A computational theory of da Vinci stereopsis

Inna Tsirlin
Centre for Vision Research, York University, Toronto, ON, Canada
Eye Movement and Vision Neuroscience Laboratory, The Hospital for Sick Children, Toronto, ON, Canada

Laurie M. Wilcox
Centre for Vision Research, York University, Toronto, ON, Canada

Robert S. Allison
Centre for Vision Research, York University, Toronto, ON, Canada

In binocular vision, occlusion of one object by another gives rise to monocular occlusions—regions visible only in one eye. Although binocular disparities cannot be computed for these regions, monocular occlusions can be precisely localized in depth and can induce the perception of illusory occluding surfaces. The phenomenon of depth perception from monocular occlusions, known as da Vinci stereopsis, is intriguing, but its mechanisms are not well understood. We first propose a theory of the mechanisms underlying da Vinci stereopsis that is based on the psychophysical and computational literature on monocular occlusions. It postulates, among other principles, that monocular areas are detected explicitly, and depth from occlusions is calculated based on constraints imposed by occlusion geometry. Next, we describe a biologically inspired computational model based on this theory that successfully reconstructs depth in a large range of stimuli and produces results similar to those described in the psychophysical literature. These results demonstrate that the proposed neural architecture could underpin da Vinci stereopsis and other stereoscopic percepts.

Introduction

When the world is viewed binocularly, the retinal images are not identical; the difference in the vantage points of the two eyes creates a disparity in the position of the imaged objects. If corresponding image points on the two retinae can be found, this positional disparity can be extracted, and the relative depth of objects can be determined using simple geometry. The difference in the vantage points of the two eyes also creates monocular areas visible to one eye only, as shown in Figure 1A. These areas arise due to physical occlusion of objects by other objects and thus are referred to as "monocular occlusions" (or "binocular half-occlusions"). Importantly, monocularly occluded areas do not have a match in the image of the other eye.

In the last two decades, it has been shown that depth can be perceived solely on the basis of monocular occlusions (Cook & Gillam, 2004; Gillam & Nakayama, 1999; Nakayama & Shimojo, 1990; Tsirlin, Wilcox, & Allison, 2010, 2012). For example, the presence of monocular occlusions in certain configurations can induce percepts of illusory occluding surfaces (Cook & Gillam, 2004; Gillam & Grove, 2004; Gillam & Nakayama, 1999). This startling phenomenon is demonstrated in Figure 2B and 2C. Depth without conventional disparity can also be perceived between two objects, one of which is monocular (see Figure 2A) and in other configurations with monocular features (Forte, Peirce, & Lennie, 2002; Pianta & Gillam, 2003; Sachtler & Gillam, 2007) (see Figure 2D). Occlusion-based depth phenomena were named "da Vinci stereopsis" by Nakayama and Shimojo (1990) in a nod to Leonardo da Vinci's early reflections on monocular occlusions (da Vinci, 1877) and to distinguish them from conventional stereopsis.

One important question that arises from the above discussion is how depth is computed in the absence of conventional positional disparity. It has been proposed that the visual system relies on occlusion geometry to estimate depth in these cases (Gillam & Nakayama, 1999; Nakayama & Shimojo, 1990). As shown in Figure 1, the line of sight from the eye that cannot see the monocular region constrains the minimum possible depth between the monocular region, or the illusory occluder, and the binocular regions. This minimum

Citation: Tsirlin, I., Wilcox, L. M., & Allison, R. S. (2014). A computational theory of da Vinci steropsis. *Journal of Vision*, *14*(7):5, 1–26, http://www.journalofvision.org/content/14/7/5, doi:10.1167/14.7.5.

doi: 10.1167/14.7.5 Received November 15, 2013; published June 9, 2014 ISSN 1534-7362 © 2014 ARVO

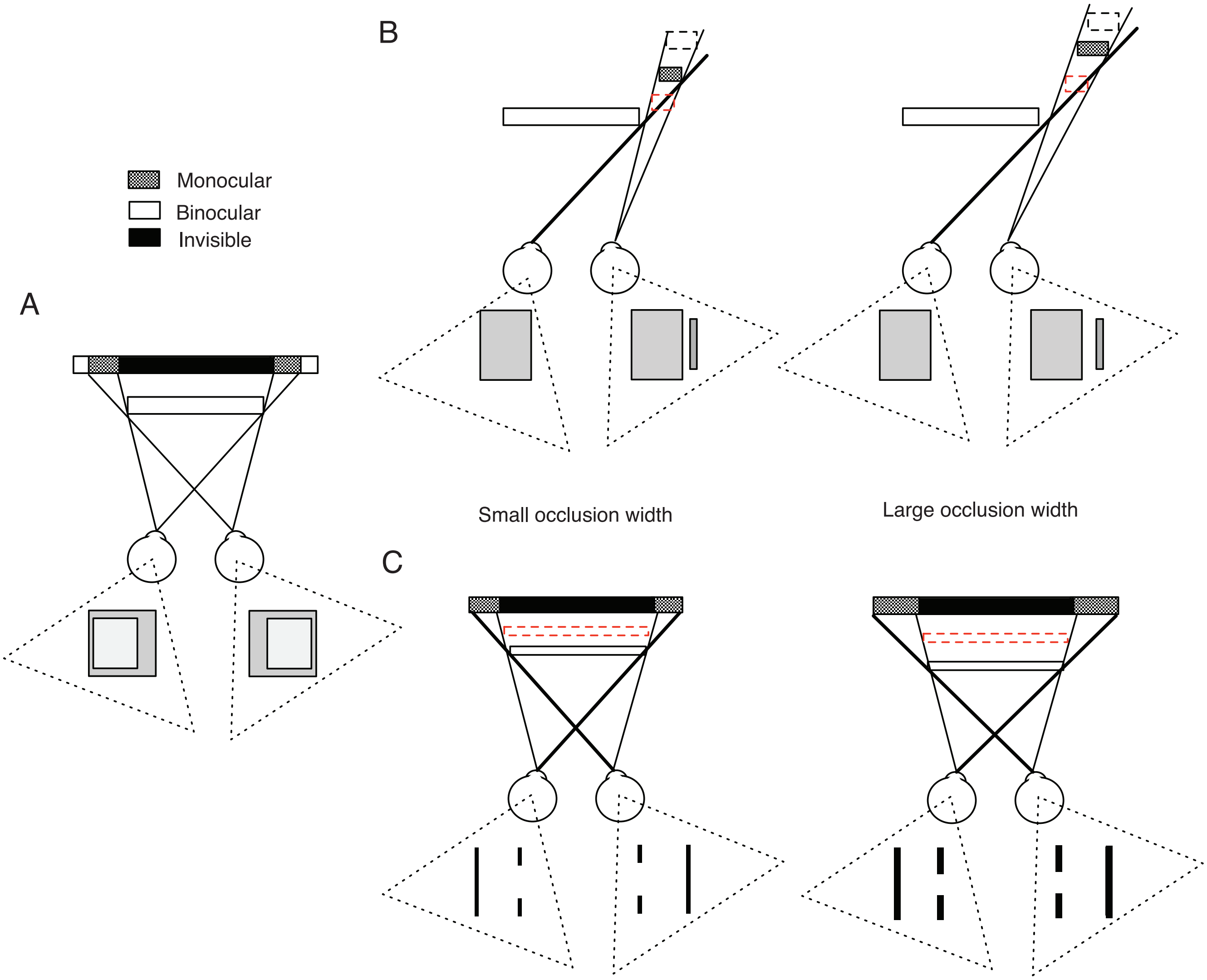

Figure 1. Monocular occlusion geometry. (A) A foreground surface occludes regions of the background in each eye. The images enclosed within the dashed triangles show what each eye is seeing. (B) In this two-object arrangement (Nakayama & Shimojo, 1990), a larger surface (rectangle) occludes a stand-alone smaller object (bar). The line of sight from the left eye (bold black line) that does not see the bar constrains the minimum possible depth of the occluded object. It cannot be located closer to the occluder (red dashed outline) because it would be seen by the left eye. It could be positioned further without violating viewing geometry (black dashed outline). Larger separations between the object and the occluder yield larger minimum possible depths between the two objects as shown in (B) (compare left-hand and right-hand schematics). (C) Similar geometric rules apply to illusory occluder stimuli, for example, that of Gillam and Nakayama (1999). The minimum possible depth of the illusory occluder on each side is constrained by the lines of sight from the eyes that do not see the occluded region. Larger occluded regions yield larger minimum possible depth between the occluded region and the illusory occluder.

possible depth is linearly related to the width of the monocular region (or the distance between the outer edges of the monocular object and the occluder) such that an increase in monocular occlusion width results in an increase in the minimum possible depth. Note that the maximum possible depth of the monocular region (or illusory occluder) is not constrained. Theoretically, the visual system could use the minimum depth constraint, computing it from the occlusion width, to position monocular regions and illusory occluders in depth. In this case, an assumption is made that the minimum possible depth is the best estimate of the depth of the monocular region (or the illusory occluder). In support of this hypothesis, it was found that increasing the width of occluded regions, thus increasing the minimum possible depth, results in an increase in the perceived depth between the occluded and the occluding surfaces both in illusory occluder stimuli (Gillam & Nakayama, 1999; Tsirlin et al., 2010; Tsirlin, Wilcox, & Allison, 2011) and in two-object arrangements (Hakkinen & Nyman, 1997; Nakayama & Shimojo, 1990; Tsirlin, Wilcox, et al., 2012).

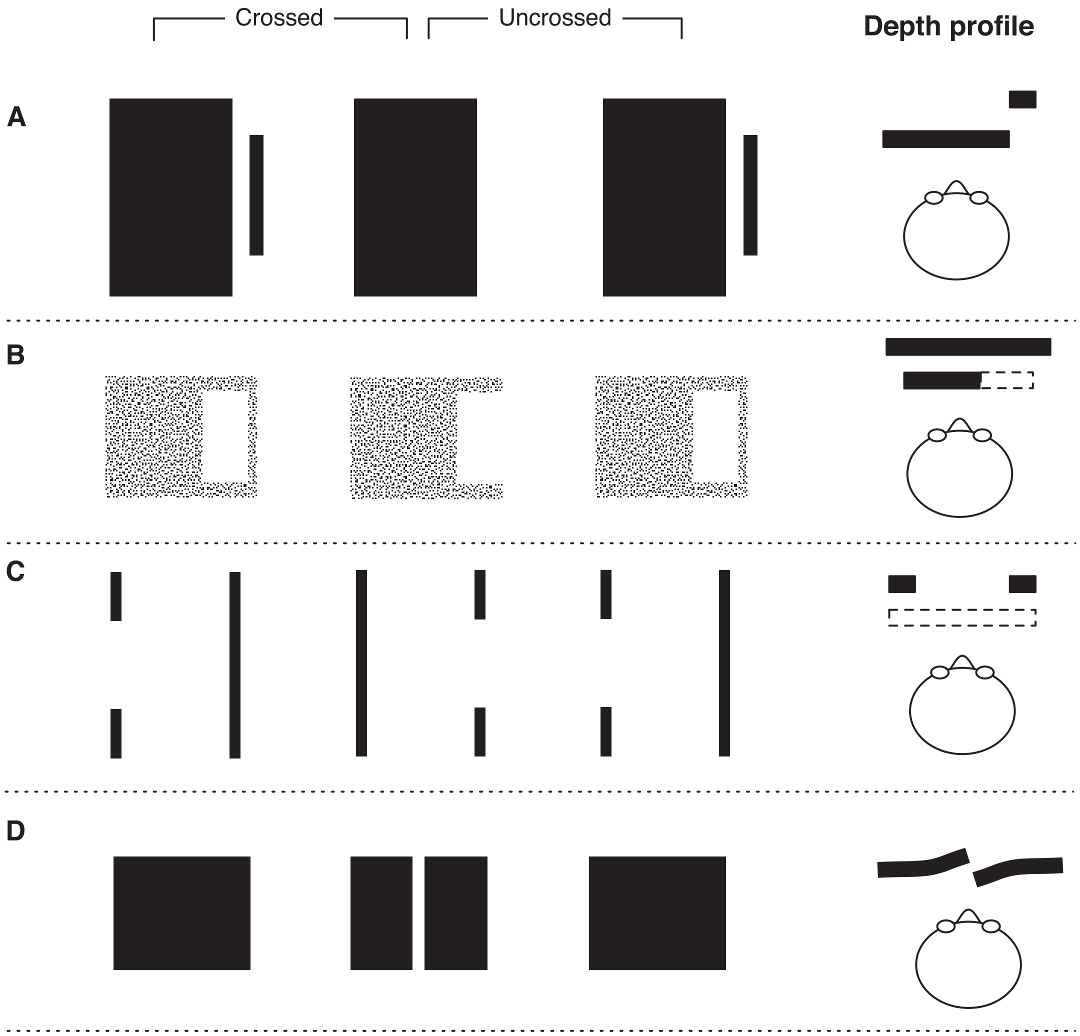

Figure 2. Depth from monocular occlusions. (A) The monocular bar is perceived to lie beyond the binocular rectangle (Nakayama & Shimojo, 1990). (B) The occluding surface has an illusory right half induced by the presence of a monocular strip of random elements (Tsirlin et al., 2010). (C) An illusory rectangular surface appears in front of the black lines (Gillam & Nakayama, 1999). (D) The monocular gap creates a percept of two surfaces bending in depth away from each other (Gillam et al., 1999). The left and the central half-images are arranged for crossed fusion and the central and the right half-images for divergent fusion.

It is also important to understand the mechanisms that could underlie da Vinci stereopsis. One possibility is that depth from occlusions is a byproduct of the activity of disparity detectors involved in stereopsis. However, computational analyses performed with model disparity detectors have shown that this is not the case (Tsirlin, 2013; Tsirlin, Allison, & Wilcox, 2012; Tsirlin, Wilcox, et al., 2012). Instead it appears that a more sophisticated set of neural mechanisms tuned to occlusion geometry is required.

Several computational models of biological vision (referred to here as "biologically inspired") addressing the mechanisms of da Vinci stereopsis have been proposed in recent years. Watanabe and Fukushima (1999) described a model in which occlusions are detected explicitly from the output of feature-based disparity detectors. The 3-D structure of the scene is reconstructed from the initial disparity and occlusion maps using traditional uniqueness and smoothness constraints (Marr & Poggio, 1976) and a novel occlusion constraint. Their occlusion constraint specifies that if a point is detected as occluded there must be an occluding point, such that points signaled as occluded, with no potential occluders, are inhibited. Hayashi, Maeda, Shimojo, and Tachi (2004) made the Watanabe and Fukushima model more biologically plausible by replacing edge-based disparity detectors with binocular energy neurons (Ohzawa, DeAngelis, & Freeman, 1990). They also added interocular inhibition and temporal dynamics, which allowed their model to predict simple binocular rivalry. Grossberg and Howe (2003) and Cao and Grossberg (2005) proposed a model of stereopsis in which monocular occlusions are not detected explicitly but instead are identified implicitly through competition between monocularly identified luminance edges and binocularly identified edges. In their model, depth in occluded areas is computed by interpolation from binocular areas or by double-duty matching. Assee and Qian (2007) have argued that the preceding models are not completely biologically plausible because they use binary (0 or 1) or discrete representations of neuronal firing rates at all (Watanabe & Fukushima, 1999) or some stages (Cao & Grossberg, 2005; Hayashi et al., 2004) of processing while real neuronal firing rates are variable and distributed. They proposed a new model in which the initial computation of disparity is made by energy neurons, and their output is then propagated to V2 neurons selective for disparity edges (von der Heydt, Zhou, & Friedman, 2000). Monocular occlusions are not explicitly detected. Instead it is assumed that a depth step signifies the presence of a monocularly occluded region. Occluded areas are assigned the depth of the further surface.

These models offer interesting hypotheses about the underlying mechanisms of da Vinci stereopsis and propose new, insightful constraints on matching; however, they also have several important drawbacks. First, as suggested by Assee and Qian (2007), three of the models use binary or discrete representations of neural firing. More importantly, none of the models explicitly use the width of the occluded regions to compute their depth (or the depth of surrounding binocular regions with a weak disparity signal). The psychophysical evidence discussed above suggests that the minimum depth constraint, computed from the occlusion width, is the most likely basis for depth perception from monocular occlusions. Moreover, judging by their architecture, these models will not be able to predict perceived depth correctly in several important classes of stimuli. For example, they will not be able to reconstruct illusory occluders in stimuli such as those shown in Figure 2 because these models require the presence of well-defined disparity around the occluded areas (e.g., textured surfaces). In the cases shown in Figure 2, monocular regions are surrounded by textureless, monochromatic areas with ambiguous disparity. Finally, the models were tested on a very limited number of monocular-occlusion test images (1–3), which makes it difficult to evaluate the models thoroughly.

The goal of our work is to develop a computational model of human depth perception that will (a) explain the computation of depth from disparity and monocular occlusions in the visual system; (b) be built on a solid theoretical basis stemming from psychophysical, physiological, and computational data; and (c) produce depth maps that closely correspond to observer percepts for a wide variety of stimuli. To this end, we first develop a theory of depth computation in da Vinci stereopsis. We then formulate and implement a biologically plausible computational model incorporating these principles. The model is tested on a large battery of stimuli, and its performance is compared with psychophysical results. Finally, we discuss the results and the theoretical implications of the neural architecture incorporated in our theory and model.

Model principles

The model proposed in this article is based on a collection of principles intended to explain the computation of depth from disparity and monocular occlusions in the visual system. Unfortunately, no physiological data is available on the mechanisms involved in the computation of depth from monocular occlusions because no one has performed single-cell recordings with monocularly occluded regions as stimuli. Thus the principles outlined below are derived from the psychophysical and computational literature. We group these principles to form a theory of Depth from Monocular Occlusion Geometry (DMOG).

Monocular occlusions are detected explicitly—To extract depth from monocular occlusions using occlusion geometry, occluded regions have to be detected first. Moreover, it has been shown that computer vision algorithms that explicitly detect (and use) monocular occlusions (e.g., Lin & Tomasi, 2004; Min & Sohn, 2008; Sizintsev & Wildes, 2007) are able to successfully recreate depth maps in complex scenes and that they perform better than algorithms that do not use occlusion detection. Thus, in this model, occluded regions are detected by several populations of specialized neurons.

The width of monocular regions is used to compute the depth in these regions and adjacent binocular areas with an unreliable disparity signal—Psychophysical findings suggest that the minimum depth constraint is used by the visual system to assign depth to monocular objects/areas and to illusory occluders (Anderson, 1994; Gillam & Nakayama, 1999; Grove & Gillam, 2007; Tsirlin et al., 2010; Tsirlin, Wilcox, et al., 2012). In order to use the minimum depth constraint to compute depth from monocular occlusions, the width of the monocular zone has to be determined. It has also been shown that the depth signal originating from monocular occlusions can affect (and capture) perceived depth in surrounding binocular regions with unreliable disparity signals (Gillam & Nakayama, 1999; Hakkinen & Nyman, 2001; Tsirlin et al., 2010). Thus, the width of the monocular regions can also be used to compute depth in adjacent binocular areas with ambiguous disparity. In the model, these operations—occlusion width computation and the resulting depth signal computa-

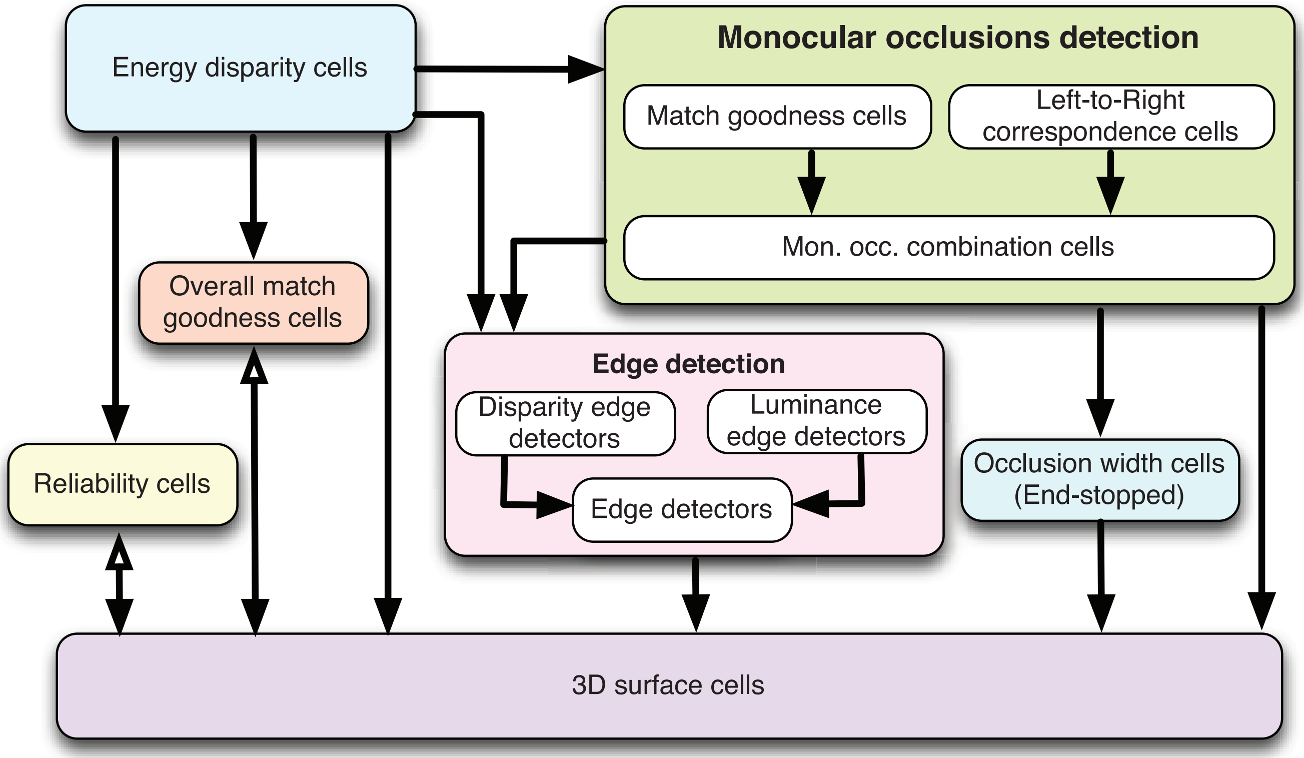
Energy disparity cells
Monocular occlusions detection
Match goodness cells
Left-to-Right correspondence cells
Mon. occ. combination cells
Overall match goodness cells
Edge detection
Disparity edge detectors
Luminance edge detectors
Edge detectors
Reliability cells
Occlusion width cells (End-stopped)
3D surface cells

Figure 3. Model overview. Filled arrows show feed-forward connections and hollow arrowheads show feedback connections.

tion and propagation—are performed by several populations of specialized neurons.

Depth from monocular occlusions and disparity is computed concurrently—Several psychophysical studies have suggested that depth from monocular occlusions and from binocular disparity is computed simultaneously in the early stages of visual processing (Gillam & Borsting, 1988; Mitsudo, Nakamizo, & Ono, 2005; Sachtler & Gillam, 2007). For example, Sachtler and Gillam (2007) showed that the minimum possible time required to process disparity-based and occlusion-based depth was very similar. Moreover, recent work with visually evoked potentials (Spang, Gillam, & Fahle, 2012) has shown that stimuli in which depth percepts are based on occlusion geometry produce cortical responses at the same latencies as similar stimuli in which depth is based on conventional disparity. These findings suggest that monocular occlusions are likely to be processed early, concurrently with binocular disparity.

Monocular camouflage is interpreted as monocular occlusion—It is possible for monocular regions to arise due to camouflage when an object is positioned in front of an identically colored surface and overlaps with it completely in one eye but not the other (see also the discussion of two-object arrangements). In this case, the monocular object should obey the same geometric constraints as monocularly occluded objects. Although theoretically possible, monocular camouflage should be quite rare in nature because the foreground and background surfaces must have identical color and luminance and the foreground has to be located completely within the boundaries of the background but only in one eye. Not surprisingly, experimental evidence shows that very little depth is seen in stimuli with configurations corresponding to monocular camouflage (the reader can appreciate this percept by cross-fusing the central and the right images of Figure 2A) and that this depth does not comply with camouflage geometry (Gillam, Cook, & Blackburn, 2003; Nakayama & Shimojo, 1990; Tsirlin, Wilcox, et al., 2012). Based on these results, it has been suggested that the visual system might not be equipped to process camouflage and instead interprets it as occlusion (see extended discussion of this issue in Tsirlin, Wilcox, et al., 2012). In our model, this principle is reflected in the way depth in occluded regions is computed from their width.

Depth signals from disparity detectors and monocular occlusion detectors interact—Tsirlin et al. (2011) and Tsirlin, Wilcox, et al. (2012) have shown that binocular disparity signals can affect depth from monocular occlusions when it is not completely constrained by occlusion geometry. On the other hand, Hakkinen and Nyman (2001) demonstrated that depth signals from monocular occlusions can influence the perceived depth of binocular areas that contain repeating texture (wallpaper patterns). This evidence suggests that the two depth signals interact in complex ways.

Model description

Based on our DMOG theory, we have designed a computational model, outlined in Figure 3. Depth processing begins with the initial disparity computation performed by a population of energy neurons. Following that, occlusions are detected by several neuronal populations, and their widths are computed. Another block of neurons detects luminance and disparity edges. Reliability and match goodness metrics that are used in the final computation of disparity (see below) are also computed from the initial disparity estimates. Finally, the output of all these neurons is used by 3-D surface neurons to iteratively reconstruct the 3-D structure of the scene. Importantly, all model neurons have variable (distributed) firing rates similar to real neurons. Below, each of the components is described in detail, and the mathematical formulations are provided for all types of neurons used in the model. For convenience, Table 2 in Appendix A summarizes all symbols and functions used in the mathematical formulation of the model.

Initial disparity computation: Energy neurons and interneural connections

The initial computation of disparity is performed using a network of disparity detectors modeled as energy neurons (DeAngelis, Ohzawa, & Freeman, 1995; Ohzawa et al., 1990). The model has been described in great detail by others (e.g., Cumming & DeAngelis, 2001), so here we provide the minimum description of the basic energy model and detailed descriptions of our modifications.

The disparity energy model postulates that simple neurons compute the sum of the left and the right images filtered with their respective receptive fields (RFs), which can be described by Gabor functions. Disparity selectivity in these neurons can be achieved through two mechanisms: (a) position shift, a shift between the positions of the RFs in the two eyes, and (b) phase shift, a shift in the phase of the Gabor in the two eyes. Because both phase and position shift mechanisms are used in the visual system (Anzai, Ohzawa, & Freeman, 1999), both mechanisms are included in our version of the model (similar to Fleet, Wagner, & Heeger, 1996, and others). We used the following Gabor filter to describe the left RF of a simple binocular neuron:

$$f_L(x, y, d_L, \varphi_L) = Gauss(x, y, d_L) \cdot \sin(x, y, d_L, \varphi_L) \tag{1}$$

$$Gauss(x, y, d_L) = \frac{1}{2\pi\sigma_x\sigma_y} \cdot e^{-\frac{1}{2}\left[\left(\frac{(x+d_L)\sin\theta + y\cos\theta}{\sigma_x}\right)^2 + \left(\frac{(x\_d_L\cos\theta - y\sin\theta)}{\sigma_y}\right)^2\right]} \tag{2}$$

$$\sin(x, y, d_L, \varphi_L) = \cos[\omega_0((x + d_L)\sin\theta) + \varphi_L] \tag{3}$$

where d_L is the position shift of the left RF, σ_x and σ_y are the horizontal and vertical Gaussian widths, θ is the preferred orientation, ω_0 is the peak preferred frequency, and φ_L is the left phase parameter. The right RF has the same definition but with a positional shift d_R and a phase shift φ_R.

In our energy neurons, the phase-shift mechanisms are used for disparities smaller than π for each preferred frequency ($d_{L/R} = 0$). For disparities larger than π, position-shift mechanisms are used ($\varphi_{L/R} = 0$). Three aspects of disparity computation should be noted. First, oriented RFs with phase-shift disparity tuning do not code strictly horizontal disparity. They code disparity orthogonal to their orientation because the phase shifts orthogonally to the neuron's orientation. This aspect is problematic because most disparities in natural viewing are horizontal due to the horizontal separation of the eyes. In contrast, position-shift neurons always encode horizontal disparity regardless of their orientation. Thus, as a simple solution to this problem, only position shifts are used for orientations other than vertical ($\varphi_{L/R} = 0$). Second, the matches are made along the same epipolar lines in the two images (epipolar constraint). Third, the shift in the phase or the position of RFs that achieves disparity tuning is performed in one eye only as was done in previous models (e.g., Chen & Qian, 2004).

The complex neurons sum the squared responses of two linear neurons S_1 and S_2 in quadrature phase

$$C_0 = S_1^2 + S_2^2 \tag{4}$$

Mathematically, the classic energy model computes a quantity that comes close to a cross-correlation between images filtered with the neuron's RFs. However, this quantity contains not only correlation information but also the monocular energy of the two images, which makes it very prone to false matches (Fleet et al., 1996; Read, 2010; Read & Cumming, 2006). Given this, disparity estimates from the output of classic energy neurons are extremely noisy. This problem can be solved by normalizing the output of the complex neurons C_0 by the sum of squared monocular energy responses of the monocular RFs of the two

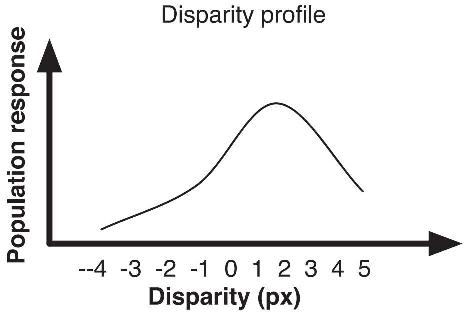

Figure 4. Disparity profile. The pooled population response $C^*_{x,y}$ for the location *x,y* is plotted as a function of disparity. In this case, the population responds maximally to disparity of 2 px.

simple neurons S_1 and S_2:

$$C = \frac{C_0}{(If_{1L})^2 + (If_{1R})^2 + (If_{2L})^2 + (If_{2R})^2}, \quad (5)$$

where $S_1 = If_{1L} + If_{1R}$ and $S_2 = If_{2L} + If_{2R}$ and I is the image falling on the RFs.

Normalization with respect to monocular energy has been proposed as a biologically plausible way to bring the responses of the energy model closer to observed human performance (Allenmark & Read, 2011; Heeger, 1992; Read, 2010).

Another way to reduce false matches in the classical energy model is to pool responses across orientations and spatial frequencies in combination with local spatial pooling (Fleet et al., 1996). In our model, orientation pooling is performed at each scale, and responses at all scales are pooled to produce the final response. Spatial pooling is performed for each orientation and each scale and on the final response. All pooling is performed by averaging the disparity profiles (weighted by a 2-D Gaussian for spatial pooling) of the different neurons (see Figure 4).

In our energy model, for each location x,y of the retinal image, there is a population of complex neurons C centered on x,y and tuned to the full ranges of disparity, orientation, and spatial frequency as specified above. The final, pooled (as specified above) response of this population for the range of disparities ($-d_m$, d_m) is labeled $C^*_{x,y}$ and represents the disparity profile of location x,y. It can be represented graphically as shown in Figure 4. The pooled response of the population to one particular disparity d is denoted $C_{x,y,d}$ (one point on the curve of Figure 4).

Monocular occlusion detection

Many techniques have been proposed in the computer vision literature to detect monocular occlusions (Egnal & Wildes, 2002). However, most require reliable estimates of disparity on both sides of the occluded region. This is an important point because in stimuli such as those shown in Figure 2 the disparity surrounding the monocular regions is ambiguous and unreliable due to the lack of texture although depth from monocular occlusions is readily perceived. Given this, we have chosen two complementary metrics for the detection of monocular occlusions that do not depend on reliable disparity estimates around the occluded regions: match goodness and left-right match correspondence. Egnal and Wildes found both techniques to be effective at detecting occlusions with left-right match correspondence outperforming match goodness. However, left-right match correspondence was inferior in areas with (only) low spatial frequency, and match goodness performed well in these areas, making these approaches complementary. These heuristics are implemented in a biologically plausible fashion here, using a distributed representation of neuronal firing.

Match goodness

Match goodness is defined here as the ratio of the strength of the maximum response of $C_{x,y,d}$ neurons for a given location to the maximum response to the given image within the whole population of disparity detectors.[1] Normally, for monocularly occluded pixels, this ratio should be lower than for binocular areas because no true match exists. Thus, $MG_{x,y}$ neurons compute the difference of the match goodness ratio at each location x,y from one, such that a higher response of these neurons indicates a higher likelihood that location x,y is occluded:

$$MG_{x,y} = \left[1 - \frac{MAX(C^*_{x,y})}{M^*}\right]_{\theta_1} \quad (6)$$

where $[x]_{\theta_1}$ indicates rectification with respect to the threshold θ_1 or with respect to 0 where $[x]_0$ is used (notation adopted from Reynolds & Heeger, 2009), such that the neuron only fires when its output is larger than the specified threshold. M^* is computed by another type of neuron and is defined as follows:

$$M^* = \max(C_{x,y,d}) \text{ for all possible } x, y \text{ and } d. \quad (7)$$

Left-right match correspondence

In classic left-right match correspondence (called left-to-right check in the computer vision literature), disparity maps are computed first using the right image as the origin and then using the left image as the origin. For binocular image locations, the match made in one direction normally corresponds to the match made in other direction. Locations for which the matches differ

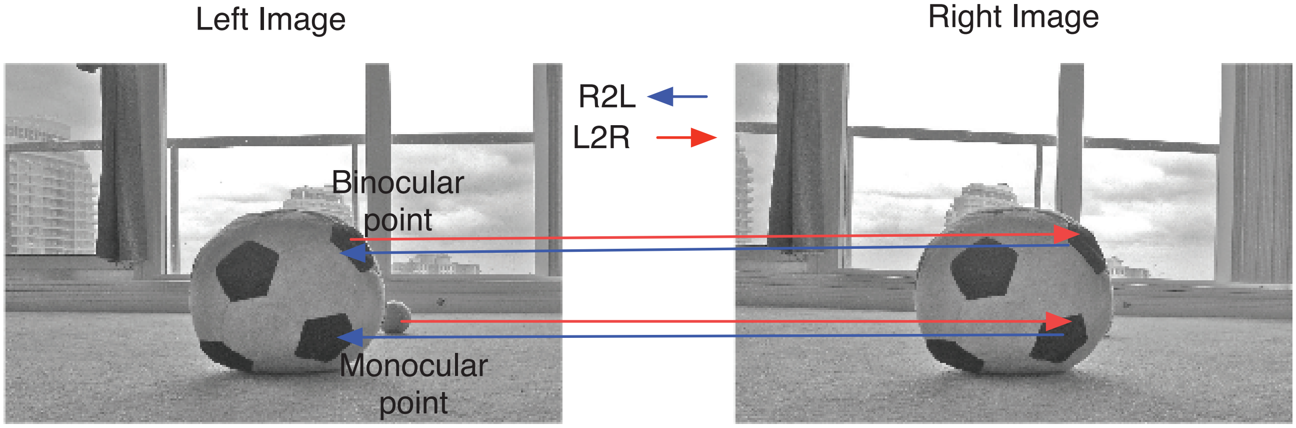

Figure 5. Left-right match correspondence example. Blue arrows indicate a right-to-left and red arrows a left-to-right disparity computation. For the binocular case, both directions match the same points in the two images. In the case of the monocular tennis ball in the left image, the left-to-right computation matches it to some binocular point *p* in the right image. However, in the right-to-left computation, *p* is matched to itself in the right eye. Thus, the match is different in the two computations, and the tennis ball in the left image is marked as monocularly occluded.

substantially are labeled as monocularly occluded. This procedure is illustrated in Figure 5.

Because the model uses a distributed representation of neuronal firing, we cannot simply compare the disparities generating the maximum response for each image location because this will entail converting neuronal output to a binary representation. Instead, disparity profiles are compared. Before describing the neurons that compute left-right match correspondence, new notation needs to be introduced. Let the response of a population of complex neurons $C^{*}_{x,y}$ with all the left eye RFs centered at location *x,y* be denoted as $C^{L*}_{x,y}$. Accordingly, a population of complex neurons $C^{*}_{x,y}$ with all the right eye RFs fixed at location *x,y* is denoted $C^{R*}_{x,y}$. The response of the $C^{L*}_{x,y}$ neurons tuned to a specific disparity is denoted $C^{L}_{x,y,d}$ for the left eye and $C^{R}_{x,y,d}$ for the right eye.

At each location *x,y,* there are two types of neurons that together produce a left-right match correspondence response. The first type, $R_{x,y,d}$ (Equation 8), computes the summed difference between the disparity profiles of neurons $C^{L*}_{x,y}$ and $C^{R*}_{x+d,y}$ for a given disparity *d* (see Figure 6). When location *x,y* has a well-defined disparity d', then the $C^{L*}_{x,y}$ disparity profile will have a peak at d', and the matching $C^{R*}_{x+d',y}$ profile will have a peak at $-d'$. Thus, the summed point-by-point difference between these profiles after they are shifted with respect to each other by $2d'$ (and normalized) will be very small. On the other hand, disparity profiles for nonmatching pixels will be different (after the appropriate shift and normalization) and will yield a relatively large difference.

The shift of the two profiles $C^{L*}_{x,y}$ and $C^{R*}_{x+d,y}$ with respect to each other is achieved simply by comparing each $C^{L}_{x,y,d'}$ response with a $C^{R}_{x+d,y,d'-2d}$ response for each disparity d'. Before the subtraction, the two profiles are normalized and cubed. The cubing is done to amplify the true peak of the profile relative to the false peaks:

$$R_{x,y,d} = \sum_{d'=d_1}^{d_2} \left| n(C^{L}_{x,y,d'})^3 - n(C^{R}_{x+d,y,d'-2d})^3 \right| \qquad (8)$$

where $d_1 = max(-d_m,\ d - d_m)$ and $d_2 = min(d_m,\ d + d_m)$ and the function $n(x)$ is a signal normalization function:

$$n(x) = \left[\frac{x}{MAX(x)}\right] \qquad (9)$$

where x is a vector.

The computation performed by $R_{x,y,d}$ neurons is illustrated in Figure 6. For convenience, the collection of $R_{x,y,d}$ responses for all d in $(-d_m, d_m)$ is denoted $R^{*}_{x,y}$.

For locations *x,y* with well-defined disparity d', the response of the neuron $R_{x,y,d'}$ should be very close to zero, indicating that both "left-to-right" and "right-to-left" computations yielded the same disparity estimate. In contrast, in monocularly occluded regions that lack well-defined disparity, all $R_{x,y,d}$ are likely to yield results much higher than zero. Thus, at each location *x,y,* out of all possible responses of $R_{x,y,d}$ neurons, the magnitude of the minimum response represents the likelihood that this location is monocularly occluded. Accordingly, $LRC_{x,y}$ neurons output the minimum $R_{x,y,d}$ response if it is higher than the threshold θ_2:

$$LRC_{x,y} = \left[MIN(R^{*}_{x,y})\right]_{\theta_2} \qquad (10)$$

Note that textureless areas would be identified as binocular by these neurons because the shape of their

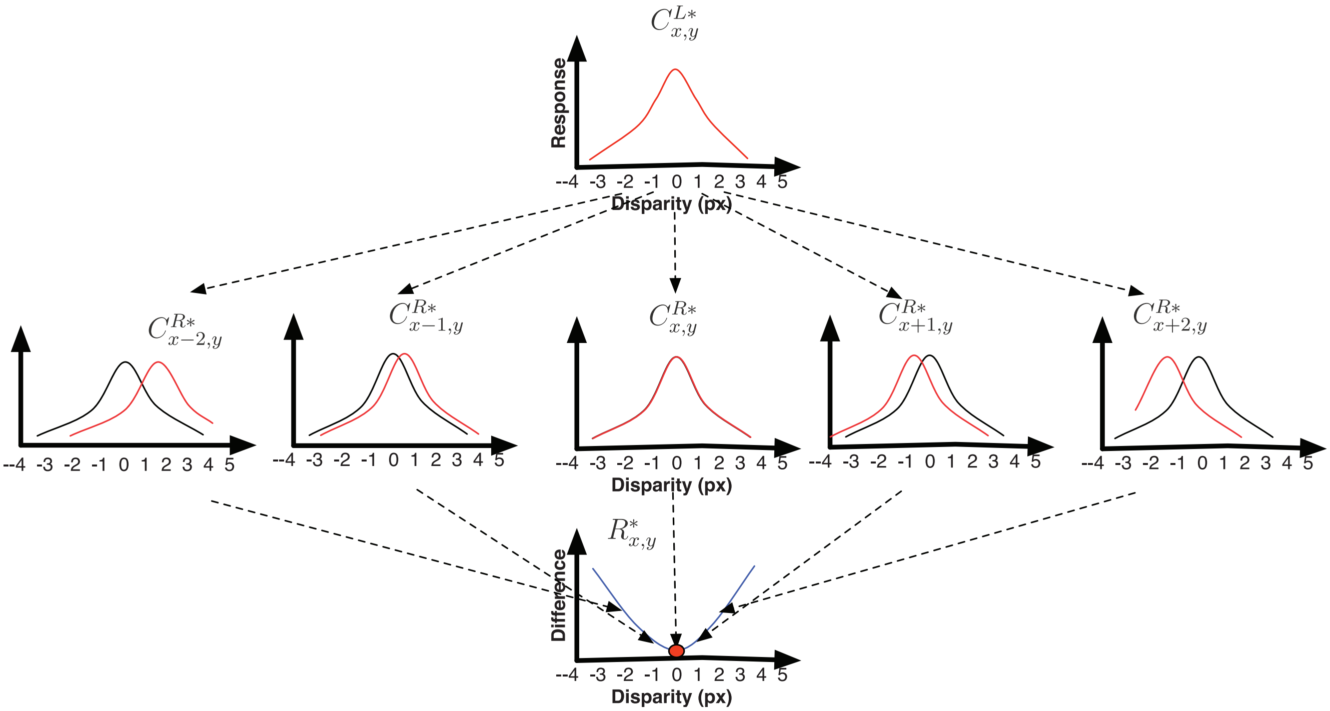

Figure 6. Biologically plausible implementation of the left-right match correspondence. For each pixel *x,y*, the difference between the left-to-right disparity profile $C^{L*}_{x,y}$, shown in red, and all the potential matches $C^{R*}_{x+d,y}$, shown in black, is computed. This computation is performed by $R_{x,y,d}$ neurons. The minimum of the $R^{*}_{x,y}$ responses, shown as the red dot on the blue curve in the bottom row, is chosen as the left-to-right check response for location *x,y*.

disparity profiles would be flat and identical (or similar) in both left-to-right and right-to-left computations, thus producing a small $LRC_{x,y}$ response.

Integrated monocular occlusion detection

The output of the two occlusion-detection mechanisms is combined by a third type of neuron that signals the presence of monocularly occluded regions. These neurons only fire when the combined input from the two mechanisms is higher than a threshold θ_3:

$$OCC_{x,y} = \left[n(MG_{x,y}) + n(LRC_{x,y})\right]_{\theta_3} \qquad (11)$$

Computing the width of monocular areas

After monocular occlusions are detected, their widths must be determined to compute the depth of monocular areas. The width computation is performed by a population of neurons that have an end-stopped architecture with a wide excitatory center (area *EC*)

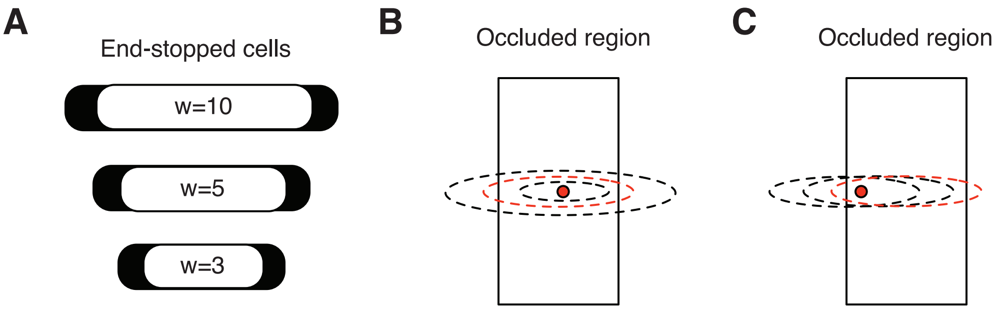

Figure 7. (A) The width of monocular occlusions is signaled by end-stopped neurons of different widths. The excitatory center changes size, and the inhibitory side lobes have the same size for all neurons. (B) At each location *x,y*, there is a population of end-stopped neurons tuned to different widths. (C) For each width, there is a population of neurons for which different parts of their excitatory centers fall on *x,y*. The largest response will be elicited from the neuron with the excitatory center that matches the size and the location of the monocularly occluded area (shown in dashed red line in B and C).

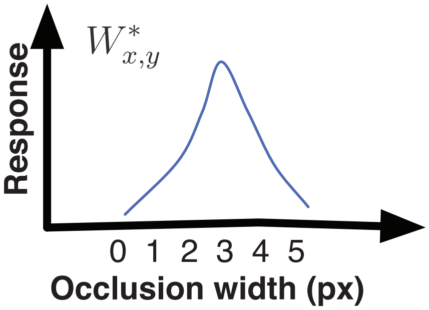

Figure 8. Occlusion width profile. The response of occlusion width detectors $W_{x,y,w}$ for the location *x,y* is plotted as a function of width. In this case, the population responds maximally to a width of 3 px.

positioned between two narrow inhibitory bands (areas *IB*) and which receive input from the OCC neurons described in Equation 11. The neurons have different widths and respond maximally when a monocular region has a width and location matching those of the excitatory center. These neurons are illustrated in Figure 7. The response of each neuron is normalized by the width of the excitatory region (this is done to ensure that large RFs do not have a larger maximum response than smaller RFs). For each possible occlusion width and for each location *x,y,* there is a population of end-stopped neurons with different parts of their excitatory centers positioned at *x,y*. The relative position of the excitatory center with respect to *x,y* is controlled with shift parameter *s*:

$$ES_{x,y,w,s} = \left[\frac{\sum_{x',y' \in EC-s} OCC_{x',y'} - \sum_{x',y' \in IB-s} OCC_{x',y'}}{w}\right]_{\theta_4} \tag{12}$$

ES neurons fire when their response is larger than a threshold θ_4. The response of the population of *ES* neurons tuned to different excitatory center locations *s* about the position *x,y* is denoted $ES^*_{x,y,w}$. The final response $W_{x,y,w}$ to each potential width *w* at each location *x,y* is equal to the response of the maximally excited neuron in $ES^*_{x,y,w}$. Because $ES^*_{x,y,w}$ neurons might respond to some locations immediately adjacent to the monocular region, the response of *W* neurons is modulated by weighted input from the OCC neurons such that the width neurons fire only if the location *x,y* was identified as occluded in the previous step:

$$W_{x,y,w} = \left[MAX(ES^*_{x,y,w}) - (\gamma_2 - OCC_{x,y} \times \gamma_1)\right]_0 \tag{13}$$

where γ_1 is the weight of the inhibitory connections from OCC neurons and γ_2 is the modulating factor of these connections. The response of the population of *W* neurons tuned to different widths at position *x,y* is denoted $W^*_{x,y}$ and referred to as the occlusion width profile (analogous to the disparity profile) (see Figure 8).

Other metrics

Reliability

In locations devoid of texture (or with a repetitive texture), a maximum $C^*_{x,y}$ response might be obtained for several disparities. Although response magnitude at these disparities might be large (high match goodness), these disparity estimates are not reliable. The reliability metric used here is a measure of how reliable or robust disparity estimates are for a particular location. It is somewhat similar to the "peak ratio metric" used to predict locations at which potential false matches could be made (Egnal, Mintz, & Wildes, 2004; Little & Gillett, 1990); however, here it is used in the final computation of the 3-D surfaces as a weight on the excitatory connections between neurons. Accordingly, here we estimate reliability as the difference between the magnitude of the largest response of the population at each location *x,y* and the magnitude of the second largest response at this location (see Figure 9).

To compute reliability, first, a population of neurons computes the difference between the maximum response in the population at *x,y* among all disparities and the maximum response of the population at *x,y* among all disparities except for disparity *d*:

$$RE_{x,y,d} = MAX(C^*_{x,y}) - MAX(C^{*-d}_{x,y}) \tag{14}$$

where $C^{*-d}_{x,y}$ is the disparity profile at location *x,y* with the response to disparity *d* taken out. If the true disparity at location *x,y* is *d′* such that $MAX(C^*_{x,y}) = C_{x,y,d'}$ then all the $RE_{x,y,d}$ for $d \neq d'$ will give an output of zero. Only the neuron $RE_{x,y,d'}$ could have a response different from zero. This neuron will output the difference between the maximum and second maximum response. Consequently, reliability is computed as the maximum of all the $RE_{x,y,d}$ responses:

$$REL_{x,y} = MAX(RE^*_{x,y}) \tag{15}$$

Overall match goodness

The overall match goodness metric used here encodes the relative strength of the response of disparity detectors at each location. It is used in the final processing stage to build a 3-D surface by serving as a weight on excitatory connections between 3-D surface neurons. Match goodness for each location *x,y* is computed as the ratio of the maximum response for a given pixel to the maximum response within the whole population:

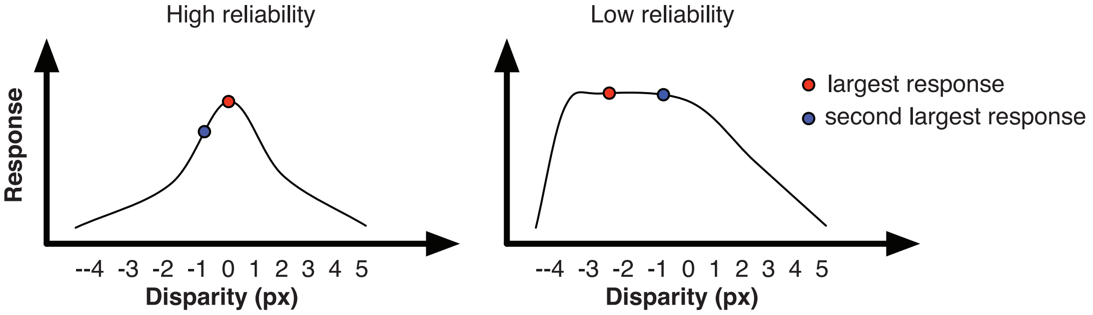

Figure 9. Computing reliability. Reliability is computed as the difference between the largest response and the second largest response of the population of neurons at each location. This difference is high when the response curve is steep as is shown in the left-hand graph. In this case, the disparity estimate can be considered reliable. On the other hand, when the curve is flat (or has more than one peak of the same magnitude) because the population is responding similarly to many different disparities, the difference between the two points is small. In this case, the disparity estimate is not reliable.

$$OMG_{x,y} = \frac{MAX(C^{*}_{x,y})}{M^{*}} \tag{16}$$

Edge detection

Disparity and luminance edges are detected and used to control the spreading of the disparity signal from one object to another. Disparity and luminance edge maps are combined to create one edge map as described below.

Disparity edges

Disparity edges are computed using simple on-off neurons. The RFs of these bipartite neurons have an excitatory half and an inhibitory half and a vertically elongated shape. They operate on the output of the occlusion detection neurons (OCC). They respond optimally when their excitatory region (area ER) is positioned on a disparity edge (OCC neurons are "on" at these locations) and their inhibitory region (area IR) is located just off this edge.

$$DE_{x,y} = \left[\sum_{x',y' \in ER} OCC_{x',y'} - \sum_{x',y' \in IR} OCC_{x',y'} \right]_{\theta_5} \tag{17}$$

Luminance edges

Luminance edge detectors $LE_{x,y}$ compute the 2-D spatial gradients in an image and so emphasize regions of high spatial frequency that correspond to edges. They are modeled after the well-known Sobel edge detector (Danielsson & Seger, 1990), and thus, the details are omitted here for brevity. The luminance edge map is computed using one of the two images depending on the direction of the disparity computation (i.e., if direction is left-to-right, then the left image is used to compute luminance edges).

Combined edges

Finally, luminance and disparity edges are added to produce a combined edge map:

$$ED_{x,y} = DE_{x,y} + LE_{x,y} \tag{18}$$

The edge map is then used to perform rough object segmentation. It is assumed here that object segmentation is performed by a higher-level process that sends feedback connections to the neurons in the early visual areas. Feedback modulation based on figure-ground relationships and object segmentation is a well-documented physiological phenomenon (Angelucci et al., 2002; Hupe et al., 1998; Schoenfeld et al., 2003). Because object segmentation is a very complex process and it is beyond the scope of this work, it is not implemented in a biologically plausible way here. For the purposes of the model formulation, the output of the higher-level neurons is denoted as $OBJ_{x,y,x',y'}$. Its output is positive when two locations x,y and x',y' belong to the same object and zero otherwise:

$$OBJ_{x,y,x',y'} = \begin{cases} 1 & \text{if } x,y \text{ and } x',y' \text{ belong to the same object} \\ 0 & \text{if } x,y \text{ and } x',y' \text{ belong to different objects} \end{cases} \tag{19}$$

Three-dimensional surface neurons: Final computation of disparity

In the final stage of the model, the 3-D surface is constructed by aggregating information from all previous stages in an iterative manner (to allow for disparity interpolation and propagation) and assigning final disparities to both monocular and binocular regions. Disparities are computed differently for locations that were identified as monocularly occluded and those that were identified as binocular. The 3-D surface neurons, $3D_{x,y,d}$, sum weighted responses of two types of neurons: $MON_{x,y,d}$, which compute the depth for monocularly occluded regions, and $BIN_{x,y,d}$, which compute the depth for binocular regions. Both types of neurons exist at each location $x,y,$ and their output is modulated by inhibitory influences such that $BIN_{x,y,d}$ is suppressed if x,y is identified as a monocular location and $MON_{x,y,d}$ is suppressed when x,y is a binocular location:

$$3D_{x,y,d} = n\Big(\big[BIN_{x,y,d} - OCC_{x,y}\times\gamma_1\big]_0 + \big[MON_{x,y,d} - (\gamma_2 - OCC_{x,y}\times\gamma_1)\big]_0\Big) \quad (20)$$

where γ_1 is the weight of the inhibitory connections from OCC neurons and γ_2 is the modulating factor of these connections. In all cases below in which the output of the $3D_{x,y,d}$ neurons is used at the first iteration, when this output is just background noise, the disparity profiles are taken instead from the corresponding $C_{x,y,d}$ complex energy neurons. In other words, for the first iteration:

$$3D_{x,y,d} = C_{x,y,d} \quad (21)$$

Reliability and overall match goodness are recomputed as specified in Equations 15 and 16 using the output of the $3D_{x,y,d}$ instead of the output of $C_{x,y,d}$ neurons after each iteration.

Depth in monocular areas

For monocularly occluded locations, equivalent disparity is derived from the width of the monocular region and the eye of origin (occlusion geometry) along with the disparity information in neighboring binocular areas. We assume that the monocular region results from occlusion (not from camouflage). The disparity estimates for occluded regions are obtained by collecting support from neighboring binocular regions via $BS_{x,y,d}$ neurons and the neighboring monocular regions via $MS_{x,y,d}$ neurons:

$$MON_{x,y,d} = BS_{x,y,d} + MS_{x,y,d} \quad (22)$$

Support from binocular areas

The contributing binocular regions are areas of set size $H \times W$ (depending on the maximum disparity the energy neurons are tuned to) to the left, right, above, and below the monocular region, labeled as NL, NR, NA, and NB, accordingly. Before being added, the outputs of $3D_{x,y,d}$ neurons in each region NX are weighted by reliability $RE_{x,y,d}$ and summed to obtain $3D_{\Sigma_{NX,d}}$ (Equation 24). Only neurons at binocular positions x,y contribute to $3D_{\Sigma_{NX,d}}$ as input from positions detected as monocular is inhibited. For an occlusion in the left eye, the response is computed as follows:

$$BS_{x,y,d} = 3D_{\Sigma_{NL,d}} + 3D_{\Sigma_{NA,d}} + 3D_{\Sigma_{NB,d}} + MAX(3D_{\Sigma_{NR,d'}} \times W_{x,y,w}) \quad (23)$$

where $d',w \in [d'+w=d]$ and $3D_{\Sigma_{NX,d}}$ are defined as follows:

$$3D_{\Sigma_{NX,d}} = \sum_{x,y\in NX} \big[3D_{x,y,d} \times RE_{x,y} - OCC_{x,y} \times \gamma_1\big]_0 \quad (24)$$

where γ_1 is the weight of the inhibitory connections from OCC neurons.

Because an occlusion arrangement is assumed, in Equation 23, the disparity signals from NL, NA, and NB (left, above, and below the monocular region) are taken without adjustment because the occluded area is assumed to be coplanar with these regions as shown in Figure 10. On the other hand, region NR, to the right of the occlusion, is assumed to be the occluding edge. Because according to occlusion geometry the disparity of the occlusion is equal to the disparity of the occluding edge plus the occlusion width (Figure 10), the $3D_{\Sigma_{NR,d}}$ input is shifted by the occlusion width before it is integrated with the rest.

Because the responses of all neurons in the model are distributed, we cannot simply shift the disparity profile $3D^*_{\Sigma_{NR}}$ by the width that generated the largest response in the width profile $W^*_{x,y}$ because that would force a switch to a binary representation. Instead, the width and the disparity profiles of the NR region have to be carefully combined in order to be added to the new disparity profile for location x,y. Each possible disparity d can be obtained through several combinations of disparities d' and widths w. For example, disparity $d = 5$ can result from $w = 3$, $d' = 2$; $w = 4$, $d' = 1$; $w = 5$, $d' = 0$; and other combinations. Consequently, for each disparity d, support is collected from all possible combinations of w and d' by multiplying $W_{x,y,w}$ with $3D_{\Sigma_{NR,d'}}$. The maximum response of these possible w and d' combinations is taken as the response to each disparity d. This part of the computation in Equation 23 is illustrated in detail in Figure 11.

Support from monocular areas

Support from neighboring monocular locations is collected by summing the disparity profiles of the locations within the support region SR that are identified as occluded. The contributions from each location within the support areas are weighted by a 2-D Gaussian, $Gauss_{x,y,\sigma_1}$ with standard deviation σ_1

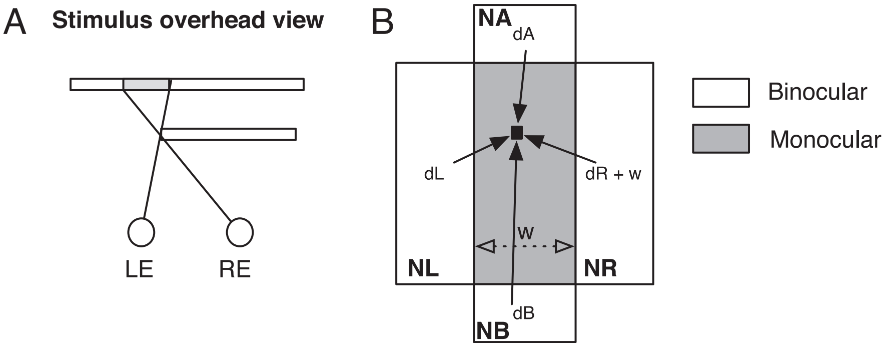

Figure 10. Collection of support from binocular areas during the computation of monocular regions' disparity. In all panels, binocular areas are shown in white and monocular areas in grey. (A) A bird's eye view of a foreground surface occluding some of the background in the right eye, creating a monocular occlusion of width *w*. (B) The black square is a point inside this monocular region surrounded by binocular regions: *NL*, *NR*, *NA*, and *NB* (left, right, above, below). Regions *NL*, *NA*, and *NB* contribute their disparity signals, *dL*, *dA*, and *dB* unaltered as it is assumed that they are coplanar with the occluded monocular region. The disparity *dR* of region *NR* is adjusted by the width *w* of the occluded region because it is assumed that *dR* is the occluding edge.

centered on x,y and by the reliability and match goodness of each particular neighboring location. Moreover, locations falling within the support neighborhood but belonging to a different object are inhibited by feedback from a higher-level process via neurons $OBJ_{x,y,x',y'}$ and do not contribute to the computation (see section "Combined edges"):

$$\begin{aligned} MS_{x,y,d} &= \sum_{x',y' \in SR} \big[3D_{x',y',d} \times Gauss_{x',y',\sigma_1} \times RE_{x',y'} \\ &\qquad \times MG_{x',y'} \times OBJ_{x,y,x',y'} \\ &\qquad - (\gamma_2 - OCC_{x,y} \times \gamma_1)\big]_0 \end{aligned} \tag{25}$$

Propagation of disparity from monocular to binocular regions

Once the final disparity profile of an occluded location is established, the disparity signal can propagate into the binocular area that is supposed to occlude the monocular region if the reliability of the occluding binocular area is low (Figure 12). Disparity from occluded areas is propagated only horizontally and is stopped when either the next edge (luminance or disparity) is reached or reliability increases beyond a certain threshold. The propagation is mediated through dedicated neurons $PROP_{x+s,y,d}$.

$$\begin{aligned} PROP_{x+s,y,d} = \big[&\max(3D_{x,y,d'} \times W_{x,y,w}) \\ &- (\gamma_2 - OCC_{x,y} \times \gamma_1) - ED_{x+s,y} \times \gamma_3 \\ &- RE_{x+s,y} \times \gamma_4\big]_0 \end{aligned} \tag{26}$$

where d', $w \in [d' + w = d]$ and γ_3 and γ_4 are the weights on the connections with edge detectors and reliability-computing neurons accordingly.

The disparity signal propagating from the monocular into the binocular areas has to be adjusted by the width of the monocular area according to occlusion geometry. As shown in Figure 12, the disparity of a right occluding edge should be equal to the disparity of the occluded region minus the width of the occluded region (assuming negative values for crossed and positive values for uncrossed disparities). The adjustment of the disparity of the occluded region by the width of the occluded region in Equation 26 is done in the same way as described in the previous section and shown in Figure 11.

Depth in binocular areas

The depth of binocular locations x,y is computed by accumulating support from the binocular locations in the surrounding region, SR, and the disparities propagated by the monocular occlusion regions via $PROP_{x+s,y,d}$ (if any). In both cases, the support is weighted by a 2-D Gaussian, $Gauss_{x,y,\sigma_2}$, with standard deviation σ_2 centered on x,y and by the reliability and match goodness of each support region. Locations falling within the support region but belonging to a different object are inhibited by feedback from a higher-level process via $OBJ_{x,y,x',y'}$ and do not contribute to the computation (see section "Combined edges"):

$$\begin{aligned} BN_{x,y,d} = &\sum_{x',y' \in SR} \big[3D_{x',y',d} \times Gauss_{x',y',\sigma_2} \\ &\qquad \times RE_{x',y'} \times MG_{x',y'} \times OBJ_{x,y,x',y'}\big] \\ &+ \sum_{x',y' \in SR} \big[PROP_{x',y',d} \times Gauss_{x',y',\sigma_2} \\ &\qquad \times \left(\frac{1}{RE_{x',y'}}\right) \times OBJ_{x,y,x',y'}\big] \end{aligned} \tag{27}$$

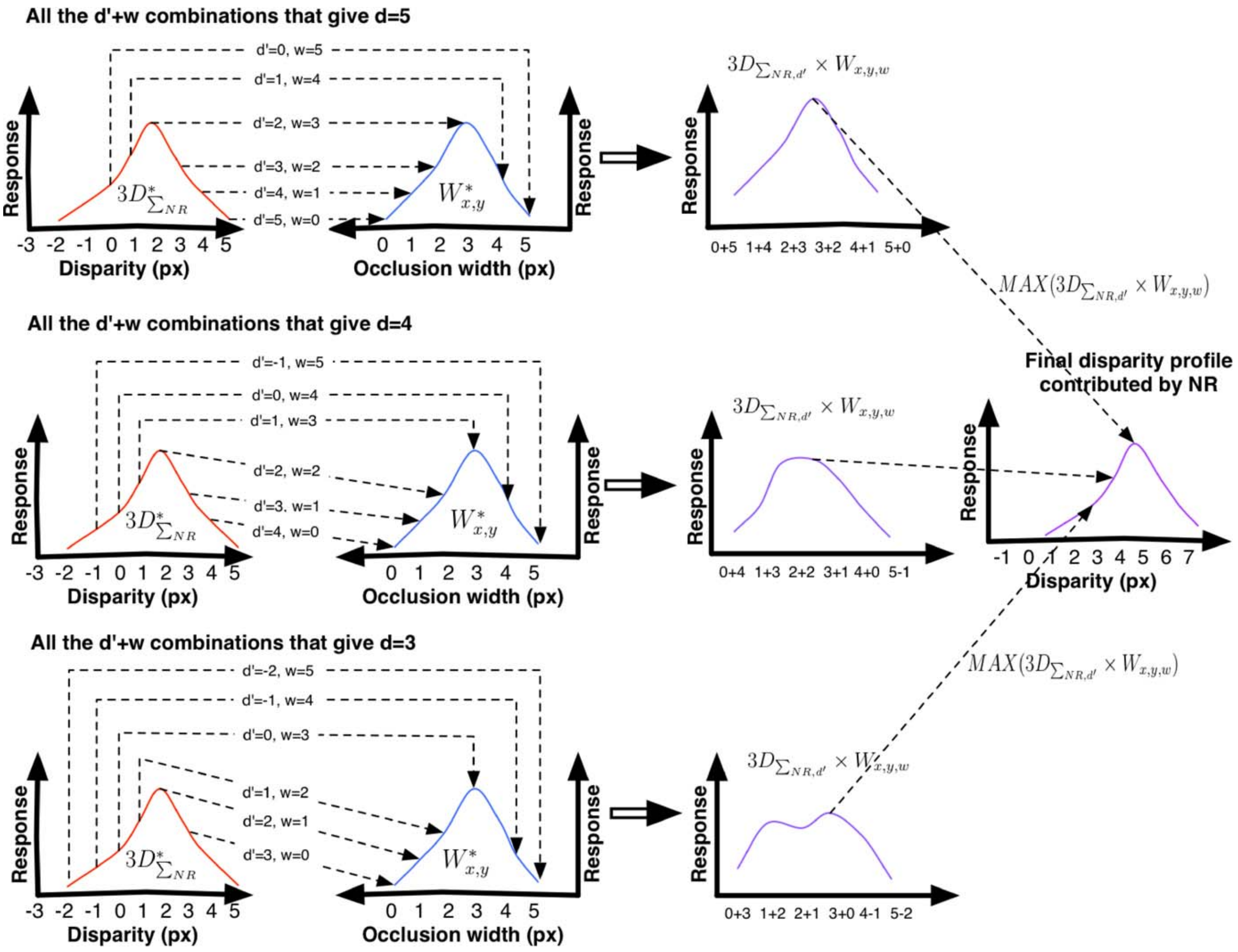

Figure 11. Using occlusion geometry to assign disparity in occluded regions. The red curves show the summed disparity profile $3D^*_{\Sigma_{NR}}$ region, which is assumed to occlude position *x,y*. The blue curves show the width profile of location *x,y*. According to occlusion geometry, the width of the occluded region must be added to the disparity of the occluding edge to produce the disparity of the occluded region. Because disparity and width representations are distributed, all possible combinations of widths and disparities must be considered. The left column shows the different combinations of widths *w* and disparities *d*′ that produce disparities $d = 5$, $d = 4$, and $d = 3$. For each possible combination, the $3D_{\Sigma_{NR,d'}}$ signal is multiplied by the appropriate $W_{x,y,w}$ signal as shown in the middle column. Then the maximum combination response for each disparity *d* is selected to produce the final disparity profile shown in the rightmost column.

Disparity assignment

Final depth maps were computed from the response of the $3D^*_{x,y}$ neurons after all model iterations were completed using the zero-bias method described in Tsirlin, Wilcox, et al. (2012). In this method, the disparity corresponding to the maximum response of the population is taken as the true disparity at each point. If more than one disparity generates the maximum response, the disparity closest to zero is used as the final disparity. This method is motivated by psychophysical studies showing that in ambiguous cases the visual system tends to prefer small disparities over larger ones (Banks & Vlaskamp, 2009; Brewster, 1847; McKee, Verghese, Ma-Wyatt, & Petrov, 2007; Prince & Eagle, 2000). Moreover, this method gave model estimates that were closest to those of observers in the experiments of Tsirlin, Wilcox, et al. (2012).

Implementation details

The model and the script for stimulus generation were implemented in MATLAB 7.10 running on Mac OS X Version 10.7 on a 2.8 GHz Intel Core MacBook Pro.

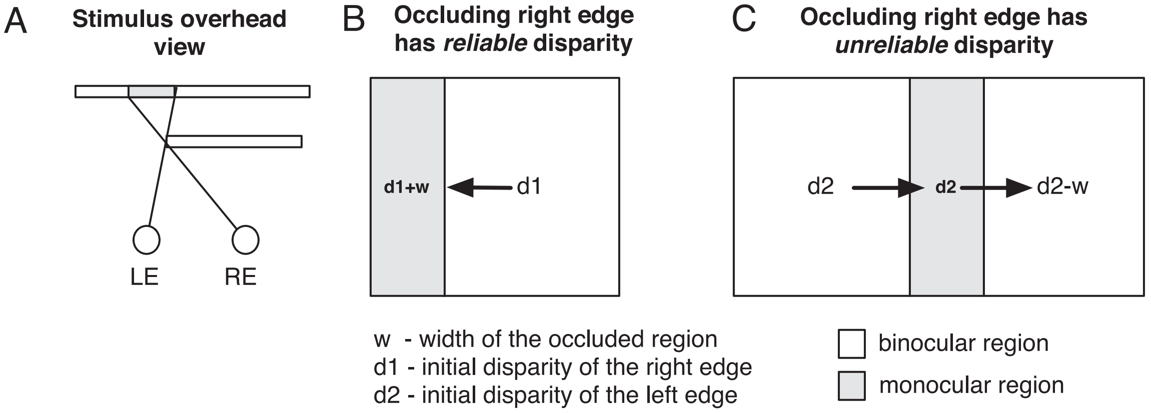
A
Stimulus overhead view
LE
RE
B
Occluding right edge has *reliable* disparity
d1+w
d1
C
Occluding right edge has *unreliable* disparity
d2
d2
d2-w
w - width of the occluded region
d1 - initial disparity of the right edge
d2 - initial disparity of the left edge
□ binocular region
▨ monocular region

Figure 12. Computing and propagating disparity in monocular regions. (A) A bird's eye view of a foreground surface occluding some of the background in the right eye, creating a monocular occlusion of width *w*. In all panels, binocular areas are shown in white and monocular areas in grey. (B) When the occluding surface has a reliable disparity, its disparity *d*1 is used to compute the disparity of the monocular region, $d1 + w$. In (C), the occluding surface does not have a reliable disparity (e.g., no texture or luminance-defined edges). In this case, the disparity of the monocular region, *d*2, is estimated using the disparity of the binocular area to the left of it. Then the disparity of the unreliable area to the left of the monocular occlusion is determined by using the occlusion width, $d2 - w$, then propagated over the right edge.

Model evaluation

Test image battery

The model was tested on a large battery of images (Figures 13 through 16) that included the majority of stimulus types used in the psychophysical studies of da Vinci stereopsis. Our choice of test stimuli was motivated by our desire to directly relate the model performance to human perception under a broad range of conditions. For each of the eight monocular occlusion stimulus types, three occlusion widths were tested to verify whether the model could predict the dependence of perceived depth on the width of the occluded region (see Introduction and Figure 1). We also added two types of richly textured stimuli that have often been used to test stereo algorithms and models, namely a dense random-dot stereogram (RDS) and a stereo photograph of a map from the Middlebury database (Scharstein & Szeliski, 2002). For the RDS stimulus, three disparities were tested. All synthetic images were generated using custom MATLAB scripts together with disparity and occlusion ground truth maps. For each of the synthetic images, we selected a region of interest at which depth was postulated to be perceived on the basis of monocular occlusions (or disparity in RDS). The rectified image and the ground truth of the map photograph were taken from the Middlebury database. In total, there were 28 images in the test battery (eight monocular occlusion stimuli × three separations + RDS stimulus × three disparities + the map photograph).

Note that, for many of the images, there are no objective ground truth maps, so the ground truth maps are either based on theory, empirical data, or both. For example, it is not clear what disparity should be assigned to textureless background surfaces in most of the synthetic images in Figures 13 through 16 because many disparities would elicit a maximum response from the population of disparity detectors. We chose to assign zero disparity to these regions because the visual system has a small disparity bias in ambiguous regions (Banks & Vlaskamp, 2009; Brewster, 1847; McKee et al., 2007;

Parameter	Value	Parameter	Value
Energy neurons (EN) preferred spatial frequencies	0.0625, 0.0877, 0.125, 0.1786, and 0.25 pixels/°	θ_5	2
EN preferred orientations	0°, 30°, 60°, 120°, 150°	γ_1	10
EN preferred disparities ($-dm$, dm)	-30 to $+30$ pixels	γ_2	8
EN RF aspect ratio	2	$H \times W$	2×15
θ_1	0.3	γ_3	10
θ_2	0.3	γ_4	10
θ_3	1	σ_1	15
θ_4	0.2	σ_2	5

Table 1. Model parameters used in testing.

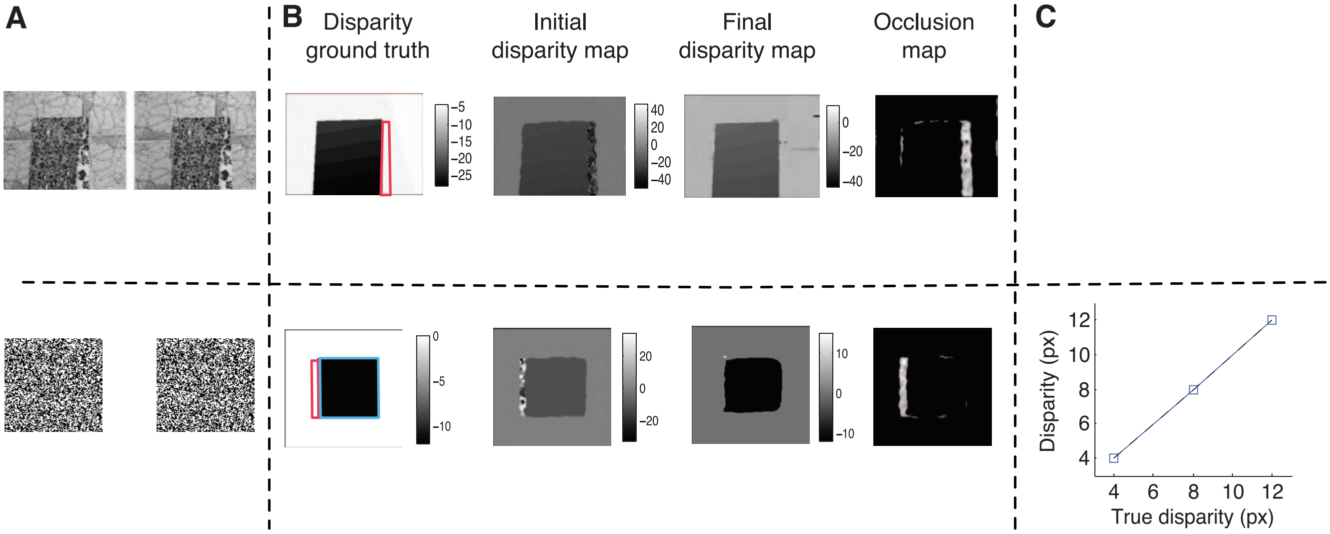

Figure 13. Model results for densely textured images. (A) Stimulus. (B) Results for disparity 8px for the RDS. (C) Mean disparity estimates of the central square of the RDS for three different disparities.

Prince & Eagle, 2000), and this preference is also reflected in the model in the form of the zero-bias disparity selection method. Moreover, it has been shown that disparity can be extrapolated from binocular features to regions lacking explicit disparity information (Takeichi, Watanabe, & Shimojo, 1992), and in the synthetic (non-RDS) stimuli used here, the binocular features have zero disparity. For the monocular gap and monocular intrusion stimuli, there are several possible ground truth maps as there are several interpretations

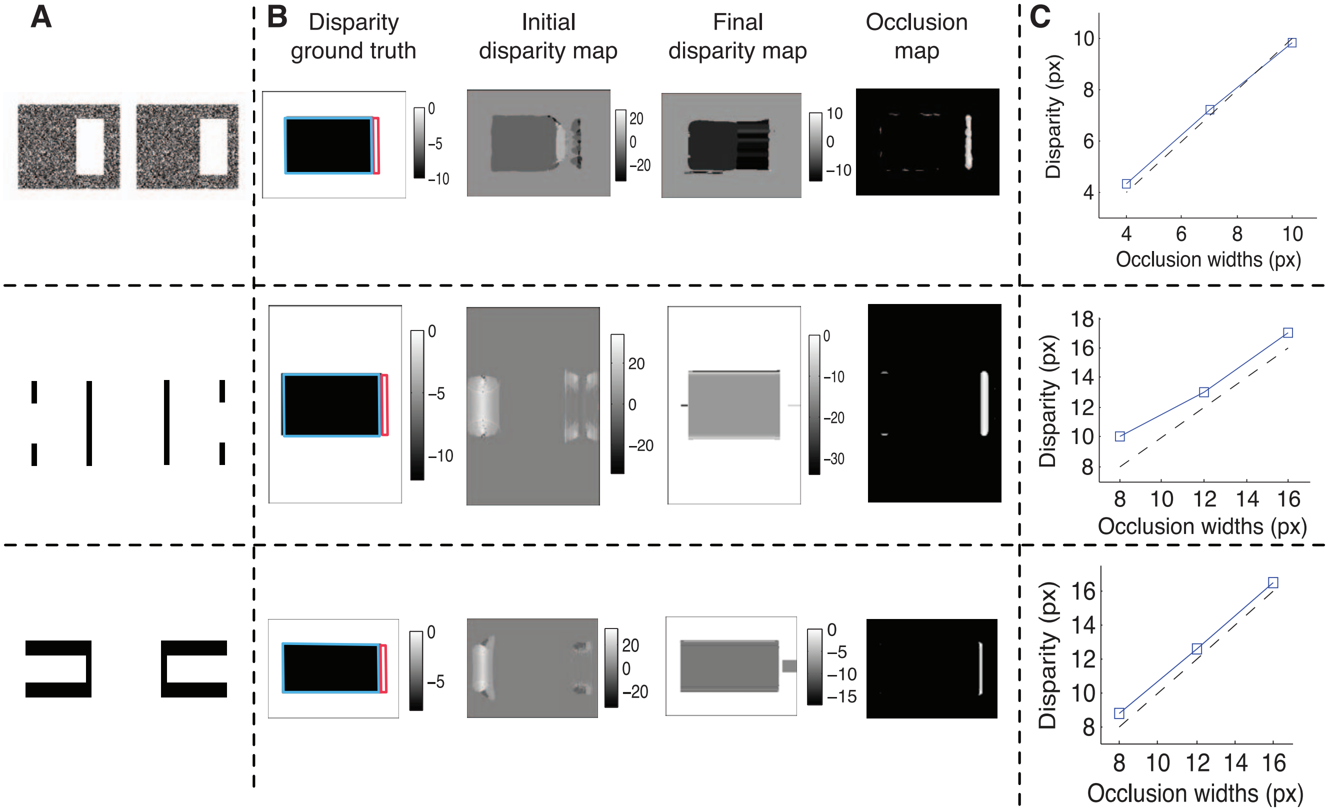

Figure 14. Model results for illusory occluder stimuli. (A) Stimulus. (B) Results for disparity −10px and occlusion width 10px for the RDS and occlusion width 12px for the other stimuli. (C) Mean disparity estimates of the region of interest (illusory occluder) for three different occlusion widths. The stereograms are arranged for crossed fusion.

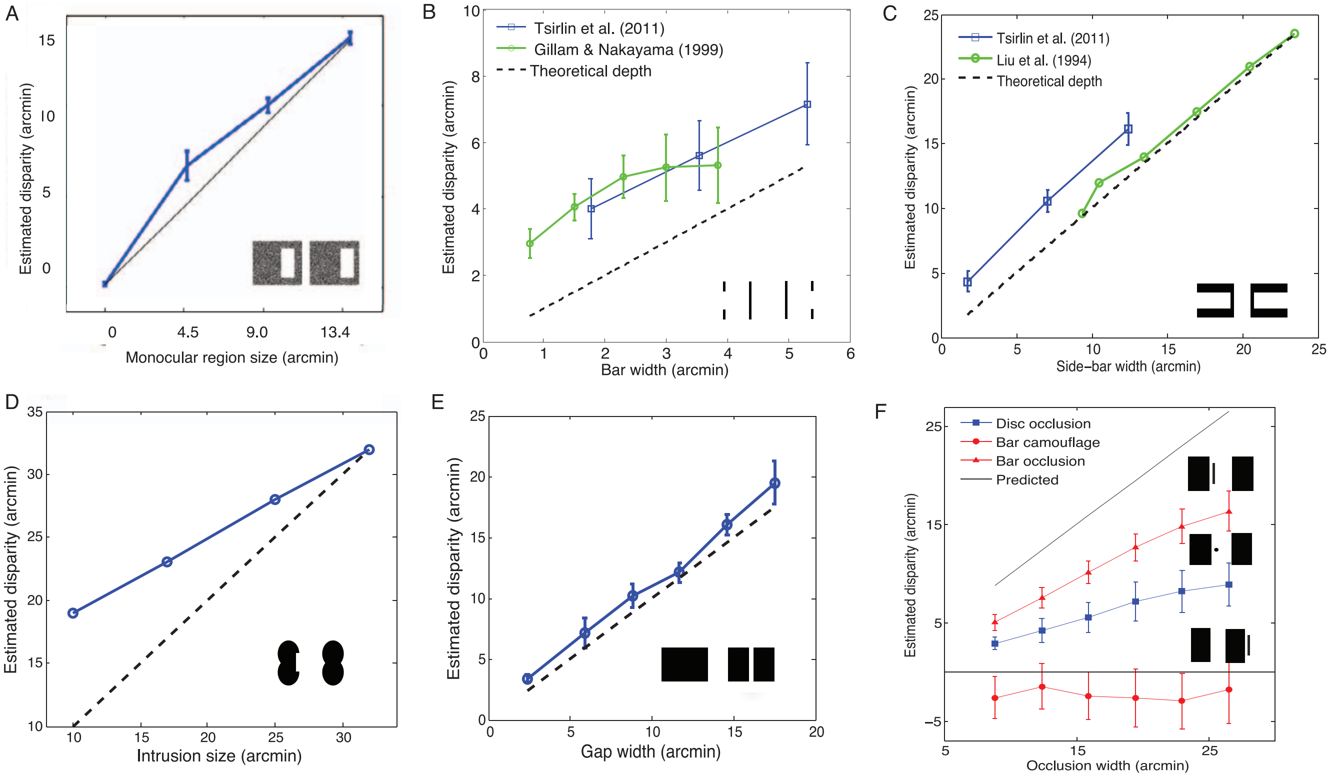

Figure 15. Psychophysical results displayed for comparison with the model. In all cases, estimated disparity (or depth) is plotted as a function of occlusion width. (A) Results for the illusory occluder stimulus of Tsirlin et al. (2010) adapted from Tsirlin et al. (2010) (means for all observers). (B) Results for the illusory occluder stimulus of Gillam and Nakayama (1999) adapted from Tsirlin et al. (2011) and Gillam and Nakayama (1999) (means). (C) Results for the illusory occluder stimulus of Liu et al. (1994) adapted from Tsirlin et al. (2011) (means) and Liu et al. (1994) (only one observer). (D) Results for the monocular intrusion stimulus adapted from Gillam et al. (1999) (means). (E) Results for the monocular gap stimulus adapted from Cook and Gillam (2004) (means). (F) Two-object arrangements with a bar and disc in occlusion configuration and with a bar in camouflage configuration. Adapted from Tsirlin, Wilcox, et al. (2012) (means).

consistent with the stimulus configuration (and observers report several different percepts). In the analysis of the model results for these stimuli, we use the ground truth map that is the closest to the model output.

Methods

The same model parameters were used to test all stimuli. The parameters are listed in Table 1 in order of their appearance in the text. The parameters pertaining to the energy model are based on previous modeling and physiology data. Several parameters, such as $H \times W$ and σ_1 and σ_2, depend on the disparity range ($-dm$, dm) supported by the model. Other parameters were estimated empirically.

To eliminate the effect of simple edge detection on model performance and to obtain the best results possible, the edge maps used in the computations were precomputed. This is a reasonable alteration because edge detection is not the focus of the model. In pilot experiments, in which the simulations were run with the edges computed by edge detectors, the results were close to those reported here but somewhat noisier in two cases with textured images (having many small edges). The overall error rate for the final depth map for the textured illusory occluder stimulus of Tsirlin et al. (2010) was larger by 2% and for the map image by 10% when the edges were computed by edge detectors. The differences for other stimuli were less than 1%. The rates of true positives and false positives in occlusion detection were exactly the same for both methods of edge estimation because occlusion detection does not depend on edge detection in the model.

Results

The maps computed by the model for each stimulus type are shown in Figures 13 through 16. All figures show the stimuli (As), detailed results of simulation trials with one of the occlusion widths (or disparities) (Bs), and plots showing the estimated mean disparity in the regions of interest for all occlusion widths (or

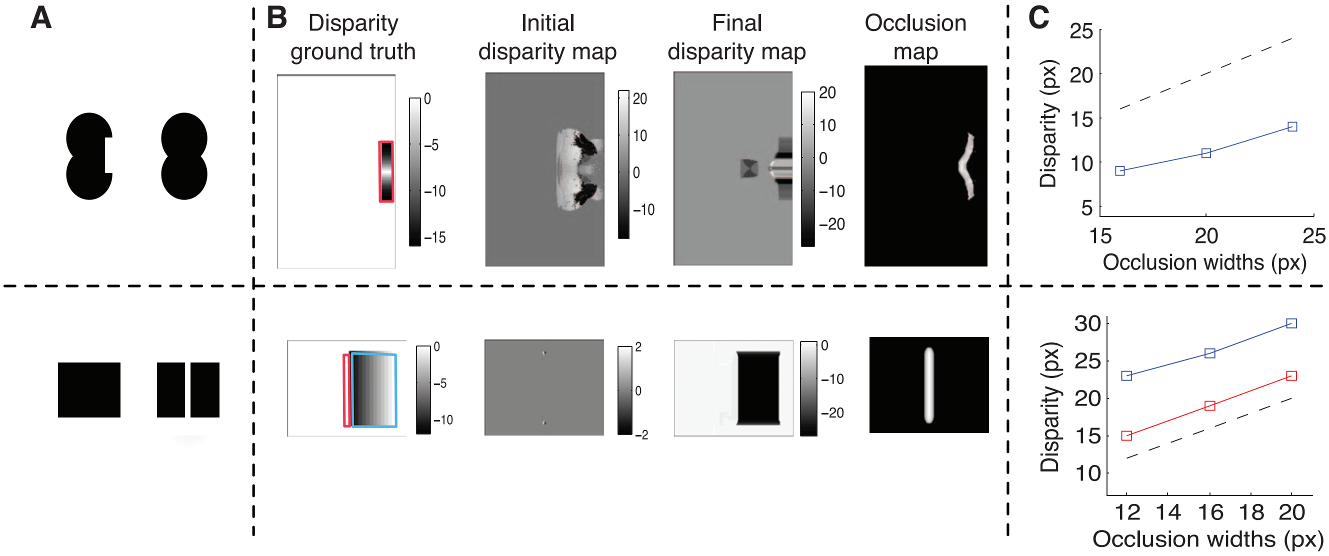

Figure 16. Model results for the monocular intrusion (top) and the monocular gap (bottom) stimuli. (A) Stimulus. (B) Results for occlusion width 25px for the intrusion stimulus and 12px for the gap stimulus. (C) Results for three different occlusion widths. The stereograms are arranged for crossed fusion.

disparities) (Cs). In each C plot, the hatched line shows the disparity predicted based on viewing geometry. Each detailed results plot B depicts:

- The ground truth map on which darker colors indicate depths closer to the observer
- The initial disparity map output by the energy neurons on which darker colors indicate depths closer to the observer
- The final disparity map output by the model on which darker colors indicate depths closer to the observer
- The computed occlusion map on which brighter colors indicate greater probability the pixel is occluded

The reported rates of true positives for occluded pixels are computed as the ratio of occluded pixels detected by the model to the total number of occluded pixels. The reported rates of false positives are computed as the ratio of binocular pixels signaled as occluded to the total number of occluded pixels. The two ratios represent an average of the ratios for the three occlusion widths/disparities unless specified otherwise.

On average, the complete model reduced the percentage of pixels with incorrect disparities by half compared to initial estimates made by the energy neurons alone. The overall rate of true positives in occlusion detection, averaged across all images, was 80%, and the rate of false positives was 30%. This is comparable to the rates of the best occlusion-detection methods evaluated in Egnal and Wildes (2002). Note that this level of performance of the DMOG model was obtained with the same set of parameters for all images. Performance could be improved further by selecting an optimal set of parameters for each individual stimulus.

Densely textured synthetic and natural stereograms

The model performed quite well with densely textured images, both synthetic and natural. Figure 13 (top) shows the model output for a photograph showing a map leaning against a textured background. The initial output of the energy neurons provided very noisy estimates in the monocularly occluded region near the right edge of the map. The DMOG model improved on this result by detecting the monocular region (85% true positives, 11% false positives) and assigning it a proper disparity.

Model results for the RDS are shown in Figure 13 (bottom). As expected, the initial disparity estimates provided by the energy model were accurate in all areas except for the monocularly occluded region to the left of the central square, which is characterized by spurious matches. The model corrected this problem by detecting the occluded pixels (80% true positives, 12% false positives) and assigning correct disparities in these locations (results were similar in RDS with an uncrossed disparity of the central square). As shown in Figure 13C, mean model estimates of the disparity of the central square corresponded to the predicted ones.

Illusory occluder stimuli

The test battery contained three examples of images in which the presence of monocular regions gave rise to illusory occluding surfaces. As expected, for the stimulus of Tsirlin et al. (2010) (see Figure 14, top), the initial estimates of the energy model for the textured areas were accurate, but the estimates for the central blank region were noisy and did not correspond to the percept of an illusory occluder. The model improved on these results by detecting the monocularly occluded region (71% true positives, 30% false positives) and propagating crossed disparity across the blank region, reconstructing the illusory occluder. Figure 14C shows that the disparity estimates for the illusory occluder increased as the occlusion width increased, and the estimates lie on the predicted line. These data are in agreement with the psychophysical data reported in Tsirlin et al. (2010) and shown in Figure 15A. As in the model results, in the psychophysical data, there is an increase in disparity estimates with an increase in occlusion width, and the estimates follow the prediction closely.

The model also performed quite well with the nontextured illusory occluder stimuli from Gillam and Nakayama (1999) and Liu, Stevenson, and Schor (1994) as shown in Figure 14 middle and bottom rows, respectively. The monocularly occluded areas were detected accurately (95% true positives, 8% false positives for both), and the illusory surfaces were fully reconstructed although the initial estimates of disparity from the energy model fell short of the ground truth. As in the psychophysical data (Gillam & Nakayama, 1999; Liu et al., 1994; Tsirlin et al., 2011), the model predicted quantitative depth in both cases as shown in Figure 14C. Moreover, the model estimations replicated both the qualitative and the quantitative aspects of observer data. This can be appreciated by comparing the model data to the psychophysical data from Tsirlin et al. (2011), Liu et al. (1994), and Gillam and Nakayama (1999) shown in Figure 15A through 15C. As can be seen in the figure, for the Liu et al. stimulus, observers' estimates were closer to the predicted disparity values (particularly in the Liu et al. data) than for the Gillam and Nakayama stimulus, in which disparity was overestimated in agreement with the model's performance. This difference is likely the result of a disparity signal present in the corners of the Liu et al. stimulus as was discussed by Gillam (1995) and as can be seen in the initial disparity map (the small dark patches in the corners of the figure). This information helps the model assign a more precise disparity estimate to the illusory surface. There were minor artifacts in which disparity was propagated beyond the illusory surface. These most likely resulted from stray pixels being identified as occluded.

Monocular intrusion and monocular gap

The results for the monocular intrusion stimulus (Cook & Gillam, 2004) are shown in Figure 16 (top). The initial estimate of the energy model showed a small curved surface on the edge of the figure eight with crossed disparity along the curviest points of the figure eight and a disparity close to zero at the midline of the figure eight. This is consistent with a one-to-one matching of the curved contour in one eye to the straight contour in the other eye as proposed by Tsirlin, Allison, et al. (2012) and Tsirlin, Wilcox, et al. (2012). However, the disparity map was noisy and did not extend to the right of the figure eight (creating an occluding surface) as it does phenomenologically when the figure is viewed stereoscopically. The model detected the narrow contour of the side of the figure eight as occluded but not the whole intrusion (true positives 54%, false positives 20%). This occurred because the match goodness of the inside area of the occlusion was quite good, and so it was not signaled as occluded. The curvature of the surface was preserved; however, the top and bottom parts of the occluder had lower disparities than those in the ground truth (partially due to the initial lower estimates given by the energy neurons). Disparity was correctly propagated toward the right of the image, reconstructing an illusory intrusion. Model estimates for this stimulus were lower than the observer estimates shown in Figure 15D. However, the model correctly predicted the increase in perceived depth with increasing intrusion width as shown in Figure 15C (figure shows the maximum disparity value within the region of interest).

The results for the monocular gap stimulus (Gillam, Blackburn, & Nakayama, 1999) are shown in Figure 16 (bottom). Here the model correctly detected the occluded area; however, its width was overestimated (93% true positives, 66% false positives), which resulted in larger disparity estimates for the side to the right of the gap. The estimated disparity also did not decrease with eccentricity toward the zero disparity right edge as it did in the ground truth. This might be an artifact of the method of final disparity selection employed in the model. Only the disparity that generates the maximum response is chosen as the true disparity of each pixel while two or more similar peaks might exist in a disparity profile. In fact, as closer examination showed, the disparity profiles in the region of interest of the monocular gap stimulus did show two peaks similar in magnitude. Thus, a method of final disparity selection that takes all peaks into consideration might produce the gradual change in disparity shown in the ground truth map. The psychophysical data (estimations of the disparity around the edges of the gap) for this stimulus is shown in Figure 15E. Comparing this figure with Figure 16C, bottom, it can be seen that the model correctly predicted the increase in disparity with increase in occlusion width; however, as discussed above, model-predicted disparities were overestimated. The overestimation in the model strictly depended on the threshold used in the match-goodness occlusion detection metric. The red line in Figure 16C shows that

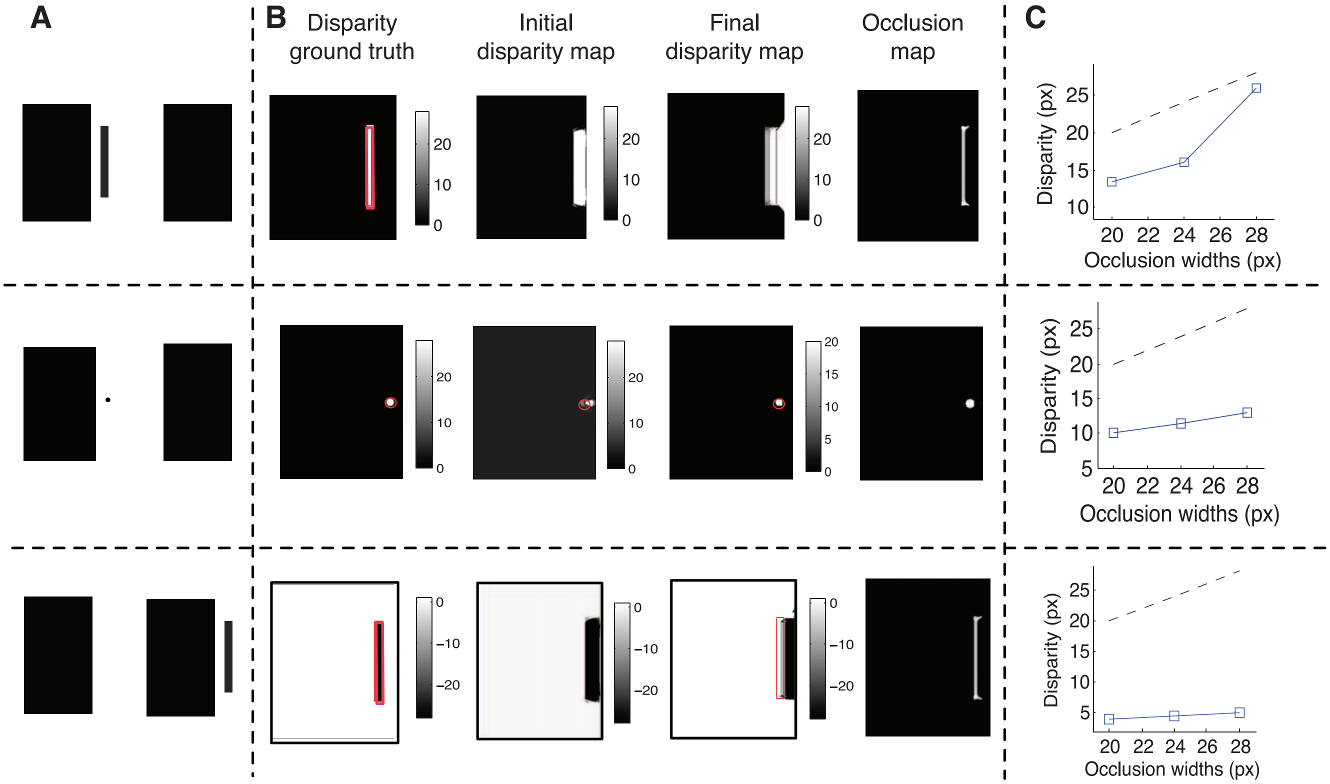

Figure 17. Model results for a two-object arrangement in occlusion configuration with a bar (top) and a disc (middle) and in a camouflage arrangement with a bar (bottom). (A) Stimulus. (B) Results for occlusion width 28 (bar) and 20 (disc). (C) Mean disparity for the region of interest (monocular object) for three different occlusion widths. The stereograms are arranged for crossed fusion.

when the threshold θ_1 (see Equation 6) was increased to 0.5, with all other parameters held constant, the model disparity estimates were much closer to the psychophysics data and the predicted disparity.

Two-object arrangements

The results for the three types of two-object arrangements with a bar (Nakayama & Shimojo, 1990) with a disc (Gillam et al., 2003; Tsirlin, Wilcox, et al., 2012) and a camouflage arrangement (Nakayama & Shimojo, 1990) are shown in Figure 17, top, middle, and bottom, respectively. In the case of the occlusion arrangement with the bar, the initial disparity map provided a fairly accurate estimate of disparity albeit encompassing an area larger than that of the bar. These estimates are based on double matching of the bar to the binocular rectangle (the uniqueness constraint is not imposed at the initial stage of the model). The monocularly occluded area detected by the model was narrower than the bar (45% true positives, 1% false positives). This occurred because the right edge of the bar has high match goodness (due to double matching), so the goodness-of-match occlusion-detection metric failed to signal these parts as occluded. The model correctly predicted an increase in quantitative depth as the occlusion width increased as shown in Figure 17C. Importantly, these disparity estimates were lower than predicted, a result that mirrors psychophysical data provided in Tsirlin, Wilcox, et al. (2012) and shown in Figure 15A (red line with triangular marks).

In the case of the disc stimulus, the outline of the disc is shown in red on all the maps to highlight that, on the initial depth map, the pocket of uncrossed disparity was located beyond the contours of the disc. However, in the final depth map, the disc was localized correctly. The occluded area was detected accurately in this case (100% true positives, 30% false positives) although it is overestimated somewhat, most likely due to the relatively low spatial resolution of the energy neurons in comparison to the small size of the disc. Figure 17C shows that the model disparity estimates increased as occlusion width increased. Importantly, disparity was underestimated quite substantially just as in the psychophysical data reported in Tsirlin, Wilcox, et al. (2012) and shown in Figure 15F (blue line with square marks). Note that, similarly to the psychophysical data (Figure 15F), the underestimation in the disc stimulus is larger than that in the bar stimulus. The disparity underestimation, in both the bar and the disc stimuli, is likely the result of two aspects of the model: (a) the monocular region detected by the model in these

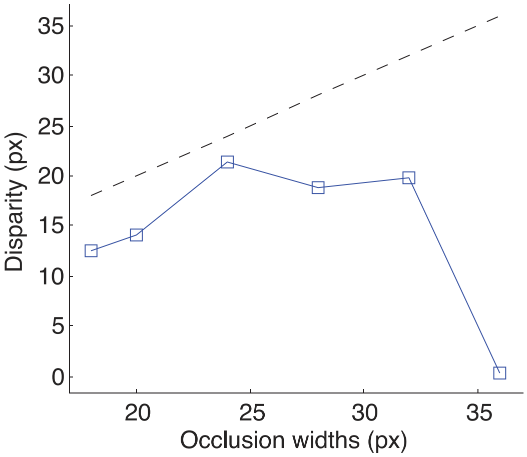

Figure 18. Mean disparity estimates for the bar stimulus computed by the model using the limited disparity range.

stimuli is restricted to the monocular object rather than the whole region between the occluder and the monocular object's outer edge, and (b) the averaging of the disparity estimates provided by the occlusion neurons and the binocular neurons in the computation of the disparity of the region of interest. The former issue is discussed in detail in the Discussion section.

The model also replicated psychophysical results obtained with camouflage arrangements shown in Figure 15A (red line with round markers). As can be seen in the figure, very little depth is perceived in this case, and there is little change in perceived depth with increase in occlusion width (Gillam et al., 2003; Nakayama & Shimojo, 1990; Tsirlin, Wilcox, et al., 2012). Column C of the bottom row of Figure 17 shows the mean disparity estimates for the monocular bar in a camouflage arrangement as computed by the model for different occlusion widths. The model estimates were very small compared to the theoretical depth and changed very little as a function of occlusion width. This occurs because the model makes the assumption that all monocular regions result from occlusion relationships. Thus, because the model detected the left side of the bar as monocular, an occluding edge was assumed to exist to the right of the monocular region. But the reliability of the area to the right of the monocular region was low due to double matching, and thus the model reconstructed an illusory occluding edge instead. As a result, the area within the boundaries of the bar (outlined in red in Figure 18) has a disparity close to zero, and the area to the right of the bar has a crossed disparity. This treatment of camouflage as occlusion is consistent with the mechanisms proposed by Tsirlin, Wilcox, et al. (2012) and to some degree by Assee and Qian (2007) to account for the absence of depth in camouflage arrangements.

We have also assessed whether the DMOG model can predict the decrease in perceived depth in two-object (occlusion) arrangements after occlusion width increases beyond a certain value (Nakayama & Shimojo, 1990). We hypothesized that this decrease is related to the size-disparity correlation. That is, stimuli of a given width can only support depth from disparity and occlusion over a specific range. In our version of the energy model, all disparities are represented at all scales due to the use of position-shift mechanisms. To test our hypothesis in the simplest way, we have modified the model to have a disparity range of ±24 px and an occlusion width range of 24 px and ran the model with two-object arrangements with a bar with occlusion widths of 18, 20, 24, 28, 32, and 36 px. The mean disparity estimates of the model for the monocular bar are shown in Figure 18. Model results had the same pattern as the psychophysical data in Nakayama and Shimojo (1990). At first, as occlusion width increased, the estimates increased as well until the maximum disparity/width represented by the population (24 px) was reached. Then, the function plateaued and eventually decreased.

Discussion

We have described a biologically plausible model of depth from disparity and monocular occlusion. This model is based on the DMOG theory—a set of principles derived from the psychophysical and computational studies of stereopsis and da Vinci stereopsis. In the DMOG model, monocular occlusions are detected explicitly and are directly used in the construction of 3-D surfaces. Depth in occluded areas is established on the basis of monocular geometry, and illusory occluding surfaces are constructed in cases in which the disparity estimates for the occluding edge have low reliability.

The proposed model offers several improvements over the existing biologically inspired models of depth from disparity and occlusion. First, the DMOG model was implemented in a biologically plausible way with a distributed representation of neuronal firing rates throughout the computation until the final selection of disparities is performed. Second, this is the first model to use the occlusion width explicitly to compute equivalent disparity in occluded areas and to proactively propagate the disparity from the occluded areas into areas of low reliability. This feature allows the model to reconstruct illusory occluders and predict depth percepts in monocular gap and monocular intrusion stimuli. Third, the model was tested on a large range of different image types, which allowed an in-depth evaluation of the model's architecture.

The DMOG model performed well on most of the test set, improving upon the initial maps provided by the disparity energy model. Importantly, in the majority of cases, it also produced disparity maps that were close to observers' percepts as reported in the psychophysical literature. It also replicated quantitative depth percepts from monocular occlusions for all stimuli in which quantitative depth has been demonstrated psychophysically. This suggests that the proposed neuronal architecture could underlie da Vinci stereopsis.

Model predictions

The DMOG model proposes that there are monocular occlusion detectors and neurons tuned to occlusion width in the visual cortex. This prediction could be tested via single-cell recordings using stimuli such as dense RDS with monocular regions. However, care has to be taken to distinguish between neurons sensitive to any uncorrelated stimuli (Poggio, Gonzalez, & Krause, 1988) (which are most likely involved in the detection of false matches) and those responding specifically to monocular occlusions.

The model also makes interesting predictions that can be tested using psychophysical methods. First, an intriguing issue arises from simulations with two-object arrangements. Theoretically, in these stimuli, the occluded area is not restricted only to the occluded object (bar or disc), but includes the space between the monocular object and the occluder. The DMOG model, in its current form, does not identify the complete textureless space between the occluded object and the occluder as occluded because its goodness of match is quite high. This results in an underestimation of the model depth estimates as seen in Figure 17. This aspect of the model might be considered a drawback; however, note that it is not clear from the psychophysical literature what area is detected as occluded by the visual system. Moreover, the underestimation of depth produced by the model is very similar to the underestimation of perceived depth in psychophysical experiments (Tsirlin, Wilcox, et al., 2012), suggesting that the visual system also labels only the monocular object as occluded. If that is indeed true, two predictions can be made about depth perception in two-object arrangements: (a) larger monocular bar widths will yield larger perceived depth even when the overall size of the monocular region remains the same, and (b) placing the two-object arrangement on a textured background should allow the identification of the complete monocular region as occluded, which will result in more accurate depth estimates.

According to the DMOG theory and model, disparity from monocular areas propagates into binocular areas when those have low reliability. Reliability is low when more than one match produces a high response from the population of energy neurons. This occurs in textureless areas such as the ones in several stimuli used in the evaluation of the DMOG model. It can also occur in areas with repeating texture (wallpaper patterns) because each element can be matched successfully to several others in the other eye. Thus, the model predicts that when a monocular region neighbors an area with a repeating pattern, the disparity computed for the monocular region can propagate into the binocular region. In fact, this is exactly what Hakkinen and Nyman (2001) found. In their experiments, they superimposed the Gillam and Nakayama (1999) stimulus (see Figure 2C) or a modified Kaniza figure, in which depth was perceived on the basis of occlusions, on top of a repetitive dot pattern. They found that the dots were "captured" by the depth signal provided by monocular occlusions.

Finally, the model predicts that a single object (e.g., a dot) presented to one eye only while the other eye views a uniform field would elicit a percept of depth through the creation of an illusory occluding edge. In fact, there is some evidence that stimuli of this type create qualitative depth percepts (Kaye, 1977; Wilcox, Harris, & McKee, 2007), named monoptic depth. Kaye and Wilcox et al. showed that, when one eye views a uniform field while another sees a small object positioned in this field, a percept of depth arises that depends on the position of the target with respect to fixation. Although our model does not incorporate a fixation (or viewing direction) constraint, it could be introduced in a form of adjustment of relative disparities with respect to an assumed fixation position. The model also predicts that the perceived depth in such stimuli would increase with the increase in the size of the monocular objects. This aspect of monoptic depth remains to be explored.

Future improvements

Several changes could be made to the model to improve its performance and to account for other psychophysical phenomena. Tsirlin et al. (2011) found that a binocular object placed next to an illusory occluding surface can bias the perceived depth of this surface. The DMOG model, in its current form, cannot account for this phenomenon because disparity signals do not spread beyond object boundaries in binocular areas. This constraint can be relaxed by allowing some support to propagate beyond object boundaries into neighboring regions with low reliability. Second, in its present form, the disparity maps in the model are computed as if the scene were observed from the point of view of one of the physical eyes. This is the convention for most algorithms and models of stereopsis; however, it contrasts with the popular notion of seeing the world from the cyclopean eye positioned in between the two physical eyes (Erkelens, Muijs, & van Ee, 1996; Ono, Wade, & Lillakas, 2002). In the future, this issue needs to be considered, taking into account the literature that explores the perceived visual direction near depth discontinuities (Erkelens et al., 1996; Ono, Lillakas, Grove, & Suzuki, 2003; Ono et al., 2002). Finally, the output of the complete model critically depends on the initial disparity estimation step. Improving the performance of the energy model will improve the performance of the complete model. One possibility could be to incorporate constraints on matching, such as smoothness and uniqueness, at the initial disparity computation stage.

Conclusions

We have proposed a unified theory of the underlying mechanisms of da Vinci stereopsis grounded in psychophysical and computational data. Based on this theory, we have constructed a computational model, which has been tested on a large battery of images, including dense natural scenes and several types of stimuli used to study monocular occlusions. These simulations show that the model is capable of producing results similar to those reported in the psychophysical literature for the majority of test cases. This suggests that the proposed neural architecture can underlay da Vinci stereopsis and serve as an integral part of binocular depth perception.

Keywords: monocular occlusions, binocular half-occlusions, stereopsis, energy model, computational model, depth perception, binocular vision

Acknowledgments

We would like to thank Jenny Read and Albert Rothenstein for stimulating discussions that helped to improve this work. This research was supported by NSERC grants to LW and RA and by NSERC doctoral scholarship to IT.

Commercial relationships: none.

Corresponding author: Inna Tsirlin.
Email: inna.tsirlin@sickkids.ca.
Address: Centre for Vision Research, York University, and Eye Movement and Vision Neuroscience Laboratory, The Hospital for Sick Children, Toronto, ON, Canada.

Footnote

[1] A similar metric called “match goodness jumps” is used in several computer vision algorithms (Egnal & Wildes, 2002). However, the “match goodness jumps” metric provides only the contours of the occluded regions, and our method detects complete occluded regions.

References

Allenmark, F., & Read, J. (2011). Spatial stereoresolution for depth corrugations may be set in primary visual cortex. *BMC Neuroscience, 12*(Suppl. 1), P263.

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

Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J.-M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, *22*(19), 8633–8646.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999). Neural mechanisms for encoding binocular disparity: Receptive field position vs. phase. *Journal of Neurophysiology, 82*(2), 874–890.

Assee, A., & Qian, N. (2007). Solving da Vinci stereopsis with depth-edge-selective V2 cells. *Vision Research, 47,* 2585–2602.

Banks, M. S., & Vlaskamp, B. (2009). The venetian-blind effect: A prior for zero slant or zero disparity? [abstract]. *Journal of Vision, 9*(8):45, http://www.journalofvision.org/content/9/8/45, doi:10.1167/9.8.45. [Abstract]

Brewster, D. (1847). XLVIII. On the knowledge of distance given by binocular vision. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 30*(202), 305–318.

Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision, 18,* 515–578.

Chen, Y., & Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural Computation, 16*(8), 1545–1577.

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

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience, 24,* 203–238.

da Vinci, L. (1877). *A treatise on painting* (Trans.). London: George Bell and Sons. Original publication, 1651.

Danielsson, P. E., & Seger, O. (1990). Generalized and separable Sobel operators. In H. Freeman (Ed.) *Machine vision for three-dimensional scences* (pp. 347–379). Waltham, MA: Academic Press.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Neuronal mechanisms underlying stereopsis: How do simple cells in the visual cortex encode binocular disparity? *Perception, 24*(1), 3–31.

Egnal, G., Mintz, M., & Wildes, R. P. (2004). A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image and Vision Computing*, *22*(12), 943–957.

Egnal, G., & Wildes, R. (2002). Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(8), 1127–1133.

Erkelens, C., Muijs, A., & van Ee, R. (1996). Binocular alignment in different depth planes. *Vision Research*, *36*(14), 2141–2147.

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.

Forte, J., Peirce, J. W., & Lennie, P. (2002). Binocular integration of partially occluded surfaces. *Vision Research, 42,* 1225–1235.

Gillam, B. (1995). Matching needed for stereopsis. *Nature, 373,* 202.

Gillam, B., Blackburn, S., & Nakayama, K. (1999). Stereopsis based on monocular gaps: Metrical encoding of depth and slant without matching contours. *Vision Research, 39,* 493–502.

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

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

Gillam, B., & Grove, P. (2004). Slant or occlusion: Global factors resolve stereoscopic ambiguity in sets of horizontal lines. *Vision Research, 44,* 2359–2366.

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

Grossberg, S., & Howe, P. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research, 43,* 801–829.

Grove, P., & Gillam, B. (2007). Global patterns of binocular image differences resolve the ambiguity between stereoscopic slant and stereoscopic occlusion. *Vision*, *19*(1), 1–13.

Hakkinen, J., & Nyman, G. (1997). Occlusion constraints and stereoscopic slant. *Perception, 26,* 29–38.

Hakkinen, J., & Nyman, G. (2001). Phantom surface captures stereopsis. *Vision Research, 41,* 187–199.

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

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

Hupe, J., James, A., Payne, B., Lomber, S., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787.

Kaye, M. (1977). Stereopsis without binocular correlation. *Vision Research, 18,* 1013–1022.

Lin, M. H., & Tomasi, C. (2004). Surfaces with occlusions from layered stereo. IEEE *Transactions on Pattern Analysis and Machine Intelligence, 26*(8), 1073–1078.

Little, J. J., & Gillett, W. E. (1990). Direct evidence for occlusion in stereo and motion. *Image and Vision Computing*, *8*(4), 328–340.

Liu, L., Stevenson, S. B., & Schor, C. M. (1994). Quantitative stereoscopic depth without binocular correspondance. *Nature, 367,* 66–68.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, *194*(4262), 283–287.

McKee, S. P., Verghese, P., Ma-Wyatt, A., & Petrov, Y. (2007). The wallpaper illusion explained. *Journal of Vision, 7*(14):10, 1–11, http://www.journalofvision.org/content/7/14/10, doi:10.1167/7.14.10. [PubMed] [Article]

Min, D., & Sohn, K. (2008). Cost aggregation and occlusion handling with WLS in stereo matching. *IEEE Transactions on Image Processing*, *17*(8), 1431–1442.

Mitsudo, H., Nakamizo, S., & Ono, H. (2005). Greater depth seen with phantom stereopsis is coded at the early stages of visual processing. *Vision Research, 45,* 1365–1374.

Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research, 30*(11), 1811–1825.

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.

Ono, H., Lillakas, L., Grove, P., & Suzuki, M. (2003). Leonardo's constraint: Two opaque objects cannot be seen in the same direction. *Journal of Experimental Psychology*, *132*(2), 253–265.

Ono, H., Wade, N., & Lillakas, L. (2002). The pursuit of Leonardo's constraint. *Perception, 31,* 83–102.

Pianta, M., & Gillam, B. (2003). Paired and unpaired features can be equally effective in human depth perception. *Vision Research, 43,* 1–6.

Poggio, T., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *Journal of Neuroscience, 8*(12), 4531–4550.

Prince, S. J., & Eagle, R. A. (2000). Weighted directional energy model of human stereo correspondence. *Vision Research*, *40*(9), 1143–1155.

Read, J. C. (2010). Vertical binocular disparity is encoded implicitly within a model neuronal population tuned to horizontal disparity and orientation. *PLoS Computational Biology*, *6*(4), e1000754.

Read, J. C., & Cumming, B. G. (2006). Does depth perception require vertical-disparity detectors? *Journal of Vision, 6*(12):1, 1323–1355, http://www.journalofvision.org/content/6/12/1, doi:10.1167/6.12.1. [PubMed] [Article]

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168.

Sachtler, W. L., & Gillam, B. (2007). The stereoscopic sliver: A comparison of duration thresholds for fully stereoscopic and unmatched versions. *Perception, 36,* 135–144.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision, 47*(1–3), 7–42.

Schoenfeld, M., Woldorff, M., Dzel, E., Scheich, H., Heinze, H.-J., & Mangun, G. (2003). Form-from-motion: MEG evidence for time course and processing sequence. *Journal of Cognitive Neuroscience*, *15*(2), 157–172.

Sizintsev, M., & Wildes, R. (2007). Computational analysis of binocular half-occlusions. In L. Harris & M. Jenkin (Eds.), *Computational vision in neural and machine systems* (pp. 221–236). Cambridge, UK: Cambridge University Press.

Spang, K., Gillam, B., & Fahle, M. (2012). Electrophysiological correlates of binocular stereo depth without binocular disparities. *PloS one*, *7*(8), e40562.

Takeichi, H., Watanabe, T., & Shimojo, S. (1992). Illusory occluding contours and surface formation by depth propagation. *Perception*, *21*(2), 177–184.

Tsirlin, I. (2013). *The perceptual consequences and neural basis of monocular occlusions*. Unpublished doctoral dissertation, York University, Toronto, Canada.

Tsirlin, I., Allison, R., & Wilcox, L. (2012). Is depth in monocular regions processed by disparity detectors? A computational analysis. *Journal of Vision, 12*(9):215, http://www.journalofvision.org/content/12/9/215, doi:10.1167/12.9.215. [Abstract]

Tsirlin, I., Wilcox, L., & Allison, R. S. (2010). Monocular occlusions determine the perceived shape and depth of occluding surfaces. *Journal of Vision, 10*(6):11, 1–12, http://www.journalofvision.org/content/10/6/11, doi:10.1167/10.6.11. [PubMed] [Article]

Tsirlin, I., Wilcox, L. M., & Allison, R. S. (2011). Disparity biasing in depth from monocular occlusions. *Vision Research*, *51*(14), 1699–1711.

Tsirlin, I., Wilcox, L. M., & Allison, R. S. (2012). da Vinci decoded: Does da Vinci stereopsis rely on disparity? *Journal of Vision, 12*(12):2, 1–17, http://www.journalofvision.org/content/12/12/2, doi:10.1167/12.12.2. [PubMed] [Article]

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

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

Wilcox, L. M., Harris, J. M., & McKee, S. P. (2007). The role of binocular stereopsis in monoptic depth perception. *Vision Research*, *47*(18), 2367–2377.

Appendix A

Symbol	Description
f_L and f_R	The left and the right RFs of a simple energy neuron
I	The image patch that falls on the RFs of the simple neurons
Gauss	The Gaussian function of the simple neuron RF
Sin	The sinusoid of the simple neuron RF
φ_L and φ_R	The left and the right phase shifts of the simple neuron RFs
σ_x and σ_y	The horizontal and vertical widths of $Gauss_{L/R}$
θ	The preferred orientation of the simple neuron RF
ω_0	The peak preferred spatial frequency of the simple neuron RF
C_0	Classical complex energy neuron
C	Complex energy neuron with a normalized response
$C_{x,y,d}$	Response of a population of C neurons with RFs at x,y and tuned to disparity d pooled over scales and orientations
$(-d_m, d_m)$	The range of disparities to which the complex energy neurons C are tuned
$C^*_{x,y}$	The response $C_{x,y,d}$ for all disparities $(-d_m, d_m)$, referred to throughout as the *disparity profile*
$MG_{x,y}$	Neurons computing the match goodness metric for occlusion detection
M^*	A neuron computing the maximal response of the whole population of disparity detectors
$\langle x\rangle_{\theta_n}/\langle x\rangle 0$	Rectification with respect to the threshold θ_n or 0, respectively
$C^{L*}_{x,y}$ and $C^{R*}_{x,y}$	The response of a population of complex neurons $C^*_{x,y}$ with all the left eye/right eye RFs fixed at location x,y
$n(x)$	A normalization function
$R_{x,y,d}$	Neurons computing the difference between disparity profiles $C^{L*}_{x,y}$ and $C^{R*}_{x+d,y}$
$LRC_{x,y}$	Neurons computing the left-right match correspondence
$OCC_{x,y}$	Monocular occlusion detectors combining the outputs of $LRC_{x,y}$ and $MG_{x,y}$
$ES_{x,y,w,s}$	End-stopped neurons, receiving input from $OCC_{x,y}$ neurons with an excitatory center of width w, which is shifted by s with respect to location x,y
$W_{x,y,w}$	Neurons computing the likelihood that x,y is located within a monocularly occluded region of size w
$W^*_{x,y}$	The response of a population of $W_{x,y,w}$ neurons tuned to different widths, referred to as the *occlusion width profile*
γ_n	The weight of inhibitory interneural connections
$C^{*-d}_{x,y}$	The disparity profile at location x,y with the response to disparity d zeroed
$RE_{x,y,d}$	Neurons computing the difference between the maximum response in $C^*_{x,y}$ and the maximum response in $C^{*-d}_{x,y}$
$RE^*_{x,y}$	The response of a population of $RE_{x,y,d}$ neurons with different disparities zeroed
$REL_{x,y}$	Neurons computing the reliability of disparity estimates at location x,y
$OMG_{x,y}$	Neurons computing the overall match goodness
$DE_{x,y}$	Disparity edge detectors
$LE_{x,y}$	Luminance edge detectors
$ED_{x,y}$	Combined edge detectors
$OBJ_{x,y,x',y'}$	Neurons signaling whether x,y and x',y' belong to the same object
$BIN_{x,y,d}$	Neurons computing final disparities for binocular locations
$MON_{x,y,d}$	Neurons computing final disparities for monocularly occluded locations
$3D_{x,y,d}$	Neurons computing final disparities by combining $BIN_{x,y,d}$ and $MON_{x,y,d}$ responses
$BS_{x,y,d}$	Neurons aggregating support from binocular regions around monocularly occluded locations
$MS_{x,y,d}$	Neurons aggregating support from monocular regions around monocularly occluded locations
NL, NA, NB, NR	Regions to the left, above, below, and to the right of a monocularly occluded region from where support is aggregated
$H \times W$	The height and the width of the support regions NL, NA, NB, NR
$3D_{\Sigma NX,d}$	Summed and weighted response to disparity d in the support region NX
$Gauss_{x,y,\sigma}$	A 2-D Gaussian function centered on x,y with a standard deviation σ
$PROP_{x+s,y,d}$	Neurons propagating disparity signals from monocularly occluded locations x,y to binocular locations $x+s$, y

Table 2. Symbols used in the article.