On the road to invariant object recognition: How cortical area V2 transforms absolute to relative disparity during 3D vision

Stephen Grossberg *, Karthik Srinivasan, Arash Yazdanbakhsh

Center for Adaptive Systems, Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA
Center of Excellence for Learning in Education, Science and Technology, Boston University, 677 Beacon Street, Boston, MA 02215, USA

1. Introduction

Cells in visual cortical area V1 are sensitive to absolute disparity (Gonzalez & Perez, 1998), which is the horizontal difference in the retinal location of an image feature in the left and right foveas after fixation. However, many cells in cortical area V2 are sensitive to relative disparity (Thomas, Cumming, & Parker, 2002), which is the difference in absolute disparity of two visible features in the visual field (Cumming & DeAngelis, 2001; Cumming & Parker, 1999); e.g., a figure and its background. Absolute disparity varies with the distance of an object from an observer. Psychophysical experiments have shown that absolute disparity can change across a visual scene without affecting relative disparity. In particular, relative disparity, unlike absolute disparity, can be unaffected by the distance of visual stimuli from an observer, or by the vergence eye movements that occur as the observer inspects objects at different depths (Miles, 1998; Yang, 2003). Thus relative disparity is a more invariant measure of an object’s depth, and hence its 3D shape, than is absolute disparity. How does the transformation from absolute to relative disparity occur between cortical areas V1 and V2 to produce such invariance?

The transformation from absolute to relative disparity is relevant to a central theme in cognitive neuroscience; namely, humans and other primates effortlessly recognize objects in the world as they move their eyes, heads, and bodies with respect to them. This flexibility implies a high degree of invariance during object recognition. Multiple cortical areas, ranging from V1, V2, and V4, through inferotemporal (IT) and prefrontal cortex (PFC), gradually build up such invariance in stages. One important early stage occurs in cortical areas V1 and V2, where binocular information from both eyes is used to code the location of objects in depth. Such stereoscopic depth perception can then support the computation of perceptual groupings, figure–ground segmentation, 3D shape representation, object motion in depth, and invariant object recognition. The transformation from absolute to relative disparity is an early step on this road to invariance.

The disparity energy model successfully simulates data about absolute disparity tuning in V1 cells (Fleet, Wagner, & Heeger, 1996; Ohzawa, 1998; Ohzawa, DeAngelis, & Freeman, 1997). This model pools inputs from a population of binocular simple cells with receptive fields (RFs) from both eyes. These responses are passed through a nonlinear rectifier that does a rectification followed by squaring. These preprocessed responses are summed by complex cells whose responses give rise to the desired absolute disparity tuning curve.

In Thomas et al. (2002), extracellular single-unit recordings in area V2 of two alert monkeys were obtained to evaluate their...
relative disparity-tuning properties. The stimulus consisted of a dynamic random dot stereogram (RDS) with a central patch and a surrounding annulus. The central patch and the surrounding annulus created a figure and ground, respectively. Such a RDS provides a cyclopean stimulus configuration; that is, the eyes were fixated when the cell responses were measured. Prior to displaying the RDS, the patch was sized and positioned to cover the receptive field (RF) of the neuron to be within the minimum response field. The disparities used for the center and surround were restricted within the range $[-1^\circ, 1^\circ]$ (see input in Fig. 1).

The disparity of the central patch at two or three different surround disparities. This ensured that the relative disparity between the center and surround was varied independently of the absolute disparity of the center. Using this protocol, they recorded from 165 neurons and found that 62 of these neurons showed significant selectivity towards relative disparity. These 62 neurons yielded sufficient data, determined from a minimum of four repetitions at each of seven disparities for each surround condition, to enable analysis. In this setting, for a neuron to have perfect relative disparity tuning, the change in direction and size of the shift of the cell tuning should match the change in surround disparity. Their analysis disclosed a gradient shift of cells tuned from absolute to relative disparity using a shift ratio metric (see Shift Ratio in Section 4), thereby showing that certain V2 cells encode for relative disparity.

To explain these data, Thomas et al. (2002) proposed an extension of the disparity energy model to account for how neurons in V2 may give rise to relative disparity from absolute disparity. Their model processes pairs of populations that code absolute disparity outputs from V1. One population of a pair consists of cell responses from the RF of the center patch in the RDS, and the other from a RF in the surround annulus. These are summed and squared. Subsequently, the outputs from monocular filters are subtracted from each pair to increase the sensitivity to relative disparity. If the particular monocular filter is not subtracted, the V2 cell response is influenced more by absolute disparity. This suggests the need for some inhibitory mechanism. The outputs of all such responses are summed by a neuron in V2 that is proposed to estimate relative disparity.

This proposal has several shortcomings. First, the model only qualitatively simulates some data from Thomas et al. (2002). Second, the model needs to somehow know which monocular filter to subtract from each population pair to compute relative disparity. The model does not provide any information as to how this may be accomplished in vivo. Third, no explanation is provided for why cells in V2 exhibit a gradient from absolute to relative disparity.

We propose a neural model that quantitatively simulates the data from Thomas et al. (2002). The model demonstrates that shunting lateral inhibition of layer 4 cells in cortical area V2 can cause a peak shift in the cell responses. This peak shift is sufficient to transform absolute disparity into relative disparity. This inhibitory circuit has previously been used to explain perceptual and neurobiological data about contrast gain control, divisive normalization, selection of perceptual groupings, and attentional focusing (Grossberg, 1999; Grossberg & Raizada, 2000). The model’s inhibitory circuit hereby links the computation of relative disparity to other visual functions and thereby suggests new ways to test its mechanistic basis.

2. Model overview and equations

As noted above, the model predicts that the transformation from absolute to relative disparity is accomplished by a peak shift in cell tuning that is caused by a network between layers 6 and 4 of V2 with spatially narrow excitatory connections and spatially broader inhibitory connections among cells that obey the membrane, or shunting, equations of neurophysiology (Hodgkin, 1964). These processes together define a shunting on-center off-surround network (Fig. 1). Such a shunting on-center off-surround network is capable of normalizing the activities of its cells (Grossberg, 1973; Heeger, 1992).

The inputs to cells in this V2 network are outputs from V1 layers 2/3 binocular complex cells that are tuned to absolute disparity. An analogous on-center off-surround network between layers 6 and 4 is known to also occur in V1 (Ahmed, Anderson, Martin, & Nelson, 1997; Callaway, 1998; McGuire, Hornung, Gilbert, & Wiesel, 1984).

These processes may be mathematically defined as follows:

1. **Model inputs.**

   The disparity $\mu_C$ of a center input is chosen in the range $[-1^\circ, 1^\circ]$. Likewise, the disparity $\mu_{S1}$ of the first surround input is chosen in the range $[-1^\circ, 1^\circ]$, as is the disparity $\mu_{S2}$ of a second surround input.

2. **V1 cell responses.**

   The responses of disparity-tuned complex cells in V1 layer 2/3 are defined in terms of the input parameters. The cells in each layer of the network are assumed to be arranged topographically along the disparity axis (see Fig. 1). In V1 layer 2/3, each cell is thus tuned to a particular absolute disparity. The Gaussian receptive field across disparity-tuned cells of the ith cell with disparity $\mu_i$ is defined by:

   $$V_i^{(1)}(\theta) = e^{-\frac{(\theta - \mu_i)^2}{2\sigma_i^2}}.$$  

   In Eq. (1), the disparity $\theta$ varies in the range $[-1^\circ, 1^\circ]$ and the tuning curve width $\sigma_i = 0.2$ is estimated from data (Gonzalez & Perez, 1998). The tuning curve width is estimated as the width of the cell at half its peak amplitude, i.e., as the Full Width Half Maximum (FWHM). The model V1 neurons used in the simulation are shown in Fig. 2. Thus a continuous array of cells tuned from near to far were used in the simulations.
3. V2 cell responses.

Activity $V^{(2)}_i(\theta)$ of the $i$th layer 4 V2 cell is defined by a membrane, or shunting, equation that receives inputs from V1 cells via a shunting feedforward on-center off-surround network:

$$\frac{dV^{(2)}_i}{dt} = -AV^{(2)}_i + (B - V^{(2)}_i)V^{(1)}_i - (C + V^{(2)}_i)\sum_j D_{ij}V^{(1)}_j. \quad (2)$$

In Eq. (2), parameter $A (=0.001)$ determines the decay rate of the cell; $B (=10)$ is the excitatory saturation point of depolarized cell activity; $V^{(1)}_i$ is the on-center input from V1, defined by Eq. (1), which inputs a narrow band of disparities to V2; $C (=3)$ is the inhibitory saturation point of hyperpolarized cell activity; and $\sum_j D_{ij}V^{(1)}_j$ describes an off-surround inhibitory input from V1 layer 6 cells with Gaussian off-surround kernel:

$$D_{ij} = D^- \frac{1}{\sqrt{2\pi} \sigma_{inh}} e^{-\frac{(\mu_j - \mu_i)^2}{2\sigma_{inh}^2}}. \quad (3)$$

For simplicity, it is assumed that V2 layer 6 cells simply relay their inputs $V^{(1)}_j$ from V1 to layer 4 via the Gaussian kernel weights $D_{ij}$. In Eq. (3), $D^-$ is the maximum value of the off-surround kernel; $\mu_j$ is the preferred disparity of the $j$th V2 cell; and $\sigma_{inh}$ scales the width of the off-surround kernel across disparity-tuned cells. In the simulations, $\mu_j$ varies in the range $[-1^\circ, 1^\circ]$ to correspond to neurons that are arranged along the disparity axis.

At equilibrium, $\frac{dV^{(2)}_i}{dt} = 0$. By Eq. (2), the $i$th V2 layer 4 cell response at equilibrium is:

$$V^{(2)}_i = \frac{BV^{(1)}_i - C \sum_j D_{ij}V^{(1)}_j}{A + V^{(1)}_i + \sum_j D_{ij}V^{(1)}_j}. \quad (4)$$

The shunting off-surround at equilibrium in Eq. (4) automatically realizes divisive normalization, unlike the disparity energy model.

3. Simulation protocol

The model input consists of points (pixels) arranged in depth in a visual field along one position, similar to the inputs used in Thomas et al. (2002); see Fig. 1. The fixation plane has a disparity of $0^\circ$. We chose one position in space to be the center dot, or patch (Fig. 1). The center dot varies from a disparity of $-1^\circ$ to $1^\circ$. Adhering to the experimental method in Thomas et al. (2002), we chose a surround dot at random.
The disparity of the surround dot also varies from $-1^\circ$ to $1^\circ$. With this center and surround fixed, each disparity-tuned cell in V1 codes for a particular absolute disparity. Disparity-sensitive cells are distributed across the network. This calculation is a valid approximation of disparity tuning (Gonzalez & Perez, 1998).

The model consists of 200 cells in each layer of the network, all arranged along the disparity axis. The disparities of the V2 cells differ by $0.01^\circ$, and thereby sweep out an interval $[-1^\circ, 1^\circ]$ of disparities.

After fixation is established, a center dot is chosen. The disparity of this center dot is assessed and the V1 tuning curve corresponding to the disparity is calculated. The disparity of the center patch is systematically varied in the range $[-1^\circ, 1^\circ]$, as in Fig. 1. To assess the nature of the response of V2 cells to surround disparities and in effect their sensitivity to relative disparity, we choose a pair of surrounds at random after the V2 cell profile is assessed at $0^\circ$ surround disparity. The surrounds are chosen from the interval $[-1^\circ, 1^\circ]$ to test possible peak shifts. For each center patch, there are 200 possible surround disparities. Thus, for each cell, keeping the center of the lateral inhibition at the cell which codes for the center patch, one of the surround disparities is chosen and is kept active until the effect of the network on this cell is found. This is repeated for the second surround as well. The need for such a protocol becomes clear to understand the shift ratio (see Shift Ratio calculation in Section 4).

4. Results

Throughout the simulations, the values of A, B, and C in Eq. (2) were held constant. Parameters $D^-$ and $\sigma_{inh}$ in Eq. (2) were varied to assess the role that the shunting inhibitory off-surround from layer 6 cells plays in the computation of relative disparity.

1. Peak shift in V2 layer 4 due to the shunting on-center off-surround network.

The activity of a model V2 layer 4 cell exhibits a peak shift in its disparity-tuning curve relative to its V1 on-center input (Fig. 3(b)). This peak shift is due to the shunting on-center off-surround network from layer 6 to layer 4 in V2.

2. Shift ratio.

The peak shift is calculated as follows: The peak activity across all V2 cells is shifted by two different V1 cells whose maximal absolute disparity sensitivities are centered at absolute disparities $\mu_{S1}$ and $\mu_{S2}$. These disparities are denoted by $\mu_{S1}$ and $\mu_{S2}$ because both of the V1 cells activate target V2 cells via their off-surrounds (S). The corresponding shifted peaks for each V2 cell $V_{i}^{(2)}$ are denoted by $p_1$ and $p_2$. The shift in peaks relative to the difference in surround disparities is called the shift ratio:

$$\text{shift ratio}_{S1S2}(V_{i}^{(2)}) = \frac{p_1 - p_2}{\mu_{S1} - \mu_{S2}}. \quad (5)$$

In particular, in Eq. (5), the peak shift of the $i$th V2 cell $V_{i}^{(2)}$ in response to the first surround input ($S1$), chosen at random, is calculated as the shift of the peak of this cell with respect to the V2 cell peak when the surround disparity is at $0^\circ$. The peak shift caused by shunting using this surround disparity is $p_{1i}$. Similarly, the peak shift of the $V_{i}^{(2)}$ cell for a second randomly chosen surround ($S2$) is calculated. The peak shift of this cell in response to this surround disparity is $p_{2i}$. The ratio of these peak shifts divided by the difference $\mu_{S1} - \mu_{S2}$ of the surround disparities defines the shift ratio in Eq. (5).

A shift ratio of zero signifies that the cell is tuned to absolute disparity. A shift ratio of one signifies that the cell is tuned to relative disparity. In V2, a gradient of shift ratios from absolute disparity to relative disparity is observed.

An exhaustive number of combinations would correspond to permutations derived from choosing two surrounds without repetition from a set of 200 cells; that is, $\binom{200}{2} = 19,900$. However, the best available data from Thomas et al. (2002) have a maximum of 91 shifts, so a random selection is justified as the only means for comparison of the summary statistics. Thus, in our simulations, for each cell we compute four shift ratios to derive a total of 1600 shifts and 800 shift ratios. The final results for shift ratios are randomly sampled without replacement to select 75 and 91 shifts, respectively (see Figs. 3(d) and 4(c)) to match the number of shifts that are computed in the experimental data (see Fig. 3(c)). Such a sampling was done to best match the experimental data and to avoid any bias in running the simulation. The network behavior measured from the shift ratio statistics for different parameter values of the inhibitory off-surround kernel is shown in Fig. 4.

The results show that, in order to achieve shift ratios that spread from 0 (absolute disparity) to 1 (relative disparity), as is observed in V2, one requires a wide surround inhibition relative to the breadth of the V1 Gaussian that represents absolute disparity tuning. Equally important is the amplitude $D^-$ of the inhibition. The amplitude of $D^-$ has to be weak relative to the maximum amplitude of the on-center input, $V_{i}^{(1)}$, in Eq. (1). If $D^-$ is small, then there is an absolute-to-relative disparity gradient in V2 similar to the data from Thomas et al. (2002). If $D^-$ is large relative to the maximum amplitude of $V_{i}^{(1)}$, then, however wide the inhibition is, the shift ratios cluster around 0, and thus lead to more absolute disparity.

It is clear from our shift ratio sampling that there is a gradient shift towards relative disparity. There are also cells in the model which have shift ratios that exceed 1 or are negative; see Fig. 4. This is possible because the surround dots are chosen at random. The best result in comparison to the data was obtained when $D^- = 0.2$ and $\sigma_{inh} = 1.0$. Thus the system requires a wide but weak inhibitory kernel.

5. Summary and discussion

The V2 shunting lateral inhibitory network operates on the outputs from the V1 network of disparity-tuned cells to transform absolute disparity into relative disparity. Model simulations provide good fits to the data in Thomas et al. (2002); see Fig. 3. The simulations go beyond the data to predict the mechanism which causes the relative disparity responses measured in V2 as well as the shift ratios observed for disparities in V2 (Figs. 3 and 4); namely, the shifts in the peak or trough of the tuning curves relative to the difference in surround disparity.

In summary, shunting lateral inhibition acting at layer 4 of cortical area V2 can quantitatively explain how absolute disparity in V1 is transformed into relative disparity in V2. This mechanism is consistent with previous theoretical analyses of perceptual and neurophysiological data which led to the conclusion that broad lateral inhibition is needed for stereoscopic depth processing. In particular, as part of a study to discriminate depth contrast (that is, to discriminate the depth of two adjacent bars), Mitchison (1993) hypothesized that center-surround disparity tuning is needed in the stereo pathway.

The current model supports that claim and predicts, in addition, that surround inhibition should be wide ($\sigma_{inh} = 1.0$) and weak ($D^- = 0.2$) in the V2 inhibitory surround that causes the peak shift (see Eqs. (2) and (3)) relative to the central activation zone — that is, the absolute disparity tuning ($\sigma_1 = 0.2$; maximum amplitude of input from $V1 = 1.0$) in the V2 inhibitory Gaussian (see Eq. (1)) — in order to transform absolute disparity into relative disparity. This set of surround inhibition parameters with respect to absolute disparity tuning best fits the shift ratio statistics of the data from V2 presented in Thomas et al. (2002); see Fig. 3(c). How
these parameters affect the V2 shift ratios is presented in Fig. 4, with Figs. 4(c) and 3(d) most resembling the data in Fig. 3(c).

The model’s peak shift mechanism to carry out this transformation adds to the list of known functions that may be traced to peak shifts due to lateral inhibition. These include the visual illusion of line neutralization (Levine & Grossberg, 1976), the peak shift and behavioral contrast that occur during reinforcement learning (Grossberg, 1975), and a peak shift that controls a vector decomposition whereby global motion appears to be subtracted from the true motion path of localized stimulus components (Grossberg, Léveillé, & Versace, in press). In the present case, the peak shift occurs within a laminar cortical network that enables the computation of relative disparity to facilitate the generation, through multiple stages of laminar cortical processing in visual cortex, of invariant representations of 3D shape to which attention can selectively be paid (Grossberg, 1999; Grossberg, Markowitz, & Cao, in press; Grossberg & Swaminathan, 2004; Grossberg & Yazdanbakhsh, 2005; Raizada & Grossberg, 2005).

Within the 3D LAMINART model (Grossberg, 1999, 2003; Grossberg & Raizada, 2000), the shunting off-surround from...
Fig. 4. Shift ratio statistics due to varying $D^-$ and $\sigma_{inh}$. Shifts towards absolute disparity or relative disparity depend on these parameters. (a) $D^-=0.5; \sigma_{inh}=1.0$. Shift towards absolute disparity. This is the usual profile of V1 disparity cells. (b) $D^-=1.0; \sigma_{inh}=0.5$. Absolute disparity is observed with a larger amplitude and narrower width of the off-surround kernel. (c) $D^-=0.2; \sigma_{inh}=1.0$. A weak amplitude modulation ($D^-=0.2$) and a wide inhibitory surround together ($\sigma_{inh}=1.0$) generate a gradient from absolute to relative disparity resembling the data. Thus the nature of the surround inhibition in V1 and V2 accounts for the type of disparity sensitivity. The parameters used in Figs. 3(d) and 4(c) are the same.

Layer 6 to 4 has earlier been shown to carry out multiple additional perceptual roles (Fig. 5). These include contrast normalization of bottom-up inputs from V1; selection of perceptual groupings via an intracortical V2 recurrent network between layers 2/3, 6, and 4; and biased competition by top-down attention from higher cortical regions via top-down circuits to layer 6 and then back up to layer 4 via its self-normalizing on-center off-surround. In particular, one target of top-down attention is layer 6 cells in V2. Thus, if there are two surrounds and attention is paid to a particular surround dot, then the peak shift may increase in the direction away from the attended surround and its size may co-vary with the attentional gain.

A shunting inhibitory network also occurs in layer 4 of V1 (Ahmed et al., 1997; McGuire et al., 1984; Tamas, Somogyi, & Buhl, 1998), where it also plays multiple roles. One recently predicted role is, in concert with other cortical circuits, to help create...
percepts of perceptual transparency (Grossberg & Yazdanbakhsh, 2005). Thus even relatively simple circuits within the hierarchy of laminar cortical circuits in the visual cortex can play an important role in the transformation of visual inputs into conscious percepts and recognized objects.

Acknowledgements

Supported in part by CELEST, an NSF Science of Learning Center (SBE-0354378), and by the SyNAPSE program of the DARPA (HR0011-09-C-0001).

References


