

# An integrative model of binocular vision: a stereo model utilizing interocularly unpaired points produces both depth and binocular rivalry

Ryusuke Hayashi <sup>a,b,\*</sup>, Taro Maeda <sup>b,c</sup>, Shinsuke Shimojo <sup>a,c</sup>, Susumu Tachi <sup>b</sup>

<sup>a</sup> *Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>b</sup> *Department of Mathematical Engineering and Information Physics, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan*

<sup>c</sup> *NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan*

Received 30 October 2003; received in revised form 26 April 2004

## Abstract

Half-occluded points (visible only in one eye) are perceived at a certain depth behind the occluding surface without binocular rivalry, even though no disparity is defined at such points. Here we propose a stereo model that reconstructs 3D structures not only from disparity information of interocularly paired points but also from unpaired points. Starting with an array of depth detection cells, we introduce cells that detect unpaired points visible only in the left eye or the right eye (left and right unpaired point detection cells). They interact cooperatively with each other based on optogeometrical constraints (such as uniqueness, cohesiveness, occlusion) to recover the depth and the border of 3D objects. Since it is contradictory for *monocularly* visible regions to be visible in *both* eyes, we introduce mutual inhibition between left and right unpaired point detection cells. When input images satisfy occlusion geometry, the model outputs the depth of unpaired points properly. An interesting finding is that when we input two unmatched images, the model shows an unstable output that alternates between interpretations of monocularly visible regions for the left and the right eyes, thereby reproducing binocular rivalry. The results suggest that binocular rivalry arises from the erroneous output of a stereo mechanism that estimates *the depth of half-occluded unpaired points*. In this sense, our model integrates stereopsis and binocular rivalry, which are usually treated separately, into a single framework of binocular vision. There are two general theories for what the “rivals” are during binocular rivalry: the two eyes, or representations of two stimulus patterns. We propose a new hypothesis that bridges these two conflicting hypotheses: *interocular* inhibition between *representations* of monocularly visible regions causes binocular rivalry. Unlike the traditional eye theory, the level of the interocular inhibition introduced here is after binocular convergence at the stage solving the correspondence problem, and thus open to pattern-specific mechanisms.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** Stereopsis; Depth; Occlusion; Interocularly unpaired point; Binocular rivalry; Interocular inhibition

## 1. Introduction

When the images projected on the left and right eyes are similar patterns with appropriate displacement (disparity), they fuse into a single vision producing an impression of depth (stereopsis). This stereoscopic process requires extraction of disparity information by establishing correspondence between image features in the two eyes.

On the other hand, if corresponding regions of the two eyes are stimulated by very dissimilar images, the patterns compete for visual dominance with each other, so that only one of them gains access to conscious perception and the other is suppressed (Levelt, 1968). Though this perceptual alternation between nonfusible dichoptic stimuli is well known as “binocular rivalry”, its mechanism remains poorly understood. Since rivalry occurs only when images have failed to fuse, it has been believed that binocular rivalry is the default outcome of the stereo mechanism when interocular matching cannot be established (Blake, 1989).

One of the elements that bridge stereopsis and binocular rivalry is the perception of half-occluded points.

\* Corresponding author. Address: MRI for Biological Cybernetics, Department of Logothetis, Spemannstrasse 38, 72076 Tübingen, Germany. Tel.: +49-7071-601-703; fax: +49-7071-601-702.

E-mail address: [ryusuke.hayashi@tuebingen.mpg.de](mailto:ryusuke.hayashi@tuebingen.mpg.de) (R. Hayashi).

As depicted in Fig. 1, when a surface occludes a more distant one, this produces regions that are partially hidden by the foreground and visible to only one eye. Though such points have no counterpart in the other eye (thus no disparity defined and may even produce false matches), psychophysical evidence indicates that the human visual system makes use of this unpaired information to reconstruct 3D structures (Gilliam & Borsting, 1988; Kaye, 1978). It has been shown that unpaired points in occlusive relations are assigned at an appropriate rear depth, whereas unpaired points that do not satisfy occlusive relations cause rivalrous perception (Nakayama & Shimojo, 1990; Shimojo & Nakayama, 1990, 1994). In addition, Nakayama and Shimojo (1990) demonstrated that unpaired points are essential for the perception of depth discontinuity and lead to the formation of occluding contours and surface (this is consistent with the observation that half-occluded zones are found at every depth discontinuity in daily visual scenes). These findings indicate that the visual system detects unpaired points somehow and makes use of occlusive relations to recover surface and occluding contour as opposed to the classical theories/models of stereopsis (e.g. Marr & Poggio, 1979) in which unpaired signals are treated merely as noise.

Several researchers have proposed stereo algorithms that use information of both disparity and unpaired points to calculate depth (Chang & Chatterjee, 1993; Geiger, Ladendorf, & Yuille, 1995; Grossberg & McLoughlin, 1997; Jones & Malik, 1992; McLoughlin & Grossberg, 1998; Nasrabadi, Clifford, & Liu, 1989; Yang & Yuille, 1995). Watanabe and Fukushima's (1999) model has an advantage in that their model extracts depth from unpaired points by using them as explicit depth cues and recovers 3D surfaces consistent with occlusion geometry, relying on the psychophysical findings of Shimojo and Nakayama (1990). Their model consists of two stages. One is a pre-processing stage for the detection of disparity and unpaired points, and the other is a 3D reconstruction stage utilizing a cooperative

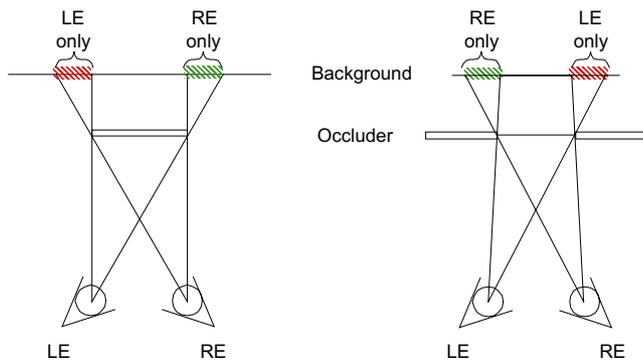


Fig. 1. Top view of two occlusion examples where parts of the background are half-occluded, and thus visible from only one eye.

algorithm (Marr & Poggio, 1979). At the latter stage, they introduced disparity detectors for representing the depth of each image point and unpaired point detectors for indicating that a point is visible only for one eye, and then they made cooperative interaction with each other depending on optogeometrical constraints.

In this paper, we propose a modification of Watanabe and Fukushima's stereo model in the following ways: We introduce (1) a physiologically plausible pre-processing stage, (2) new optogeometrical constraints such as interocular inhibition, and (3) temporal dynamics in sub-circuits, all of which were missing in Watanabe and Fukushima's original model. Furthermore, we will show that our modified model reproduces binocular rivalry when unfusable images are input. While stereopsis and binocular rivalry should have close interactions in binocular vision processing, there is no explicit computational model, to our knowledge, that successfully simulates both binocular phenomena in a single framework, providing an ecologically and optogeometrically valid interpretation for binocular rivalry. Our model of binocular vision can account for (a) stereopsis, (b) the depth of unpaired points and (c) binocular rivalry, all in a single framework. What is unique in our model is that (c) came out "for free" as an emergent outcome of the computation for (a) and (b), shedding light on the uncertain mechanism of binocular rivalry.

## 2. Method

### 2.1. Pre-processing stage

Our model consists of two computational stages as depicted in Fig. 2. In the first stage, paired and unpaired points are detected tentatively by filtering binocular input.

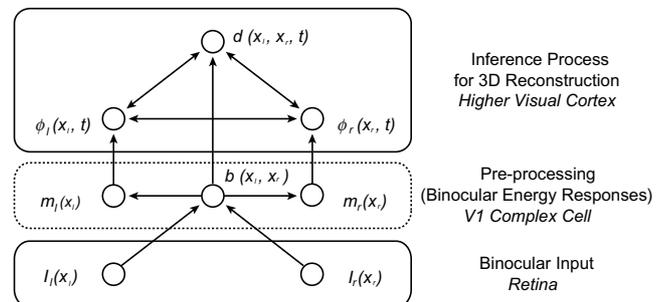


Fig. 2. The framework of our stereo model. It consists of two computational layers: one is a pre-processing stage resembling the function of disparity selective neurons in V1, modeled by binocular energy model. Initial detection of the disparity map and unpaired points is achieved at this stage. The other is the inference processing stage for 3D structures, where depth detection cells and unpaired point detection cells interact with each other based on physical constraints to estimate 3D structure of objects.

The site for the preliminary processing of binocular images has been attributed to binocular neurons in the primary visual cortex (e.g. Poggio & Fischer, 1977). The response-profiles of these disparity selective neurons can be fitted with a simple filtering model, named *the binocular energy model* (Ohzawa, DeAngelis, & Freeman, 1990). In addition, the binocular energy model can also fit neural responses to rivalrous stimuli (Cumming & Parker, 1997), and psychophysical studies suggest that the model is also a plausible implementation of disparity detection in the human visual system (Hayashi, Miyawaki, Maeda, & Tachi, 2003; Neri, Parker, & Blake-more, 1999). Furthermore, theoretical studies indicate that the model can code the disparity information regardless of the Fourier phases of input patterns (Qian, 1994), and pooled responses of binocular energy neurons across orientations, phases, and spatial frequencies provide an unambiguous representation of disparity (Fleet, Wagner, & Heeger, 1996). In this way, it has been shown that the binocular energy model captures many aspects of neural behavior in both the primary visual cortex and psychophysical behavior and provides a computational framework for representing disparity map from stereograms. Therefore, it is very reasonable to use this currently accepted model of disparity processing as our pre-processing.

The binocular energy model used here is our own implementation of the model described in Ohzawa et al. (1990) and Fleet et al. (1996). In our model, the input from each eye is convolved with the Gabor function, and the binocular sum for each filtered eye input is then squared and summed to generate the output of binocular energy responses. Here, disparity tunings are introduced by the positional shift between receptive fields (RF) in the two eyes. For simplification, we neglect vertical displacement between the two eyes' images, and thus consider only matching along horizontal axis.

Fig. 3a shows an example of binocular energy responses to a random dot stereogram (RDS) that includes unpaired points at depth discontinuities. Whereas the result shows strong and selective excitation at paired regions (thus representing their disparity successfully), various disparity selective cells are broadly activated at unpaired points. Hayashi et al. (2003) found psychophysical evidence supporting the broad activation of binocular cells in response to interocularly unpaired stimuli. These results indicate that (1) the binocular energy model alone is not sufficient to reproduce depth from unpaired points, suggesting the requirement of further processing and (2) it may be possible that unpaired points are detected by monitoring broad activations of multiple disparity selective cells receiving input from a same retinal area. We therefore assume that while disparity selective cells reflect strong and selective activation of binocular energy responses, the detectors of unpaired points are aligned at each retinal position in

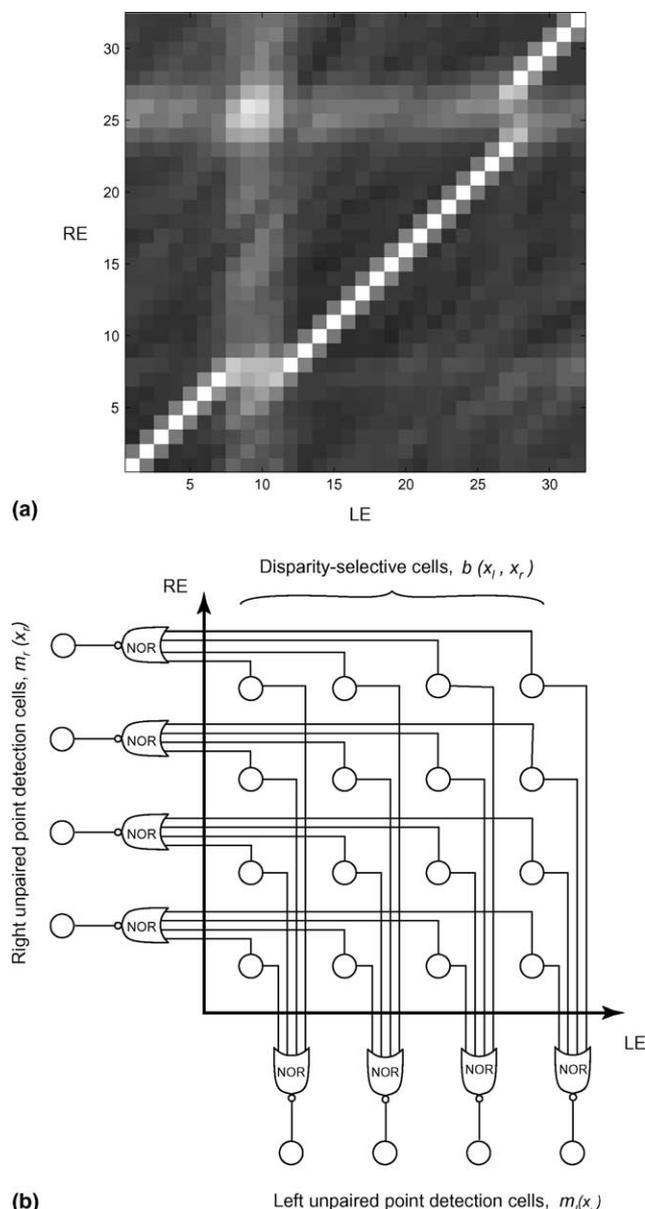


Fig. 3. (a) Pooled binocular energy responses to RDS in which the center area has near depth (four pixels crossed disparity) relative to background. The horizontal axis indicates the RF position of disparity selective cells in the left eye, and the vertical axis shows the RF position in the right eye. Each pixel represents the activity of the cells (increasing from black to white). It is successful in representing the depth of paired points where disparity can be defined, while disparity selective cells are merely broadly activated at unpaired regions. (b) Proposed mechanism of unpaired point detection. Unpaired points where disparity are not defined can be detected by converging outputs of disparity selective cells within a ocular dominance column.

each eye and check for the absence of strong activation among binocular energy responses within a corresponding ocular dominance column, as depicted in Fig. 3b. If we define the *normalized pooled response* of the binocular energy model as  $Cx(\mathbf{x}_l, \mathbf{x}_r)$ , then we can define a *disparity selective cell* ( $b(\mathbf{x}_l, \mathbf{x}_r)$ ), coding matching

between  $\mathbf{x}_l = (x_l, y)$  in the left eye and  $\mathbf{x}_r = (x_r, y)$  in the right eye) as follows:

$$b(\mathbf{x}_l, \mathbf{x}_r) = f[Cx(\mathbf{x}_l, \mathbf{x}_r) - h], \quad (1)$$

where  $f[x]$  is the Heaviside function

$$f[x] = \begin{cases} 1 & (x > 0), \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

and  $h$  is a threshold constant. Then  $m_l(\mathbf{x}_l)$ , representing whether or not a point at  $\mathbf{x}_l$  in the left eye (and  $m_r(\mathbf{x}_r)$  at  $\mathbf{x}_r$  in the right eye) is interocularly unpaired, can be formulated as follows:

$$m_l(\mathbf{x}_l) = 1 - f\left[\sum_{\mathbf{x}_r} b(\mathbf{x}_l, \mathbf{x}_r)\right], \quad (3)$$

$$m_r(\mathbf{x}_r) = 1 - f\left[\sum_{\mathbf{x}_l} b(\mathbf{x}_l, \mathbf{x}_r)\right]. \quad (4)$$

## 2.2. Inference process stage for 3D reconstruction

As Watanabe and Fukushima (1999) proposed, we hypothesize that there are unpaired point detection cells for the left and right eye in addition to depth detection cells. These three different types of cells, receiving input from the pre-processing stage, cooperatively interact with each other to reconstruct 3D structures based on physical and geometrical rules, much like Marr and Poggio's cooperative algorithm (1979). Here, surface formation and unpaired point discrimination are solved at the same time and depth detection cells code the depth of both unpaired points and paired points. That is, if a certain retinal point in the left eye is unpaired and perceived at the background depth, the left unpaired point detection cell *and* the depth detection cell turn ON simultaneously to represent this situation. We formulate the dynamics of depth detection cells and unpaired point detection cells according to Watanabe and Fukushima's model (1999) except for the choice of constraints.

The output of a depth detection cell at time  $t$  ( $d(\mathbf{x}_l, \mathbf{x}_r, t)$ ) will be ON when two points in the left and right eye coordinates ( $\mathbf{x}_l$  and  $\mathbf{x}_r$ ) are matched or when constraints indicate that an unpaired point lies at that depth. The dynamics of depth detection cells is described by the following equations:

$$d(\mathbf{x}_l, \mathbf{x}_r, t) = f[u(\mathbf{x}_l, \mathbf{x}_r, t) - h], \quad (5)$$

$$\frac{\partial}{\partial t} u(\mathbf{x}_l, \mathbf{x}_r, t) = -u(\mathbf{x}_l, \mathbf{x}_r, t) + b(\mathbf{x}_l, \mathbf{x}_r) + \alpha_u U_u + \alpha_o U_o + \alpha_c U_c, \quad (6)$$

where  $u(\mathbf{x}_l, \mathbf{x}_r, t)$  is the membrane potential of the cell at time  $t$ ,  $b(\mathbf{x}_l, \mathbf{x}_r)$  is the input from the previous stage,  $\alpha_*$  are positive constants, and  $U_*$  represent inputs from

other depth detection and unpaired point detection cells. These inputs implement the following three constraints:

1. uniqueness ( $U_u$ ),
2. occlusion ( $U_o$ ),
3. cohesiveness ( $U_c$ ).

In a similar way, we can formalize unpaired point detection cells as follows:

$$\phi_l(\mathbf{x}_l, t) = f[v_l(\mathbf{x}_l, t) - h], \quad (7)$$

$$\frac{\partial}{\partial t} v_l(\mathbf{x}_l, t) = -v_l(\mathbf{x}_l, t) + m_l(\mathbf{x}_l) + \beta_{lo} V_{lo} + \beta_{lc} V_{lc} + \beta_{li} V_{li}, \quad (8)$$

where  $\phi_l(\mathbf{x}_l, t)$  is the output of an unpaired point detection cell, representing whether a point at  $\mathbf{x}_l$  in the left eye is interocularly unpaired or not at time  $t$ , and  $v_l(\mathbf{x}_l, t)$  is the membrane potential of the cell.  $m_l(\mathbf{x}_l)$  is input from the previous stage.  $\beta_*$  are positive constants and  $V_{i*}$  represent inputs from other depth detection and unpaired point detection cells based on physical constraints. Here we introduce three constraints for unpaired point detection cells as follows:

1. occlusion ( $V_{lo}$ ),
2. cohesiveness ( $V_{lc}$ ),
3. interocular inhibition ( $V_{li}$ ).

Likewise, the output  $\phi_r(\mathbf{x}_r, t) = f[v_r(\mathbf{x}_r, t) - h]$  represents whether a point at  $\mathbf{x}_r$  in the right eye is interocularly unpaired or not.

Next, we will review physical constraints that govern 3D surfaces in the real world and then implement the constraints as the connections in a cooperative neural network. Although we will describe specific formulations for the listed constraints in detail, there are a number of choices of equation form. The aim of the next section is to propose a qualitative stereo model that outputs results consistent with the requirements of physical constraints. Therefore, equations defined and parameter settings made in the followings function as a proof-of-concept rather than as a description of the specific neural implementation in the real brain.

## 2.3. Constraints for depth detection cells

### 2.3.1. Uniqueness

Marr and Poggio (1979) established a uniqueness constraint based on the fact that "Each point from each image may be assigned at most one disparity", and implemented this as inhibition between depth detection cells aligned on the same line of sight. This implementation, however, breaks down in a situation known as Panum's limiting case depicted in Fig. 4a. If a point in one eye can be matched with only one point in the other

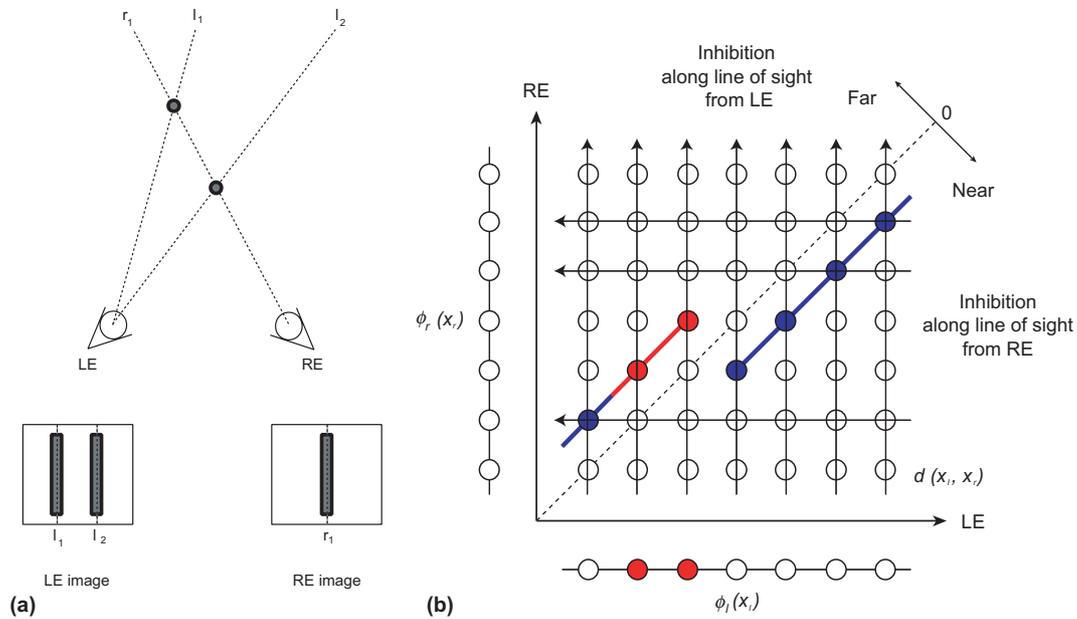


Fig. 4. (a) Top view of Panum's limiting case. (b) Schematic view of the uniqueness constraint, modified from that proposed by Watanabe and Fukushima (1999). The horizontal axis indicates the RF position of depth detection cells and unpaired point detection cells in the left eye and the vertical axis shows the RF position in the right eye. Red dots in the array of unpaired point detection cells (under the horizontal axis) indicate that the corresponding retinal location is interocularly unpaired. Blue dots in the matrix of depth detection cells represent the depth of interocularly paired dots and red dots represent the depth of interocularly unpaired dots. Black arrows depict inhibitory interactions along line of sight based on uniqueness constraints. Note that such inhibition is not applied between depth detection cells representing occluding points and occluded unpaired points (along the line of sight from the right eye crossing the unpaired points). (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

eye, two depths could not be seen in this case. However, this is not psychophysically true. We can indeed observe two dots in two different depths. Watanabe and Fukushima (1999) solved this contradiction by viewing Panum's limiting case as an example of an occlusion: one paired point and one unpaired point occluded by the former. They modified the implementation of uniqueness constraint to allowing two depth detection cells along one line of sight turn ON simultaneously if the far point is interocularly unpaired (see Fig. 4b). We therefore formulate uniqueness constraint term  $U_u$  as done in Eq. (12) of (Watanabe & Fukushima, 1999) (see Appendix A, Eq. (A.1)).

2.3.2. Occlusion

Since half-occluded unpaired points have to be invisible from one eye and visible from the other eye, the occlusion constraint requires that an occluding paired point is assigned at the location between one eye and an occluded unpaired point in the other eye (see Fig. 5). So, if a left unpaired point  $\phi_l(x_l, t)$  is localized at a depth  $d(x_l, x_r, t)$ , this constraint provides excitatory inputs to depth detection cells coding potential occluders of the left unpaired point. If the left unpaired point has its occluder, this serves to reinforce interpretations consistent with occlusion. We formulate occlusion constraint term  $U_o$  as in Eq. (13) of (Watanabe & Fukushima, 1999) (see Appendix A, Eq. (A.2)).

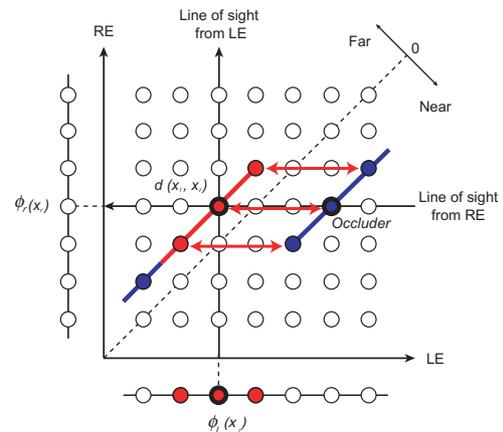


Fig. 5. Schematic view of the occlusion constraint proposed by Watanabe and Fukushima (1999). In this figure, the unpaired point detection cell  $\phi_l(x_l, t)$  is activated, coding the fact that the corresponding retinal location is interocularly unpaired and visible only from the left eye. The depth of the unpaired point is now represented by activation of a depth detection cell  $d(x_l, x_r, t)$  (red dot). If there is an interocularly paired occluder hiding the unpaired point from the view of the right eye (blue dots in this case), then the interpretation as occluded unpaired points will be facilitated. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

2.3.3. Cohesiveness

The cohesiveness (or smoothness, continuity) constraint is derived from the fact that disparity varies

smoothly almost everywhere (Marr & Poggio, 1979). In order to implement this general rule, Marr and Poggio (1979) set excitatory connections between cells selective to similar disparities in neighboring region of space. If we define disparity gradient between two stereo pairs,  $(\mathbf{x}_l, \mathbf{x}_r)$  and  $(\xi_l, \xi_r)$  as  $\Gamma = \left| \frac{\Delta d}{\Delta x_c} \right|$ , where  $\Delta d = (\xi_l - x_l) - (\xi_r - x_r)$ ,  $\Delta x_c = \{(\xi_l - x_l) + (\xi_r - x_r)\}/2$ , it is known that the human visual system cannot fuse two stereo pairs properly if  $\Gamma \geq 1$  (Burt & Julesz, 1980). Pollard, Mayhew, and Frisby (1985) implement a cohesiveness constraint in their model by setting a limit of 1 on allowable disparity gradients for psychophysically plausible implementation and for the ability to deal with a wide range of surfaces. Here we chose a weight function for this excitatory connection in a similar way:

$$w(\xi_l - x_l, \xi_r - x_r) = \exp\left(-\frac{1}{2\sigma_\Gamma^2} \Gamma^2\right) \exp\left(-\frac{1}{2\sigma_{x_c}^2} \Delta x_c^2\right), \tag{9}$$

where parameters  $\sigma_\Gamma$  and  $\sigma_{x_c}$  are chosen to make  $w$  almost zero when  $\Gamma \geq 1$  and when two points are far from each other.

Whatever function is chosen for the cohesiveness constraint, such a smoothing process should be terminated at the boundaries that are discontinuous in depth. Otherwise, this surface interpolation scheme smooths over edges and fails in reconstructing an object's shape. One solution is to exploit information that interocularly unpaired regions provide. As mentioned above, psychophysical experiments showed that the presence of unpaired points indicates depth discontinuity and leads to the perceptual formation of occluding contours (Nakayama & Shimojo, 1990). We therefore implement a mechanism that cuts filling-in excitatory connections at the location of unpaired points, which is not included in Watanabe and Fukushima's model (1999). The idea of a scheme for cutting a smoothing process is similar to the notion of "line process" (Geman & Geman, 1984; Koch, Marroquin, & Yuille, 1986). As depicted in Fig. 6a, when observing the depth of unpaired regions located at a depth discontinuity, we can find general rules of surface interpolation. The first rule is that "left (right) unpaired points receive filling-in inputs only from the left (right) side, and not from the right (left) side". In addition, if we define the *border point* of the left (right) unpaired area as the point next to the right (left) edge of the unpaired area, then we can describe the second rule as "a smoothing process from the left (right) side is terminated at the border points of the left (right) unpaired area". It is noteworthy that the border point of an unpaired area corresponds to the occluding contour of the objects and is perceived at the same depth as the occluding object. The border point of the left and right unpaired area ( $\mu_l(\mathbf{x}_l, t)$  and  $\mu_r(\mathbf{x}_r, t)$  respectively) can be formulated as follows:

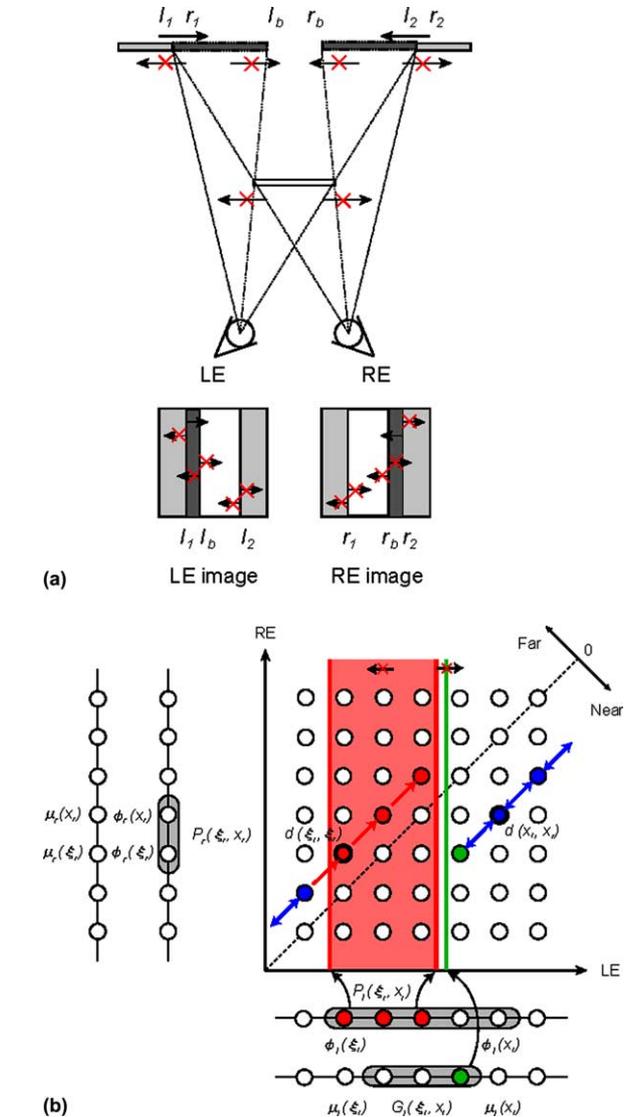


Fig. 6. (a) Possible direction of the smoothing process around depth discontinuities. This figure depicts a top view of an occlusion example. In order to estimate the depth of unpaired points as the background and form object contours, the smoothing process should work only toward the left at left unpaired points (and toward the right at right unpaired points) and be terminated at the border of the two surfaces. Black arrows indicate the direction of filling-in process and red 'x's represent the potential filling-in processes which are not executed. (b) Schematic view of the cohesiveness constraint. Excitatory connections are put between depth detection cells coding similar disparity within their neighborhood (blue double-headed arrows), while such smoothing process is only one-way at unpaired points (red single-headed arrows). All depth detection cells within the red, unpaired, area receives filling-in inputs only from one-side (left to right, in this case), terminating at the border points of unpaired points (green dots). Filling-in processes never cross the green line. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

$$\mu_l(\mathbf{x}_l, t) = \phi_l(x_l - 1, y, t)(1 - \phi_l(x_l, y, t)), \tag{10}$$

$$\mu_r(\mathbf{x}_r, t) = \phi_r(x_r + 1, y, t)(1 - \phi_r(x_r, y, t)). \tag{11}$$

Let us define function  $P_l$  indicating whether there is a left unpaired point between  $\mathbf{x}$  and  $\xi$  and function  $G_l$  indicating whether there is a border point of left unpaired area between  $\mathbf{x}$  and  $\xi$  as follows (and  $P_r$  and  $G_r$  in a same way).

$$P_l(\xi_1, \mathbf{x}_1, t) = f \left[ \sum_{\xi_1 \leq s \leq \mathbf{x}_1} \phi_l(s, y, t) \right], \quad (12)$$

$$G_l(\xi_1, \mathbf{x}_1, t) = f \left[ \sum_{\xi_1 < s < \mathbf{x}_1} \mu_l(s, y, t) \right]. \quad (13)$$

Then we can formulate the cohesiveness constraint term  $U_c$  by the following equation (see Fig. 6b for schematic view of this implementation):

$$U_c = \sum_{\xi_1 < \mathbf{x}_1 \& \xi_r < \mathbf{x}_r}^{\text{left}} [w(\xi_1 - \mathbf{x}_1, \xi_r - \mathbf{x}_r)(1 - P_r(\xi_r, \mathbf{x}_r, t)) \times \{(1 - P_l(\xi_1, \mathbf{x}_1, t)) + \phi_l(\mathbf{x}_1, t)(1 - G_l(\xi_1, \mathbf{x}_1, t))\} \times d(\xi_1, \xi_r, t)] + \sum_{\xi_1 > \mathbf{x}_1 \& \xi_r > \mathbf{x}_r}^{\text{right}} [w(\xi_1 - \mathbf{x}_1, \xi_r - \mathbf{x}_r)(1 - P_l(\xi_1, \mathbf{x}_1, t)) \times \{(1 - P_r(\xi_r, \mathbf{x}_r, t)) + \phi_r(\mathbf{x}_r, t)(1 - G_r(\xi_r, \mathbf{x}_r, t))\} \times d(\xi_1, \xi_r, t)] \quad (14)$$

The first term indicates the excitatory inputs from the left side and the second term indicates those from the right side. In the first term, filling-in input from the left side will be terminated if a right unpaired point intervenes between  $d(\mathbf{x}_1, \mathbf{x}_r, t)$  and  $d(\xi_1, \xi_r, t)$ . Otherwise, the filling-in process works (1) if no left unpaired point intervene between  $\mathbf{x}_1$  and  $\xi_1$  ( $P_l(\xi_1, \mathbf{x}_1, t) = 0$ ) or (2) if  $\mathbf{x}_1$  is a left unpaired point ( $\phi_l(\mathbf{x}_1, t) = 1$ ) and no border point intervenes between  $\mathbf{x}_1$  and  $\xi_1$  ( $G_l(\xi_1, \mathbf{x}_1, t) = 0$ ).

#### 2.4. Constraints for unpaired point detection cells

In the following, we will discuss constraints for left unpaired point detection cells only. The dynamics of right unpaired point detection cells can be formulated in the same way.

##### 2.4.1. Occlusion

As described above, the occlusion constraint requires that an unpaired point should be occluded from one eye and visible (not occluded) from the other eye. Therefore, if a left unpaired point  $\phi_l(\mathbf{x}_1, t)$  is localized at a certain depth  $d(\mathbf{x}_1, \mathbf{x}_r, t)$  and the point does not have its occluder from the right eye, the occlusion constraint enhances the activity of  $\phi_l(\mathbf{x}_1, t)$ .  $V_o$  is the term for occlusion constraint and is formulated in a similar way as Eq. (18) in Watanabe and Fukushima (1999) (see Appendix A, Eq. (A.5)).

##### 2.4.2. Cohesiveness

Cohesiveness constraint is derived from the assumption that small, isolated, unpaired point areas are considered as noise rather than occluded surface. We can implement this constraint as  $V_c$  by excitatory connections among neighboring unpaired points of same eye as Eq. (19) in Watanabe and Fukushima (1999) (see Appendix A, Eq. (A.6)).

##### 2.4.3. Interocular inhibition

A new constraint introduced here is the *interocular inhibition constraint*. This is to prevent both left and right unpaired point detection cells from turning ON at the same time ( $\phi_l(\mathbf{x}_1, t) = 1$  and  $\phi_r(\mathbf{x}_r, t) = 1$ ) while the corresponding depth detection cell is ON ( $d(\mathbf{x}_1, \mathbf{x}_r, t) = 1$ ). Since we defined “unpaired point detection cells for the left (right) eye” so as to represent that the

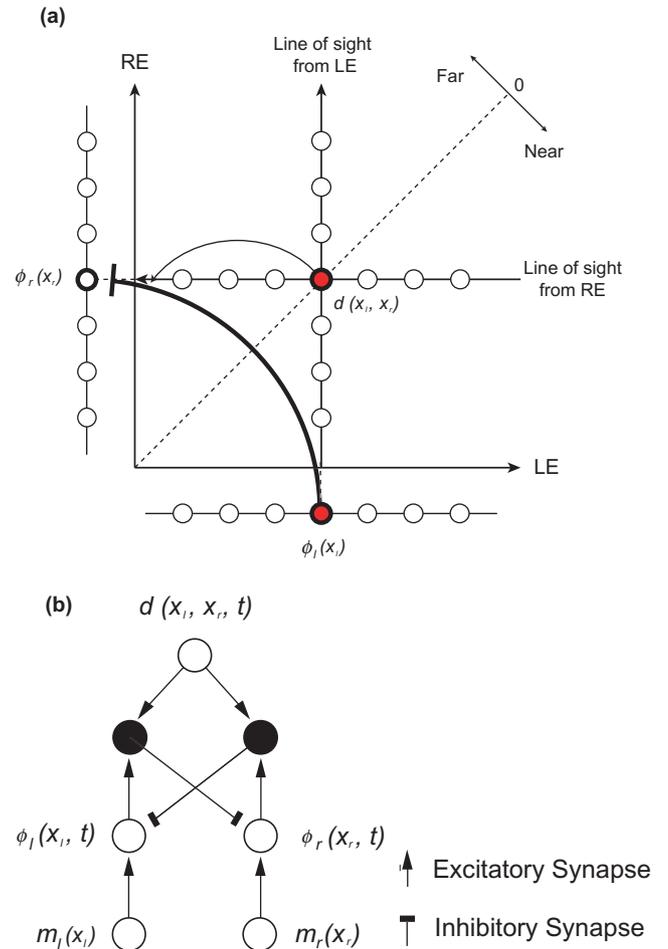


Fig. 7. (a) Schematic view of interocular inhibition constraint. If a retinal location  $\mathbf{x}_1$  in the left eye is interocularly unpaired ( $\phi_l(\mathbf{x}_1, t)$  is ON) and perceived at a certain depth ( $d(\mathbf{x}_1, \mathbf{x}_r, t)$  is ON), then the retinal location  $\mathbf{x}_r$  in the right eye should not be interocularly unpaired at the same time (thus  $\phi_r(\mathbf{x}_r, t)$  should be inhibited). (b) Implementation of interocular inhibition constraint as a reciprocal inhibition circuit between unpaired point detection cells for the left and the right eyes.

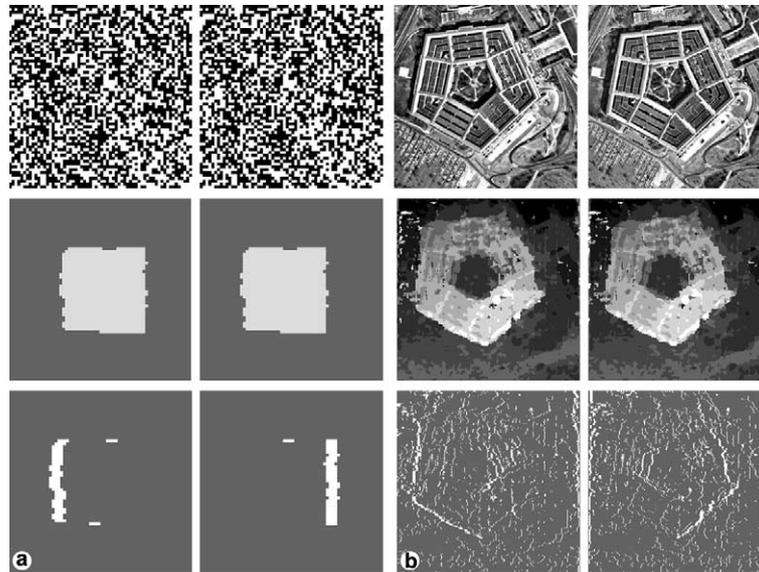


Fig. 8. Outputs of our stereo model in response to stereo images. The top row indicates input images, the middle row indicates the disparity map (output of depth detection cells, bright = near, dark = far), and the bottom row represents output map of unpaired point detection cells (white = ON, gray = OFF). Left column depicts inputs and outputs in left-eye coordinates and right column in right-eye coordinates. (a) Random dot stereogram (RDS). (b) Stereo photo images.

corresponding retinal point is unmatched with any points in the other eye (that is “only visible” for the left (right) eye), it is contradictory for an object localized at the depth of  $d(\mathbf{x}_l, \mathbf{x}_r, t)$  to be “only visible for the left eye” and “only visible for the right eye” simultaneously. Thus, if an unpaired point for the left eye  $\phi_l(\mathbf{x}_l, t)$  is assigned at a certain depth  $d(\mathbf{x}_l, \mathbf{x}_r, t)$ , then the unpaired point detection cell for the right eye  $\phi_r(\mathbf{x}_r, t)$ , corresponding to the right eye retinal point for  $d(\mathbf{x}_l, \mathbf{x}_r, t)$ , should be inhibited (see Fig. 7a). We can implement this rule as reciprocal inhibition circuit between unpaired point detection cells for the left and right eyes, depicted in Fig. 7b, and formulate  $V_{li}$  as follows.

$$V_{li} = - \sum_{x_r} \phi_r(\mathbf{x}_r, t) d(\mathbf{x}_l, \mathbf{x}_r, t). \quad (15)$$

### 3. Results

Since there is a limit of disparity which can be fused (known as Panum’s fusional area), we consider matching only within a certain range of disparities in the following simulation. Also, to simplify the convergence process for the system of equations, we used a winner-take-all selection process to create discrete firing patterns in each of the 10 iterations of the model. For each paired retinal position, the depth detection cell with the greatest gross input along both lines of sight is chosen to fire, based on the uniqueness constraint. For each unpaired retinal position, the depth detection cell with the greatest gross input is chosen to fire. We chose the parameters as follows:

$$\alpha_u = 2, \quad \alpha_o = 5, \quad \alpha_c = 5, \quad \sigma_r = 0.3, \\ \sigma_c = 4.5, \quad \beta_{l/ro} = 2, \quad \beta_{l/rc} = 2, \quad \beta_{l/ri} = 10$$

The top row of Fig. 8a is an example of input RDS image in which the center square pops out from the background. The outputs of our model in response to this RDS are shown as disparity map of depth detection cells (the middle row) and the output map of unpaired point detection cells (the bottom row), in the left-eye coordinates (the left column) and in the right-eye coordinates (the right column). As shown in the figure, the model reproduces the disparity of the center square, detects unpaired points located at the vertical edge of the square, and assigns them the depth of the background. One can see similar results when we apply our model to photo stereo images (Fig. 8b). Thus, our model seems to reconstruct 3D structures from unpaired points reasonably well. (The performance of representing the depth of a da Vinci stereogram (Nakayama & Shimojo, 1990) in our model is due to the implementation of Watanabe & Fukushima (1999). See a more detailed explanation about how to assign unique depths to unpaired regions in the case of da Vinci stereopsis in Watanabe & Fukushima (1999).)

### 4. Introducing temporal dynamics into the interocular inhibition constraint

As mentioned at the previous section, we introduced the interocular inhibition constraint using a reciprocal inhibition circuit between left and right unpaired point detection cells ( $\phi_l(\mathbf{x}_l, t)$  and  $\phi_r(\mathbf{x}_r, t)$ ), gated by a depth

detection cell ( $d(\mathbf{x}_l, \mathbf{x}_r, t)$ ). According to theoretical analysis, any mutual inhibition circuit will oscillate to produce alternative activation of each subunit if (1) the activity of each subunit attenuates with time because of adaptation, (2) there is equivalent constant input to the both subunits and (3) there is enough inhibition weight between two subunits (Matsuoka, 1984, 1985, 1987). To reflect this, we implement the oscillation of the interocular reciprocal inhibition circuit as follows: if  $\phi_l(\mathbf{x}_l, t)$  and  $\phi_r(\mathbf{x}_r, t)$  receive constant input from the pre-processing stage and one of them is activated (say  $\phi_l(\mathbf{x}_l, t)$  is ON,  $\phi_r(\mathbf{x}_r, t)$  is OFF) while  $d(\mathbf{x}_l, \mathbf{x}_r, t)$  is ON, then swap the activity of unpaired point detection cell ( $\phi_l(\mathbf{x}_l, t)$  turns OFF,  $\phi_r(\mathbf{x}_r, t)$  turns ON) after a certain duration. Although we explicitly oscillate outputs for simplification, instead of implementing adaptation dynamics in each subunit, the oscillation itself is the nature of the stereo model with adaptation effect if inhibition weight is chosen properly.

We found interesting results when we input interocularly incoherent images. In these inputs, the square central areas are independent random dot patterns between the two eyes, while surrounding areas are the same for both eyes. Fig. 9a and b shows the temporal changes of outputs of depth detection cells and unpaired point detection cells respectively. The model shows an unstable output in which unpaired point detection cells

for the left and right eye turn ON and OFF alternatively like mosaic patches. These results mean each part of the images is visible only for one eye at a time and such eye dominance alternates in sequence, and thus reproduces the perception of binocular rivalry (Levelt, 1968). It is noteworthy that even in this binocular rivalry situation, the right (left) edge of the unmatchable area tend to be perceived as a left (right) unpaired zone allocated at the far depth relative to the surrounds and such outputs are stable throughout time, which is consistent with psychophysical findings (He & Ooi, 2002). This is because our model interprets the edge of unmatchable areas as regions half-occluded by interocularly paired surrounds due to the occlusion and cohesiveness constraints. Such stable areas during binocular rivalry are limited to areas near the surrounds because these constraints work locally.

The alternation of unpaired point detection cells does not occur when we input RDS whose unpaired regions satisfy natural occlusion geometry (“the interocularly valid” case in Shimojo & Nakayama (1990)) even though the same time fluctuation mechanism is present. Consequently, our results suggest that binocular rivalry is an erroneous output of a stereo mechanism that estimates the depth of half-occluded unpaired points on the basis of occlusion geometry.

## 5. Discussion

We proposed a modification of Watanabe and Fukushima’s stereo model (1999) that uses several constraints derived from natural optogeometry to recover depth from both interocularly paired and unpaired points. In our model, active detection of interocularly unpaired points helps recover depth information around occluding contours. What is unique here is that binocular rivalry comes as an emergent outcome of the stereo computation when we input two unmatchable images into the two eyes.

The key mechanism that causes this integration of stereopsis and binocular rivalry is the reciprocal inhibition circuits between left and right unpaired point detection cells, gated by depth detection cells. This interocular inhibition is necessary for solving the stereo problem because it is contradictory for an object to be “visible *only* for the left eye” and “visible *only* for the right eye” simultaneously.

Moreover, the interocular inhibition is inherent for representing both the depth and monocular visibility in a single framework. Here, we assume that inputs from both eyes will become accessible to conscious perception if inputs are interocularly matchable (both left and right unpaired point detection cells are OFF for a corresponding depth detection cell). Imagine the case when an object is located at a certain depth and is visible only

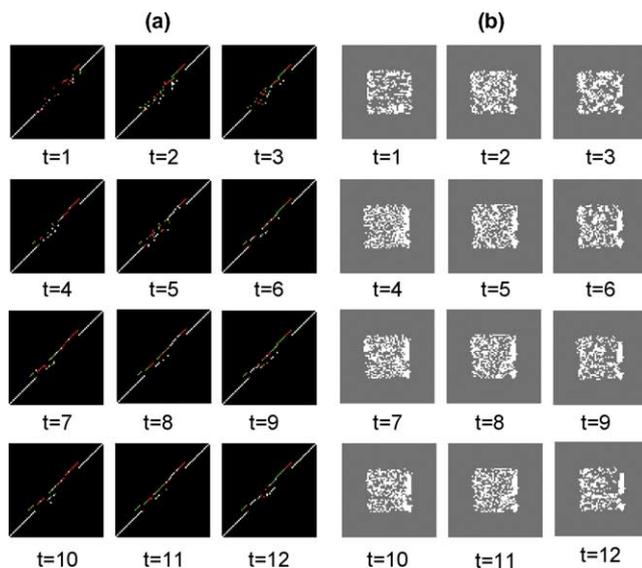


Fig. 9. Outputs for uncorrelated RDS at each iteration time, exhibiting binocular rivalry. (a) View of depth detection cells’ diagram. Red is for left unpaired points, green for right unpaired points, and white is for paired points. The upper left part of interocularly uncorrelated area, for example, is allocated as left unpaired points (red) in one frame ( $t = 4$ ) then turns into right unpaired points (green) in another frame ( $t = 6$ ) and alternates between the two states. (b) Outputs of left unpaired point detection cells in left-eye coordinate (white = ON, gray = OFF). Left unpaired regions alternate over time in mosaic-like patches.

from the left eye. Our stereo model represents this situation by firing both a depth detection cell and a left unpaired point detection cell. Without interocular inhibition, unpaired point detection cells for one eye would not suppress the input from the other eye, so the right eye input would superimpose on the left eye input, interfering with the perception of “monocularly visible object”.

### 5.1. Relationship between the present model and previous models of rivalry

In our model, detection of unpaired points, which provides a significant step towards depth computation, causes binocular rivalry as a by-product when input images are inconsistent with occlusion geometry, triggering equal activation of both left and right unpaired point detection cells. In this sense, our binocular rivalry model is a variant of “fusion theory” (Blake, 1989; Blake & O’Shea, 1988); “rivalry is the default outcome when stereopsis or fusion fails to occur”. Although classical binocular rivalry theories and models are based on the assumption that interocularly unmatchable stimuli cause rivalry, Nakayama and Shimojo (1990) showed that this is not always true; unpaired points could be stably perceived at a certain depth if they satisfy an occlusive relationship. Therefore, the failure of fusion is not sufficient for binocular rivalry. On the other hand, our stereo-model reproduces binocular rivalry only if unpaired inputs do not satisfy an occlusion configuration, as consistent with Nakayama and Shimojo’s findings. In addition, the interocular inhibition constraint that is necessary for our model to represent the depth from unpaired points provides the reason why interocular suppression occurs during binocular rivalry. Classical “fusion theories” explain that interocular suppression exists to keep a single unified vision in spite of two unfused inputs. But this explanation cannot account for why two dissimilar inputs from the two eyes are not “superimposed” to achieve single vision as in the case of low contrast patterns or transient presentation.

“Suppression theory” (Wolfe, 1986, 1988), on the other hand, expresses doubt about the predominance of stereopsis over binocular rivalry at all because stereopsis and rivalry can coexist (in the case of color rivalry, in particular (Treisman, 1962)). Before addressing this issue, we wish to claim that binocular color rivalry can be separable from contour based binocular rivalry as anatomical and physiological studies indicate that chromatic information and luminance (achromatic) information are processed through parallel visual pathways (Livingstone & Hubel, 1988). Consistent with this hypothesis, Hastorf and Myro (1959) reported that luminance-defined contours and colors can become

dissociated and mis-bound during binocular rivalry; The contour seen by the left eye can be dominant with the color seen by the right eye, and vice versa. Furthermore, Smith, Levi, Harwerth, and White (1982) measured the spectral sensitivity of the eye during the dominance and suppression phases of binocular rivalry and found that interocular suppression is dominated by the luminance channel rather than chromatic channel. It is also considered that luminance channel dominates stereoscopic vision because equiluminance presentation degrades stereoscopic depth perception (Livingstone & Hubel, 1988; Lu & Fender, 1972). Therefore, it is conceivable that coexistence of color rivalry and stereopsis (Treisman, 1962) reflects parallel processing of stereo and color information rather than that of stereo and rivalry. As for coexistence of contour rivalry and stereopsis, it was reported that coexistence is possible only if the disparity information and the rivalry contours are registered in different spatial frequency channels (Julesz & Miller, 1975; Mayhew & Frisby, 1976). However, recent studies indicate that rivalry can be ongoing in one portion of the visual field while stereoscopic depth is seen elsewhere, but both rivalry and depth are not experienced at the same spatial location simultaneously (Blake, Yang, & Wilson, 1991). Consequently, we believe that our model accounting for the predominance of stereopsis over contour rivalry is plausible for human binocular vision.

Blake (1989) proposed that binocular “EXCLUSIVE OR” cells play an important role in rivalry, in which cells are similar to our “unpaired point detection” cells in the sense that both cells represent the interocularly unmatchable stimuli. However, Blake’s “XOR” cells are assigned for each orientation rather than each eye, thus require selective connections between orientation selectivity and eye selectivity. Our unpaired point detection cells are made from simpler neural circuits (converged outputs from each ocular-dominance column), and also derive from stereo processing principles rather than accounting for binocular rivalry per se. Wolfe (1986) provided the first general binocular vision model that accounts for both stereo and rivalry. However, his model did not provide how to solve stereo-problem and how to decide which retinal input is interocularly unpaired. Furthermore, he assumed that stereo and rivalry pathways are completely independent and binocular perception is just a weighted summation of both processes. As a result, it is difficult for his hypothesis to account for findings showing that binocular rivalry and stereopsis are tightly linked with each other, such as Nakayama and Shimojo (1990).

In the following, we will review two general theories for what is rivaling during binocular rivalry and discuss how to make these superficially contradicting theories compatible. Then we will speculate on the neural substrate of binocular rivalry.

## 5.2. Eye theory vs. pattern theory

There are two general theories as to what is suppressed during binocular rivalry. One possibility is that visual information is suppressed by inhibitory interactions between left and right eyes at a relatively low-level process (“eye theory”). The alternative hypothesis is that binocular rivalry reflects a competition between different pattern representations in a relatively high-level process (“pattern theory”), categorizing binocular rivalry as just one of several phenomena related with the viewing of bistable figures, such as a Necker cube or Rubin’s vase-face stimulus (Andrews, 2001).

Several lines of evidence support low-level processing’s role in binocular rivalry. Since suppression during binocular rivalry operates non-selectively over a broad range of probe stimuli (Blake & Camisa, 1979; Fox & Check, 1972; Wales & Fox, 1970) and observers immediately experience a switch in dominance when the two rival stimuli are swapped between the two eyes (Blake, Westendorf, & Overton, 1980), it was argued that a region of an eye is suppressed rather than information about a particular set of stimulus features. Binocular rivalry occurs independently in patches of the visual field and individual regions dominated by one eye’s input are scaled in size proportional to the magnification factor of the striate cortex (Blake, O’Shea, & Halpern, 1988), implying that the suppression operates over small domains in early vision, which again argues for the ocular origin of rivalry. Finally, several fMRI studies indicate that neural activity correlated with rivalrous perception is measurable within the primary visual cortex (Lee & Blake, 2001; Polonsky, Blake, Braun, & Heeger, 2000) and even show that interocular competition mediates binocular rivalry (Tong & Engel, 2001). Based on these lines of evidence, many models based on the oscillating circuit involving reciprocal feedback inhibition between “pure monocular neurons” in V1 or LGN (thus prior to the point of binocular combination) have been proposed to simulate the temporal dynamics of binocular rivalry (Blake, 1989; Lehky, 1988).

On the other hand, many studies support a stimulus feature suppression mechanism, relying on the fact that suppression during binocular rivalry is not purely monocular and involving higher level cortical activities where monocular information is lost. Single-unit recordings have shown that neurons whose activity correlates with perception during rivalry are not monocularly but binocularly driven (Sengpiel, Blakemore, & Harrad, 1995) and that the extent of rivalry-related modulations increase in successive stages of early visual areas (V1, V4 and MT, (Leopold & Logothetis, 1996)). It is only in higher visual areas such as inferotemporal cortex, that a greater proportion of neurons show activity that reflects the ongoing alternations in perceptual dominance (Logothetis & Schall, 1989). Early

human fMRI studies yield signals highly correlated with observers’ perceptual reports in higher-level brain (Lumer, Friston, & Rees, 1998; Tong, Nakayama, Vaughan, & Kanwisher, 1998). It is also reported that visual context can influence the predominance of a figure during binocular rivalry (Alais & Blake, 1999; Kovacs, Papathomas, Yang, & Feher, 1997). Moreover, Logothetis, Leopold, and Sheinberg (1996) used a new stimulus paradigm in which the rivaling patterns were repeatedly exchanged between the two eyes and found smooth and slow perceptual alternations in spite of continuous fast alternation of the stimulus in each eye. (Though Lee and Blake (1999) found that pattern rivalry occurs only within a limited range of spatial and temporal parameters, otherwise eye rivalry dominates, thus the issue remains unresolved.) These psychophysical findings indicate that binocular rivalry is not due to complete suppression of one monocular channel but that dominance can be distributed between the eyes. Therefore, pure monocular cells in the LGN or layer 4 of V1 are unlikely to provide the neural substrate for the suppression underlying binocular rivalry. Rather rivalry may result from competition after binocular integration.

Despite extensive research, therefore, the issue as to what exactly is rivaling during binocular rivalry (eye vs. pattern) has remained highly controversial. What makes two theories contradictory is the assumption that any form of eye suppression has to occur within pure monocular channel, before binocular integration.

Here we propose a new hypothesis that bridges these two conflicting theories. It is that *interocular inhibitions between unpaired point detection cells for the left and right eye cause binocular rivalry*. In our model, interocularly unpaired regions are initially detected from the pooled activities of disparity selective neurons (simulated by the binocular energy model) whose RFs share a particular retinal location in one eye (i.e. neurons which share input from a particular ocular dominance column in V1). Then, unpaired point detection cells for the left and right eyes inhibit each other based on the interocular inhibition rule at the inference process stage for 3D reconstruction. Unlike the traditional eye competition theory, however, the interocular inhibition proposed here occurs not at the level of pure monocular neurons but *after* binocular convergence where eye of origin information is still retained.

Consequently, while our hypothesis is consistent with psychophysical findings supporting eye competition mechanisms, it is also compatible with psychophysical and single-unit studies indicating that the neural basis of binocular rivalry is binocularly driven. Since physiological evidence indicates a continuous, rather than a discrete, gradient from purely monocular to binocular neurons in the primary visual cortex (Hubel & Wiesel, 1962), it is reasonable to assume that eye of origin

information is not represented only by pure monocular neurons but by population of neurons within an ocular dominance column with a varying degree of monocularity. Also, inhibitory interactions between adjacent ocular dominance columns are suggested by physiological study (Buzas, Eysel, Adorjan, & Kisvarday, 2001; Kisvarday et al., 2002; Sengpiel et al., 1995). The mechanism of coding “interocular unpairedness” may thus trigger rivalry suppression by gating, or gain controlling, the output signal from the ocular dominance column to higher-level processing. Also, the same mechanism could selectively inhibit particular ocular dominance columns or monocular regions in V1 through feedback projections, resulting in the periodic fluctuations in fMRI signal of V1 reported by Tong and Engel (2001).

Furthermore, the process coding monocularly visible points requires global depth information and knowledge of occlusion geometry, leading to the formation of surface and occluding contour of objects (Nakayama & Shimojo, 1990). In this sense, the processing of unpaired points is modulated by visual context and may be closely related with object recognition with which pattern specific mechanisms or other cognitive factors would be concerned. Therefore, our hypothesis is open to additional mechanisms to explain pattern dominance effects and thus *not inconsistent* with pattern competition theory. (However, we do not implement further pattern specific mechanism here to duplicate pattern competition phenomena.)

### 5.3. The neural substrate of binocular rivalry

It is apparent that multiple neural operations are involved in pattern perception during rivalry and each of these operations is implemented by neural events distributed throughout the visual pathways (Blake & Logothetis, 2002). We speculate what competes during rivalry are two exclusive reentrant circuits including earlier and higher levels and feedforward and feedback pathways. Our results point to one of the underlying processes of binocular rivalry (presumably earlier level but not as early as pure monocular channel) that “knows” in which eye interocularly unpaired stimuli are imaged and uses such eye of origin information for 3D perception. It triggers bistable outputs as an emergent error when incompatible images are projected on the two eyes, thus leading to binocular rivalry.

Recent physiological studies suggest that V2 cells play an important role in the processing of contour detection (von der Heydt, Peterhans, & Baumgartner, 1984; Lee & Nguyen, 2001), border ownership detection (Zhou, Friedman, & von der Heydt, 2000), relative depth (von der Heydt, Zhou, & Friedman, 2000; Thomas, Cumming, & Parker, 2002) and the integration of contour / segmentation of surfaces based on contextual

depth information such as occlusion (Bakin, Nakayama, & Gilbert, 2000). All of these processes are tightly linked with the implementation of the inference process for 3D structure depicted here. As V2 is adjacent to V1, V2 is an adequate location to receive convergent output from an ocular dominance column and modulate its activity through a feedback pathway. We speculate that V2 cells are crucially involved in the coding of interocularly unpaired points as well as the figure-ground segregation process. Although our hypothesis assumes that binocular rivalry is triggered at the level of binocular neurons in V1 and V2, it does not preclude the possibility of interactions from other higher visual areas, via reentrant circuits.

## 6. Conclusion

We propose a stereo-model that reconstructs 3D structures using not only disparity information but also interocularly unpaired points. Furthermore, when we input incompatible two images to the model, it outputs a continuous struggle between regions that are visible only from the left eye and regions that are visible only from the right eye, reproducing binocular rivalry. The results lead us to a new hypothesis regarding what is “rivaling” during binocular rivalry: the representation of unpaired points causes interocular inhibition at a stage after the convergence of binocular information. Our hypothesis bridges two theories about rivals (eye vs. pattern) and is consistent with most of the psychophysical and physiological findings related with binocular rivalry. More importantly, our theory derives not from a model explaining binocular rivalry per se but from a framework of stereopsis that is fundamental to the functioning of our visual system. From our viewpoint, binocular rivalry arises from an erroneous output of a stereo-mechanism that estimates “the depth of half-occluded unpaired points”.

## Appendix A

### Uniqueness constraint for depth detection cells

$$\begin{aligned}
 U_u = & - \sum_{\xi_l > x_l}^{\text{near}} (1 - \phi_l(\mathbf{x}_l, t) + \phi_r(\mathbf{x}_r, t)) d(\xi_l, \mathbf{x}_r, t) \\
 & - \sum_{\xi_l < x_l}^{\text{far}} (1 - \phi_l(\xi_l, t) + \phi_r(\mathbf{x}_r, t)) d(\xi_l, \mathbf{x}_r, t) \\
 & - \sum_{\xi_r < x_r}^{\text{near}} (1 - \phi_r(\mathbf{x}_r, t) + \phi_l(\mathbf{x}_l, t)) d(\mathbf{x}_l, \xi_r, t) \\
 & - \sum_{\xi_r > x_r}^{\text{near}} (1 - \phi_r(\xi_r, t) + \phi_l(\mathbf{x}_l, t)) d(\mathbf{x}_l, \xi_r, t). \quad (\text{A.1})
 \end{aligned}$$

## Occlusion constraint for depth detection cells

$$\begin{aligned}
 U_o = & \sum_{\xi_l < x_l}^{\text{far}} \phi_l(\xi_l, t) (1 - N_r(\xi_l, \mathbf{x}_r, t)) d(\xi_l, \mathbf{x}_r, t) \\
 & + \phi_l(\mathbf{x}_l, t) N_r(\mathbf{x}_l, \mathbf{x}_r, t) \\
 & + \sum_{\xi_r > x_r}^{\text{far}} \phi_r(\xi_r, t) (1 - N_l(\mathbf{x}_l, \xi_r, t)) d(\mathbf{x}_l, \xi_r, t) \\
 & + \phi_r(\mathbf{x}_r, t) N_l(\mathbf{x}_l, \mathbf{x}_r, t), \quad (\text{A.2})
 \end{aligned}$$

$$N_l(\mathbf{x}_l, \mathbf{x}_r, t) = f \left[ \sum_{\xi_r < x_r}^{\text{near}} d(\mathbf{x}_l, \xi_r, t) \right], \quad (\text{A.3})$$

$$N_r(\mathbf{x}_l, \mathbf{x}_r, t) = f \left[ \sum_{\xi_l > x_l}^{\text{near}} d(\xi_l, \mathbf{x}_r, t) \right]. \quad (\text{A.4})$$

## Occlusion constraint for left unpaired point detection cells

$$V_{lo} = \sum_{x_r} N_r(\mathbf{x}_l, \mathbf{x}_r, t) d(\mathbf{x}_l, \mathbf{x}_r, t). \quad (\text{A.5})$$

## Cohesiveness constraint for left unpaired point detection cells

$$V_{lc} = f \left[ \sum_{\xi_l \in \varepsilon} \phi_l(\xi_l, t) - h \right]. \quad (\text{A.6})$$

## References

- Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research*, 39, 4341–4353.
- Andrews, T. J. (2001). Binocular rivalry and visual awareness. *Trends in Cognitive Sciences*, 5, 407–409.
- Bakin, J. S., Nakayama, K., & Gilbert, C. D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *Journal of Neuroscience*, 20, 8188–8198.
- Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review*, 96, 145–167.
- Blake, R., & Camisa, J. C. (1979). On the inhibitory nature of binocular rivalry suppression. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 315–323.
- Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience*, 3, 13–21.
- Blake, R., & O’Shea, R. P. (1988). “Abnormal fusion” of stereopsis and binocular rivalry. *Psychological Review*, 95, 151–154.
- Blake, R., O’Shea, R. P., & Halpern, L. (1988). Spatial extent of binocular rivalry suppression. *Investigative Ophthalmology and Visual Science*, 29(Suppl.), 411.
- Blake, R., Westendorf, D., & Overton, R. (1980). What is suppressed during binocular rivalry? *Perception*, 9, 223–231.
- Blake, R., Yang, Y., & Wilson, H. R. (1991). On the coexistence of stereopsis and binocular rivalry. *Vision Research*, 31, 1191–1203.
- Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, 208, 615–617.
- Buzas, P., Eysel, U. T., Adorjan, P., & Kisvarday, Z. F. (2001). Axonal topography of cortical basket cells in relation to orientation, direction, and ocular dominance maps. *Journal of Comparative Neurology*, 437, 259–285.
- Chang, C., & Chatterjee, S. (1993). Ranging through Gabor logons—a consistent, hierarchical approach. *IEEE Transactions of Neural Networks*, 4, 827–843.
- Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, 389, 280–283.
- Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research*, 36, 1839–1857.
- Fox, R., & Check, R. (1972). Independence between binocular rivalry suppression duration and magnitude of suppression. *Journal of Experimental Psychology*, 93, 283–289.
- Geiger, D., Ladendorf, B., & Yuille, A. (1995). Occlusions and binocular stereo. *International Journal of Computer Vision*, 14, 211–226.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gilliam, B., & Borsting, E. (1988). The role of monocular regions in stereoscopic display. *Perception*, 17, 603–608.
- Grossberg, S., & McLoughlin, N. P. (1997). Cortical dynamics of three-dimensional surface perception: binocular and half-occluded scenic images. *Neural Networks*, 10, 1583–1605.
- Hastorf, A. H., & Myro, G. (1959). The effect of meaning on binocular rivalry. *American Journal of Psychology*, 72, 393–400.
- Hayashi, R., Miyawaki, Y., Maeda, T., & Tachi, S. (2003). Unconscious adaptation: a new illusion of depth induced by stimulus features without depth. *Vision Research*, 43, 2773–2782.
- He, Z. J., & Ooi, T. L. (2002). Figural contours and border-ownership constraint in binocular rivalry. *Journal of Vision*, 2, 465a.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *Journal of Physiology*, 160, 106–154.
- Jones, D. J., & Malik, J. (1992). A computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing*, 10, 395–410.
- Julesz, B., & Miller, J. E. (1975). Independent spatial-frequency-tuned channels in binocular fusion and rivalry. *Perception*, 4, 125–143.
- Kaye, M. (1978). Stereopsis without binocular correlation. *Vision Research*, 18, 1013–1022.
- Kisvarday, Z. F., Ferecsko, A. S., Kovacs, K., Buzas, P., Budd, J. M. L., & Eysel, U. T. (2002). One axon-multiple functions: Specificity of lateral inhibitory connection by large basket cells. *Journal of Neurocytology*, 31, 255–264.
- Koch, C., Marroquin, J., & Yuille, A. (1986). Analog “neural” networks in early vision. *Proceedings of the National Academy of Sciences of the United States of America*, 83, 4263–4267.
- Kovacs, I., Papathomas, T. V., Yang, M., & Feher, A. (1997). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 15508–15511.
- Lee, S. H., & Blake, R. (1999). Rival ideas about binocular rivalry. *Vision Research*, 39, 1447–1454.
- Lee, S. H., & Blake, R. (2001). V1 activity is reduced during binocular rivalry. *Journal of Vision*, 2, 618–626.
- Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 1907–1911.
- Lehky, S. R. (1988). An astable multivibrator model of binocular rivalry. *Perception*, 17, 215–228.
- Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys’ percepts during binocular rivalry. *Nature*, 379, 549–553.
- Levelt, W. J. M. (1968). *On binocular rivalry*. Paris, Mouton: The Hague.
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240, 740–749.

- Logothetis, N. K., Leopold, D. A., & Sheinberg, D. L. (1996). What is rivaling during binocular rivalry? *Nature*, *380*, 621–624.
- Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, *245*, 761–763.
- Lu, C., & Fender, D. H. (1972). The interaction of color and luminance in stereoscopic vision. *Investigative Ophthalmology*, *11*, 482–490.
- Lumer, E. D., Friston, K., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*, 1930–1934.
- Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London Series B: Biological Sciences*, *B204*, 301–328.
- Matsuoka, K. (1984). The dynamic model of binocular rivalry. *Biological Cybernetics*, *49*, 201–208.
- Matsuoka, K. (1985). Sustained oscillations generated by mutually inhibiting neurons with adaptation. *Biological Cybernetics*, *52*, 367–376.
- Matsuoka, K. (1987). Mechanisms of frequency and pattern control in the neural rhythm generators. *Biological Cybernetics*, *56*, 345–353.
- Mayhew, J. E. W., & Frisby, J. P. (1976). Rivalrous texture stereogram. *Nature*, *264*, 53–56.
- McLoughlin, N. P., & Grossberg, S. (1998). Cortical computation of stereo disparity. *Vision Research*, *38*, 91–99.
- Nakayama, K., & Shimojo, S. (1990). Da vinci stereopsis; depth and subjective occluding contours from unpaired image points. *Vision Research*, *30*, 1811–1825.
- Nasrabadi, M. N., Clifford, S. P., & Liu, Y. (1989). Integration of stereo vision and optical flow by using an energy-minimization approach. *Journal of the Optical Society of America*, *A6*, 900–907.
- Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, *401*, 695–698.
- 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.
- Pollard, S. B., Mayhew, J. E. W., & Frisby, J. P. (1985). PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, *14*, 449–470.
- Polonsky, A., Blake, R., Braun, J., & Heeger, D. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neuroscience*, *3*, 1153–1159.
- Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, *6*, 390–404.
- Sengpiel, F., Blakemore, C., & Harrad, R. (1995). Interocular suppression in the primary visual cortex: a possible neural basis of binocular rivalry. *Vision Research*, *35*, 179–195.
- 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, E. L., Levi, D. M., Harwerth, R. S., & White, J. M. (1982). Color vision is altered during the suppression phase of binocular rivalry. *Science*, *218*, 802–804.
- Thomas, O. M., Cumming, B. G., & Parker, A. J. (2002). A specialization for relative disparity in V2. *Nature Neuroscience*, *5*, 472–478.
- Tong, F., & Engel, S. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature*, *411*, 195–199.
- Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759.
- Treisman, A. (1962). Binocular rivalry and stereoscopic depth perception. *Quarterly Journal of Experimental Psychology*, *14*, 23–37.
- von der Heydt, R., Peterhans, R., & Baumgartner, G. (1984). Illusory contours and cortical neurons responses. *Science*, *224*, 1260–1262.
- von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, *40*, 1955–1967.
- Wales, R., & Fox, R. (1970). Increment detection thresholds during binocular rivalry suppression. *Perception and Psychophysics*, *8*, 90–94.
- Watanabe, O., & Fukushima, K. (1999). Stereo algorithm that extracts a depth cue from interocularly unpaired points. *Neural Networks*, *12*, 569–578.
- Wolfe, J. M. (1986). Stereopsis and binocular rivalry. *Psychological Review*, *93*, 269–282.
- Wolfe, J. M. (1988). Parallel ideas about stereopsis and binocular rivalry: a reply to Blake and O'Shea (1988). *Psychological Review*, *95*, 155–158.
- Yang, Y., & Yuille, A. L. (1995). Multilevel enhancement and detection of stereo disparity surfaces. *Artificial Intelligence*, *78*, 121–145.
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, *20*, 6594–6611.