A Bayesian Approach to the Stereo Correspondence Problem

Jenny C. A. Read
jenny.read@physiol.ox.ac.uk
University Laboratory of Physiology, Oxford, OX1 3PT, U.K.

I present a probabilistic approach to the stereo correspondence problem. Rather than trying to find a single solution in which each point in the left retina is assigned a partner in the right retina, all possible matches are considered simultaneously and assigned a probability of being correct. This approach is particularly suitable for stimuli where it is inappropriate to seek a unique partner for each retinal position—for instance, where objects occlude each other, as in Panum's limiting case. The probability assigned to each match is based on a Bayesian analysis previously developed to explain psychophysical data (Read, 2002). This provides a convenient way to incorporate constraints that enable the ill-posed correspondence problem to be solved. The resulting model behaves plausibly for a variety of different stimuli.

1 Introduction

A fundamental problem facing the visual system is how to extract information about a three-dimensional world from a two-dimensional retinal image. One clue in this task is provided by retinal disparity: the difference in the position of an object's images in the left and right eyes, arising from the horizontal displacement of the eyes in space. Evidently, this calculation depends on correctly matching each feature in the left retinal image with its counterpart in the right retina. This task has become known as the correspondence problem. In a simple visual scene such as that illustrated in Figure 1, this presents little difficulty. However, in a more complicated visual scene, the correspondence problem may be highly complex and indeed is ill posed. In general, there are many possible solutions of the correspondence problem, each implying a different arrangment of physical objects. Despite this, the visual system is capable of arriving, almost instantaneously, at a judgment of disparity across a scene. This implies that the visual system must be using additional constraints to select a solution.

Bayesian probability theory (Knill & Richards, 1996) provides a natural way of framing these constraints. Bayesian models of perception typically envisage an observer attempting to deduce information about the visual scene, S, given an image I. In the context of stereopsis, S represents the location of objects in space, and I represents the pair of retinal images.

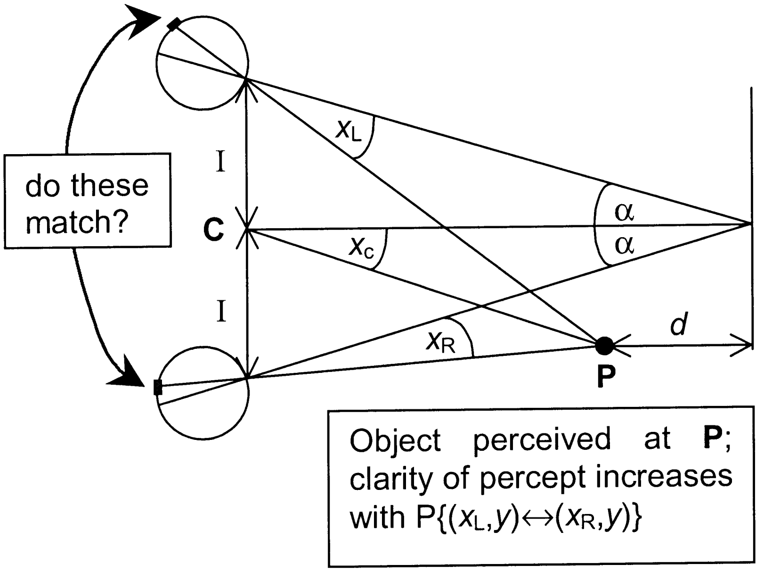
do these match?
I
C
I
x_L
x_C
x_R
α
α
F
d
P
Object perceived at P; clarity of percept increases with P{(x_L,y)↔(x_R,y)}

Figure 1: How probability relates to perception in physical space. The eyes (circles) are fixating on object **F**, whose image thus falls at the fovea in both retinas. Object **P** is in front of the fixation point, and thus has positive (crossed) disparity $\delta = x_L - x_R$, where x_L, x_R are the horizontal positions of the image of **P** in the left and right retinas. The distance d describes how far the object at **P** is in front of the fixation point **F**, while the angle x_c describes how far it is to the right of **F**. $2I$ is the interocular distance, and α is the vergence angle. Under the approximation that the fixation point is sufficiently distant and all objects viewed are sufficiently close to it that the angles α, x_L, x_R, and x_c are all small, it can be shown that $x_c \approx (x_L + x_R)/2$ and $d \approx (x_L - x_R)I/2\alpha^2$. Each potential match between a point (x_L, y) in the left retina and a point (x_R, y) in the right retina implies a percept of an object at the corresponding location in space P, with luminance depending on the mean of the light intensities recorded in the two retinas. The strength of the perception is presumed to increase monotonically with the probability $\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\}$ assigned to the match (zero probability = no perception).

Due to noise within the system, a given configuration of objects does not necessarily produce the same image on successive presentations. However, we assume that the imaging system and its limitations is well characterized, so that the brain knows the likelihood $\mathrm{P}(I \mid S)$ of obtaining an image I given a particular scene S. Furthermore, we assume that the brain has its own a priori estimate of the probability $\mathrm{P}(S)$ that a particular scene occurs. Then Bayes' theorem allows us to deduce the posterior probability of a particular scene S, given that we receive the image I:

$$\mathrm{P}(S \mid I) = \mathrm{P}(I \mid S)\mathrm{P}(S)/\mathrm{P}(I).$$

For instance, in the double-nail illusion, human observers presented with two nails in the midsagittal plane report a clear perception of two nails with zero disparity in the frontoparallel plane (Krol & van der Grind, 1980; Mallot & Bideau, 1990). This solution is preferred over the physically correct solution, even though other cues, such as size and shading, mean that the latter presumably has the higher likelihood $\mathrm{P}(I \mid S)$. One way of explaining this within a Bayesian framework is to postulate a prior preference for small disparities. Then the solution in which both nails have zero disparity is assigned a higher posterior probability than the other solution, in which one nail has crossed disparity and the other uncrossed. The zero-disparity solution is thus the one perceived.

A second advantage of a probabilistic approach is its ability to handle ambiguity. Several existing models of the correspondence problem (Dev, 1975; Nelson, 1975; Marr & Poggio, 1976, 1979; Grimson, 1981; Pollard, Mayhew, & Frisby, 1985; Sanger, 1988) employ a uniqueness constraint. That is, they seek a unique match in the right eye for every point in the left, and vice versa. This could be implemented within a Bayesian scheme by taking the correct match at every point to be that which has the highest posterior probability. However, although computationally convenient in avoiding false matches, a uniqueness constraint is clearly not satisfied in practice. Parts of the visual scenes are often occluded from one or the other eye; for instance, a stereogram consisting of a disparate target superimposed on a zero-disparity background may contain regions that have no match in the other eye. Conversely, occluding stimuli may require one point in the left image to be matched with two points in the right image. Experimental evidence suggests that the human visual system does indeed produce double matches in this situation (McKee, Bravo, Smallman, & Legge, 1995). An algorithm avoiding the uniqueness constraint is capable of deriving the correct solution (McLoughlin & Grossberg, 1998). Finally, although algorithms incorporating the uniqueness constraint have been able to solve stereograms incorporating transparency (Qian & Sejnowski, 1989; Pollard & Frisby, 1990), the uniqueness constraint certainly appears less than ideal for such stimuli (Westheimer, 1986; Weinshall, 1989).

A probabilistic theory is well suited to such situations. Let $\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\}$ denote the probability that the point (x_L, y) in the left retina corresponds to the point (x_R, y) in the right retina, that is, both points are images of the same object. High probability can be assigned to several matches for a given object without necessarily having to implement a uniqueness constraint by deciding on only one match as being "correct."

I adopt the following working hypothesis of how the match probability relates to perception, illustrated in Figure 1. If the point (x_L, y) in the left retina corresponds to the point (x_R, y) in the right, then both must be viewing an object located at an angle $x_c = (x_L + x_R)/2$ to the straightahead direction, and at a distance $d = (x_L - x_R)I/2\alpha^2$ in front of the fixation point (see Figure 1). I propose that a nonzero match probability $\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\}$ implies a perceptual experience of an object at the corresponding point in space, with luminance equal to the mean luminance of the retinal images, $[I_L(x_L, y) + I_R(x_R, y)]/2$. I propose that the strength of the perceptual experience depends on the probability assigned to the match, $\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\}$, with higher probabilities creating a clearer perception.

In previous work (Read, 2002), I developed a computational model based on these Bayesian ideas. This model was designed to explain the results of psychophysical experiments involving two-interval forced-choice discrimination of the sign of stereoscopic disparity (crossed versus uncrossed) (Read & Eagle, 2000). It was tested only with stimuli whose disparity was constant across the entire image, and the model reported the most likely value of this global disparity. This procedure was appropriate for modeling the results of our two-interval forced-choice discrimination experiments, and the simulations captured the key features of the psychophysical data across a variety of spatial frequency and orientation bandwidths. However, the model actually calculated internally a detailed probabilistic disparity map of the stimulus, assigning a probability to every potential match. For the previous article, this information was then compressed into a single estimate of stimulus disparity. Now I wish to probe the probabilistic disparity map in more detail, testing the model on stimuli whose disparity varies across the image, to see whether the model can reconstruct disparity in a way that qualitatively captures the behavior of human subjects.

2 Methods

2.1 Structure of the Model.

The model is described in full mathematical detail in Read (2002), and only an outline is given here. Following earlier work by Qian (1994) and Prince and Eagle (2000a), the model was designed to incorporate the known physiology presumed to underlie binocular vision. Thus, rather than applying a Bayesian analysis to the original retinal images, the retinal images are first processed by model simple cells, which are assumed to be linear with an output nonlinearity of half-wave rectification (Movshon, Thompson, & Tolhurst, 1978; Anzai, Ohzawa, & Freeman, 1999). The receptive fields of these cells are Gabor functions (the product of a gaussian and a cosine), described by the spatial frequency and orientation to which they are tuned, and by the bandwidth of this tuning. For simplicity, all simple cells in my model have a spatial frequency bandwidth of 1.5 octaves and an orientation bandwidth of 30 degrees, defined as the full width at half height of the tuning curve. This is in accordance with psychophysical and physiological evidence for the bandwidths of channels in the visual system (Mansfield & Parker, 1993; de Valois, Albrecht, & Thorell, 1982; de Valois, Yund, & Hepler, 1982; de Valois & de Valois, 1988). These model simple cells then feed into disparity-tuned model complex cells, whose response is simulated using the energy model developed by Adelson & Bergen (1985), Fleet, Wagner, & Heeger (1996), Ohzawa, De Angelis, & Freeman (1990, 1997), and analyzed by Qian (1994). All simple cells feeding into a single complex cell are assumed to be tuned to the same spatial frequency and orientation; they differ only in the phase of their Gabor receptive field and in their position on the retina. This model uses quadrature pairs of simple cells—simple cells whose receptive fields differ in phase by $\pi/2$. The difference between the positions of the receptive fields in left and right retina defines the disparity to which the complex cell is tuned. My model employs only tuned-excitatory complex cells, which respond maximally to stimuli at their preferred disparity.

2.2 The Bayesian Analysis. A complex cell of this form can be used to test the hypothesis that the region of the stimulus within the complex cell's receptive field has the disparity that the complex cell is tuned to. If this were the case, then the part of the image falling within the complex cell's left retinal receptive field would be expected to be identical to the part of the image viewed by the right retinal receptive field. Thus, the response of the binocular complex cell could actually be predicted from a knowledge of the firing rates of simple cells with receptive fields in one eye only. This observation forms the basis of the Bayesian analysis carried out by the present computational model.

The model assumes that the left and right retinal images are independently degraded by noise. Any other sources of noise within the visual system are neglected. The retinal noise means that even if the stimulus does have exactly the disparity that the complex cell is tuned to, there will nevertheless usually be some slight discrepancy between the actual firing rate of the complex cell and that predicted by considering the response of simple cells with receptive fields in just one eye. However, the distribution of this discrepancy can be calculated analytically given the amplitude of the retinal noise. Hence, one can deduce the probability of obtaining the observed binocular complex cell firing rate, given the firing rate of the monocular simple cells from one eye, on the assumption that the disparity the complex cell is tuned to is that actually present in the stimulus. According to Bayes' theorem, this calculation can then be inverted to arrive at the probability that the stimulus really does have the disparity that the complex cell is tuned to, given the observed firing rates of the binocular complex cell itself and the monocular simple cells that feed into it.

This calculation applies only to the patch of the stimulus falling within the complex cell's receptive field. In addition, since the complex cell is tuned to a particular spatial period λ and orientation θ, it applies only to that part of the Fourier spectrum of the stimulus that falls within the complex cell's bandpass region. I thus refer to this probability as the local single-channel match probability (Read, 2002):

$$\mathrm{P}_{\lambda\theta}\{(x_L, y) \leftrightarrow (x_R, y)\}.$$

This is the probability that the region of the left retinal image centered on (x_L, y) corresponds to the region of the right image centered on (x_R, y). Note that both RFs are assumed to have the same vertical position y: I include model complex cells tuned to horizontal disparities only. λ, θ are the orientation and spatial period to which the complex cell is tuned. I include several different orientation tunings (0°, 30°, 60°, 90°, 120°, 150°) ranging from horizontal to vertical, and several different spatial frequencies (1, 2, 4, 8, 16 cycles per image) designed to cover the full range of frequencies visible to humans (de Valois & de Valois, 1988).

Precisely what is meant by "local" in this context depends on the channel under consideration. The probability analysis within each channel implements a local smoothness constraint—that is, it assumes the stimulus disparity is constant across the complex cell's receptive field. This is a region in each retina whose extent scales with the spatial period λ to which the complex cell is tuned to (for the bandwidths employed here, it is approximately $0.277 \times 0.506\lambda$).

I now average the local single-channel match probability over all spatial frequency and orientation channels to arrive at a local match "probability" to which all channels contribute:

$$\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\} = \Sigma_{\lambda\theta}\mathrm{P}_{\lambda\theta}\{(x_L, y) \leftrightarrow (x_R, y)\}.$$

This averaging process is purely heuristic. A full mathematically valid treatment would require the joint probability of obtaining a particular set of firing rates from the entire population of complex cells. The motivation for this approach is discussed in Read (2002). Thus, $\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\}$ is strictly not a probability, although I shall refer to it as such. It is regarded as an estimate of the probability that the position (x_L, y) in the left retina corresponds to the position (x_R, y) in the right retina, that is, that the stimulus disparity in the vicinity of this region is $\delta = x_L - x_R$. The "vicinity" represents an average over the receptive fields of the different channels, whose area and orientation are different for each channel. This local match probability $\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\}$ forms the probabilistic disparity map, describing how likely each potential match $(x_L, y) \leftrightarrow (x_R, y)$ is.

2.3 Details of the Model Parameters. The Bayesian prior, describing the a priori probability $\mathrm{P}\{\delta\}$ that a stimulus has the disparity δ, is taken to have the form

$$\mathrm{P}\{\delta\} = [D^2 + (\delta - D/2)^2]^{-3/2} + [D^2 + (\delta + D/2)^2]^{-3/2}.$$

This closely resembles a gaussian function but is less sharply peaked at the origin and decays less steeply (Read, 2002). D is a scale parameter, effectively describing which disparities are counted as "small." In the original article (Read, 2002), D was one of two free parameters that were systematically adjusted in order to produce a good fit to experimental results, the other being the level of retinal noise. I found it necessary to postulate a very low noise level, just 0.075% of the contrast of the binary random dot patterns used here and in the previous article. Then a very tight prior of 2.4 arcmin was necessary in order to reproduce the observed decline in performance as stimulus disparity increased. Although the model was constructed on the presumption that the brain introduces rather little noise and that the major noise affecting the calculation arises at the inputs (for instance, spontaneous photoisomerizations, photon shot noise, and the limitation of the eye's optics), the fitted noise level is extremely low and may not be realistic. The low noise was found to be necessary in order to prevent the model finding incorrect "reversed-phi" (Anstis & Rogers, 1975) matches in dense anticorrelated stereo stimuli, which would disagree with human psychophysics (Julesz, 1971; Cogan, Lomakin, & Rossi, 1993; Cumming, Shapiro, & Parker, 1998). With correlated stimuli, good matches to human psychophysics could be obtained with much higher noise levels. The very low noise level may thus be an artifact of other inadequacies of the model—for instance, its failure to incorporate any non-Fourier mechanisms that might suppress false matches in anticorrelated stimuli. Alternatively, it is possible that such low noise levels are achieved by averaging uncorrelated noise over a population. In this view, a single unit in the model would represent a small local population of identical physiological units, in which noise tends to be averaged away. All simulations presented here use the value of noise and prior scale-length D fitted to psychophysical data (Read, 2002).

The model was originally developed to account for the results of psychophysical experiments (Read & Eagle, 2000) in which the stimuli were 128 × 128 pixels at a distance of 127 centimeters, subtending an angle $1.7^\circ \times 1.7^\circ$. Accordingly, the model retina was originally constructed to be 128 × 128 pixels, where each pixel represents an angle of 0.8 arcmin on the retina. In this article, I also use a more detailed model retina of 256 × 256 pixels. This is constructed to represent the same visual angle of $1.7^\circ \times 1.7^\circ$, meaning that each pixel now represents 0.4 arcmin. The pixel values of the spatial periods of the model's channels and the prior scale length D were accordingly doubled.

The computer memory and runtime required for simulations depend on the number of different simple and complex cells included. The simulations presented here use 128 different horizontal positions of receptive field (RF) centers. This dense sampling makes the model highly sensitive to variations in disparity across the stimulus. The simulations use simple cell RFs at just one vertical position in the model retina. Note that this does not mean that the model uses information only from a single horizontal strip across the image, since the RFs themselves extend over a wide region of the image.

2.4 Perception. My hypothesis is that the local match probability $P\{(x_L, y) \leftrightarrow (x_R, y)\}$ determines which matches are consciously perceived. Several different matches may be perceived for the same feature. In section 3, I show plots of $P\{(x_L, y) \leftrightarrow (x_R, y)\}$, plotted against x_L and x_R for a particular horizontal section through the retina. These plots imply a distribution of probability in space, for a particular horizontal plane in front of the observer. Lines of constant disparity $\delta = x_L - x_R$ run diagonally up across the (x_L, x_R) plot; these correspond to frontoparallel lines in space. Lines of constant $x_c = (x_L + x_R)/2$ run diagonally down across the (x_L, x_R) plot; these correspond to radial lines from the observer, at an angle x_c to straight ahead.

Incorporating a prior preference for small disparities, as implied by psychophysical data, inevitably means that lower posterior probability will be assigned to matches with nonzero disparities than those with zero, even if the matches are equally valid and thus have the same likelihood. But in our perceptual experience, disparate regions within Panum's fusional limit are perceived as clearly as regions with zero disparity. This may imply that the relationship between probability and perception saturates, so that all probabilities above a certain threshold cause the same clarity of perception. This article postulates only that clarity of perception increases (not necessarily strictly) monotonically with probability.

2.5 Maximal Complex Cell Response. For comparison, I also consider an extension of the model of Qian (1994) to multiple spatial frequency and orientation channels. Qian's model extracts, for each line of constant $x_c = (x_L + x_R)/2$, the disparity $\delta = x_L - x_R$ for which the complex cell firing rate is maximal. This yields a disparity map giving the best disparity as a function of position x_c across the image:

$$\delta^{best}_{\lambda\theta}(x_c) = \operatorname*{argmax}_{\delta} C_{\lambda\theta}(x_c, \delta)$$

One simple way of extending this to multiple spatial frequency and orientation channels would be to sum the complex cell responses across all channels and extract the disparity, for each x_c, at which this summed response is maximal:

$$\delta^{best}(x_c) = \operatorname*{argmax}_{\delta} \sum_{\lambda,\theta} C_{\lambda\theta}(x_c, \delta)$$

However, I found that this method gave noisy results; it was poor at extracting the disparity of the target region in a random dot stereogram. Better results were obtained by extracting the maximum within each channel independently and combining these by constructing a "maxima field" $M(x_c, \delta)$, where the value of $M(x_c, \delta)$ is the number of channels that had $\delta^{best}_{\lambda\theta}(x_c) = \delta$. This "maxima field" shares several properties with the Bayesian approach already described. It is independent of how we choose to relate firing rates across different channels (whether by making all simple cells respond with the same firing rate to their optimal sinusoidal grating, or some other method). In the Bayesian approach, this was achieved by converting all firing rates into the common language of probability—here, by allowing each channel to signal only the position where its response is maximum rather than the value of that maximum. Similarly, Qian's approach is capable of matching a single point x_L in the left retina with several different points x_R in the right retina. The form of uniqueness constraint it imposes is that along each line of sight x_c, there should be only one disparity δ. In the present version, this constraint is imposed on each channel separately. A conventional disparity map, such as those plotted by Qian (1994), could then be extracted from $\mathrm{M}(x_c, \delta)$ by defining the disparity perceived at each x_c to be that where $\mathrm{M}(x_c, \delta)$ is maximal. To achieve good matches to human psychophysics, some weighting function would have to be imposed that favored small disparities (Cleary & Braddick, 1990; Prince & Eagle, 2000a, 2000b; Read & Eagle, 2000); this complication is neglected here.

3 Results

3.1 Random Dot Patterns. I begin by investigating the model's response to binary random dot patterns, where the false-matching problem is at its most acute. I have previously (Read, 2002) demonstrated that the model performs close to 100% when tested with binary random dot stereograms in a front-back discrimination task. Now, I investigate the probability field in more detail and examine whether the model is capable of extracting the details of how disparity varies acoss the image.

First I consider a central disparate region superimposed on a zero-disparity background. Figure 2 shows the images used in the simulation (A–C), and the results obtained (D–F), considering just the $y = 0$ horizontal cross-section through the image. Figure 2C shows the total luminance of the left and right images, at $y = 0$ (horizontal line in Figures 2A and 2B) and different x positions. Figure 2D shows the total response of complex cells, summed over all channels. Figure 2E shows the maxima field $\mathrm{M}(x_c, \delta)$, obtained by extending the model of Qian (1994) to multiple spatial-frequency and orientation channels. Figure 2F shows the probability field obtained with the Bayesian approach. Both models succeed in extracting the disparate region. The lines mark regions of each monocular image that have no match in the other eye due to occlusion. Naturally enough, the models cannot assign these a clear disparity, although the imposed smoothness leads to a slight tendency to continue the disparity of adjacent binocularly viewed regions into the occluded region.

3.2 The Double Nail Illusion. Here, I consider the model's response to the double nail illusion. The model was presented with the images in Figures 3A and 3B. The images in the left and right eyes are identical (apart

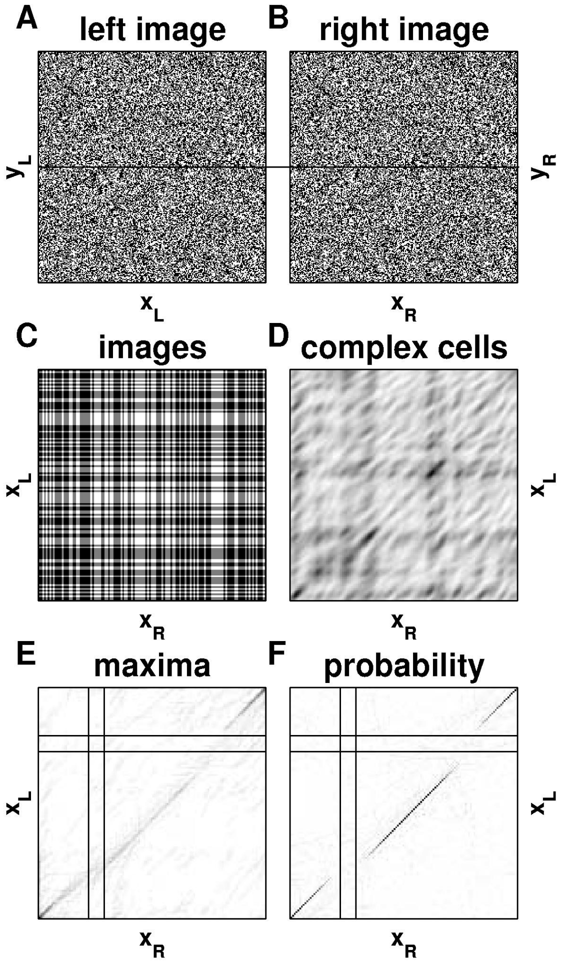

Figure 2: (A, B) The random dot stereogram used in obtaining the model results shown in the lower two plots. The stereogram is 256 × 256 pixels, representing 1.7° × 1.7° of visual angle. It contains a target region of 128 × 128 pixels, with disparity 18 pixels, centered on a background with zero disparity. The dots' luminance relative to background is ±1300 times the standard deviation of the model's retinal noise. (C) The sum of the left and right images, $\mathrm{I}_L(x_L, y)+\mathrm{I}_R(x_R, y)$, at the vertical position y marked with the line in *A* and *B*. (D) The response of complex cells, summed over all spatial frequency and orientation channels. The grayscale at (x_L, x_R) represents the total response of the population of complex cells with left- and right-eye RFs centered on (x_L, y) and (x_R, y), respectively. (E) The maxima field $\mathrm{M}(x_c, \delta)$ obtained by generalizing the model of Qian (1994). Within each channel, for each x_c, the disparity where the complex cell response was maximal contributes 1 to $\mathrm{M}(x_c, \delta)$. (F) The probabilistic disparity map, in which the grayscale at (x_L, x_R) represents the probability $\mathrm{P}\{(x_L, y) \leftrightarrow (x_R, y)\}$ that (x_L, y) matches (x_R, y).

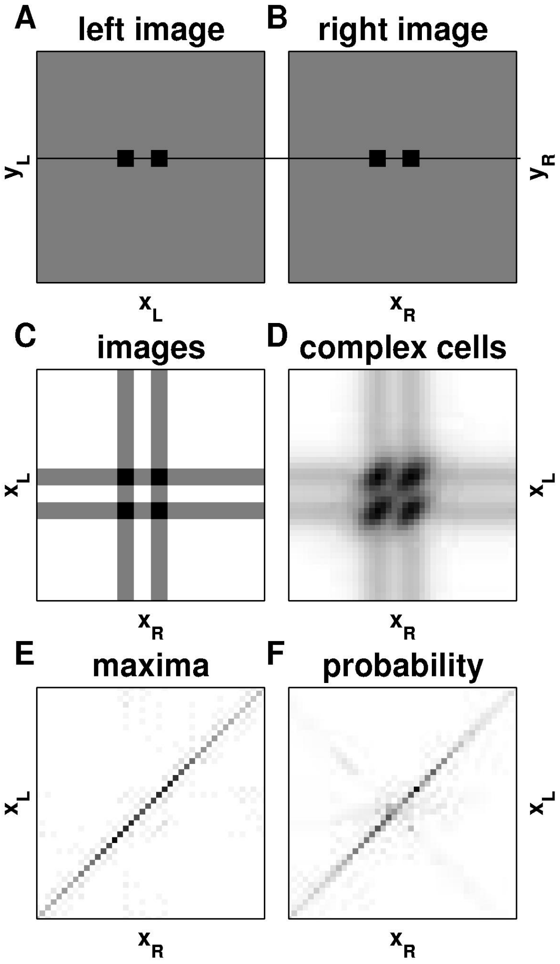

Figure 3: (A, B) Stimuli for the double-nail illusion. Each image contains two dots, 3 pixels (2.4 arcmin) square, with luminance 1300 times the noise. Only the center 40 pixels are shown. The remaining panels concern the horizontal plane containing the two objects (horizontal line across image plots). Details as for Figure 2.

from the noise), each containing two objects positioned at $x_L = -3$ pixels, $x_R = +3$ pixels. Depending on the correspondence made, this can be interpreted as two bars with disparity 0, positioned at $x_c = \pm 3$ pixels (2.4 arcmin) in the frontoparallel plane, or as two bars with $x_c = 0$ and disparity ± 6 (4.8 arcmin). These four matches are apparent in the plots of total luminance (see Figure 3C) and in the complex cell response (see Figure 3D).

Because neither the Bayesian nor the maxima model imposes a uniqueness constraint demanding a one-one match between x_L and x_R, they could potentially find all four matches. However, in fact the zero disparity match is favored throughout the image, as is apparent from the dark stripe along the line of zero disparity, $x_L = x_R$, in Figures 3E and 3F. Since the background is devoid of features, any region of the background matches equally well with any other region. The smoothing implied by the receptive fields and (in the Bayesian case) the prior preference for small disparities ensure that the background is assigned zero disparity.

So far there is little reason to prefer the more elaborate Bayesian analysis to the simpler maxima analysis based on the model of Qian (1994) and Qian and Zhu (1997). I therefore turn now to a stimulus where the Bayesian model performs better.

3.3 Occluding Bars/Panum's Limiting Case. I consider a famous stimulus that provides a challenge to many existing models of stereopsis (Howard & Rogers, 1995; Qian, 1997) because it violates the uniqueness constraint. The visual scene is supposed to consist of two vertical bars, one centered on $x_c = 4$ and disparity $+6.4$ arcmin (here, 8 pixels) and the other centered on $x_c = -4$ and disparity -6.4 arcmin. Both bars fall at the same position in the left retina, $x_L = 0$, while falling at different positions in the right retina, $x_R = \pm 8$ pixels. Physically, in the left eye, the nearer bar occludes the farther bar, whereas in the right eye, both are visible. Thus, the correct match requires the position $x_L = 0$ in the left retina to be matched both with position $x_R = -8$ *and* with position $x_R = 8$ in the right retina.

The images and the results of the simulation are shown in Figure 4. The probability field in Figure 4F shows high probability at disparities $\delta = \pm 8$ pixels, as indicated by the dark stripes along the lines drawn at these disparities. Thus, the model successfully finds both correct matches. Because it does not enforce a uniqueness constraint, it is able to match the single bar in the left eye simultaneously with both bars in the right eye.

Although the bars themselves are only 3 pixels wide, the probability is nonzero for several pixels along the lines $\delta = \pm 8$, even though the prior preference for small disparities means that blank regions of the image are normally assigned zero disparity. These constant disparity stripes reflect the smoothness constraint built into the model. The lowest-frequency simple cells have very large receptive fields, so the match implied by the bars can potentially influence the probability assigned to very distant matches. The prior preference for zero disparity is thus opposed by the assumption that adjacent points have the same disparity. Under the hypothesis that the probability plotted here underlies perception, the intepretation is that an isolated disparate object tends to produce the perception of being embedded in a disparate frontoparallel plane. This is consonant with perceptual experience and could potentially help the brain reconstruct smooth surfaces from discrete disparate stimuli (Grimson, 1982).

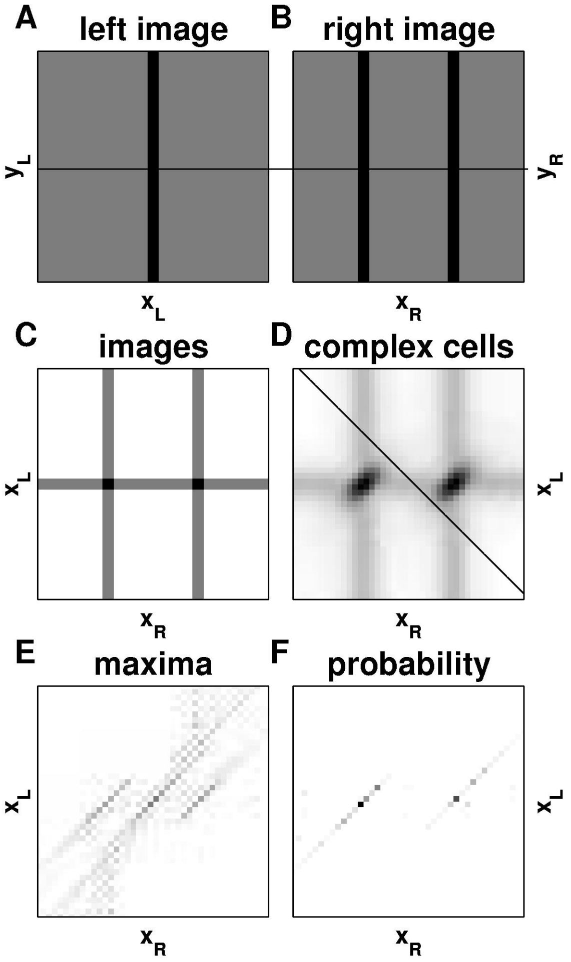

Figure 4: (A, B) Stimuli for Panum's limiting case. The right image contains two bars and the left image one. Each bar is 3 pixels (2.4 arcmin) wide, with luminance 13 in arbitrary units. Only the center 40 pixels are shown. Remaining panels as in Figure 2.

In contrast, the maxima field in Figure 4E still shows clear remnants of the cruciform structure of the raw complex cell response. This leads to a predicted perception of a "ghost" object in between the two objects actually present. (In the stimuli used, this is at the fixation point; in general, wherever the bars are with respect to fixation, the ghost will lie exactly between them.) The origin of this ghost is clear if one traces along the line $x_c = 0$, marked

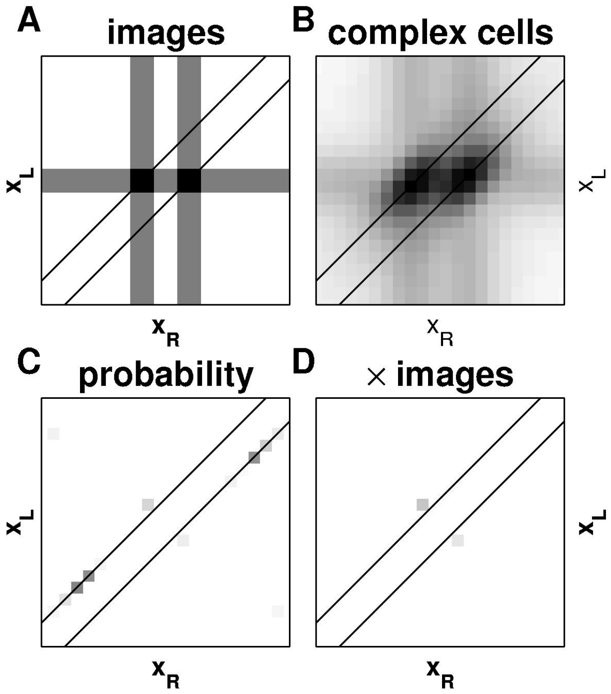

Figure 5: Stimuli as for Figure 4 except that the bars have disparity ±2 pixels. Only the center 20 pixels are shown. In all panels, the diagonal lines indicate disparity ±2 pixels. (A) The sum of the left and right images, $I_L(x_L, y)+I_R(x_R, y)$ at $y = 0$. (B) The response of complex cells, summed over all spatial frequency and orientation channels. (C) The probabilistic disparity map $P\{(x_L, y) \leftrightarrow (x_R, y)\}$. (D) Probability weighted by image intensity: $P\{(x_L, y) \leftrightarrow (x_R, y)\} \times [I_L(x_L, y) + I_R(x_R, y)]$. This is intended to approximate the perceptual experience. Thus, the model predicts a perception of two bars, disparity slightly greater than 2 pixels.

with a line in the plot of summed complex cell responses (see Figure 4D). Within each channel, the maximum complex cell response along this line occurs at zero disparity.

If the bars' disparity is reduced, the Bayesian model experiences a repulsion illusion. Figure 5 shows the results obtained with bars of disparity ±2 pixels (1.6 arcmin). Now, the model perceives the bars at slightly different positions in the left eye and at slightly larger disparities than veridical. This is especially apparent when we plot the probability field weighted by the combined images, $[I_L(x_L, y) + I_R(x_R, y)]$, Figure 5D, in an attempt to show the spatial location of the visual objects perceived by the model.

This illusion occurs when the separation between the images in the left

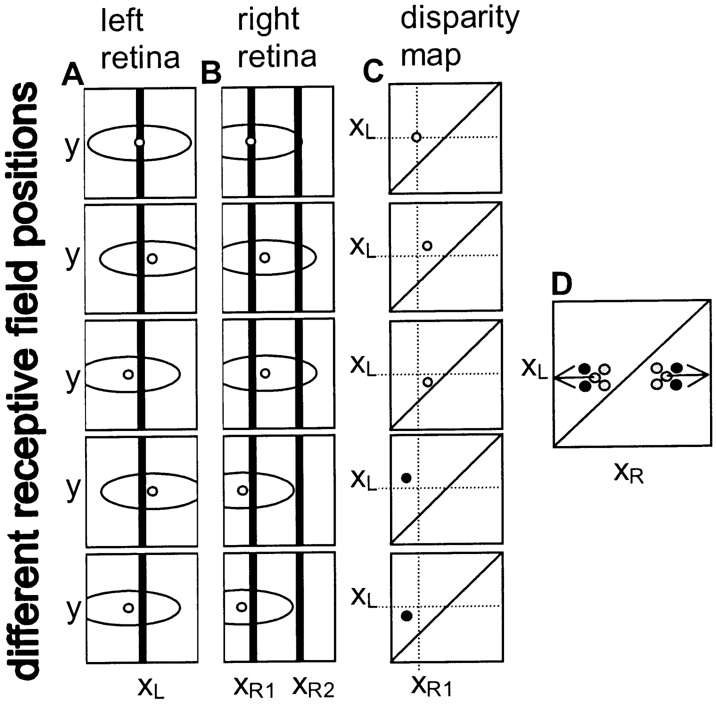

Figure 6: Why the bars in Panum's limiting case tend to repel at small separations. The five rows show complex cells tuned to five different values of disparity δ and x_c. (A, B) The positions of the complex cell receptive fields in left and right retinas. (C) The corresponding position on the disparity map (circle = RF center). The shading in the disparity map indicates the probability assigned to the match. (D) The effect of all this on the probability map assigned to the stimulus. The bars tend to "repel" each other, being shifted away from each other along both δ and x_c.

and right eyes becomes smaller than the RF size in the majority of complex cells. Figure 6 explains why. Figures 6A and 6B represent the left and right retinas, with the positions of the bar stimuli in each eye indicated. The left retina contains a single bar at horizontal position x_L; the right retina, two bars at positions x_{R1} and x_{R2}. In addition, each plot contains an example complex cell RF. The circle represents the center of the RF in each retina, and the oval indicates the extent of the RF. The five rows of plots differ in the positions of the complex cell RFs. Figure 6C shows where the RF centers are located in the (x_L, x_R) disparity space familiar from the previous figures.

The top row shows a complex cell tuned to a correct match, in which the bar in the left eye is correctly identified with a bar in the right eye (clearly there is another correct match, given by the pairing with the other bar in the right eye, which is not shown here). However, although this match is in fact correct, the complex cell shown will accord it low probability. This is because the large RF of this cell extends over both bars in the right retina. The RFs in either eye are thus not stimulated equally, meaning that the calculated probability is low. The total match probability summed over all channels will be relatively low, as represented by the pale shading assigned to this correspondence in the disparity map in Figure 6C.

The same effect occurs for the cells shown in the second row, which are tuned to the same disparity but a mean retinal position x_c slightly to the right, and for those in the third row, which are tuned to the same mean retinal position x_c as those in the first row, but a smaller disparity. Again, the complex cell reports low probability.

The fourth row shows a cell that has the same x_c as the first row but larger disparity; the fifth row shows a cell that has the same disparity but smaller x_c. Here, exactly one bar falls in each eye's RFs. Since the RFs are symmetric, the complex cell receives equal stimulation in both eyes, and so signals high probability. This is represented by the dark shading in the disparity map.

Figure 6D summarizes the results from the five rows. As in Figures 2 through 5, the grayscale at the point (x_L, x_R) represents the probability that the point x_L is the correct match for x_R. The correct match is not assigned the highest probability; matches with larger disparity or lower x_c are considered more likely than the correct match. Thus, the bars are perceived shifted away from each other both in the front-back (δ) and the left-right (x_c) direction. This effect is due to channels whose RFs are longer horizontally than the separation between the bars' images in the right eye. The repulsion is therefore due predominantly to channels tuned to low spatial frequencies or orientations close to horizontal. Similar repulsion illusions have been reported with human subjects (Ruda, 1998; Badcock & Westheimer, 1985) and have been reproduced with a non-Bayesian model also based on energy-model complex cells (Qian, 1997; Mikaelian & Qian, 2000).

At smaller bar separations still, human observers display an apparent "pooling" effect: the bars appear to attract rather than repel each other, and the actual disparity perceived is an average of the disparities of each bar. Qian's model reproduces this effect. This model, however, cannot display such pooling. This is because when two bars falls in one eye's RF and only one in the other, the two eyes' RFs are very unequally stimulated; the model naturally assigns a very low probability to such a match. This deficiency might be addressed by including some form of local contrast normalization between the left and right images.

4 Discussion

This article discusses one way of applying a probabilistic approach to the solution of the correspondence problem. Rather than seeking a unique match for every point in each retina, I propose assigning a probability to each potential match between left and right retinal positions. This is interpreted as the probability that an object exists in front of the viewer at the spatial location implied by the match. This probability is assumed to underlie perception: zeros of the probability field imply no perception of an object at that location, whereas nonzero values of probability imply a perception, whose clarity is presumed to increase with increasing probability, of an object with luminance given by the mean of the values in the left and right retinas (see Figure 1).

There are many ways in which the brain might attempt to assign a probability to each potential match. The approach adopted here was motivated by a desire to incorporate as much as possible of the known physiology of binocular vision. The available physiological and psychophysical evidence suggests that initial processing takes place within channels tuned to a particular spatial frequency and orientation. Accordingly, the match probability is initially calculated within a single channel, using a Bayesian analysis based on the outputs of disparity-tuned complex cells and the simple cells that feed into them. This analysis introduces the two key constraints employed by the model in order to overcome the ill-posed nature of the correspondence problem. First, the disparity is smoothed over the RF dimensions in each channel. Second, the Bayesian prior is used to enforce a preference for small disparities. Information from different channels is then combined by averaging the probability reported from each individual channel. This means that information from all spatial scales is handled simultaneously; the model employs no coarse-to-fine hierarchy (Mallot, Gillner, & Arndt, 1996).

Bayesian stereo algorithms have a long history in the computer vision literature (Szeliski, 1990; Geiger, Ladendorf, & Yuille, 1992; Chang & Chatterjee, 1992; Scharstein & Szeliski, 1998). This article differs from these in two major respects. First, it attempts to incorporate the known physiology of disparity-tuned cells in primary visual cortex. Thus, it applies Bayesian ideas to the results of previous workers who used linear filters as a matching primitive (Lehky & Sejnowski, 1990; Jones & Malik, 1992; Sanger, 1988; Qian, 1994). The probability field is then derived (albeit with a number of simplifications and approximations) from the statistics of retinal images degraded by gaussian noise and processed by simple and complex cells. In contrast, previous Bayesian models have usually been postulated to take a Gibbs form: exp(-potential/temperature), where the potential is some cost function incorporating punishment for disparity discontinuities, poor luminance matches, and so on.

Second, this article postulates that the probability field directly underlies perception. Previous studies incorporating Bayesian ideas have generally used it as a tool to arrive at a single match between points in the left and right eyes, explicitly including a form of uniqueness constraint (often allowing for occlusion—each point in the left image must match at most one point in the right image). The approach here was designed to allow for the possibility of multiple matches. The model is closest to that developed by Qian and coworkers (Qian, 1994; Zhu & Qian, 1996; Qian & Zhu, 1997; Qian & Andersen, 1997) and Prince and Eagle (2000a). It differs from their work in incorporating multiple spatial frequency and orientation channels and in processing the complex cell output to assign a probability to each match.

The model presented here has a number of serious limitations, which means that it can be regarded only as a preliminary model of stereo correspondence. It was initially developed to model psychophysical data obtained in a front-back depth discrimination task in which images were presented for 130 ms. Although it succeeds admirably in this task, the percept implied by the model lacks the clarity and sharpness of human perceptions. For instance, the ragged outline of the target region perceived in the random dot stereogram of Figure 2 does not accord with the sharp, square outline that humans viewing this figure perceive. This is perhaps not surprising for a model based on the known physiology of primary visual cortex, which incorporates no hierarchy of interactions or iterative processes that sharpen the performance of many other stereo models (cf. Marr & Poggio, 1976; Geiger et al., 1992; Scharstein & Szeliski, 1998). Thus, this model may represent an initial step in stereoscopic perception, providing a sketch of the 3d visual scene, which is then processed by still higher visual areas.

The model makes use of a prior probability assigned to each disparity but does not address how the brain could arrive at such a judgment. It does not include any mechanism for updating the prior on the basis of visual experience or information from nonvisual sources. A full Bayesian model would require such a mechanism, based on experimental evidence of how performance develops with training.

The model fails to show the correct response with sparse anticorrelated stereograms. Anticorrelated stereograms are those in which the polarity of one eye's image has been inverted, with black pixels becoming white and vice versa. Dense examples of such stimuli, such as binary random dot patterns, produce little or no impression of depth in most human subjects, although some display a slight tendency to see depth in the opposite direction to that implied by the disparity of the stereogram. The model reproduces this behavior (Read, 2002). However, sparse anticorrelated stimuli give human observers the impression of depth in the veridical direction (von Helmholtz, 1909; Cogan, Kontsevich, Lomakin, Halpern, & Blake, 1995), presumably because boundaries are extracted and matched even though the polarity of the boundaries is opposite. The model presented here assigns low probability to all anticorrelated potential matches and thus has no mechanism for reproducing this behavior.

Further, the model currently includes no form of contrast normalization between left and right eyes. It seems likely that some such normalization exists, because human observers can fuse stereograms even where considerable contrast differences exist between the monocular images. The model cannot, because any contrast differences between left and right eyes mean that even the correct matches are considered improbable. In addition, the model fails to reproduce the "disparity pooling" observed with disparate bars at very small separations. A suitable scheme of local contrast normalization might be able to correct both of these discrepancies, although considerable work might be required to find a suitable implementation.

Despite these limitations, the model is able to produce plausible solutions of the correspondence problem for a range of stimuli. It extracts the appropriate disparity map in dense random dot stereograms containing several target regions with different disparity. In the double-nail illusion, it perceives only two of four possible matches, in agreement with human observers. Models that impose a strict uniqueness constraint cannot handle Panum's limiting case, whereas the model presented here performs correctly with this stimulus. This is an advantage of the Bayesian approach over models with a similar physiological basis but a different method of extracting disparity, such as that of Qian (1994) and Qian and Zhu (1997). Finally, it has already been shown (Read, 2002) that the model accurately reproduces psychophysical functions in a front-back discrimination task, for both correlated and anticorrelated random noise stereograms with a range of spatial frequency and orientation bandwidths. Thus, the model combines physiological plausibility with wide explanatory power.

Acknowledgments

I am supported by a Mathematical Biology Training Fellowship from the Wellcome Trust. The simulations presented in this article were run with support from the Oxford Supercomputing Centre. I thank Bruce Cumming for helpful discussions.

References

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

Anstis, S. M., & Rogers, B. J. (1975). Illusory reversal of visual depth and movement during changes of contrast. *Vision Research*, *15*, 957–961.

Anzai, A., Ohzawa, I., & Freeman, R. (1999). Neural mechanisms for processing binocular information I. Simple cells. *Journal of Neurophysiology*, *82*(2), 891–908.

Badcock, D. R., & Westheimer, G. (1985). Spatial location and hyperacuity: Flank position within the centre and surround zones. *Spatial Vision*, *1*(1), 3–11.

Chang, C., & Chatterjee, S. (1992). A deterministic approach for stereo disparity calculation. In G. Sandin (Ed.), *Computer vision: ECCV'92* (pp. 425–433). New York: Springer-Verlag.

Cleary, R., & Braddick, O. (1990). Direction discrimination for band-pass filtered random dot kinematograms. *Vision Research*, *30*(2), 303–316.

Cogan, A. I., Kontsevich, L. L., Lomakin, A. J., Halpern, D. L., & Blake, R. (1995). Binocular disparity processing with opposite-contrast stimuli. *Perception*, *24*(1), 33–47.

Cogan, A. I., Lomakin, A. J., & Rossi, A. F. (1993). Depth in anticorrelated stereograms: Effects of spatial density and interocular delay. *Vision Research*, *33*(14), 1959–1975.

Cumming, B., Shapiro, S. E., & Parker, A. J. (1998). Disparity detection in anticorrelated stereograms. *Perception*, *27*(11), 1367–1377.

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

de Valois, R. L., & de Valois, K. K. (1988). *Spatial vision*. Oxford: Oxford University Press.

de Valois, R., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 531–544.

Dev, P. (1975). Perception of depth surfaces in random-dot stereograms: A neural model. *International Journal of Man-Machine Studies*, *7*, 511–528.

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

Geiger, D., Ladendorf, B., & Yuille, A. (1992). *Occlusions and binocular stereo*. In G. Sandini (Ed.), *Computer vision: ECCV'92* (pp. 425–433). New York: Springer-Verlag.

Grimson, W. E. (1981). A computer implementation of a theory of human stereo vision. *Philosophical Transactions of the Royal Society, London, B: Biological Sciences*, *292*(1058), 217–253.

Grimson, W. E. (1982). A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society, London, B: Biological Sciences*, *298*(1092), 395–427.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford: Oxford University Press.

Jones, D. G., & Malik, J. (1992). A computation framework for determining stereo correspondence from a set of linear spatial filters. In G. Sandini (Ed.), *Computer vision: ECCV'92* (pp. 395–410). New York: Springer-Verlag.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.

Knill, D., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Krol, J. D., & van de Grind, W. A. (1980). The double-nail illusion: Experiments on binocular vision with nails, needles, and pins. *Perception*, *9*(6), 651–669.

Lehky, S. R., & Sejnowski, T. J. (1990). Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *Journal of Neuroscience*, *10*(7), 2281–2299.

Mallot, H., & Bideau, H. (1990). Binocular vergence influences the assignment of stereo correspondences. *Vision Research*, *30*(10), 1521–1523.

Mallot, H. A., Gillner, S., & Arndt, P. A. (1996). Is correspondence search in human stereo vision a coarse-to-fine process? *Biological Cybernetics*, *74*(2), 95–106.

Mansfield, J. S., & Parker, A. (1993). An orientation-tuned component in the contrast masking of stereopsis. *Vision Research*, *33*(11), 1535–1544.

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

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society, London, B: Biological Sciences*, *204*(1156), 301–328.

McKee, S. P., Bravo, M. J., Smallman, H. S., & Legge, G. E. (1995). The "uniqueness constraint" and binocular masking. *Perception*, *24*(1), 49–65.

McLoughlin, N. P., & Grossberg, S. (1998). Cortical computation of stereo disparity. *Vision Research*, *38*(1), 91–99.

Mikaelian, S., & Qian, N. (2000). A physiologically-based explanation of disparity attraction and repulsion. *Vision Research*, *40*(21), 2999–3016.

Movshon, J., Thompson, I., & Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology*, *283*, 53–77.

Nelson, J. I. (1975). Globality and stereoscopic fusion in binocular vision. *Journal of Theoretical Biology*, *49*(1), 1–88.

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

Ohzawa, I., DeAngelis, G., & Freeman, R. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology*, *77*(6), 2879–2909.

Pollard, S. B., & Frisby, J. P. (1990). Transparency and the uniqueness constraint in human and computer stereo vision. *Nature*, *347*(6293), 553–556.

Pollard, S. B., Mayhew, J. E., & Frisby, J. P. (1985). PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, *14*(4), 449–470.

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

Prince, S. J. D., & Eagle, R. A. (2000b) Stereo correspondence in one-dimensional Gabor stimuli. *Vision Research*, *40*(9), 913–924.

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

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, *18*(3), 359–368.

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

Qian, N., & Sejnowski, T. (1989). Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation. In *Pro-*

ceedings of the 1988 Connectionist models summer school (pp. 435–443). London: Harcourt.

Qian, N., & Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research*, *37*(13), 1811–1827.

Read, J. C. A. (2002). A Bayesian model of stereo depth/motion direction discrimination. *Biological Cybernetics, 86(2)*, 117–136.

Read, J. C. A., & Eagle, R. A. (2000). Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Research*, *40*(24), 3345–3358.

Ruda, H. (1998). The warped geometry of visual space near a line assessed using a hyperacuity displacement task. *Spatial Vision*, *11*(4), 401–419.

Sanger, T. (1988). Stereo disparity computation using Gabor filters. *Biological Cybernetics*, *59*, 405–418.

Scharstein, D., & Szeliski, R. (1998). Stereo matching with nonlinear diffusion. *International Journal of Computer Vision*, *28*, 155–174.

Szeliski, R. (1990). Bayesian modelling of uncertainty in low-level vision. *International Journal of Computer Vision*, *5*, 271–301.

von Helmholtz, H. (1909). *Handbuch der physiologischen Optik*. Hamburg, Germany: Voss.

Weinshall, D. (1989). Perception of multiple transparent planes in stereo vision. *Nature*, *341*(6244), 737–739.

Westheimer, G. (1986). Panum's phenomenon and the confluence of signals from the two eyes in stereoscopy. *Philosophical Transactions of the Royal Society, London, B: Biological Sciences*, *228*(1252), 289–305.

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

Received April 3, 2001; accepted October 15, 2001.