Vision Research 50 (2010) 181–192

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The changing disparity energy model

Qiuyan Peng, Bertram E. Shi *

Dept. of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 7 April 2009
Received in revised form 19 October 2009

Keywords:
Stereo disparity
Changing disparity
Stereomotion
Energy model
Motion in depth

ABSTRACT

Changing disparity is a possible cue for stereomotion perception. We propose the changing disparity energy model, a physiologically plausible model for neurons tuned to changing disparity. This model combines the disparity and motion energy models commonly used to model cortical neuron outputs. The model outputs are consistent with psychophysical experiments indicating that stereomotion speed discrimination thresholds for dynamic random dot stereograms are higher than for random dot stereograms. Thus, these experimental results are not necessarily strong evidence for the existence of an inter-ocular velocity difference cue. The model also predicts a relationship between the speed discrimination threshold ratio and the dot density.

1. Introduction

Stereomotion, also known as motion in depth (MID), refers to object motion either approaching or receding from the observer. There are two visual cues that are commonly used to account for the perception of motion in depth: changing disparity (CD) and Inter-Ocular Velocity Difference (IOVD) (Rashbass & Westheimer, 1961a, 1961b). The CD cue assumes that information is first combined between the two eyes to yield a disparity estimate. The temporal evolution of this disparity estimate is then used to estimate the change in disparity, which in turn leads to the perception of motion in depth. The IOVD cue assumes that monocular velocity estimates are first extracted. The difference between the velocity estimates then leads to a perception of motion in depth. Psychophysical evidence suggests that both play a role in stereomotion perception (Brooks, 2002; Brooks & Stone, 2004). These models often assume that measurements of velocity and/or disparity are available, but do not address how they might be computed, represented, nor combined.

Although the motion and disparity energy models can be combined to yield model neurons that jointly encode disparity and motion (Qian, 1994), these model neurons are not tuned to motion in depth, but rather fronto-parallel motion (Chen, Wang, & Qian, 2001). It has been suggested that this joint encoding can explain the Pulfrich phenomenon, where an inter-ocular time delay leads to a perception of depth in a stimulus moving in a fronto-parallel direction (Anzai, Ohzawa, & Freeman, 2001; Pack, Born, & Livingstone, 2003; Qian & Andersen, 1997). Read and Cumming argue that non-directionally selective space–time separable filtering can also account for the Pulfrich effect (Read & Cumming, 2005a, 2005c), and that neurons that are simultaneously tuned to direction and disparity are rare in V1 (Read & Cumming, 2005b). However, recent work has questioned the separable model (Qian & Freeman, 2009).

Recently, Sabatini et al. have shown that introducing imbalances in ocular dominance into the joint disparity/motion model results in neurons tuned to motion in depth (Sabatini & Solari, 2004; Sabatini, Solari, Cavalleri, & Bisio, 2003). We consider this model an example of the IOVD mechanism. In the limit when the ocular dominance index goes to unity, the model first computes the opponent motion energy for each eye separately, and then differences the left and right eye motion energy in a second stage.

Here, we describe a physiologically plausible model for constructing retinotopic arrays of neurons that are selective for changing disparity. This model has two stages: the first based on the disparity energy model and the second based on the motion energy model. Both of these models are commonly used to describe the response properties of neurons in the visual cortex (Adelson & Bergen, 1985; Ohzawa, DeAngelis, & Freeman, 1990; Watson & Ahumada, 1985). This model takes stereo video pairs as input. It outputs model neuron responses that are tuned to a particular changing disparity. Due to its close relationship with other energy models, we refer to our model as the changing disparity (CD) energy model. An initial description and characterization of this model has appeared in (Guo & Shi, 2008).

We can characterize our model using the same stimuli used for psychophysical experiments on motion in depth perception. Two commonly used stimuli are random dot stereograms (RDS), where the disparity between the dot pattern varies over time, and

* Corresponding author. Fax: +852 2358 1485.
E-mail addresses: eepeng@ust.hk (Q. Peng), eebert@ee.ust.hk (B.E. Shi).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.11.012

dynamic random dot stereograms (DRDS), where both the disparity and the dot pattern changes over time. RDS stimuli contain both CD and IOVD cues. DRDS stimuli contain only the CD cue, since the changing dot pattern eliminates coherent monocular motion.

Using these stimuli, we show that our model is consistent with recent psychophysical experiments. Brooks and Stone (2004) showed that speed discrimination thresholds are higher for DRDS stimuli than for RDS stimuli by factors ranging from 1.4 to 2.4. Intuitively, a neuronal mechanism based on extracting CD should exhibit no difference in speed discrimination thresholds for the two inputs. Thus, the improved discriminability for RDS stimuli appears to support the hypothesis that both CD and IOVD mechanisms are used in stereomotion perception. Surprisingly, when we replicated these experiments with the CD energy model, we find a similar difference in speed discrimination thresholds. Thus, by itself the threshold difference does not necessarily support the joint use of both CD and IOVD mechanisms for MID perception. The model is also consistent with other psychophysical experiments. For example, we observe no significant change in speed discrimination threshold for stimuli directly receding or moving obliquely in depth (Brooks & Stone, 2004, 2006).

The paper is organized as follows. Section 2 reviews the disparity energy and motion energy models, and shows how they can be combined into the changing disparity energy model. Section 3 describes how we simulate the model and our parameter choices. Section 4 describes the results of running this model on RDS and DRDS stimuli. It shows that the model is consistent with prior physical experiments, and generates a testable prediction regarding the dependency of the speed discrimination threshold on the dot density. Finally, Section 5 concludes with a summary and expanded explanation of these results.

2. Model description

In this section, we start by briefly reviewing the disparity energy and motion energy models. We then describe how they are combined to yield the CD energy model.

2.1. Disparity energy model

The disparity energy model was proposed to account for the responses of disparity selective complex cells in the primary visual cortex (Ohzawa et al., 1990). As shown in Fig. 1a, monocular inputs are first combined spatially using receptive field (RF) profiles described by Gabor functions. The RF outputs are then combined binocularly by summing and squaring signals from the left and right eyes. The preferred disparity of a disparity energy neuron depends upon both phase and position shifts between the monocular RF profiles. Here, we consider only phase shifts.

We define $U_e(x, y, t)$ as the 2D image intensity over space (x, y) and time t. The subscript $e \in \{L, R\}$ indexes the left and right eyes. We assume that the left and right input images are rectified so that each pixel in the right image corresponds to a pixel in the left image on the same scan line. We define the disparity d to be the difference between the locations of corresponding points in the left and right eyes:

$$\begin{aligned} U_L(x,y,t) &= U(x,y,t) \\ U_R(x,y,t) &= U(x+d,y,t) = U_L(x+d,y,t) \end{aligned} \tag{1}$$

A positive disparity indicates that the environmental point is closer than the fixation point.

At each spatial location and time, two linear monocular cell outputs are obtained by convolving the input image with a pair Gabor functions which are 90° out of phase. Here, we assume vertically oriented Gabor functions. Mathematically, we express the two outputs as the real and imaginary parts of a single complex number:

$$V_e(x,y,t) = \int\int \mathcal{N}(\xi,\eta|0,\mathbf{C}) e^{j\Omega_x \xi} \cdot U_e(x-\xi, y-\eta, t) d\xi d\eta \tag{2}$$

where $\mathcal{N}(\xi,\eta|0,\mathbf{C})$ is the 2D Gaussian kernel with mean 0 and covariance matrix C, which we assume to be a diagonal matrix with elements σ_x^2 and $\sigma_y^2 = (2\sigma_x)^2$ so that it is longer in the vertical than in the horizontal direction, and Ω_x is the spatial frequency of the Gabor filter.

The disparity energy is the squared modulus of the sum of the left monocular cell output and the right monocular cell output with a phase shift $\Delta\psi$

$$E_d(x,y,t,\Delta\psi) = |V_L(x,y,t) + e^{j\Delta\psi} V_R(x,y,t)|^2 \tag{3}$$

Qian (1994) has shown that the disparity energy can be approximated by

$$E_d(x,y,t,\Delta\psi) \approx (2 + 2\cos(\Delta\psi + \Omega_x d))|V_L(x,y,t)|^2 \tag{4}$$

We define a phase-tuned population of disparity energy neurons as a set of neurons with the same monocular RF centers, but tuned to different disparities by phase shifts varying between $-\pi$ and $+\pi$. Fig. 1c shows the construction of five neurons from a phase-tuned population, and how the population responses vary as the input stimulus disparity changes. The peak location in the response varies with the input disparity, and can be approximated by (Qian, 1994)

$$\Delta\psi_{\text{peak}} \approx -\Omega_x d \tag{5}$$

The standard disparity energy model considers only the input at the current instant in time, neglecting the effect of past inputs through the temporal dynamics of the neuron. This assumption is intuitively appealing here, since the changing disparity cue assumes that the visual system estimates the disparity first at each instant in time, and examines how that estimate changes over time. Nonetheless, because we are dealing with time varying inputs, it is possible that conclusions we draw from using the standard disparity energy model may be invalid because we have neglected the temporal dynamics of the neuron. Real disparity selective neurons in V1 have finite temporal kernels (DeAngelis, Ghose, Ohzawa, & Freeman, 1999; DeAngelis, Ohzawa, & Freeman, 1993; Ohzawa, DeAngelis, & Freeman, 1996).

To address this concern, we also examine the performance of a model that includes temporal dynamics by convolving the input image by a spatio-temporal kernel in computing the linear monocular cell outputs. Specifically, we replace Eq. (2) with

$$\begin{aligned} V_e(x,y,t) = \iiint \mathcal{N}(\xi,\eta|0,\mathbf{C}) e^{j\Omega_x \xi} \cdot \mathcal{G}(\delta|\alpha_1,\tau_1) \\ \cdot U_e(x-\xi, y-\eta, t-\delta) d\xi d\eta d\delta \end{aligned} \tag{6}$$

where $\mathcal{G}(t|\alpha,\tau)$ is a low pass temporal envelope whose shape is the probability density function of the Gamma distribution:

$$\mathcal{G}(t|\alpha,\tau) = \frac{1}{\Gamma(\alpha)\tau^\alpha} t^{\alpha-1} e^{-t/\tau_{u(t)}} \tag{7}$$

In this equation, α and τ are constants determining the skew and time constant, $\Gamma(\alpha)$ is the standard Gamma function for normalization, and $u(t)$ is the unit step function. The subscript "1" on the parameters in Eq. (6) indicate that these are associated with the first stage of the model. This spatio-temporal receptive field profile is the same as that used by (Chen et al., 2001; Qian, 1994) to model the responses of disparity tuned neurons to time varying input, except we have eliminated the sinusoidal modulation in the temporal receptive field. The sinusoidal modulation is introduced

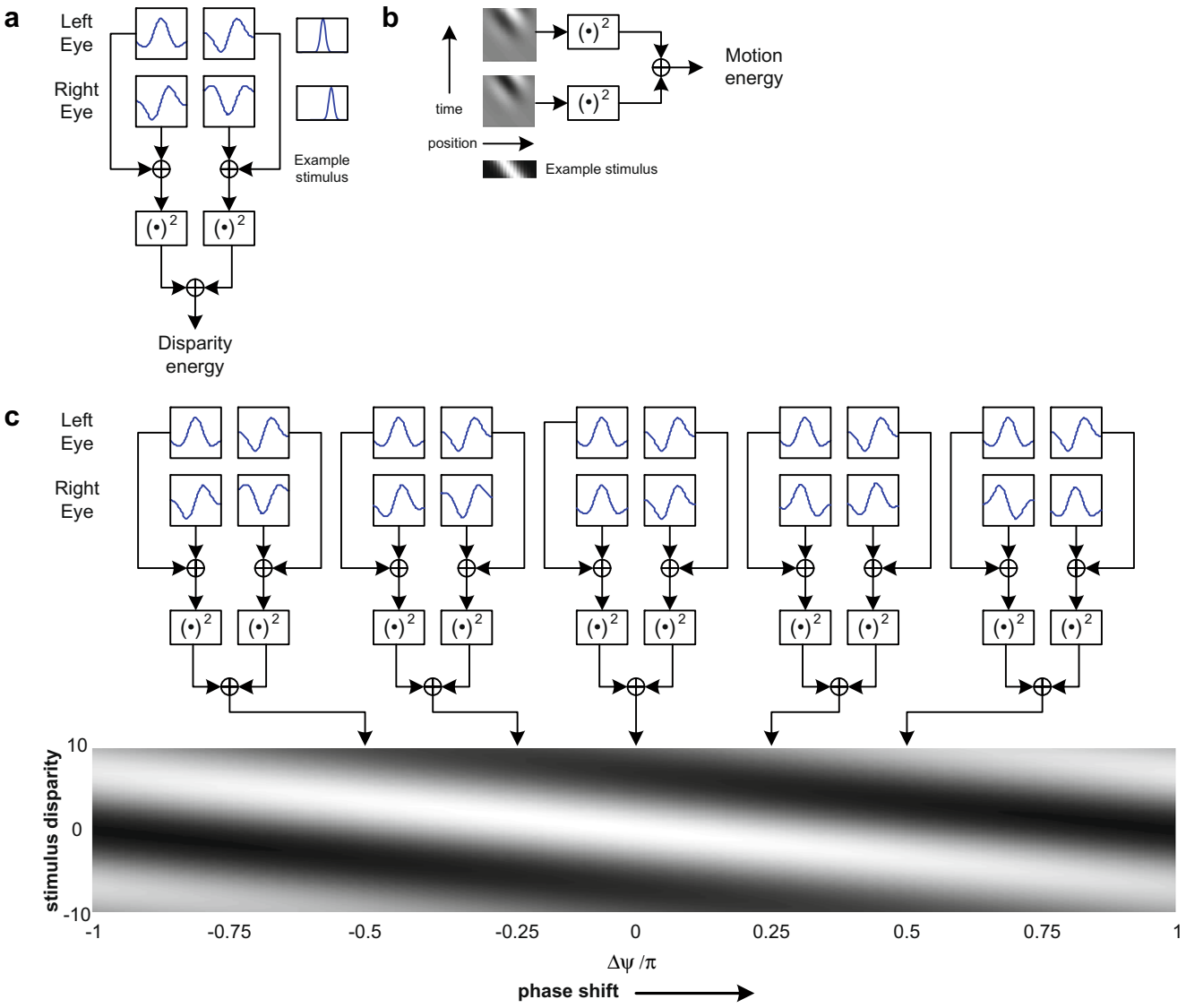

Fig. 1. Model architecture. (a) Disparity energy model. The four boxes on the left show the monocular RF weights as a function of spatial position. Each row corresponds to a different eye. The RFs in different columns differ in phase by 90°. A phase shift between the left-eye and right-eye neurons tunes this disparity energy neuron to negative disparities. (b) Motion energy model. The two boxes on the left show the receptive field weights as a function of space and time. The upper row refers to the current time while the lower one the previous time. The RF profiles differ by a 90° phase shift to form a quadrature pair whose outputs are squared and summed to derive the motion energy. (c) A phase-tuned population of disparity energy neurons. The lower image represents the outputs of the phase-tuned population as the input stimulus disparity changes. Each horizontal cross-section represents the responses of a phased tuned population of disparity energy neurons to an input stimulus with fixed disparity. Different cross-sections correspond to different disparities. The image intensity represents the response magnitude (white for higher response and black for lower response).

to in order to model joint disparity/motion selectivity. We do not use it here, since the first stage of a changing disparity model should be purely disparity selective. Intuitively, the temporal filtering we have introduced adds local temporal integration of the input to the response of the disparity energy neuron.

We also examine the effect of normalization, a common component of models of cortical neurons (Albrecht & Geisler, 1991; Carandini, Heeger, & Movshon, 1997; Heeger, 1992). In our experiments that included normalization, we divide the responses of all disparity energy neurons in the population by the average response taken globally across phase disparity and locally across space and time. The global average across disparity weights all neurons in the phase population equally. The average across space weights neural responses by a circularly symmetric Gaussian with variance σ_n^2. The average across time weights each neuron in the past according to the Gamma distribution envelope with parameters α_n and τ_n.

2.2. Motion energy model

The motion energy model (Adelson & Bergen, 1985; Watson & Ahumada, 1985) is based on the observation that a velocity of motion corresponds to an orientation in space–time. The motion energy model combines the outputs of two phase quadrature spatio-temporal filters that are tuned to detect this orientation by squaring and summing, as shown in Fig. 1b.

As with the disparity energy model, we can represent the filter outputs as the real and imaginary parts of a single complex number. The complex valued spatio-temporal convolution kernel is separable into the product of spatial and temporal functions. The spatial filter is usually modeled using a spatial Gabor filter. For simplicity and consistency with Fig. 1c and the following description of the CD energy model, we assume only a single spatial dimension. Defining $U(x, t)$ as the input image intensity over space and time, the output of the spatial Gabor filter is

$$V(x,t) = \int \mathcal{N}(\xi|0,\sigma_x^2) e^{j(\Omega_x \xi)} U(x-\xi,t) d\xi \quad (8)$$

where $\mathcal{N}(\xi|0,\sigma_x^2)$ is a 1D Gaussian with mean 0 and variance σ_x^2 and Ω_x is the spatial frequency of the Gabor filter. The kernel of the temporal can be described by the Gamma probability density function modulated by a complex exponential (Chen et al., 2001). The output of the temporal filter is

$$W(x,t) = \int h(\delta|\alpha,\tau,\Omega_t) V(x,t-\delta) d\delta \quad (9)$$

where

$$h(t|\alpha,\tau,\Omega_t) = \mathcal{G}(t|\alpha,\tau) e^{j\Omega_t t} \quad (10)$$

Here, $\mathcal{G}(t|\alpha,\tau)$ is the Gamma envelope given in Eq. (7) and Ω_t is the center frequency of the temporal filter.

If the image input is a sine wave grating whose spatial frequency matches the center frequency of the spatial Gabor (Ω_x), then the motion energy output is maximized when the velocity of the grating is

$$v^{pref} \approx -\frac{\Omega_t}{\Omega_x} \quad (11)$$

Since the spatial filter passes mainly frequency components around Ω_x, the velocity v^{pref} is often thought of as being the preferred velocity of the filter. However, strictly speaking the filter is not velocity tuned. For sine wave gratings, the velocity that leads to maximum response decreases with the spatial frequency of the grating. For more general inputs, the velocity that leads to maximum response changes with the spatial frequency content.

The opponent energy is often used to model tasks involving discrimination between two opposite directions of motion. The opponent energy is the difference between the outputs of two motion energy neurons tuned to the same spatial frequency but opposite directions of motion. The difference in direction tuning can be obtained by using temporal filters with opposite temporal frequencies (Adelson & Bergen, 1985). The opponent energy eliminates responses due to stationary input.

2.3. Changing disparity energy model

The changing disparity energy model is based on the observation that a changing disparity causes the location of the peak response in a phase-tuned population to change over time. In the same way that image motion leads to a space time orientation (Fig. 1b), changing disparity leads to a phase-time orientation (Fig. 1c). This similarity enables us to develop an energy model for changing disparity by cascading the disparity energy model with the motion energy model.

Assume that we have the outputs of a phase tuned disparity energy neuron population at a particular retinal location (x, y): $E_d(x, y, t, \Delta\psi)$. We extract an orientation in phase-time by using a phase-time filter that is the cascade of a phase filter with a temporal filter. The output of the phase filter is given by

$$N(x,y,t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\Delta\psi} E_d(x,y,t,\Delta\psi) d(\Delta\psi) \quad (12)$$

This is very similar to Eq. (8) with three key differences. First, the integration over phase ($\Delta\psi$) replaces the integration over space (x). Second, we eliminate the Gaussian envelope which was used to limit the spatial extent of the RF. The envelope is not needed here because phase is naturally limited between $-\pi$ and $+\pi$. Third, we replace the spatial frequency Ω_x by unity to match the natural phase periodicity in 2π. The temporal filter is exactly the same as that described previously:

$$Y(x,y,t) = \int h(\delta|\alpha_2,\tau_2,\Omega_t) N(x,y,t-\delta) d\delta \quad (13)$$

where $h(t|\alpha,\tau,\Omega_t)$ is given in Eq. (10). By analogy with the motion energy filter, this filter is tuned to a preferred phase velocity of $-\Omega_t$.

The changing disparity energy is defined as the squared magnitude of the output of the phase-time filter, $|Y(x,y,t)|^2$. As a final step, we perform a moderate amount of spatial pooling by convolving the output with a circularly symmetric 2D Gaussian in x and y with standard deviation $2\sigma_x$. Spatial pooling is commonly applied in energy models, in part to account for the larger receptive fields of complex cells in comparison with simple cells, and in part to improve performance (Fleet, Wagner, & Heeger, 1996; Heeger, 1987; Zhu & Qian, 1996).

When presented with a stimulus moving in depth, the peak of the response in the population shifts over time. If the stimulus is a sinusoidal grating with spatial frequency Ω_x, the shift is equal to the change in disparity times the spatial frequency Ω_x (Eq. (5)). Since the spatial filter passes frequencies around Ω_x, we will refer to

$$v_d^{pref} \approx \frac{\Omega_t}{\Omega_x} \quad (14)$$

as preferred rate of changing disparity. As with the motion energy model, we define the opponent energy to be the difference between the changing disparity energy for two opposite preferred rates of changing disparity by choosing temporal frequencies of opposite sign: Ω_t for a neuron tuned to approaching MID, i.e. increases in disparity, and $-\Omega_t$ for a neuron tuned to receding MID, i.e. decreases in disparity.

In our experiments, we will examine the performance of three versions of the CD energy model, which differ according to the temporal dynamics of the disparity energy stage. Model I, the simplest, assumes that the disparity energy is computed based only upon the current frame input. This is equivalent to the assumption that the temporal receptive field profiles of the disparity energy neurons are Dirac delta functions. Model II assumes that the disparity energy neurons have temporal receptive fields given by the probability density function of the Gamma distribution in Eq. (7). Model II reduces to Model I, when the time constant τ is zero. Model III adds the normalization step to Model II.

3. Methods

In our simulations, we discretize space and time into pixels and frames. One arcmin of visual angle corresponds to one pixel. One second corresponds to 120 frames.

Unless otherwise noted, we use the following model parameters. For the spatial receptive fields in the disparity energy stage, we use vertical orientations with center spatial frequency tuning of 3.75 cycles per degree, consistent with the range of 3–5 cycles per degree for macaque cortical cells serving the fovea (De Valois, Albrecht, & Thorell, 1982). We choose the spatial bandwidth to be 1.95 octaves, consistent with the range of 0.5–2.5 octaves for V1 neurons (Dayan & Abbott, 2001). The aspect ratio is two. In discrete space, this corresponds to $\Omega_x = 2\pi/16$ radians per pixel, $\sigma_x \approx 5$ pixels, and $\sigma_y \approx 10$ pixels.

For the experiments including the temporal integration into the disparity energy neurons, we use temporal filter parameters $\alpha_1 = 1$ and values of τ_1 ranging between 20 and 100 ms. In cases where the value is fixed, we use $\tau_1 = 30$ ms. For the experiments including normalization, the parameters range from $\sigma_n = \sigma_x$ to $\sigma_n = 5\sigma_x$ for the spatial averaging and $\tau_n = 20$ ms to $\tau_n = 100$ ms for the temporal averaging. In cases where these values are fixed, we use nominal values of $\sigma_n = 3\sigma_x$ and $\tau_n = 80$ ms.

For the temporal filters of the second stage, we choose the shape parameter $\alpha = 1$, the time constant $\tau = 51.8$ ms, and the center temporal frequency $\Omega_t = 2\pi(4\text{ Hz})$. The resulting filters have a bandwidth of 2.93 octaves, consistent with the peak frequency around 4 Hz and attenuation at around 10 Hz found in cortical cells (Dayan & Abbott, 2001). In discrete time, this corresponds to $\Omega_t = 2\pi/30$ radians per frame and $\tau = 6.22$ frames. With these parameters, the preferred changing disparity rate is 1.067 deg/s (0.533 pixels per frame). We compute the outputs of two changing disparity energy models tuned to approaching and receding stimuli, as well as their difference (the opponent energy).

We choose the parameters of the random dot stimuli for consistency with the experiments reported in (Brooks & Stone, 2004, 2006). Dot sizes are 3 arcmin in diameter (3 pixels), which is similar to the dot size 2.5 × 3.7 arcmin used in the experiments. The images cover 2.1° by 2.1° of visual space (128 × 128 pixels). To avoid boundary effects, we do not consider the responses of neurons whose receptive field centers closer than 0.27° ($16 \approx 3\sigma_x$ pixels) from the boundary. Dot densities vary between 1% and 50%. For the DRDS stimuli, the dot pattern changes at 120 Hz (every frame).

In replicating the psychophysical experiments, we use stimuli with depth trajectories that are both direct with identical but opposite speeds in the two monocular images or oblique with non-identical monocular speeds. Rates of change in disparity range between −2 and 2 deg/s (−1 and 1 pixel/frame) in units of 0.2 deg/s, which covers the range of stimuli (−1 to 1 deg/s) used in the experiments.

In order to compute tuning curves, we average the changing disparity energy over 100 ms ($13 \approx 2\tau$ frames) after the binocular images coincide at fixation. For each changing disparity rate, we collect the time averaged responses from a 17 × 17 rectangular array of 289 neurons whose RF centers are spaced by 5 arcmin ($5 \approx \sigma_x$ pixels) from each other, and at least 16 arcmin ($16 \approx 3\sigma_x$ pixels) from the image boundary over 10 trials. We fit the 2890 data points to the Gamma distribution using Maximum Likelihood Estimation (MLE). We define the tuning curve to be the means of the estimated distributions as a function of the changing disparity rate. The opponent energy tuning curve is the difference between the means of changing disparity energy units tuned to approaching and receding stimuli.

We compute a putative speed discrimination threshold difference as the ratio of the slopes of the opponent energy curves near the origin. This ratio is based upon the assumption that discrimination is done by determining whether the opponent energy exceeds a threshold, *E*th. For rates of changing disparity near zero, the opponent energy is approximately linear with the changing disparity. For a given threshold on the opponent energy, the corresponding changing disparity rate discrimination threshold is inversely proportional to the slope. Thus, the speed discrimination threshold ratio is the ratio of the slopes a:

$$\frac{v^{\text{th}}_{\text{DRDS}}}{v^{\text{th}}_{\text{RDS}}} = \frac{a_{\text{RDS}}}{a_{\text{DRDS}}} \tag{15}$$

We compute the slopes by applying a linear fit to the opponent energy curves in the range from $-\frac{1}{2}v_{\text{d}}^{pref}$ to $\frac{1}{2}v_{\text{d}}^{pref}$, where the opponent energy curves are approximately linear.

4. Results

In the first set of experiments described here, we examine the behavior of Model I, the simplest changing disparity energy model. This case is the most straightforward extension of the disparity energy model to changing disparity model, and thus serves to elucidate the key characteristics of the model. We examine the tuning characteristics of a changing disparity energy neuron, as well as the difference in its responses to RDS and DRDS stimuli. We show that the properties of the model are robust to changes in the tuning parameters of the model, as well as the stimulus trajectory.

In the second set of experiments, we examine Model II, which incorporates local temporal integration into the disparity energy neurons. This makes the model more physiologically plausible, since real V1 disparity selective neurons have finite temporal kernels (DeAngelis et al., 1999, 1993; Ohzawa et al., 1996). However, we find that this simple extension has some drawbacks, which are ameliorated by the addition of normalization in Model III. In a final experiment, we examine the effect of varying the dot density in the stimuli.

4.1. Model I

4.1.1. Tuning characteristics

The tuning characteristic of a Model I CD energy neuron tuned to approaching motion in response to RDS stimuli is illustrated in Fig. 2a. The CD energy model exhibits preference to the motion in depth but is relatively less sensitive to the lateral motion. The maximum energy response appears along the lines $v_d = v_L - v_R = 1$ deg/s, as evident in the cross-sectional plot for the case $v_L + v_R = 0$ deg/s shown in Fig. 2b. On the other hand, cross-sections in the orthogonal direction ($v_d = v_L - v_R = 1$ deg/s) are much flatter, as shown in Fig. 2c.

4.1.2. Comparison between RDS and DRDS stimuli

By applying the Gamma distribution fitting to the 2890 data points, we estimate the means of the changing disparity energies at each stimulus changing disparity rate for both RDS and DRDS. The estimated means are plotted in Fig. 3a and b where the error bars indicate the upper and lower bound of the 95% confidence intervals. The opponent energies, which are the difference between the estimated means of the approaching and receding changing disparity energies, are plotted in Fig. 3c.

Note that the slope of RDS curve for changing disparity rates near zero is larger than DRDS curve. The speed discrimination threshold ratio for DRDS versus RDS stimuli obtained in these experiments is about 2.0, which is in the ratio range from 1.4 to 2.4 in Brooks' experiments, although a bit larger than their average value of 1.7 (Brooks & Stone, 2004).

The difference in speed discrimination thresholds for DRDS and RDS is maintained for other choices for model parameters that are consistent with cortical physiology. Plots of the speed discrimination threshold ratios for different choices of spatial and temporal frequencies in the model, shown in Fig. 4a and b, are relatively constant. For the data in Fig. 4a, we vary the spatial filter frequency between 3 and 5 cycles per degree ($2\pi/20, 2\pi/18, 2\pi/16, 2\pi/14$ and $2\pi/12$ radians per pixel), the range observed for cortical cells serving the macaque fovea (De Valois et al., 1982). The spatial bandwidth and aspect ratio are held constant at 1.95 octaves and 2. The temporal tuning parameters are unchanged. The threshold ratio varies between 1.8 and 2.1. For the data in Fig. 4b, we vary the temporal filter frequency in a range from 3 Hz to 6 Hz ($2\pi/40, 2\pi/38, 2\pi/36, 2\pi/34, 2\pi/32, 2\pi/30, 2\pi/28, 2\pi/26, 2\pi/24, 2\pi/22, 2\pi/20$ radians per frame). The relative temporal bandwidth is held constant. The spatial tuning parameters are unchanged. The threshold ratio varies between 2.0 and 2.1, indicating that the ratio presented here is robust to model parameters. In both cases, the range of threshold ratios is within the range of 1.4–2.4 reported by Brooks and Stone (2004).

4.1.3. The effect of varying 3D trajectory

The tuning characteristics of the Model I CD energy neuron shown in Fig. 2a–c indicate that the changing disparity energy is relatively insensitive to lateral motion. This suggests that the per-

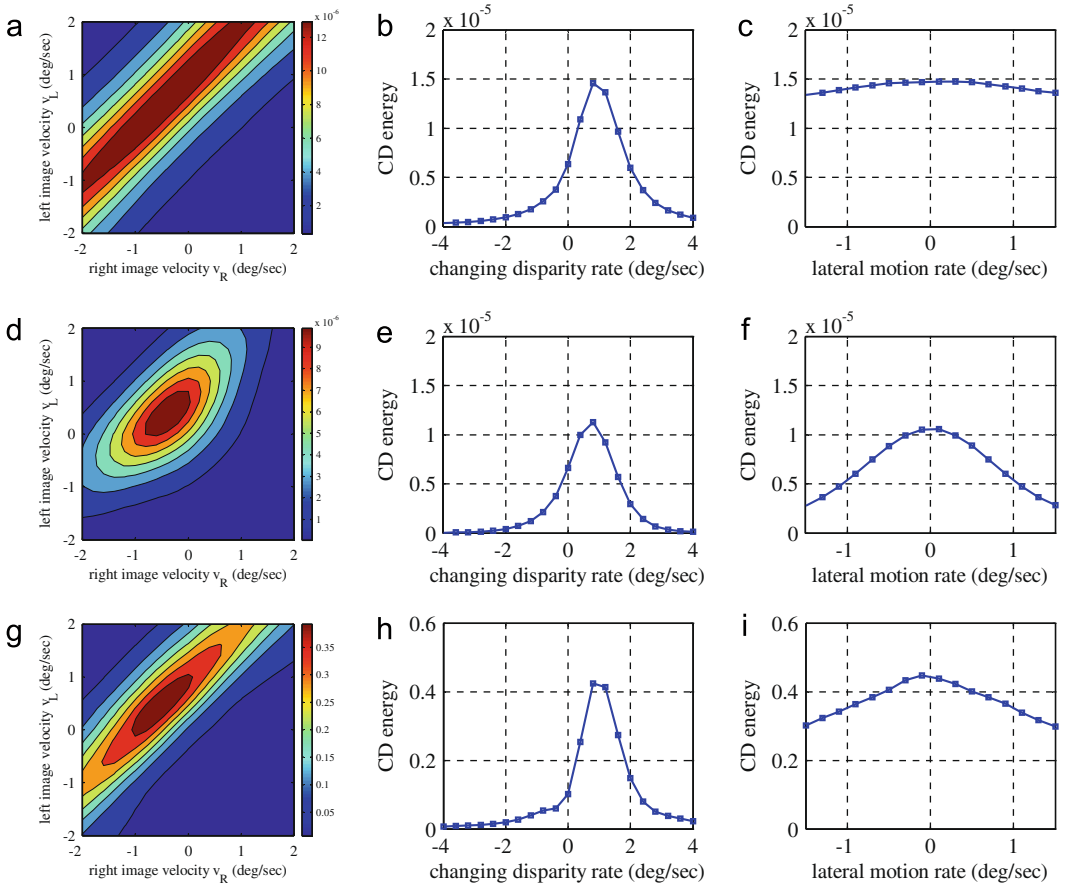

Fig. 2. Tuning characteristics of the CD energy neuron. 1st row: Model I; 2nd row: Model II; 3rd row: Model III. (a) Contours of the energy generated by the approaching changing disparity neuron in response to RDS stimuli by jointly varying v_L and v_R. Each diagonal cross-section from bottom-left to upper-right corresponds to a set of inputs of the same changing disparity rate but various lateral motion rates. Each diagonal cross-section from bottom-right to upper-left corresponds to a set of inputs of the same lateral motion rate but various changing disparity rates. (b) Cross-section of (a) along the line $v_L + v_R = 0$ deg/s, versus the changing disparity rate v_d. The maximum energy response appears when $v_d = v_L - v_R = 1$ deg/s. (c) Cross-section of (a) along the peak line $v_d = 1$ deg/s, versus the lateral motion rate. (d–f) Tuning characteristics of Model II CD energy neuron, shown in a similar format as in (a–c). All stimuli and model parameters are the same as those for (a–c), except that a temporal dynamic of the disparity energy neuron is incorporated. The roll off in (f) is much sharper than in (c). (g–i) Tuning characteristics of Model III CD energy neuron, shown in a similar format as in (a–c). All stimuli and model parameters are the same as those for (a–c), except that both temporal dynamic and normalization are included. The tuning characteristics are more similar to (a–c) than those shown in (d–f). Only one trial of data (289 data points) is used in this figure.

formance of the model outputs on the stereomotion speed discrimination task for stimuli with oblique trajectories will be similar to the outputs for stimuli with direct trajectories.

To verify this, we perform simulations where the model parameters are exactly the same as the standard settings, but with lateral motion added to the input images. Five types of stimuli are used, as Fig. 8 in (Brooks & Stone (2004)). Direct stimuli (D) feature identical but opposite monocular speeds, while oblique stimuli feature a ratio of monocular velocities of −1:2 (Hit L), −2:1 (Hit R), 1:2 (Miss L) or −2:−1 (Miss R). Hit stimuli have opposite monocular directions. Miss stimuli have the same monocular directions. The different trajectories correspond to different 1D cross-sections of the (v_L, v_R) space, as illustrated in Fig. 5e.

Fig. 5a and b plots the changing disparity opponent energy for the different stimulus trajectories as a function of the changing disparity rate, which is the projection of the distance along the cross-section in Fig. 5e onto the line $v_L = -v_R$. We find that the curves for different trajectories do not vary for either RDS or DRDS stimuli. In particular, the slope of the curve near the origin, which is used to determine v_d^{th}, is nearly the same in all cases. Consequently, the speed discrimination threshold ratios for the five trajectories are all 2.0. This is consistent with the idea that the changing disparity energy model is relatively insensitive to fronto-parallel motion. The result is also consistent with Brooks' observation that the direct or oblique stimuli show no significant difference in stereomotion speed discrimination thresholds (Brooks & Stone, 2004).

The slight drop in the response for the Miss stimuli at larger input changing disparity rates is due to the fact that the model is not totally invariant to lateral motion, as shown in Fig. 2c. The cross-sections for the Miss stimuli pass through the upper-right and lower-left regions in Fig. 2a or Fig. 5e.

4.2. Models II and III

4.2.1. Model II: adding temporal integration

Fig. 2d–f shows the tuning characteristics of a Model II CD energy neuron that has temporal dynamics with time constant $\tau_1 = 30$ ms and is tuned to approaching motion. Compared with Model I neuron (Fig. 2a–c), the Model II neuron exhibits a preference to the same motion in depth, but is less selective and has an increased sensitivity to lateral motion. The cross-sectional plot for $v_L + v_R = 0$ deg/s in Fig. 2e shows that the maximum energy response still appears along the line where $v_d = v_L - v_R = 1$ deg/s. Fig. 2f plots the cross-section of the tuning characteristic that passes through the peak line in the orthogonal direction. In comparison with the Model I neuron (Fig. 2c), the roll off is much sharper. Stimuli with higher lateral velocities have faster temporal variations, which are reduced by the temporal integration

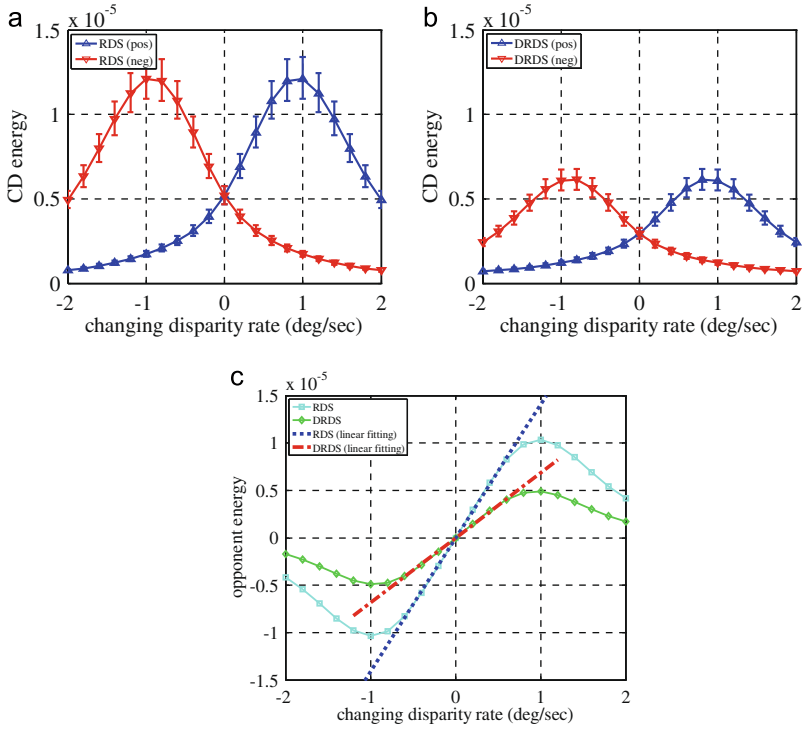

Fig. 3. Energy responses of the Model I CD energy neuron at each stimulus changing disparity rate for both RDS and DRDS. The estimated means of energies are plotted on the curves and the error bars indicate the upper and lower bound of the 95% confidence intervals obtained by Gamma distribution fitting. (a) Energy responses for RDS. (b) Energy responses for DRDS. In both (a) and (b), blue curves represent the responses of neurons tuned to approaching MID (positive changing disparity), while red curves represent the responses of neurons tuned to receding MID (negative changing disparity). (c) The opponent energies for both RDS and DRDS, represented by the cyan and green curves respectively. Linear fitting is applied to approximate the slope near the origin (blue dotted line for RDS and red dash-dot line for DRDS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

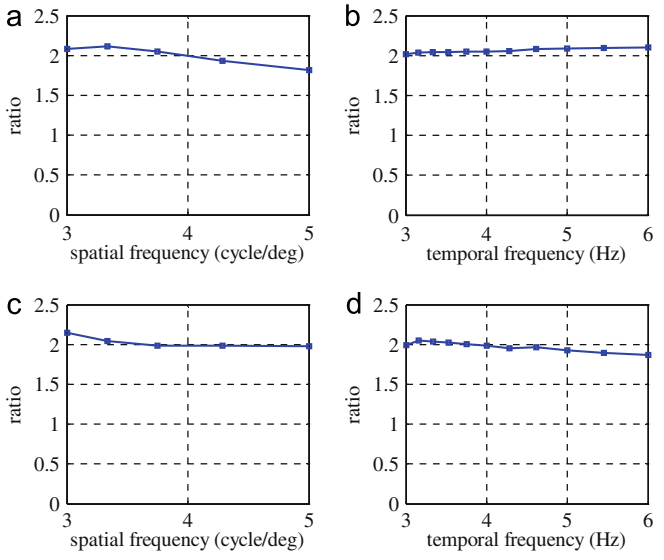

Fig. 4. Stereomotion speed discrimination threshold ratio of DRDS versus RDS, tested under different model parameters. (a) Results for Model I tuned to various spatial filter frequencies ranging from $2\pi/20$ to $2\pi/12$ rad/pixel to cover the typical cortical cells' spatial tuning of 3–5 cycle/deg. The ratio varies between 1.8 and 2.1. (b) Results for Model I tuned to various temporal frequencies ranging from $2\pi/40$ to $2\pi/20$ rad/frame to cover the typical cortical cells' temporal tuning of a few Hertz. The ratio varies between 2 and 2.1. (c,d) Analogous results for Model III. All stimuli and model parameters are the same as those for (a,b), except that both temporal dynamic and normalization are included. The threshold ratio of Model III CD energy neuron exhibits a similar invariance to changes in spatial frequency and temporal frequency.

introduced by Model II. The blue curve in Fig. 6a plots the speed discrimination threshold ratio as a function of the time constant of the temporal dynamics. The ratio increases rapidly, quickly exceeding the range of ratios observed in psychophysical experiments (Brooks & Stone, 2004). Because this model exhibits such high sensitivity and results in predictions at odds with the experimental results, we do not consider it in the remainder of the experiments.

4.2.2. Model III: adding temporal integration and normalization

The rapid increase in the threshold ratio with time constant can be offset by including normalization (Model II), where the spatial and temporal smoothing parameters used in computing the normalization factor are $\sigma_n = 3\sigma_x$ and $\tau_n = 80$ ms. Although the threshold ratio still increases, it remains within the range of values observed in psychophysical experiments (Brooks & Stone, 2004) for all time constants between 0 and 100 ms, as shown by the red curve in Fig. 6a. Increasing the amount of temporal and/or spatial averaging of the normalization factor increases the speed discrimination threshold ratio. Fig. 6b plots the threshold ratio as a function of σ_n as τ_n varies, where the amount of temporal integration by the disparity energy neuron is fixed at $\tau_1 = 30$ ms. The ratios remain within the range of values observed psychophysically over a wide range of averaging parameters.

Fig. 2g–i shows the tuning characteristics of the Model III CD energy neuron that includes both temporal integration and normalization ($\tau_1 = 30$ ms, $\sigma_n = 3\sigma_x$, $\tau_n = 80$ ms). The tuning characteristics of the Model III neuron are more similar to those of the Model I neuron (Fig. 2a–c) than are the tuning characteristics of the Model II neuron (Fig. 2d–f). In particular, the roll off for lateral motion (Fig. 2i) is much shallower (compared with Fig. 2f), although not as shallow as that for Model I (Fig. 2c). Because the tuning curves are so similar, it should not be surprising that the threshold ratio of Model III exhibits a similar invariance to changes in spatial

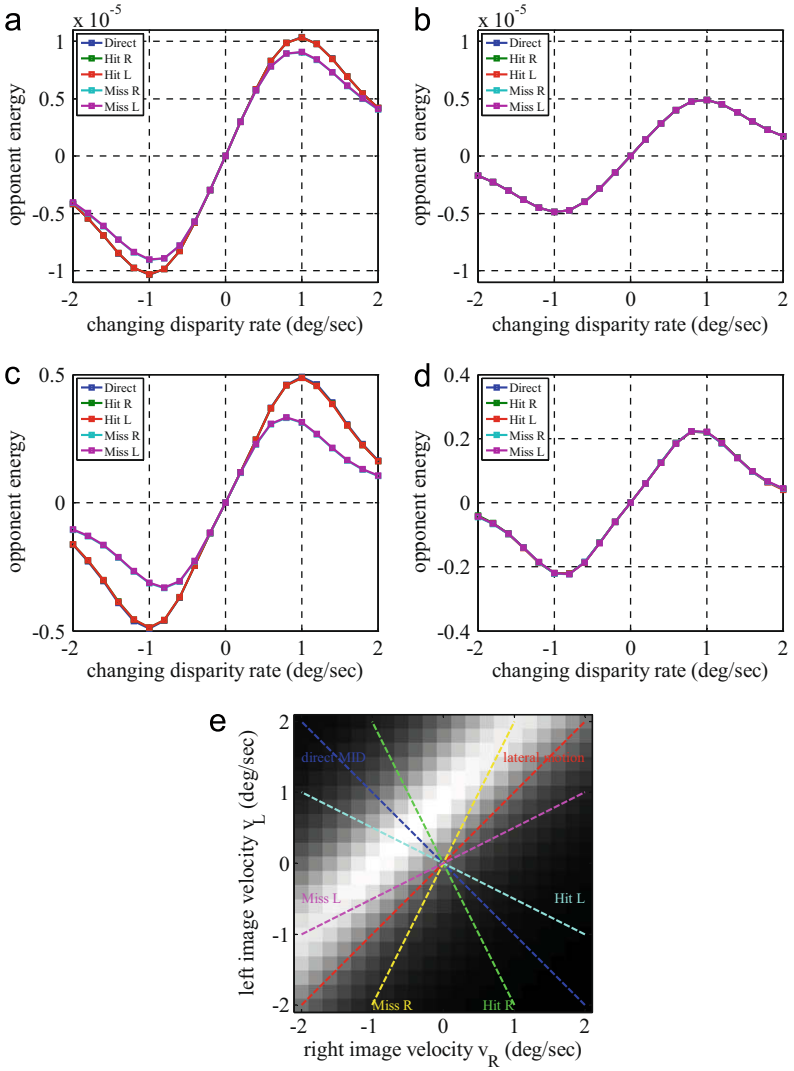

Fig. 5. Opponent energy outputs of Model I CD energy neuron tested using stimuli of direct or oblique trajectories. The changing disparity rate is the projection of the distance along the cross-section onto the line $v_L = -v_R$ in (e). (a) Opponent energies for RDS. (b) Opponent energies for DRDS. In both cases, the curves for various trajectories almost coincide. (c,d) The analogous opponent energies for Model III. All stimuli and model parameters are the same as those for (a,b), except that both temporal dynamic and normalization are included. The slope of the opponent CD energy curve near the origin is also unchanged with stimulus trajectory. (e) Different depth trajectories correspond to different 1D cross-sections of the (v_L, v_R) space. Take the approaching neurons as an example, six cross-sections of Fig. 2a are illustrated here: direct motion in depth ($v_L = -v_R$), pure lateral motion ($v_L = v_R$), Hit R ($v_L = -2v_R$), Hit L ($v_R = -2v_L$), Miss R ($v_L = 2v_R$) and Miss L ($v_R = 2v_L$).

frequency (Fig. 4c) and temporal frequency (Fig. 4d) as the Model I neuron (Fig. 4a and b). In addition, the slope of the opponent CD energy curve near the origin is also unchanged with stimulus trajectory (Fig. 5c and d). For the DRDS stimulus, the opponent CD energy curves are essentially identical for the different trajectories. For the RDS stimulus, the positive peak and negative trough in the opponent CD energy for the Miss trajectories occur at lower velocities (~0.8 deg/s) and have lower magnitude relative to the peak/trough for Hit and Direct trajectories than observed in Model I. This is due to the faster roll off of the tuning characteristic for lateral motions.

4.2.3. The effect of varying dot density

One of the advantages of building an explicit model that operates directly on binocular image sequences is that we can use the model to make predictions about the effect of varying other stimulus parameters. In the psychophysical experiments discussed above (Brooks & Stone, 2004), the dot density was held constant at 50%. In this final set of experiments, we examine the effect of changing the dot density on the threshold ratio. Except for dot density, other settings (e.g., model parameters, stimulus contrast and velocity) are kept the same as in our previous experiments. We consider only Model III.

Fig. 7 shows the output of the CD energy neurons tuned to approaching or receding stimuli as a function of stimulus velocity for inputs with different temporal coherence (RDS versus DRDS) and dot density (50% and 1%). At lower dot density, the RDS and DRDS curves increase in magnitude. The fewer dots in the stimulus decrease the magnitude of the raw disparity energy, but this appears to be offset by a relatively larger decrease in the normalization factor. Estimating the discrimination threshold by the slope of the opponent energy, we find that the threshold ratio increases to 3.1 at 1% density compared with 2.0 at 50% density. The relationship between the threshold ratio and the dot density depends upon the amount of spatial and temporal averaging used in computing the normalization factor, with more dependency upon the amount of spatial averaging. Fig. 6c shows the variation of the threshold ratio as the dot density varies from 1% to 50% for differing values of the spatial averaging parameter σ_n. For $\sigma_n > 2\sigma_x$, the threshold ratio increases with dot density. For smaller σ_n, the threshold ratio de-

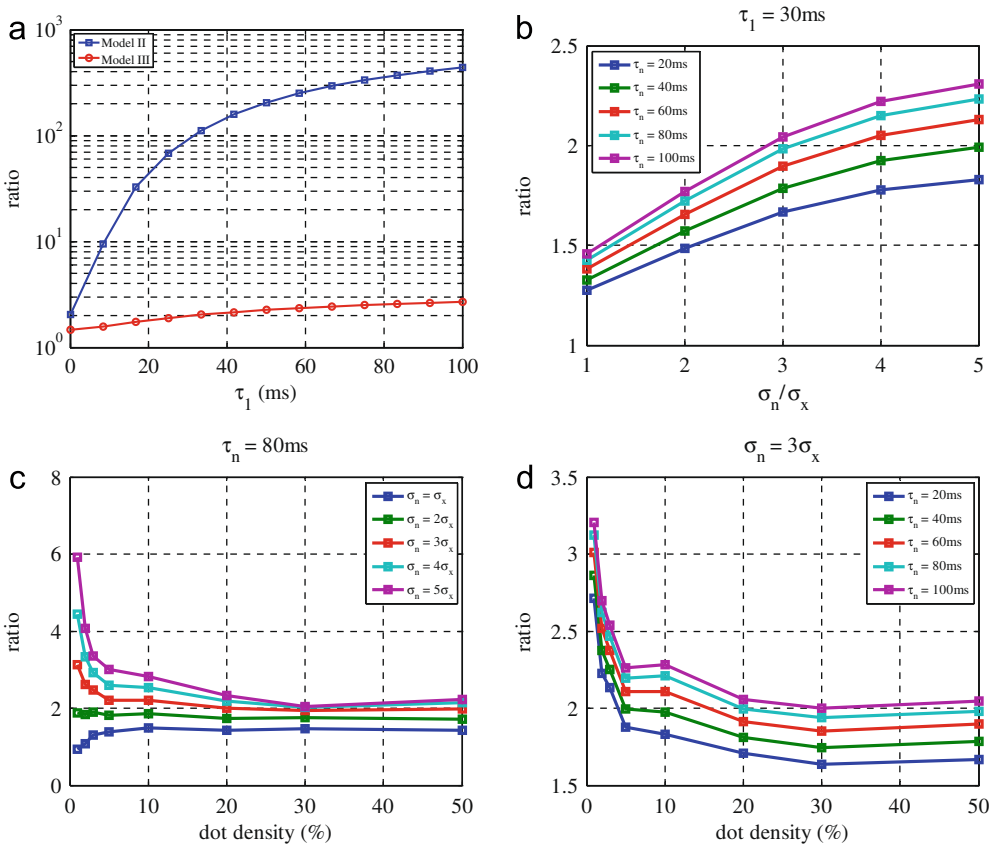

Fig. 6. The dependency of stereomotion speed discrimination threshold ratio on the model structure, model parameters and stimuli parameters. (a) The threshold ratios of Model II and Model III, as functions of the time constant (τ_1) of the temporal dynamics at the disparity stage. For Model III (red curve), the temporal averaging parameter for normalization factor is fixed at τ_n = 80 ms. (b) The threshold ratio of Model III plotted as a function of the extent of spatial averaging (σ_n) used in computing the normalization factor, and for different amounts of temporal averaging (τ_n). (c) The threshold ratio of Model III varies with dot density. The variation depends upon the extent of spatial averaging (σ_n). Temporal averaging τ_n is fixed at 80 ms. (d) The threshold ratio of Model III as a function of dot density for different amounts of temporal averaging (τ_n). The extent of spatial averaging σ_n is fixed at $3\sigma_x$. In both (c) and (d), τ_1 is fixed at 30 ms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

creases with dot density. On the other hand, the qualitative dependency of the threshold ratio on dot density does not vary much with the amount of temporal averaging τ_n. As shown in Fig. 6d, increasing τ_n does not change the shape, but simply results in an upward shift in the ratio.

5. Discussion

We have presented a biologically plausible changing disparity energy model whose outputs are consistent with prior psychophysical experiments showing higher stereomotion speed discrimination thresholds for DRDS than for RDS stimuli (Brooks & Stone, 2004). Threshold ratios obtained by varying the model parameters over ranges consistent with cortical physiology are all within the range of threshold ratios reported experimentally. Previously, this higher speed discrimination threshold for DRDS than for RDS has been taken as support for the use of an IOVD mechanism in the biological perception of stereomotion. Our finding that this threshold ratio is higher even in our changing disparity energy model suggests that the higher threshold by itself does not necessarily provide strong support for the use of an IOVD mechanism. Adding local temporal integration into the receptive field profiles of the disparity energy neuron stage increases the threshold ratios beyond those observed psychophysically. However, incorporating normalization keeps the threshold ratios within the psychophysical range. A testable prediction generated by our model is that threshold ratios may vary as the dot density used in the RDS and DRDS stimuli decreases. In our model, the variation in the threshold indicates the extent of the spatio-temporal window used in computing the normalizing factor.

5.1. Higher speed discrimination threshold

We can give an intuitive explanation for the higher threshold observed in our model by examining the complex-valued output of the phase filter, $N(x, y, t)$, which is the input to the temporal filter. This output rotates in the complex plane as the image disparity changes over time. The temporal filter in the second stage is selective to a particular direction and speed of rotation. Since the RDS stimulus is temporally coherent, the magnitude of $N(x, y, t)$ remains fairly constant as it rotates, providing a consistent input to the temporal filter. On the other hand, since DRDS is temporally incoherent, dots may appear, disappear, and reappear frequently within the spatial receptive field. This causes the magnitude of $N(x, y, t)$ to change much more rapidly and unpredictably.

The average energy in the phase filter output for both the RDS and DRDS stimuli is approximately the same. However, the amplitude modulation for the DRDS stimuli means that this energy is spread over a much wider frequency range than for the RDS stimuli. Fig. 8a–d compares the power spectral density of $N(x, y, t)$ for RDS and DRDS inputs at four different rates of changing disparity. We estimate the power spectral density using maximum likelihood estimation assuming a Gamma distribution. The data is calculated by applying a 32-point FFT with zero padding to 13 frames ob-

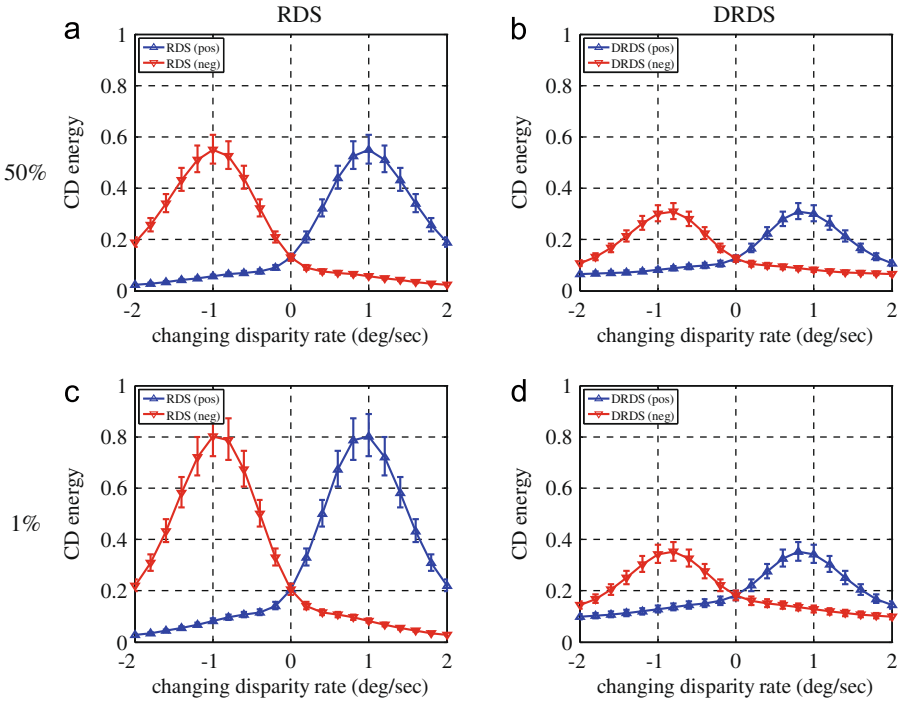

Fig. 7. Responses of Model III CD energy neurons tuned to approaching MID (blue curves) and receding MID (red curves) for different combinations of stimulus type: RDS/DRDS (left/right column), and dot density: 50%/1% (upper/lower panel). Error bars indicate the upper and lower bound of the 95% confidence intervals obtained by Gamma distribution fitting. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tained at each of the neurons on the 17 × 17 pixel grid used in our other experiments over 10 trials. The spread in power spectral density reduces the changing disparity energy output for DRDS stimuli, since much of the energy in $N(x, y, t)$ is blocked by the bandpass temporal filter of the second stage, whose frequency response magnitude is shown in Fig. 8e.

5.2. Temporal integration

At first glance, we might expect that the threshold differences between the RDS and DRDS stimuli to be reduced by the addition of the temporal integration. Intuitively, by smoothing out the rapid temporal variations in the DRDS stimulus, the temporal integration makes it appear more like an RDS stimulus. In fact, our experiments reveal that the opposite is true. Adding temporal integration increases the threshold ratio, and even a moderate amount of integration increases the threshold ratio beyond those observed experimentally. While it is true that local temporal integration smoothes out the rapid temporal variations in the DRDS stimulus, this also decreases the energy in the signal being passed to the second stage. This reduces the CD energy, which in turn reduces the slope of the opponent CD energy, thus leading to dramatically increased threshold ratios.

5.3. Normalization

Our results with Model III show that including normalization dramatically reduces the increase in the threshold ratio, keeping it within the bounds observed psychophysically over a range of biologically plausible time constants (0–100 ms) in the disparity energy stage. Normalization is a common component of models of cortical neurons (Albrecht & Geisler, 1991; Carandini et al., 1997; Heeger, 1992). The normalization factor we use pools information not only over disparity, but also over a spatial neighborhood of the cell, consistent with evidence that surround

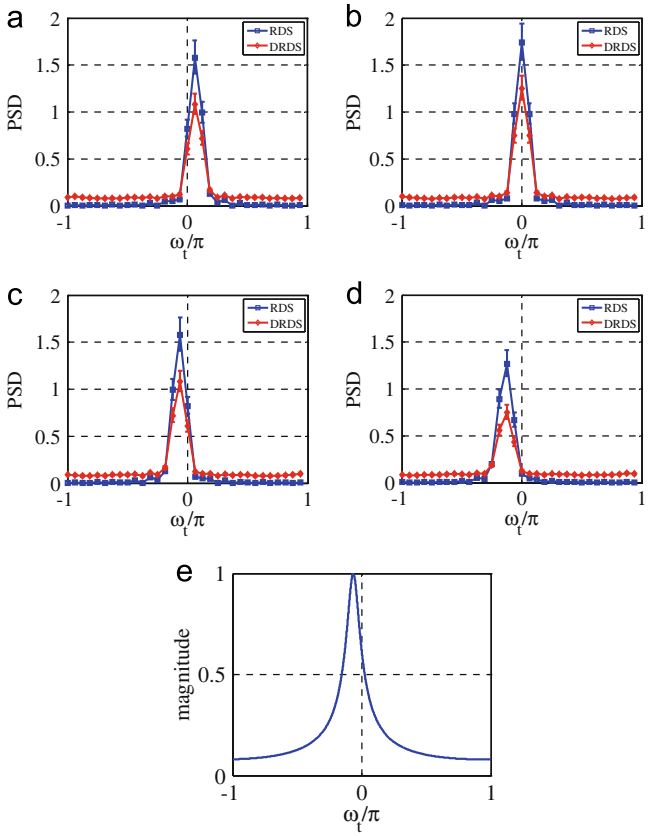

Fig. 8. Explanation of the higher threshold ratio. (a–d) Power spectrum analysis of the phase filter output $N(x, y, t)$, by simulating Model III CD energy neuron. Results of four stimuli moving at changing disparity rate of (a) −1, (b) 0, (c) 1, (d) 2 deg/s are shown here as examples. Error bars indicate the upper and lower bound of the 95% confidence intervals obtained by Gamma distribution fitting. (e) The frequency response magnitude of the temporal filter in the second stage shows that the filter blocks all but a narrow range of frequencies.

suppression is divisive (Muller, Metha, Krauskopf, & Lennie, 2003). Functionally, it has also been shown to improve estimation and validation of disparity estimates (Tsang & Shi, 2007, 2008). Intuitively, the dramatic reduction in response to the DRDS stimulus due to temporal averaging described above will be experienced by all neurons in a population of neurons tuned to different disparities and different neighboring spatial locations. Thus, we can compensate for this overall reduction in a consistent way by normalizing by the local time average across these neurons.

5.4. Changing dot density

A testable prediction suggested by our model is the relationship between the speed discrimination threshold ratio and the dot density, and its strong qualitative dependence upon the size of the spatial neighborhood used in computing the normalization factor. It has been reported that the diameter of the suppressive surround in V1 neurons typically varies from about twice the size over which they summate to about 13 times that (Levitt & Lund, 2002). In our model, when the size of the normalization neighborhood is twice that of the surround ($\sigma_n = 2\sigma_x$), there is no variation in the threshold ratio with dot density. However, for larger normalization neighborhoods, the threshold ratio increases as dot density decreases. Thus, if our model is representative of the actual computations being performed by the brain, then the dependency is indicative of the size of the spatial surrounds of the disparity selective neurons involved in the computation of the changing disparity cue.

5.5. Cascade model structure

The cascade of a disparity selective stage and a subsequent motion-like processing stage and more importantly, the detailed connectivity between them is another strong prediction of our model. Although we cannot yet find evidence in the literature for the proposed pattern of connectivity, there is evidence that stereomotion processing is performed in a cascade structure as suggested by our model. Although areas V1 and MT do contain neurons selective for disparity or motion, electrophysiological studies have not found strong evidence for stereomotion selective cells in these areas (Maunsell & Van Essen, 1983b). However, electrophysiological studies of higher areas have found neurons that appear to be selective for stereomotion in both cat (Akase, Inokawa, & Toyama, 1998; Cynader & Regan, 1978, 1982; Toyama, Komatsu, Kasai, Fujii, & Umetani, 1985) and monkey (Poggio & Talbot, 1981; Zeki, 1974). Recent fMRI studies have found evidence for stereomotion selectivity in human MT+ (Rokers, Cormack, & Huk, 2009) and areas anterior to it (Likova & Tyler, 2007). The area MT+ in human is often thought of as being analogous to both areas MT and MST in macaque, suggesting that perhaps area MST or an area anterior to it may be involved in stereomotion processing. Area MT has strong projections to area MST (Desimone & Ungerleider, 1986; Maunsell & Van Essen, 1983a), suggesting that plausible locations for the first disparity selective processing stage of our model would be in areas V1 or MT, and plausible locations for the second stage of processing would be in area MST or further anterior. Consistent with this idea, Likova and Tyler (2007) suggest that "the stereomotion effect is 'computed' from the temporal modulations of disparity specific signals" from disparity selective populations in the primate motion area. Rokers et al. (2009) suggest that area "V1 likely serves to extract the building blocks for three-dimensional motion computations that are ultimately performed in extrastriate areas." Evidence from a study using event related potentials also supports the idea of a bottom-up computation for motion in depth (Lamberty, Gobbelé, Schoth, Buchner, & Waberski, 2008).

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A – Optics Image Science and Vision, 2*(2), 284–299.

Akase, E., Inokawa, H., & Toyama, K. (1998). Neuronal responsiveness to three-dimensional motion in cat posteromedial lateral suprasylvian cortex. *Experimental Brain Research, 122*(2), 214–226.

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience, 7*(6), 531–546.

Anzai, A., Ohzawa, I., & Freeman, R. D. (2001). Joint-encoding of motion and depth by visual cortical neurons: Neural basis of the Pulfrich effect. *Nature Neuroscience, 4*(5), 513–518.

Brooks, K. R. (2002). Interocular velocity difference contributes to stereomotion speed perception. *Journal of Vision, 2*(3), 218–231.

Brooks, K. R., & Stone, L. S. (2004). Stereomotion speed perception: Contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *Journal of Vision, 4*(12), 1061–1079.

Brooks, K. R., & Stone, L. S. (2006). Stereomotion suppression and the perception of speed: Accuracy and precision as a function of 3D trajectory. *Journal of Vision, 6*(11), 1214–1223.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17*(21), 8621–8644.

Chen, Y., Wang, Y., & Qian, N. (2001). Modeling V1 disparity tuning to time-varying stimuli. *Journal of Neurophysiology, 86*(1), 143–155.

Cynader, M., & Regan, D. (1978). Neurones in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *Journal of Physiology, 274*, 549–569.

Cynader, M., & Regan, D. (1982). Neurons in cat visual cortex tuned to the direction of motion in depth: Effect of positional disparity. *Vision Research, 22*(8), 967–982.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience. Computational and mathematical modeling of neural systems. Computational neuroscience*. Cambridge, MA: MIT Press [pp. 60–74].

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*(5), 545–559.

DeAngelis, G. C., Ghose, G. M., Ohzawa, I., & Freeman, R. D. (1999). Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *Journal of Neuroscience, 19*(10), 4046–4064.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology, 69*(4), 1091–1117.

Desimone, R., & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *Journal of Comparative Neurology, 248*(2), 164–189.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research, 36*(12), 1839–1857.

Guo, X., & Shi, B. (2008). A two stage energy model exhibiting selectivity to changing disparity. In: Advances in neural networks – ISNN (2008) [pp. 47–54].

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A – Optics Image Science and Vision, 4*(8), 1455–1471.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*(2), 181–197.

Lamberty, K., Gobbelé, R., Schoth, F., Buchner, H., & Waberski, T. D. (2008). The temporal pattern of motion in depth perception derived from ERPs in humans. *Neuroscience Letters, 439*(2), 198–202.

Levitt, J. B., & Lund, J. S. (2002). The spatial extent over which neurons in macaque striate cortex pool visual signals. *Visual Neuroscience, 19*(4), 439–452.

Likova, L. T., & Tyler, C. W. (2007). Stereomotion processing in the human occipital cortex. *Neuroimage, 38*(2), 293–305.

Maunsell, J. H., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology, 49*(5), 1127–1147.

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*(5), 1148–1167.

Muller, J. R., Metha, A. B., Krauskopf, J., & Lennie, P. (2003). Local signals from beyond the receptive fields of striate cortical neurons. *Journal of Neurophysiology, 90*(2), 822–831.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science, 249*(4972), 1037–1041.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1996). Encoding of binocular disparity by simple cells in the cat's visual cortex. *Journal of Neurophysiology, 75*(5), 1779–1805.

Pack, C. C., Born, R. T., & Livingstone, M. S. (2003). Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. *Neuron, 37*(3), 525–535.

Poggio, G. F., & Talbot, W. H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *Journal of Physiology, 315*, 469–492.

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation, 6*(3), 390–404.

Qian, N., & Freeman, R. D. (2009). Pulfrich phenomena are coded effectively by a joint motion–disparity process. *Journal of Vision, 9*(5) [24.1–16].

Qian, N., & Andersen, R. A. (1997). A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. *Vision Research, 37*(12), 1683–1698.

Rashbass, C., & Westheimer, G. (1961a). Disjunctive eye movements. *Journal of Physiology, 159*, 339–360.

Rashbass, C., & Westheimer, G. (1961b). Independence of conjugate and disjunctive eye movements. *Journal of Physiology, 159*, 361–364.

Read, J. C., & Cumming, B. G. (2005a). All Pulfrich-like illusions can be explained without joint encoding of motion and disparity. *Journal of Vision, 5*(11), 901–927.

Read, J. C., & Cumming, B. G. (2005b). Effect of interocular delay on disparity-selective v1 neurons: Relationship to stereoacuity and the Pulfrich effect. *Journal of Neurophysiology, 94*(2), 1541–1553.

Read, J. C., & Cumming, B. G. (2005c). The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth. *Journal of Vision, 5*(5), 417–434.

Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nature Neuroscience, 12*(8), 1050–1055.

Sabatini, S. P., & Solari, F. (2004). Emergence of motion-in-depth selectivity in the visual cortex through linear combination of binocular energy complex cells with different ocular dominance. *Neurocomputing, 58–60*, 865–872.

Sabatini, S. P., Solari, F., Cavalleri, P., & Bisio, G. M. (2003). Phase-based binocular perception of motion in depth: Cortical-like operators and analog VLSI architectures. *EURASIP Journal of Applied Signal Processing*(1), 690–702.

Toyama, K., Komatsu, Y., Kasai, H., Fujii, K., & Umetani, K. (1985). Responsiveness of Clare-Bishop neurons to visual cues associated with motion of a visual stimulus in three-dimensional space. *Vision Research, 25*(3), 407–414.

Tsang, E. K. C., & Shi, B. E. (2007). *Estimating disparity with confidence from energy neurons*. *Neural information processing systems*. British Columbia, Canada: Vancouver.

Tsang, E. K. C., & Shi, B. E. (2008). Normalization enables robust validation of disparity estimates from neural populations. *Neural Computation, 20*(10), 2464–2490.

Watson, A. B., & Ahumada, A. J. Jr., (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A – Optics Image Science and Vision, 2*(2), 322–341.

Zeki, S. M. (1974). Cells responding to changing image size and disparity in the cortex of the rhesus monkey. *Journal of Physiology, 242*(3), 827–841.

Zhu, Y. D., & Qian, N. (1996). Binocular receptive field models, disparity tuning, and characteristic disparity. *Neural Computation, 8*(8), 1611–1641.