyman (1996) used a stimulus configuration similar to that of Nakayama and Shimojo (1990), and found no difference between valid and invalid cases; both show the same distance dependence. We attempted to understand this apparent conflict of data by examining whether there are multiple *valid* geometric interpretations for what has been called the "invalid" configuration. For completeness, let us first consider the geometric configuration for what Nakayama and Shimojo (1990) call the "valid" cases in Fig. 2. Fig. 6a depicts the three-dimensional scene. The opaque, featureless regions on background are indicated by the rectangles delineated by dashed lines. When an observer is fixating the near, central surface, the half-images in Fig. 6b are generated, which represent the two valid cases of Fig. 2. Now, when the depth order of the near surface and the featureless regions of the background is reversed, as shown in Fig. 6c, the resulting half-images (Fig. 6d) represent the two "invalid" cases of Fig. 2, assuming that an observer is now fixating the far, central surface. Therefore, there is a valid geometric interpretation of the "invalid" cases. Note that Fig. 6c still follows Nakayama and Shimojo's original rule that a right(left)-eye-only region is to the right(left) of a near surface. However, when the near surface is featureless, the resulting stimuli (Fig. 6e) are identical to what they classified as the "invalid" stimuli (Fig. 2). Featureless occluders are also used to create stimuli with perceived phantom surfaces (Anderson, 1994; Gillam & Nakayama, 1999; Liu et al., 1997).

Nakayama and Shimojo (1990) also considered a valid interpretation for their "invalid" case called "silhouette camouflage geometry." However, the configuration is relatively complex involving transparent surfaces, and one could argue that it is less likely to happen in the real world and thus less likely to be used by the brain (Nakayama & Shimojo, 1990; but see Cook & Gillam, 2004; Ehrenstein & Gillam, 1998). Here we show that there is a simple valid interpretation for the "invalid" cases using only opaque surfaces. Indeed, the configuration in Fig. 6c is probably more likely to occur in the real world than that in Fig. 6a. Fig. 6c occurs whenever the occluding surface is featureless while Fig. 6a occurs when the far surface is featureless except at the locations which happened to be monocularly occluded by a near surface.

Based on the above considerations, we suggest that the lack of depth in Nakayama and Shimojo's (1990) "invalid" configuration is not because it is invalid but because there is a valid, flat interpretation with no depth between the monocular region and the central binocular region (Fig. 6c). Fig. 6d shows another valid interpretation of the same "invalid" case. Here, a depth exists between the monocular region and the central binocular region. This configuration does not contradict double matching or the psychophysical results of Hakkinen and Nyman (1996) who found the same distance dependence for both valid and "invalid" cases. We therefore suggest that Hakkinen

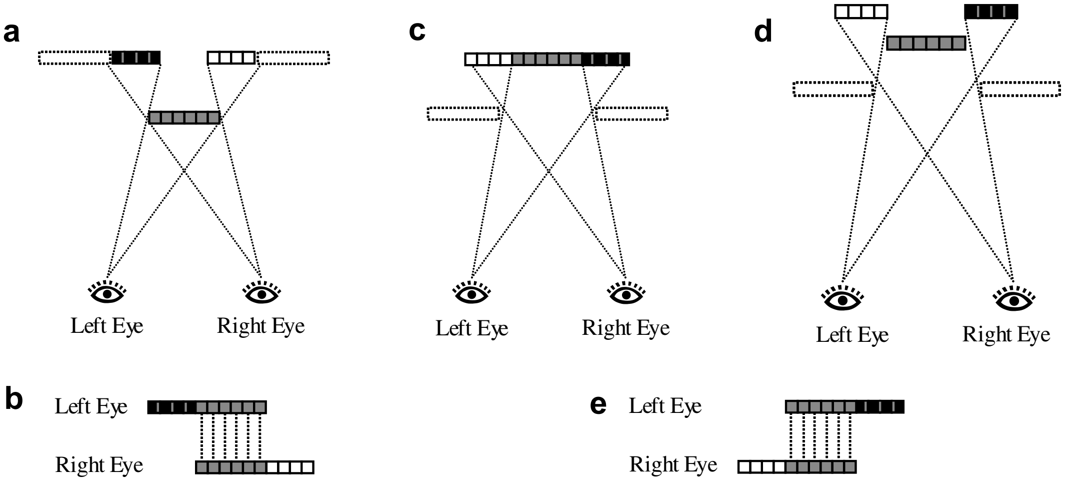

Fig. 6. Valid geometric interpretations for both valid and "invalid" stimuli in Fig. 2. (a) Scene that generates the two valid cases in Fig. 2. (b) The retinal half-images of the scene in (a). (c) Scene that generates the two "invalid" cases in Fig. 2. (d) Another scene that generates the two "invalid" cases in Fig. 2. (e) The retinal half-images of the scenes in (c and d). Rectangles delineated by dotted lines in (a), (c), and (d) represent featureless opaque surfaces. Other figure conventions are the same as in Fig. 1.

and Nyman's subjects and Nakayama and Shimojo's subjects perhaps used different interpretations and thus reported different results. We further speculate that when subjects use an interpretation shown in Fig. 6c, the double matching is overruled by this interpretation so that there is little perceived depth between the monocular and the central binocular regions.

In the above, we considered a special case involving featureless surfaces as used by Nakayama and Shimojo (1990). In the general case where the surfaces are textured, can the "invalid" configuration really be invalid? Although we do not know the general answer that covers every possibility, we examined a stimulus used by Shimojo and Nakayama (1990) in a later study. The authors generated a random-dot stereogram with a central near surface and a surround far surface. Both surfaces are textured with dots and the stereogram has two valid monocular regions as in Fig. 1a and b. These valid regions are labeled as L1 and R2 in Fig. 7b. They then created an "invalid" monocular region in the left eye's image by copying the monocular region on the left side of the front surface (L1 in Fig. 7b) and pasting it to the right side (L3 in Fig. 7b). They found that the "invalid" monocular region L3 generates binocular rivalry instead of depth. Is the rivalry due to the invalid nature of the stimulus? Fig. 7a shows that there is actually a valid interpretation of the stimulus. Note that in the alignment of the two eyes' images in Fig. 7b, the fixation depth is assumed to be at the background plane because in the experiment, subjects were asked to judge the depth of image patches on the background. Also note that when the "invalid" monocular region L3 is created in the left eye, it simultaneously creates a monocular region in the right eye (R3 in Fig. 7b) next to the valid monocular region R2 in the right eye. This happens because L3 overrides the original dot pattern corresponding to R3, thus rendering R3 monocular with no binocular correspondence (Fig. 7b). Therefore, L3 and R3 are at the same spatial location but their dot patterns are unrelated (Fig. 7b). Fig. 7a shows that this situation can arise from looking through a small aperture on the background. In fact, Howard (1995) and Tsai and Victor (2000; Tsai and Victor, 2005) have studied similar configurations and found that in addition to rivalry, they also generate a weak sense of depth perception, termed the sieve effect.

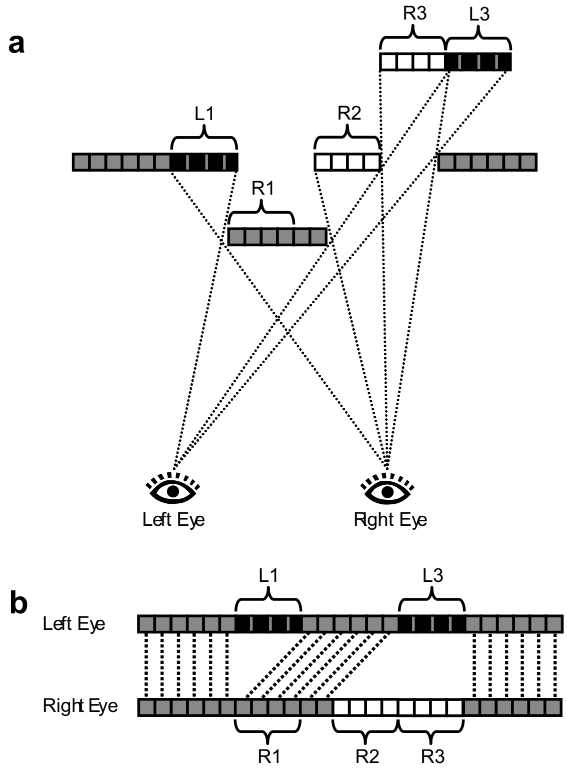

Fig. 7. A valid geometric interpretation for an "invalid" random-dot stimulus in Shimojo and Nakayama (1990). The scene depicted in (a) generates the half-images in (b) used in Shimojo and Nakayama (1990). Figure conventions are the same as in Fig. 1.

Therefore, we conclude that rivalry in Shimojo and Nakayama (1990) random-dot stimulus is not due to the invalid nature of the stimulus as there is a valid interpretation. Instead, it is probably due the fact that the left and right image patches at the same spatial location are unrelated. Rivalry does not happen for what Shimojo and Nakayama (1990) call the valid monocular regions, such as L1 in Fig. 7b, presumably because the aligned region in the other eye, R1, *is* correlated with a patch to the right of L1.

3.3. Solving da Vinci stereopsis with V2 disparity-boundary-selective cells

The above considerations indicate that at least some of the configurations thought to be invalid actually have simple, valid interpretations, and that the special case of monocular occlusion with featureless surface areas used by Nakayama and Shimojo (1990) can be explained by double matching in our coarse-to-fine disparity energy model for conventional stereopsis, and thus may not qualify as da Vinci stereopsis. We therefore focused our da Vinci stereopsis model on the general case of monocular occlusion with textured surfaces. As we mentioned in Section 1 (Fig. 1b), monocular cells are generally not suited for finding the monocular regions which are binocularly defined and are characterized by the lack of correlated regions in the other eye. We showed previously that V1 disparity energy units compute binocular correlation (Qian & Zhu, 1997) and that the computation is much enhanced via a coarse-to-fine algorithm (Chen & Qian, 2004). We therefore first examined whether our coarse-to-fine algorithm could be adopted to locate the monocular regions as areas of weak binocular correlation. We found that the method works to some degree but the results (not shown) are not robust. Since monocular regions always occur at depth discontinuities, it is natural to use V2 disparity-boundary-selective cells to improve V1 computation. We constructed these V2 cells from V1 cells via feedforward connections so that V2 cells have various combinations of preferred

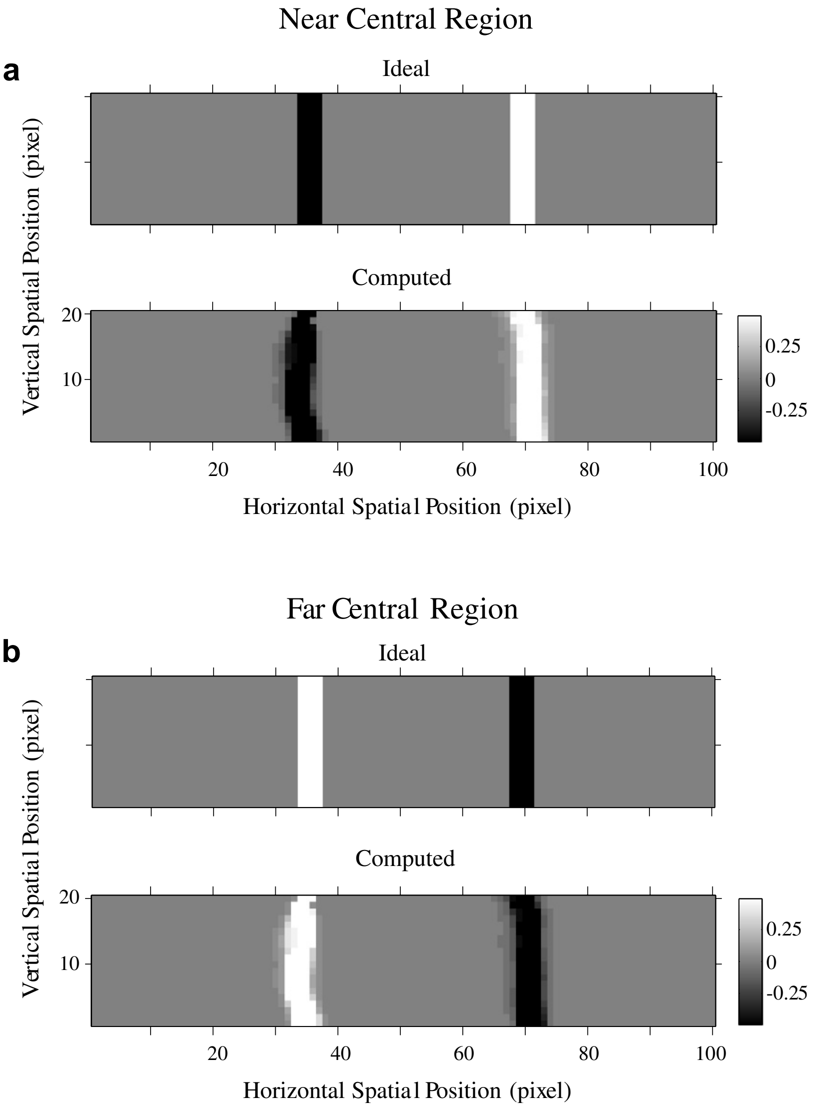

Fig. 8. Ocularity (eye-of-origin) maps computed at the V2 stage of our model for two random-dot stereograms. (a) Simulation results from a stereogram whose middle third had a near disparity of -4 pixels and the remaining two thirds had a 0 disparity. (b) Simulation results from a stereogram whose middle third had a far disparity of 4 pixels and the remaining two thirds had a 0 disparity. In each part, the top panel shows the ideal ocularity map, with gray color representing binocular regions, and black and white colors representing left- and right-eye-only monocular regions, respectively. The bottom panel shows the raw computed ocularity map where a continuous gray scale represents the normalized difference between the two preferred disparities of the most responsive cell at each image location. Gray color represents 0 difference, and black and white colors represent negative and positive differences, respectively. A threshold can be used to convert each raw map shown here into a specific ocularity map for comparison with the corresponding ideal map.

disparities on the left and right portions of their RFs, and we used V2 population responses to determine both the location and ocularity of monocular regions (see Section 2).

An example of our simulations on a random-dot stereogram is shown in Fig. 8a. The middle third of the stereogram has a near disparity of -4 pixels, and the remaining two thirds have 0 disparity. The top row of Fig. 8a shows the ideal ocularity map; the black and white colors represent the left-eye-only and right-eye-only monocular regions, respectively, and gray color represents binocular regions. The width of the monocular regions equals the disparity magnitude of the near surface. The bottom row is a representation of the computed ocularity map. As we explained in Section 2, ocularity of each image location can be determined by finding the most responsive cell of the two-dimensional V2 population response plot for that location (Fig. 4) and calculating the difference between the two preferred disparities of the cell according to Eq. (9). A zero difference means the location is binocular, whereas positive and negative values indicate a right-eye-only and left-eye-only monocular location, respectively. In the bottom row of Fig. 8a, we used a gray scale to represent the disparity difference of the most responsive V2 cell, normalized by the largest difference in the map. The gray color represents 0 difference for binocular locations. A progressively more white (black) color represents a larger positive (negative) difference, indicating that the location is more likely to be left-eye-only (right-eye-only).

Table 1
Accuracy of ocularity map computation with different thresholds

Threshold	Fraction of incorrect pixels for near stimulus	Fraction of incorrect pixels for far stimulus
0	0.0745	0.0745
0.05	0.0525	0.0525
0.1	0.048	0.048
0.4	0.0455	0.0455
0.75	0.054	0.054
0.9	0.0745	0.0745

Visually, the ideal and computed ocularity maps are in good agreement. To make the comparison more quantitative, we used a threshold for the disparity difference to commit continuous values in the computed map to the discrete ocularity representation used in the ideal map. If the absolute value of the disparity difference was below the threshold, the point was considered as binocular; otherwise, it is monocular with ocularity determined by the sign of the difference. The second column of Table 1 shows the fractions of misclassified pixels under various threshold values in the first column. For a threshold value between 0.05 and 0.4, only 5% of the points are incorrectly classified. The error rate remains low for an even wider range of threshold values.

Fig. 8b shows another example of our simulations with the same format. The random-dot stereogram used is iden-

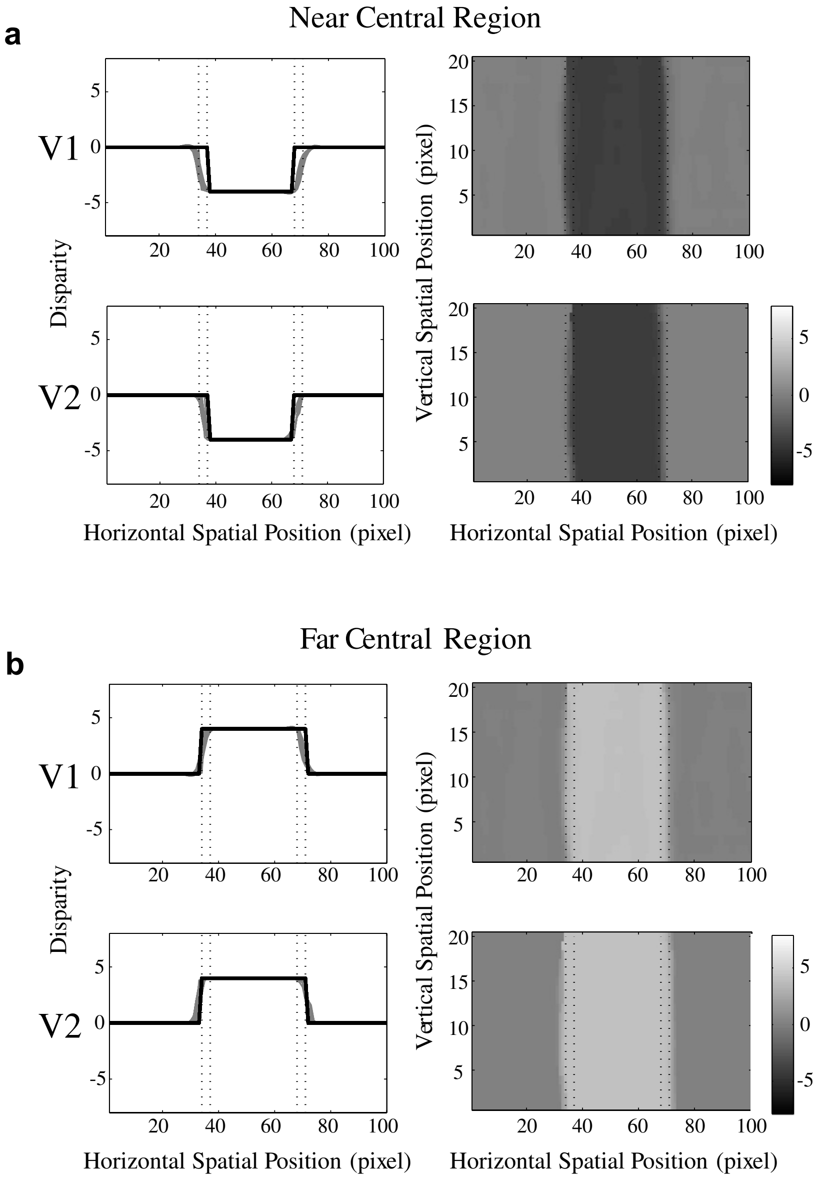

Fig. 9. Disparity maps computed at both the V1 and V2 stages of our model for the same random-dot stereograms used in Fig. 8. (a) Results from the stereogram whose middle third has a near disparity of −4 pixels. (b) Results from the stereogram whose middle third has a far disparity of 4 pixels. In each part, the first and second rows show the V1 and V2 results, respectively. Plots in the first column show disparity as a function of the horizontal image position. The black traces are the ideal disparity maps; the monocular regions (marked by dotted vertical lines) take the disparity values of the background surfaces. The gray traces are the computed disparities. Plots in the second column are gray scale representation of computed disparities as a function of both horizontal and vertical image positions.

tical to that used for Fig. 8a except that the middle third of the stereogram has a *far* disparity of +4 pixels. The fractions of misclassified points under various thresholds are shown in the third column of Table 1.

In addition to ocularity maps, we also computed disparity maps at both V1 and V2 levels. The results for the two stereograms used in Fig. 8 are shown in Fig. 9. Fig. 9a is for the stereogram with a near, −4 pixel disparity. The top and bottom rows show the disparity maps extracted from V1 and V2 responses, respectively. In each row, the left panel shows disparity as a function of the horizontal position of the stereogram. The solid curve represents the ideal disparity map where the disparity of the monocular regions (delineated by dotted lines) takes the value of the far background surface. The gray curves represented the computed disparities, with one curve for each vertical position of the stereogram. The right panel of each row of Fig. 9a is a gray scale representation of the same computed disparity map; near and far disparities are represented by black and white colors, respectively.

It can be seen from Fig. 9a that the disparity map computed from the V1 level is already quite good. The errors mainly occur at the two monocular regions. The V2 level reduces these errors by shifting the disparity in the monocular regions toward the 0 disparity of the background, thus making the computed map in better agreement with the ideal map. This happens because we assume that each V2 cell signals the farther of its two preferred disparities (da Vinci disparity) to later stages (see Section 2). To quantify the improvement, we computed the mean absolute errors between the ideal and computed V1 and V2 maps. The errors were 0.27 and 0.14 pixels for V1 and V2, respectively. Even at the V2 level, the computed map still differs from the ideal map as the image points in the monocular regions do not all take the far disparity exactly. Interestingly, Shimojo and Nakayama (1994) measured human subjects' perceived depth of a probe dot in the monocular region, and Fig. 3 in their paper shows that the agreement between the perceived depth and the ideal map is also not precise.

Fig. 9b shows results for the stereogram with a *far* disparity of +4 pixels in the middle portion of the stereogram. The format of presentation is the same as that for Fig. 9a. The ideal map follows the same rule as stated above that the monocular regions take the disparity of the far surface. Note, however, that in Fig. 9a, the far disparity is 0 pixel on the left and right portions of the stereogram, whereas in Fig. 9b, the far disparity is 4 pixel in the middle portion of the stereogram. This explains why the ideal map in Fig. 9b has a wider central surface than that in Fig. 9a. The errors of computed disparity maps were 0.11 and 0.08 pixels for V1 and V2, respectively. The improvement by the V2 level is not as large as that in Fig. 9a mainly because the V1 map here is in better agreement with the ideal map.

A major parameter in our model is the number of V1 cells connected to a V2 cell (Fig. 3), and was set to 4 in the above simulations. To explore the dependence on this parameter, we varied the number from 2 to 12 by an increment of 2 cells. The results for the stereograms with −4 (near) or 4 (far) pixels of disparity are summarized in Fig. 10. The percent error improvement is the difference between the V1 error (in disparity map computation) and the V2 error divided by the V1 error, averaged over 10 stereograms for each disparity. There is substantial improvement at the V2 stage when the number of V1 cells connected to a V2 cell is not larger than 8. Note that there is greater improvement for the near disparity than for the far disparity. This asymmetry results from a bias in the da Vinci disparity rule which always assigns the further of the two depths to a monocular region. The rule makes the V1 stage perform better for the far disparity case, which leaves less room for the V2 stage to improve. We also generated similar plots for disparity magnitudes of 2 and 6 pixels (results not shown) and found that the V2 stage improves disparity computation for all cases when the number of V1 input cells to a V2 cell is not larger than 6.

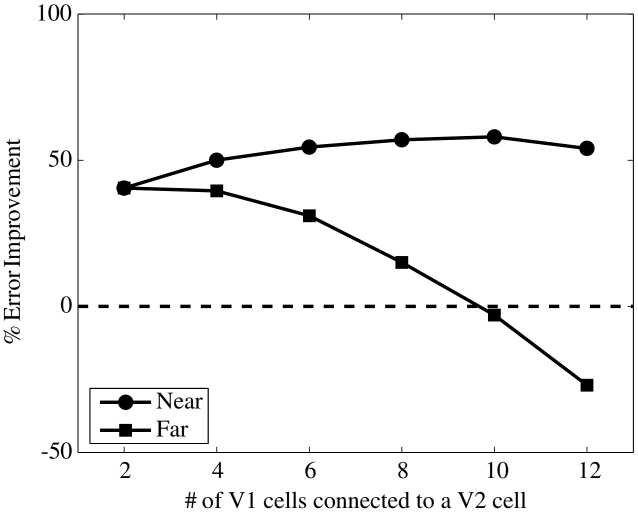

Fig. 10. Effect of varying the number of V1 cells connected to a V2 cell. The percent error improvement is the difference between the V1 error (in disparity map computation) and the V2 error divided by the V1 error, averaged over 10 stereograms. Circles and squares represent results for stimuli with −4 (near) and 4 (far) pixels of disparity, respectively.

3.4. Comparison with V2 physiolgy

Finally, we attempted to reproduce some key electrophysiological properties of depth-edge-selective V2 cells described by von der Heydt et al. (2000). Fig. 11a shows the responses of two recorded V2 cells to a random-dot stereogram with a near central square region against a background. The responses were plotted as a function of the relative horizontal position between the RFs and the stereogram. Each cell responded well only when its RF was aligned with one of the two vertical edges of the central square in the stereogram. To simulate these results, we considered two of the V2 cells in our model. The first V2 cell

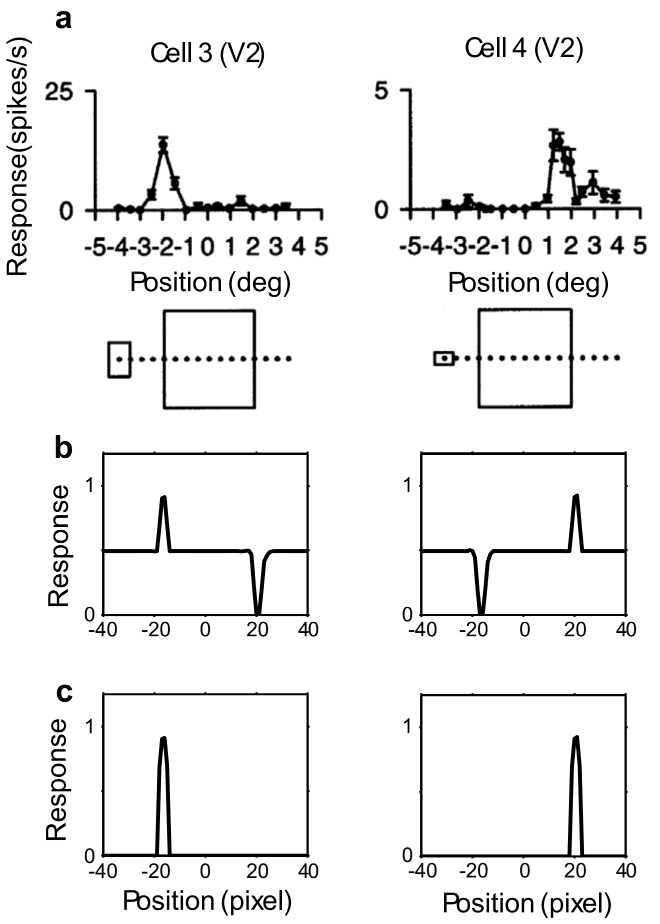

Fig. 11. Simulations of disparity-boundary-selectivity in V2. (a) Recordings of two real disparity-boundary-selective V2 cells reprinted from von der Heydt et al. (2000) with permission from Elsevier. Responses are shown as a function of the relative position between the cells' RFs (small rectangles) and the central figures (large rectangles) of random-dot stereograms. Each cell is tuned to a specific disparity edge in the stereograms. (b) Disparity-boundary-selectivity for two of our model V2 cells. (c) A half-maximum threshold is applied to (b) to eliminate the intermediate responses.

receives inputs from V1 cells with a preferred disparity of −2 pixels on the left of its RF and 2 pixels on the right of its RF. The second V2 cell prefers the opposite depth order, with a preferred disparity of 2 pixels on the left and −2 pixels on the right. We then stimulated the cells with a random-dot stereogram with a central disparity of 2 pixels and a background disparity of −2 pixels. The responses of the model V2 cells as a function of the relative position between the cells' RFs and the stereogram are shown in Fig. 11b. As expected from how the cells are constructed, each cell responds maximally to only one of the two depth edges in the stereogram, similar to the real edge-selective V2 recorded by von der Heydt et al. (2000). However, the model cells also have an intermediate response at binocular regions of the stimulus. This is because each model V2 cell has two preferred disparities (Fig. 3), and its response is half-maximum when its RF "sees" only one of the two preferred disparities. For example, consider the V2 cell preferring −2 and 2 pixels of disparities on the left and right sides of its RF, respectively. If this cell covers a disparity edge with −2 pixels on the left side and 2 pixels on the right side, both of its preferred disparities match the stimulus disparities and it fires maximally. On the other hand, if the depth order of the edge is reversed, neither side of the cell's RF matches the preferred disparity and it fires minimally. Finally, if the cell covers a uniform disparity of either −2 or 2 pixels, one of its preferred disparities matches the stimulus disparity and it fires at an intermediate, half-maximum level. This intermediate response can be eliminated with a threshold. The results, shown in Fig. 11c, are in better agreement with the data in Fig. 11a.

von der Heydt et al. (2000) also measured two-dimensional disparity-tuning plots of V2 edge-selective cells by using random-dot stereograms with different combinations of surround and figure disparities, all presented at a fixed position with the depth edges aligned with the cells' RF centers. Plots for two example cells are shown in Fig. 12a, where the horizontal and vertical axes represent surround and figure disparities, respectively, and the area of each filled dot represents the magnitude of the response. To replicate these results, we considered two model V2 cells with preferred disparity combinations of 2 and 0 pixels, and 0 and 2 pixels for the left and right subregions of the RFs. We positioned the centers of our model V2 RFs to within 0.5 pixels of the disparity edges of the stimuli. The two-dimensional disparity-tuning plots for these model cells are shown in Fig. 12b, and they resemble those of real cells recorded by von der Heydt et al. (2000). Note that the tuning plots of our model cells may not be as sharp or as precise as one would predict from the population response plots (Fig. 4) or from the preferred disparities of the cells. This is due to the fact that for the disparity energy model, there is always a larger variability in disparity-tuning plots than in population response plots for random-dot stereograms (Chen & Qian, 2004). For the same reason as in the response curves of Fig. 11b, an intermediate response was prevalent and can be removed by the same half-maximal threshold used above. The thresholded results shown in Fig. 12c are in better agreement with the physiological data.

Since we found it helpful to use a response threshold equal to that of the half-maximal responses of our model V2 cells in order to better reproduce the observed responses by von der Heydt et al. (2000), we repeated our computations of ocularity and disparity maps using thresholded model V2 cells. Fig. 13 shows the results on the same stereogram used in Figs. 8a and 9a with a near central region. The results are comparable to those obtained without the response threshold. With a normalized disparity difference threshold value of 0.5, we observe that 6% of the points are classified incorrectly. Moreover, the V2 stage continues to outperform the V1 stage at disparity computation (Fig. 13b, where conventions are the same as in Fig. 9) as it assigns a more appropriate disparity to the monocular

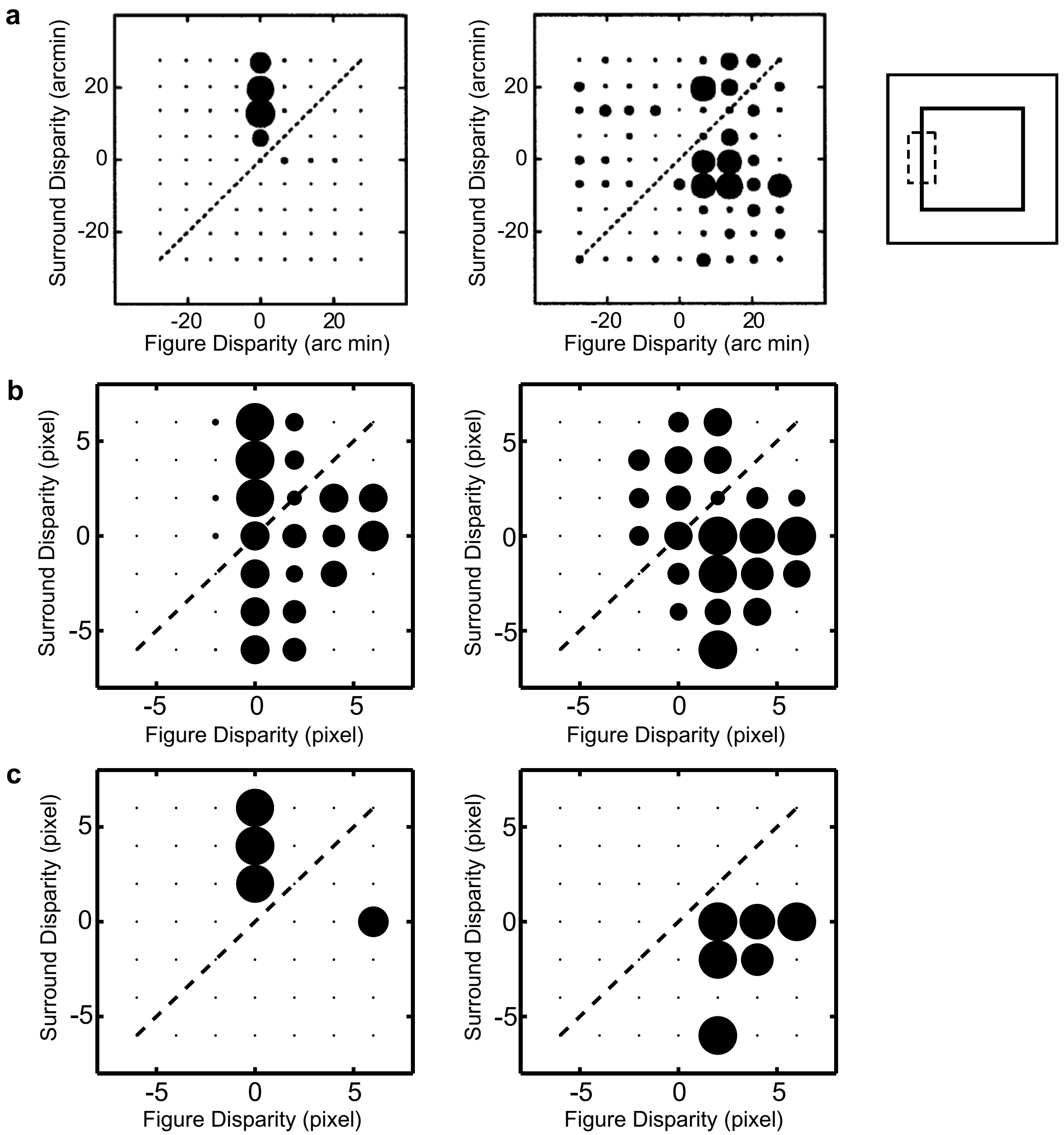

Fig. 12. Simulations of two-dimensional disparity tuning of disparity-boundary-selective V2 cells. (a) Recordings of two real disparity-boundary-selective V2 cells reprinted from von der Heydt et al. (2000) with permission from Elsevier. The RF of each cell was always aligned with a disparity boundary in a random-dot stereogram (shown in the schematic to the right in the top row). Each plot shows the responses of a single V2 cell to various combinations of the foreground and background disparities. (b) Disparity tuning for two of our model V2 cells. (c) A half-maximum threshold is applied to (b) to eliminate the intermediate responses.

regions. The mean absolute error was 0.16 pixels for the V2 stage with the half-maximum threshold, compared with an error of 0.14 pixels without the threshold in Fig. 9a, and an error of 0.27 pixels at the V1 stage.

3.5. Discussion

In this study, we investigated a few related issues of da Vinci stereopsis. The first concerns the perceived depth of a monocular bar in the special stimulus configuration used by Nakayama and Shimojo (1990), where the binocular background regions are completely featureless (see Fig. 2). The perceived depth was found to be dependent on the lateral separation between the bar and the rectangle. Although the observation was initially interpreted as evidence for da Vinci stereopsis, Gillam et al. (2003) later suggested that it may be explained by double matching as in Panum's limiting case. We applied the disparity energy model for conventional, disparity-based stereopsis to this problem, and found that the coarse-to-fine version of the

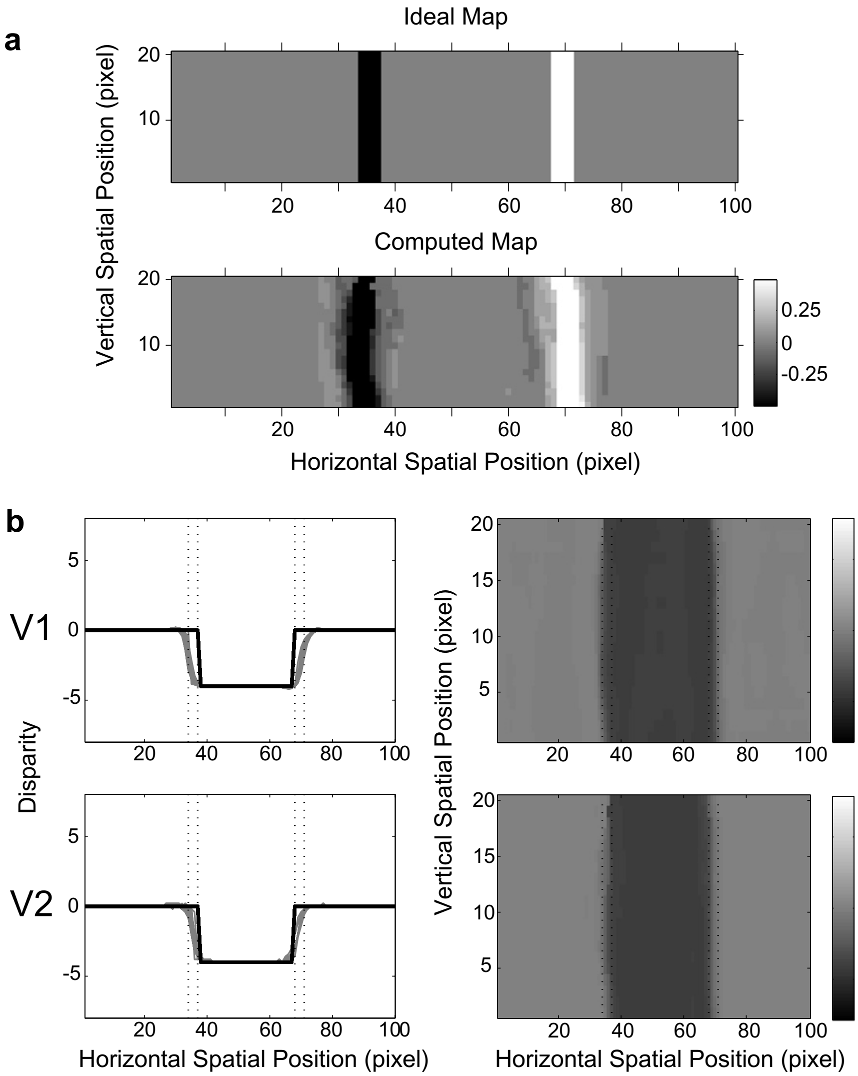

Fig. 13. Ocularity map and disparity map computations of our model with the half-maximum threshold used in Figs. 11 and 12. The stereogram is the same as that in Figs. 8a and 9a, with a near disparity of −4 pixels for the middle third of the image. (a) Computed ocularity map at the V2 stage of the model. (b) Computed disparity maps at the V1 and V2 stages of the model. The figure conventions are the same as in Figs. 8 and 9.

model explains double matching and the distance dependence of the perceived depth. Our work thus lends computational support to Gillam et al.'s interpretation. Gillam et al. (2003) also confirmed an obvious prediction of their interpretation: when the bar is replaced by a disc to avoid double matching, the distance-dependent depth is not observed.

The second issue is about the distinction between valid and invalid monocular regions as classified by Nakayama and Shimojo (1990) and shown in Fig. 2. Nakayama and Shimojo observed the distance-dependent depth of the monocular bar only for the valid cases. This appears to contradict the double matching interpretation which does not distinguish between the valid and invalid cases. However, Hakkinen and Nyman (1996) used stimuli similar to those of Nakayama and Shimojo, and found distance-dependent depth for both valid and invalid cases. The main difference between stimuli used in the two psychophysical studies is that Hakkinen and Nyman introduced a binocular rectangle above the one in Fig. 2 and shortened the monocular bar. Since this new rectangle does not change the occlusion relationship between the original binocular rectangle and the monocular bar, it should not change the validity of the monocular bar. Thus, the two psychophysical studies appear to contradict each other. To help resolve this problem, we showed in this paper that there is actually more than one valid interpretation for the "invalid" cases. A subject's perception may depend on his/her individual preferences to a particular interpretation and on how that interpretation interacts with double matching computation.

The final issue we examined is how to solve da Vinci stereopsis with a simple and physiologically plausible mechanism. We demonstrated that a feedforward model based upon the V1 binocular energy model (Ohzawa et al., 1990; Qian, 1994), extended with a coarse-to-fine algorithm

(Chen & Qian, 2004), and enhanced with the addition of a layer of disparity-edge-selective V2 cells is able to determine both the location and eye-of-origin of the monocular regions. This information could be used by later stages of the visual processing pathway to infer occlusion geometry. For the simple task of assigning a monocular region the depth of the binocular background, we assumed that when a V2 cell fires, it signals evidence for the further one of its two preferred disparities to the next stage. The assumption is motivated by the psychophysical observation that monocular regions appear to have the depth of the far background, and currently there is no physiological evidence for or against it. The disparity map so computed at the V2 stage is more accurate than that at the V1 stage.

The main advantages of our model are: (1) it uses a distributed disparity representation without converting it into a binary representation at any stage, and is thus more physiologically plausible than previous da Vinci stereopsis models, and (2) our model is much simpler than previous models which often rely on implementing complicated constraints. Another key feature of our model is the absence of monocular cells; only binocular cells are needed to compute ocularity maps, as well as disparity maps. As we mentioned in Section 1, monocular regions for da Vinci stereopsis are binocularly defined. They are generally characterized by the absence of correlated regions in the other eye (Fig. 1b) and should thus be detected by binocular cells that can sense binocular correlation (or the lack of it). In our model, multi-scale binocular energy responses in V1 provide an initial measure of binocular correlation (Qian & Zhu, 1997). This is then made much more robust by V2 disparity boundary cells. In the real V1, different binocular cells have different degrees of balance between the two eyes and they may all contribute to da Vinci stereopsis. However, more balanced cells are better binocular correlators and may play a more dominant role. Strictly monocular cells cannot sense binocular correlation at all and thus cannot contribute to monocular region detection in the general case where the binocular background surfaces are textured (Fig. 1b). Monocular cells are not found beyond V1. Even in V1, they are probably much rarer than commonly believed since many cells not responding to one eye nevertheless show non-linear binocular interactions when both eyes are stimulated together (Ohzawa & Freeman, 1986a, 1986b; Poggio & Fischer, 1977). Thus, our suggestion and demonstration of using binocular cells to solve da Vinci stereopsis remove a major restriction on possible physiological mechanisms. The non-linear binocular cells mentioned above, though not included in our current model, may also contribute to da Vinci stereopsis. On the other hand, monocular cells may be important for the related phenomena of phantom surface perception from monocular gaps (Gillam & Nakayama, 1999) since the monocular gaps are featureless and could be located by comparing monocular activities (cf. Fig. 1c).

We used excitatory feedforward connections to construct V2 cells from V1 cells. This is perhaps the simplest way to create the required disparity-boundary-selectivity, and we certainly do not mean to exclude potential roles of feedback/recurrent connections or inhibitory inputs via interneurons in shaping V2 responses or stereo computation. It may well be that our feedforward model only provides a first-order approximation to V2 disparity boundary tuning in the manner that Hubel-Wiesel's feedforward model approximates orientation mechanism in V1 (Teich & Qian, 2006). Feedback/recurrent connections and inhibitory inputs may enhance V2 disparity boundary tuning in the manner that such connections in V1 can make orientation tuning sharper and contrast invariant. For example, Mexican-hat type of recurrent excitation and inhibition could be introduced among V2 cells tuned to different disparity steps so that a given V2 cell's response to the nonpreferred disparity steps would be suppressed and hence its selectivity enhanced. Such interactions could also provide a contrast-invariant implementation of the half-maximal threshold for removing the intermediate responses in Fig. 11. The half-maximal threshold has to be adjusted according to the stimulus contrast since the intermediate responses grow with the contrast. If the threshold is implemented by V2 inhibitory cells, then the adjustment is automatic because the inhibitory activities also grow with the contrast. Alternatively, in Fig. 3a, other V1 cells in the same columns as the four connected to the V2 cell could send feedforward inhibitory inputs to the same V2 cell to enhance its disparity-edge-selectivity and achieve contrast invariance of the selectivity.

A couple of testable predications can be made from our model. The first is the specific connectivity pattern from V1 to V2 suggested by our model. The connectivity suggests that each V2 disparity-boundary-selective cell should have two preferred disparities in its RF (Fig. 3b). Secondly, our model V2 cells fail to respond reliably when the monocular gap introduced between the binocular regions of an image is larger than the maximum disparity the system can compute. Thus, the model predicts that subjects' perception of conventional stereopsis from disparity and da Vinci stereopsis from monocular regions should degrade at roughly the same stimulus disparity between figure and ground.

Gillam and Borsting (1988) reported that monocular regions having the same dot density as the rest of a random-dot stereogram facilitate stereopsis. These findings were extended by Grove and Ono (1999) who additionally demonstrated that depth is perceived more quickly when the texture of a monocular region matches the texture of a far background, rather than that of a near surface in random-dot stereograms. Our model does not contain a component for perceptual latencies and thus cannot be applied to explain these findings directly. Nevertheless, we found that our model is able to compute both ocularity and disparity maps even when monocular regions have dot densities (including zero dot density or blank regions) different from the rest of a random-dot stereogram (results not shown), as has been demonstrated with human subjects. Perhaps the latency effects could be accounted for by a grouping process in higher visual areas not currently implemented in our model.

In summary, we have designed a model for da Vinci stereopsis that determines the location and eye-of-origin of monocular regions and improves disparity map computation by using edge-selective V2 cells. The model is simple and physiologically plausible, and does not use any monocular cells. Our analysis also casts doubts on the distinction between valid and invalid monocular regions, and suggests that da Vinci stereopsis studies should focus on the general stimuli whose binocular background regions are not featureless.

Acknowledgments

We thank Drs. Aniruddha Das, Ralph Freeman, Mickey Goldberg, Ken Miller, Ken Nakayama, Shinsuke Shimojo, Rudiger von der Heydt, and the two reviewers for their helpful discussions and suggestions. This work was supported by NIH research Grant R01 EY016270 and training Grant T32 EY013933, and by a Gatsby Pilot Project grant.

References

Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature, 367*, 365–368.

Assee, A., & Qian, N. (2006). Solving da Vinci stereopsis with V2 disparity-boundary-selective cells. Society for Neuroscience Meeting.

Burkhalter, A., & Van Essen, D. C. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *Journal of Neuroscience, 6*, 2327–2351.

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.

Chen, Y., & Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural Computation, 16*, 1545–1577.

Cook, M., & Gillam, B. (2004). Depth of monocular elements in a binocular scene: The conditions for da Vinci stereopsis. *Journal of Experimental Psychology. Human Perception and Performance, 30*, 92–103.

Ehrenstein, W. H., & Gillam, B. J. (1998). Early demonstrations of subjective contours, amodal completion, and depth from half-occlusions: "Stereoscopic experiments with silhouettes" by Adolf von Szily (1921). *Perception, 27*, 1407–1416.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. *The Journal of Comparative Neurology, 201*, 519–539.

Gillam, B., & Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception, 17*, 603–608.

Gillam, B., & Nakayama, K. (1999). Quantitative depth for a phantom surface can be based on cyclopean occlusion cues alone. *Vision Research, 39*, 109–112.

Gillam, B., Cook, M., & Blackburn, S. (2003). Monocular discs in the occlusion zones of binocular surfaces do not have quantitative depth—A comparison with Panum's limiting case. *Perception, 32*, 1009–1019.

Grove, P. M., & Ono, H. (1999). Ecologically invalid monocular texture leads to longer perceptual latencies in random-dot stereograms. *Perception, 28*, 627–639.

Hakkinen, J., & Nyman, G. (1996). Depth asymmetry in da Vinci stereopsis. *Vision Research, 36*, 3815–3819.

Hayashi, R., Maeda, T., Shimojo, S., & Tachi, S. (2004). An integrative model of binocular vision: A stereo model utilizing interocularly unpaired points produces both depth and binocular rivalry. *Vision Research, 44*, 2367–2380.

Howard, I. P. (1995). Depth from binocular rivalry without spatial disparity. *Perception, 24*, 67–74.

Julesz, B. (1971). *Foundations of cyclopean perception*. University of Chicago Press.

Liu, L., Stevenson, S. B., & Schor, C. M. (1997). Binocular matching of dissimilar features in phantom stereopsis. *Vision Research, 37*, 633–644.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194*, 283–287.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological sciences, 204*, 301–328.

McLoughlin, N. P., & Grossberg, S. (1998). Cortical computation of stereo disparity. *Vision Research, 38*, 91–99.

Menz, M. D., & Freeman, R. D. (2003). Stereoscopic depth processing in the visual cortex: A coarse-to-fine mechanism. *Nature Neuroscience, 6*, 59–65.

Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research, 30*, 1811–1825.

Ohzawa, I., & Freeman, R. D. (1986a). The binocular organization of complex cells in the cat's visual cortex. *Journal of Neurophysiology, 56*, 243–259.

Ohzawa, I., & Freeman, R. D. (1986b). The binocular organization of simple cells in the cat's visual cortex. *Journal of Neurophysiology, 56*, 221–242.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science, 249*, 1037–1041.

Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology, 40*, 1392–1405.

Poggio, G. F., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *Journal of Neuroscience, 8*, 4531–4550.

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation, 6*, 390–404.

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron, 18*, 359–368.

Qian, N., & Sejnowski, T. J. (1989). Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation. In *Proceedings of the 1988 Connectionist Models Summer School* (pp. 435–443). Morgan Kaufmann.

Qian, N., & Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research, 37*, 1811–1827.

Shimojo, S., & Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Research, 30*, 69–80.

Shimojo, S., & Nakayama, K. (1994). Interocularly unpaired zones escape local binocular matching. *Vision Research, 34*, 1875–1881.

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex, 11*, 1182–1190.

Teich, A. F., & Qian, N. (2006). Comparison among some models of orientation selectivity. *Journal of Neurophysiology, 96*, 404–419.

Tsai, J. J., & Victor, J. D. (2000). Neither occlusion constraint nor binocular disparity accounts for the perceived depth in the 'sieve effect'. *Vision Research, 40*, 2265–2276.

Tsai, J. J., & Victor, J. D. (2005). Binocular depth perception from unpaired image points need not depend on scene organization. *Vision Research, 45*, 527–532.

von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research, 40*, 1955–1967.

Watanabe, O., & Fukushima, K. (1999). Stereo algorithm that extracts a depth cue from interocularly unpaired points. *Neural Networks, 12*, 569–578.

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *Journal of Physiology, 277*, 273–290.

Zhaoping, L. (2002). Pre-attentive segmentation and correspondence in stereo. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 357*, 1877–1883.

Zitnick, C. L., & Kanade, T. (2000). A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 675–684.