Pergamon

PII: S0042-6989(96)00331-8

Vision Res., Vol. 37, No. 13, pp. 1811–1827, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

Physiological Computation of Binocular Disparity

NING QIAN,*† YUDONG ZHU*

Received 2 July 1996; in revised form 12 November 1996

We previously proposed a physiologically realistic model for stereo vision based on the quantitative binocular receptive field profiles mapped by Freeman and coworkers. Here we present several new results about the model that shed light on the physiological processes involved in disparity computation. First, we show that our model can be extended to a much more general class of receptive field profiles than the commonly used Gabor functions. Second, we demonstrate that there is, however, an advantage of using the Gabor filters: similar to our perception, the stereo algorithm with the Gabor filters has a small bias towards zero disparity. Third, we prove that the complex cells as described by Freeman *et al.* compute disparity by effectively summing up two related cross products between the band-pass filtered left and right retinal image patches. This operation is related to cross-correlation but it overcomes some major problems with the standard correlator. Fourth, we demonstrate that as few as two complex cells at each spatial location are sufficient for a reasonable estimation of binocular disparity. Fifth, we find that our model can be significantly improved by considering the fact that complex cell receptive fields are, on average, larger than those of simple cells. This fact is incorporated into the model by averaging over *several* quadrature pairs of simple cells with nearby and overlapping receptive fields to construct a model complex cell. The disparity tuning curve of the resulting complex cell is much more reliable than that constructed from a *single* quadrature pair of simple cells used previously, and the computed disparity maps for random dot stereograms with the new algorithm are very similar to human perception, with sharp transitions at disparity boundaries. Finally, we show that under most circumstances our algorithm works equally well with either of the two well-known receptive field models in the literature. © 1997 Elsevier Science Ltd.

Stereo vision Binocular disparity Spatial pooling Complex cells Computer modeling

INTRODUCTION

We see the world as three-dimensional even though the input to our visual system, the light intensity distributions on our retinas, has only two spatial dimensions. It is well known that the third dimension, the relative depth of objects in the world, can usually be inferred from a variety of visual cues present in the retinal images. One such cue is binocular disparity, defined as the difference between the locations (relative to the corresponding foveas) of the two retinal projections of a given point in space. How the brain computes this disparity, and thus achieves stereoscopic depth perception, has been the subject of many studies, and numerous computational models for stereo vision have been proposed in the past. We recently proposed a new algorithm for computing disparity maps from stereograms (Qian, 1994a) which differs from previous models in that it is solely based on known physiological properties of real binocular cells in the brain (Ohzawa *et al.*, 1990; Freeman & Ohzawa, 1990; DeAngelis *et al.*, 1991). Here we provide some further analyses of our model along with computer simulations. These results, we believe, give us a better understanding of the physiological process involved in computing binocular disparity. In particular, we demonstrate that by incorporating an additional piece of physiological data into our model, we can greatly improve the quality of the computed disparity maps. The results reported here have been presented previously in abstract form (Qian & Zhu, 1995).

THE MODEL

We briefly review our stereo model (Qian, 1994a) in this section. Our model is based on the physiological and modeling studies of Freeman and coworkers (Freeman & Ohzawa, 1990; Ohzawa *et al.*, 1990; DeAngelis *et al.*, 1991). These investigators found that the left and right spatial receptive field profiles of a binocular simple cell in cat's primary visual cortex can be described by two Gabor functions with the same Gaussian envelopes but different phase parameters in the sinusoidal modulations. For horizontal disparity computation, only the horizontal dimension of cells' receptive fields is relevant. The left

*Center for Neurobiology and Behavior, Columbia University, 722 W. 168th Street, New York, NY 10032, U.S.A.
†To whom all correspondence should be addressed [*Tel* 212-960-2213; *Fax* 212-960-2561; *Email* nq6@columbia.edu].

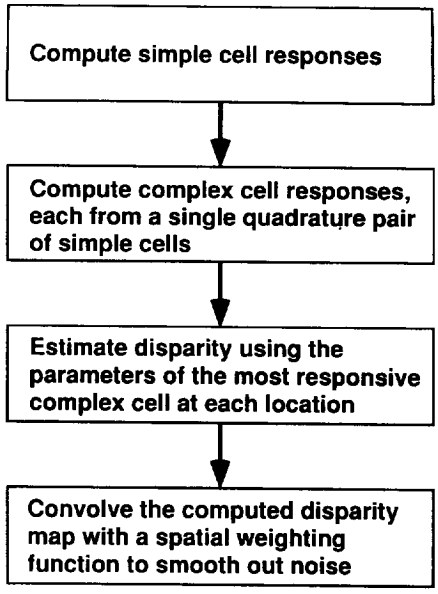

FIGURE 1. Steps used in our original algorithm (Qian, 1994a) for computing disparity maps from stereograms. For a given stereogram, we first compute, at each location, the responses of a family of simple cells with appropriately chosen parameters. We then compute complex cell responses, each from a single quadrature pair of the simple cell responses. After that the parameters of the complex cell with maximum responses are found through a parabolic interpolation, and are then used to estimate the disparity according to Eq. (7). Finally, because the disparity map so obtained is usually noisy, a smoothing step has to be applied to average out noise. We will show later in this paper that this *ad hoc* final step can be removed if the complex cell responses are obtained by pooling several, instead of a single, quadrature pairs. Note that the parabolic interpolation is used in order to reduce the number of model complex cells needed in our simulations. It is not meant to be a step used in the brain, which does not need this step because it has a large number of cells tuned to various disparities.

and right receptive field profiles of a simple cell centered at $x = 0$ are then given by:

$$f_l(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)\cos(\omega_0 x + \phi_l) \tag{1}$$

$$f_r(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)\cos(\omega_0 x + \phi_r) \tag{2}$$

where σ and ω_0 are the Gaussian width and the preferred spatial frequency* of the receptive fields; ϕ_l and ϕ_r are the left and right phase parameters.

Freeman and coworkers (Freeman & Ohzawa, 1990; Ohzawa *et al.*, 1990) found that to a good approximation the response of a simple cell can be determined by first filtering, for each eye, the retinal image by the corresponding receptive field profile, and then adding the two contributions from the two eyes:

$$r_s = \int_{-\infty}^{+\infty} dx\,[f_l(x)I_l(x) + f_r(x)I_r(x)] \tag{3}$$

where $I_l(x)$ and $I_r(x)$ are the left and right retinal images of the stimulus. They further showed that the response of a complex cell can be modeled by summing the squared outputs of a quadrature pair (Adelson & Bergen, 1985; Watson & Ahumada, 1985; Ohzawa *et al.*, 1990; Qian, 1994a) of such simple cells:

$$r_q = (r_{s.1})^2 + (r_{s.2})^2. \tag{4}$$

Through mathematical analysis we found that under the assumption that stimulus disparity D is significantly smaller than the width of the receptive fields (about 2σ), the response of a model complex cell to the disparity is given by (Qian, 1994a):

$$r_q \approx c^2|\tilde{I}(\omega_0)|^2\cos^2\left(\frac{\Delta\phi}{2} - \frac{\omega_0 D}{2}\right), \tag{5}$$

where

$$\Delta\phi \equiv \phi_l - \phi_r \tag{6}$$

is the phase parameter difference between the left and right receptive fields, c is a constant, and $|\tilde{I}(\omega_0)|^2$ is the Fourier power of the stimulus patch (under the receptive field) at the preferred spatial frequency of the cell. According to Eq. (5), a complex cell's preferred disparity is determined by its receptive field parameters according to:

$$D_{\text{pref}} \approx \frac{\Delta\phi}{\omega_0}, \tag{7}$$

which is the relative shift between the sinusoidal modulations of the left and right receptive fields of the constituent simple cells. Using this relationship we were able to compute disparity maps from random dot stereograms using a population of model complex cells without employing any non-physiological procedures such as explicit matching of fine stimulus features (Qian, 1994a). The stimulus disparity is identified with the preferred disparity of the most responsive complex cells in the population. The steps used in the computation are summarized in Fig. 1.

Note that the periodic function of D in Eq. (5) is an approximation. A more accurate derivation of the complex cell response to broad-band stimuli (Zhu & Qian, 1996) reveals that the side peaks in the disparity tuning curve rapidly decay to zero and that the main peak (the preferred disparity) of the complex cell with preferred spatial frequency ω_0 is always located within the range $[-\pi/\omega_0, \pi/\omega_0]$. An intuitive explanation of this constraint on preferred disparities is given in Fig. 2. Because of this restriction, the family of complex cells with spatial frequency ω_0 can only code disparities in the range $[-\pi/\omega_0, \pi/\omega_0]$ (Qian, 1994a; Zhu & Qian, 1996; Smallman & MacLeod, 1994). It is, however, incorrect to conclude that our algorithm can only compute small disparities (Fleet *et al.*, 1996) because cells in the visual cortex are tuned to a wide range of spatial frequencies (DeValois *et al.*, 1982; Shapley & Lennie, 1985), and those cells with small preferred frequencies can compute large stimulus disparities. A prediction is that a stimulus with a sharp frequency spectrum centered at Ω can only generate perceived disparity within the range $[-\pi/\Omega, \pi/\Omega]$ because it predominantly activates cells with the

*Note that ω is an angular spatial frequency with the units radians per degree. It is related to the ordinary spatial frequency f (in cycles per degree) by $\omega = 2\pi f$. We prefer to use ω for notational simplicity.

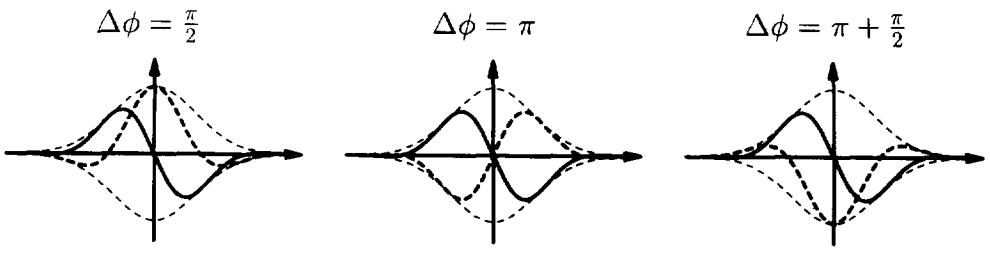

FIGURE 2. An intuitive explanation of why the preferred disparity of a complex cell with preferred frequency ω_0 is limited within the range $[-\pi/\omega_0, \pi/\omega_0]$ under the phase-difference model for receptive fields (Ohzawa *et al.*, 1990). Three binocular receptive field profiles with the phase difference $\Delta\phi$ equal to $\pi/2$, π and $\pi + \pi/2$ are shown. In all three panels, the left receptive field profiles are shown by solid lines and the right profiles by dashed lines. The Gaussian envelopes of the receptive fields are indicated by thin dashed lines. When $\Delta\phi$ is less then π (left panel), the resulting complex cell will be tuned to a disparity equal to the distance between the two positive peaks ($\Delta\phi/\omega_0$). When $\Delta\phi$ is over π (right panel), however, the two negative peaks become more similar to each other and the cell has an effective $\Delta\phi$ smaller than π. The maximum peak separation occurs when $\Delta\phi$ equals π (middle panel). Therefore, the preferred disparity of the complex cell is always smaller than π/ω_0. Similarly, the preferred disparity of the cell is also always larger than $-\pi/\omega_0$.

preferred frequency $\omega_0 = \Omega$. This so-called size–disparity correlation has been observed psychophysically (Smallman & MacLeod, 1994; Schor & Wood, 1983).

GENERALIZATION

The complex cell response expression [Eq. (5)] was previously derived with the specific assumption of using the Gabor filters as the simple cell receptive field profiles (Qian, 1994a). We have now shown that the same equation can be derived under some very general assumptions. Specifically, if the left and right receptive field profiles $f_l(x)$ and $f_r(x)$ of a simple cell differ by a phase difference $\Delta\phi$ and if the frequency tuning of the receptive field profiles is significantly sharper than the frequency spectrum of the input stimulus, the complex cell response constructed from such simple cells to stimulus disparity D is approximately given by:

$$r_q \approx c^2 |\tilde{I}(\omega_0)|^2 \cos^2\left(\frac{\Delta\phi}{2} - \frac{\omega_0 D}{2}\right), \tag{8}$$

where constant c is defined as:

$$c \equiv 4\int_0^\infty d\omega |\tilde{f}_l(\omega)|. \tag{9}$$

and ω_0 is the preferred spatial frequency of the cell. The details of the derivation are presented in the Appendix. We conclude that our stereo algorithm works with a rather general class of receptive field profiles, including the Gabor functions (see also Qian & Andersen (1997)). The general derivation of Eq. (8) also enables an easy estimation of the error term associated with the equation. The error is found to be proportional to the variance (width) of the frequency tuning function of the receptive fields (see the Appendix).

The above assumption that the frequency tuning of the receptive fields are significantly sharper than the Fourier spectra of the retinal stimulus is usually a good one because most visual cortical cells are well tuned to spatial frequencies (DeValois *et al.*, 1982; Shapley & Lennie, 1985) while the natural environment is rich in complex textures and sharp boundaries and therefore tends to produce images with broad spectra. However, in the rare case when the visual system is looking at a sine wave grating this assumption is clearly violated. In general, if the retinal image has a Fourier spectrum sharper than the frequency tuning of the cells, then the preferred frequency of the cell (ω_0) in Eqs (5), (7) and (8) should be replaced by the dominant spatial frequency Ω of the image (Zhu & Qian, 1996; Qian & Andersen, 1997). The preferred disparity of a given cell [Eq. (7)] will then become $\Delta\phi/\Omega$, which is different for different stimulus frequencies. Consequently, if one uses a single family of cells with a fixed preferred frequency ω_0 to estimate stimulus disparity according to Eq. (7), the results will not be accurate unless the dominant stimulus frequency matches the preferred frequency of the cells. This, however, does not pose a serious problem for the real visual system, except for the stimulus with very high or low frequencies, because the brain contains cells tuned to a wide range of frequencies and the cells with the highest responses are those whose preferred frequencies do match those of the stimuli.

ZERO DISPARITY BIAS

Although the result in the previous section shows that one does not have to use the Gabor functions as the front end filters in our stereo vision model, there are good reasons to do so. The main reason, of course, is that the Gabor filters have been found to describe the spatial receptive field profiles of real primary visual cortical cells very well (Marcelja, 1980; Jones & Palmer, 1987; Ohzawa *et al.*, 1990; Freeman & Ohzawa, 1990; DeAngelis *et al.*, 1991) [but see Stork & Wilson (1990) for a different point of view]. There is, however, a hitherto unrecognized advantage of using Gabor filters as simple cell receptive field profiles in disparity computation: within the framework of our stereo model, the DC components of the Gabor filters generate a small bias towards zero disparity. This bias is considered desirable because it naturally explains the perceptual observation that when we are looking at a degenerate pattern with uniform luminance along the horizontal dimension, we

see zero disparity.* Without the DC components and the associated bias, the results would be indeterminant as the responses of all filters would be zero and any disparity values would be equally valid.

Specifically, it can be shown that for a binocular stimulus with a horizontally uniform light intensity distribution

$$I_l(x) = I_r(x) = a, \tag{10}$$

the response of the simple cell with binocular receptive fields given by Eqs (1) and (2) is:

$$r_{s,1} = a\int_{-\infty}^{+\infty} dx[f_l(x) + f_r(x)]$$

$$= a\sqrt{2\pi\sigma^2}e^{-\omega_0^2\sigma^2/2}(\cos\phi_l + \cos\phi_r), \tag{11}$$

the response of the simple cell forming a quadrature pair with the cell in Eq. (11) is:

$$r_{s,2} = a\sqrt{2\pi\sigma^2}e^{-\omega_0^2\sigma^2/2}(\sin\phi_l + \sin\phi_r), \tag{12}$$

and the complex cell response constructed from the quadrature pair of the simple cells is therefore given by:

$$r_q = a^2 8\pi\sigma^2 e^{-\omega_0^2\sigma^2}\cos^2\frac{\Delta\phi}{2}. \tag{13}$$

Note that no approximations are used in deriving the above three equations. Since Eq. (13) predicts that among the population of complex cells, the one with $\Delta\phi = 0$ gives the maximum response, the disparity reported by the cells is zero, consistent with our perception. The reason that the bias is at zero disparity is because the cell tuned to zero disparity has the largest DC component. The bias also makes the computed disparity maps from stimuli with unambiguous disparities slightly less accurate. The error introduced depends on how the strength of the disparity signal in the stimulus [the amplitude of the cosine function in Eq. (8)] compares with the strength of the bias [the amplitude of the cosine function in Eq. (13)]. In our computer simulations on random dot stereograms, the bias is always less than the small fluctuations in the computed disparity surfaces caused by the stochastic nature of the stereograms.

HOW DO BINOCULAR CELLS COMPUTE DISPARITY?

Since binocular disparity is defined as a relative shift between the corresponding left and right image patches, one may expect intuitively that a cross correlation type of operation should be a natural choice for solving the problem. Indeed, correlation-based stereo algorithms have been proposed previously in the machine vision community (Hannah, 1974; Panton, 1978). On the surface, however, it is not clear how the cells in our model compute disparity and whether our physiological algorithm is related to cross-correlation. We investigate this issue in this section. Since the simple cells in the algorithm simply add the contributions from their left and right receptive fields (Ohzawa *et al.*, 1990; Qian, 1994a) instead of multiplying them, they are clearly not related to cross-correlation. The complex cells in our algorithm, on the other hand, are modeled by summing the squared outputs of a quadrature pair of simple cells (Ohzawa *et al.*, 1990; Qian, 1994a). If the disparity tuning behavior of the complex cells is largely determined by the cross terms of the squaring operation, then these cells are doing something similar to a cross-correlation. We now show that this is indeed the case.

To simplify the following presentation, let us first rewrite simple cell response expression Eq. (3) as:

$$r_s = L + R \tag{14}$$

where

$$L \equiv \int_{-\infty}^{+\infty} dx f_l(x) I_l(x) \tag{15}$$

$$R \equiv \int_{-\infty}^{+\infty} dx f_r(x) I_r(x) \tag{16}$$

are the filtered left and right retinal images (by the corresponding receptive fields), respectively. With these definitions, the response of the complex cell constructed from a quadrature pair of simple cells can then be written as

$$r_q = (r_{s,1})^2 + (r_{s,2})^2 \tag{17}$$

$$\equiv (L_1 + R_1)^2 + (L_2 + R_2)^2 \tag{18}$$

$$= L_1^2 + L_2^2 + R_1^2 + R_2^2 + 2L_1 \times R_1 + 2L_2 \times R_2 \tag{19}$$

where the subscripts 1 and 2 refer to the two simple cells in the quadrature pair. It can be shown (see the Appendix) that under the same general assumptions for deriving Eq. (8) in "Generalization", the four square terms in the above equation approximately sum to a constant and the disparity tuning behavior of the cell is determined by the last two cross terms. Equation (19) can thus be written as:

$$r_q \approx \text{const.} + 2L_1 \times R_1 + 2L_2 \times R_2. \tag{20}$$

Therefore, the complex cell essentially sums up two related cross-products between the band-pass filtered left and right retinal images, resembling cross-correlation type of operation. In this sense, our model is related to a class of stereo algorithms using complex image phases (Sanger, 1988; Fleet *et al.*, 1991) since those algorithms are also in some ways related to cross-correlation.

However, we would like to emphasize that although the complex cells are doing something similar to cross-correlation, they are quite different from the standard cross-correlators. The standard cross-correlation opera-

*One can easily convince him/herself of this claim by looking at a horizontally uniform pattern generated on a computer monitor or a uniformly painted wall (at an appropriate distance so that the fine features on the wall are not detectable). If a uniform pattern is presented within a dark boundary region, the patch may sometimes appear slightly behind the boundary. However, this depth effect is most likely caused by occlusion instead of stereo vision *per se*.

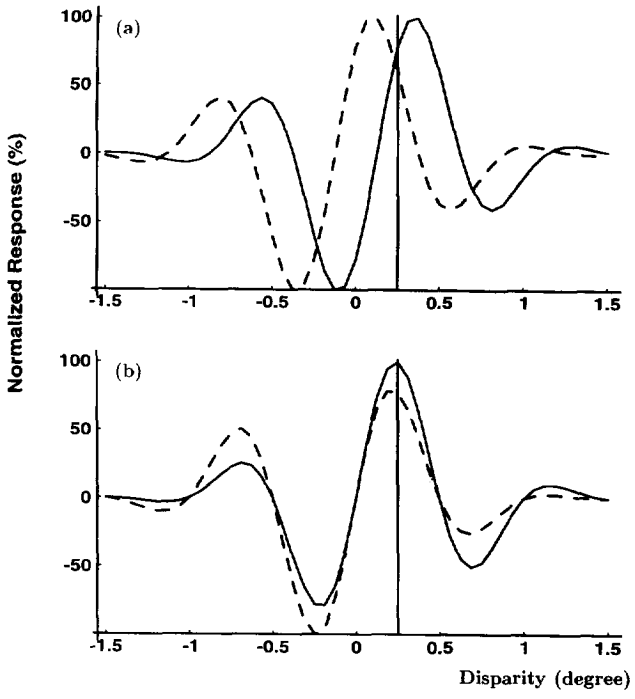

FIGURE 3. Normalized disparity tuning curves to line stimuli with (a) a single cross product term in Eq. (20); and (b) with both cross terms in Eq. (20). Note that there are negative responses at some disparities because we have omitted the unimportant constant term in Eq. (20). For each case, two sets of line stimuli covering the same disparity range but with different lateral locations (−0.125 and 0.125 deg) with respect to the cells' receptive field center were used to obtain two different tuning curves. The main peak locations of the tuning curves using a single cross term depend on the line positions (or equivalently, the Fourier phases), while those using both cross terms do not. The expected location of the main peak according to Eq. (7) is indicated by the vertical lines. The following set of simple cell parameters was used in the simulations: $\omega_0/2\pi = 1$ cycle/degree, $\sigma = 0.25$ deg, and $\Delta\phi = \pi/2$. Sixteen pixels were used to represent 1 deg in the simulations.

tion between the left and right images of a stereogram is defined as:

$$r(d) = \int_{-\infty}^{+\infty} dx I_l(x) I_r(x + d). \tag{21}$$

This expression differs from Eq. (20) in a few important aspects. First, the left and right images in Eq. (20), but not in Eq. (21), are band-pass filtered by the cell's receptive fields before being multiplied. Second, there are two cross-terms in Eq. (20) while only one in Eq. (21). Finally, there is an integration in Eq. (21) across the whole image patches while it is just a product in each cross-term in Eq. (20). We believe that these differences are essential for the complex cells to overcome some of the major problems with the standard cross-correlator. The main problem with the standard cross-correlator is that it is very sensitive to small distortions of the images since distortions will misalign corresponding image pixels. A closely related problem is that one has to use a large number of correlators with different d values in Eq. (21) for disparity computation. This problem becomes worse when one wants to have an algorithm with hyperacuity as d will then have to take sub-pixel increments. Both of these problems can be solved by band-pass filtering, which smoothes the images at a given spatial scale. The smoothing makes the algorithm insensitive to small image distortions so long as the distortions are smaller than the spatial extent of the smoothing operation. As we will show in the next section, as a consequence of the band-pass filtering, as few as two complex cells at each location are sufficient for a reasonable estimation of binocular disparity at that location. Of course, one can also modify the standard cross-correlator by using the band-pass filtered version of the left and right retinal images in Eq. (21). However, the integration in Eq. (21) is computationally far more expensive than the simple products in Eq. (20).

Although band-pass filtering solves the above-mentioned problems with the standard cross-correlation, it also introduces a new problem not present before: the response of a single cross-term in Eq. (20) is sensitive to Fourier phases of input stimulus as well as to disparity. That is why two related cross-terms from a quadrature pair of simple cells need to be added in Eq. (20) to remove the stimulus phase dependence (see the Appendix). The computer simulations demonstrating the importance of adding the two cross-terms in Eq. (20) are presented in Fig. 3. This figure shows that a single cross-product between the filtered left and right retinal images is not sufficient for reliable disparity coding because the peak location of its disparity tuning curve strongly depends on the stimulus Fourier phase.

COMPUTING DISPARITY WITH TWO COMPLEX CELLS

It is easy to show with Eq. (8) that as few as two independent complex cells at each spatial location are sufficient for estimating the disparity at that location. Assume that the two complex cells are constructed from simple cells with their phase parameter differences equal to $\Delta\phi_1$ and $\Delta\phi_2$, respectively. If the responses of these two cells are r_1 and r_2, then the disparity at the location is given by (see the Appendix):

$$D \approx \frac{-1}{\omega_0} \arcsin \frac{r_2 - r_1}{\sqrt{a^2 + b^2}} - \frac{\delta;}{\omega_0}, \tag{22}$$

where

$$a = r_2 \cos \Delta\phi_1 - r_1 \cos \Delta\phi_2, \tag{23}$$

$$b = r_2 \sin \Delta\phi_1 - r_1 \sin \Delta\phi_2, \tag{24}$$

$$\delta = \arctan \frac{a}{b}. \tag{25}$$

We have performed some computer simulations with two complex cells at each location to compute binocular disparity using the above equations. The procedure is similar to that outlined in Fig. 1 except that we now only need two quadrature pairs of simple cells and Eq. (22) is used in the third step for disparity estimation. An example of our simulations is shown in Fig. 4 together with a simulation with eight complex cells at each location used previously (Qian, 1994a). There is no

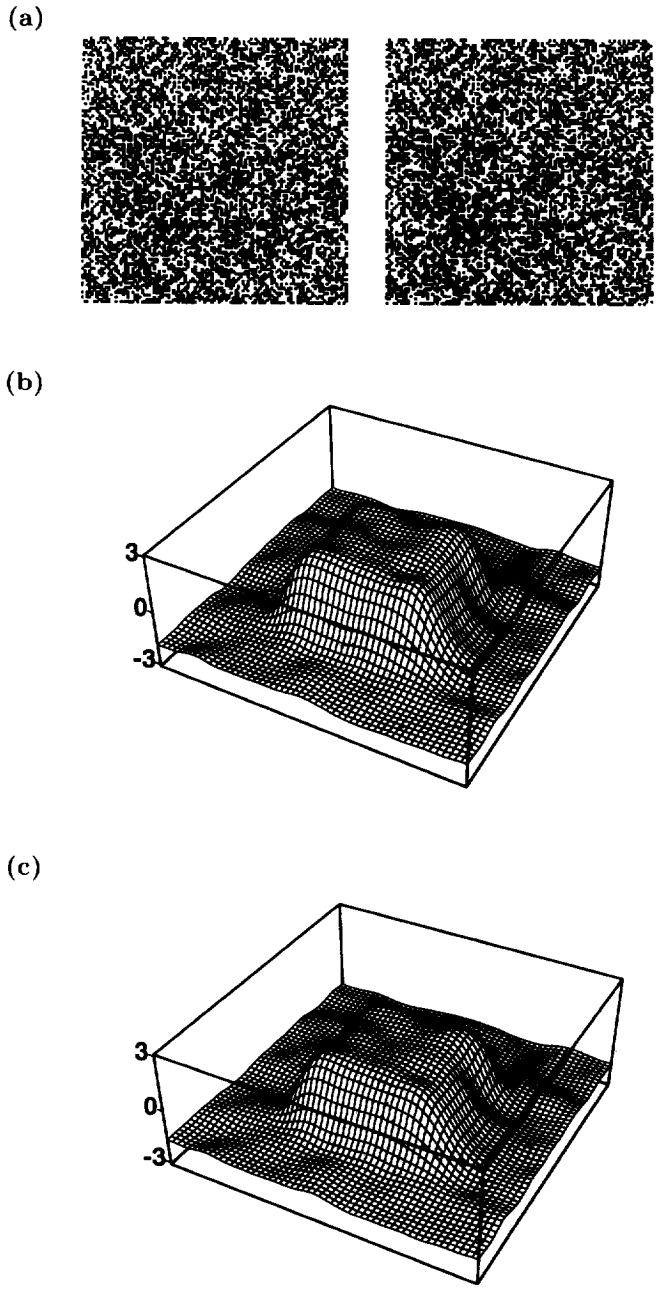

FIGURE 4. (a) A 110×110 random dot stereogram with a dot density of 50% and dot size of 1 pixel. The central 50×50 area and the surround have disparities of 2 and -2 pixels, respectively. When fused with uncrossed eyes the central square appears further away than the surround. (b) The disparity map of the stereogram computed with eight complex cells at each location using the method outlined in Fig. 1. For all cells, $\omega_0/2\pi = 0.125$ cycle/pixel and $\sigma = 4$ pixels, giving a frequency bandwidth (defined at half peak amplitude) of 1.14 octave (Qian *et al.*, 1994). The eight complex cells had their $\Delta\phi$ parameters uniformly distributed in $[-\pi, +\pi]$ starting at $-\pi$. They were constructed from 16 simple cells, eight of which had their (ϕ_l, ϕ_r) parameters equal to $(-6\pi/8, 2\pi/8)$, $(-5\pi/8, \pi/8)$, $(-4\pi/8, 0)$, $(-3\pi/8, -\pi/8)$, $(-2\pi/8, -2\pi/8)$, $(-\pi/8, -3\pi/8)$, $(0, -4\pi/8)$ and $(\pi/8, -5\pi/8)$, respectively. The remaining eight simple cells formed quadrature pairs with the first eight and their (ϕ_l, ϕ_r) parameters were $(-2\pi/8, 6\pi/8)$, $(-\pi/8, 5\pi/8)$, $(0, 4\pi/8)$, $(1\pi/8, 3\pi/8)$, $(2\pi/8, 2\pi/8)$, $(3\pi/8, \pi/8)$, $(4\pi/8, 0)$ and $(5\pi/8, -\pi/8)$, respectively. The resulting eight complex cells were tuned to disparities -4, -3, -2, -1, 0, 1, 2, and 3 pixels, respectively. With the current set of parameters, the cells tuned to -4 and +4 pixels were identical, and because of the parabolic interpolation used in locating the peaks of responses, the actual disparity range covered by the cells was [-4 pixels, +4 pixels]. (c) The disparity map of the same stereogram computed with two complex cells at each location. The two cells were picked from the eight cells used in (a) that were tuned to -1 and +1 pixel of disparity. The method is the same as that shown in Fig. 1, except that the third step is replaced by Eq. (22). The distance between two adjacent sampling lines in (b) and (c) represents a distance of 2 pixels in (a). Negative and positive values indicate near and far disparities, respectively.

significant difference between the two simulation results. Although it is not known whether the real visual system uses only two complex cells at each location and frequency band to compute binocular disparity, this result does demonstrate how efficiently complex cells encode binocular disparity.

It can be seen from the general derivation of Eq. (8) that the reason that only two complex cells are needed for

disparity computation is the band-pass filtering. Intuitively, after filtering the images through the filters with preferred frequency ω_0, the outputs contain Fourier power mainly at ω_0 and can therefore be approximately represented by only two samples based on Shanon's sampling theorem. This gain of efficiency is accompanied by the occurrence of side peaks around the main peak in a cell's disparity tuning curve, which in turn, requires that the cells with preferred frequency ω_0 only code disparity within the range $[-\pi/\omega_0, \pi/\omega_0]$ to avoid ambiguity (Qian, 1994a).

IMPROVING THE MODEL WITH SPATIAL POOLING FOR COMPLEX CELL RESPONSES

Our stereo vision algorithm can be significantly improved by taking into account the additional physiological fact that the receptive field sizes of real complex cells are, on average, larger than those of the simple cells at the same eccentricity (Hubel & Wiesel, 1962; Schiller *et al.*, 1976). We proposed recently (Qian & Zhu, 1995; Zhu & Qian, 1996) that this fact can be incorporated into the model by averaging *several* quadrature pairs of simple cells with nearby and overlapping receptive fields (and with otherwise identical parameters) to construct a model complex cell. Mathematically, this spatial pooling process for obtaining the complex cell response is given by:

$$r_c = r_q * w \tag{26}$$

where r_q is the response of a single quadrature pair given by Eq. (4), w is a spatial weighting function, and $*$ denotes the spatial convolution operation. In our simulations, the weighting function w was chosen to be a symmetric two-dimensional (2D) Gaussian. We show below that the disparity tuning curve of the resulting complex cell (r_c) is much more reliable than that constructed from a *single* quadrature pair (r_q) of simple cells used previously. This in turn improves the quality of the computed disparity maps from stereograms.

To understand the effect of the spatial pooling, we need a more accurate expression for the response of a single quadrature pair. As we have shown elsewhere (Zhu & Qian, 1996), with Eqs (1) and (2) as the simple cell receptive field profiles, the quadrature pair response to a stimulus with Fourier transform $|\tilde{I}(\omega)|e^{i\theta_I(\omega)}$ and disparity D is exactly given by

$$r_q = 8\pi\sigma^2 \int_0^\infty \int d\omega d\omega' |\tilde{I}(\omega)||\tilde{I}(\omega')| \times e^{-(\omega-\omega_0)^2\sigma^2/2} e^{-(\omega'-\omega_0)^2\sigma^2/2} \times \cos\left[\frac{\theta_I(\omega') - \theta_I(\omega)}{2} + \frac{(\omega-\omega')D}{2}\right] \times \cos\left[\frac{1}{2}(\Delta\phi - \omega D)\right] \cos\left[\frac{1}{2}(\Delta\phi - \omega' D)\right]. \tag{27}$$

According to this expression, the response of a quadrature pair depend on the difference of the Fourier phases of the input stimulus measured at two different frequencies ($\theta_I(\omega')-\theta_I(\omega)$). The integrand contains two Gaussian factors that are significantly large only when both ω and ω' are close to ω_0. If we approximate the Gaussian functions as the Dirac delta functions centered at ω_0 and carry out the integrations, Eq. (27) then reduces to the approximate complex cell response expression in Eq. (5), which is independent of stimulus Fourier phases.

This means that the complex cell constructed from a single quadrature pair is only approximately independent of the stimulus Fourier phase. The approximation is a good one for simple patterns such as lines, bars or gratings. For these patterns, their Fourier phases are continuous functions of frequency. Since the two Gaussian terms effectively make $\omega' - \omega$ very small, they also make $\theta_I(\omega') - \theta_I(\omega)$ close to zero. We can therefore neglect the θ dependence in Eq. (27) for these stimuli by assuming

$$\cos\left[\frac{\theta_I(\omega') - \theta_I(\omega)}{2} + \frac{(\omega-\omega')D}{2}\right] \approx \text{const.} \tag{28}$$

However, $\theta_I(\omega)$ is not a smooth function of ω for stimuli such as random dot patterns, and this is when the pooling step for computing complex cell responses becomes important. In this pooling step the responses of several quadrature pairs with nearby receptive fields (and with otherwise identical parameters) are averaged. The response expressions [Eq. (27)] for the different quadrature pairs are identical except for the $\theta_I(\omega)$ functions,

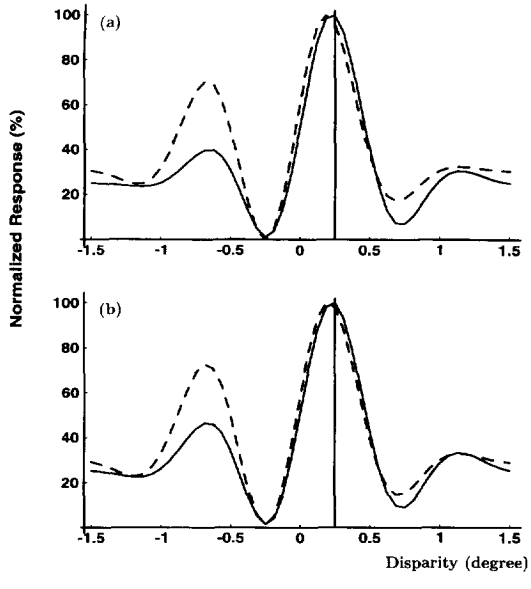

FIGURE 5. Normalized disparity tuning curves to line stimuli of the model complex cells (a) without spatial pooling; and (b) with spatial pooling. For each model cell, two sets of line stimuli covering the same disparity range but with different locations on the cell's receptive fields were used to obtain two different tuning curves. The peak locations of the tuning curves to the two sets of lines are very similar regardless of whether the spatial pooling is used. The expected location of the main peak according to Eq. (7) is indicated by the vertical lines. The parameters used in this simulation were identical to those used in Fig. 3. The σ_w of the spatial weighting function used in the pooling step of (b) was 4 pixels.

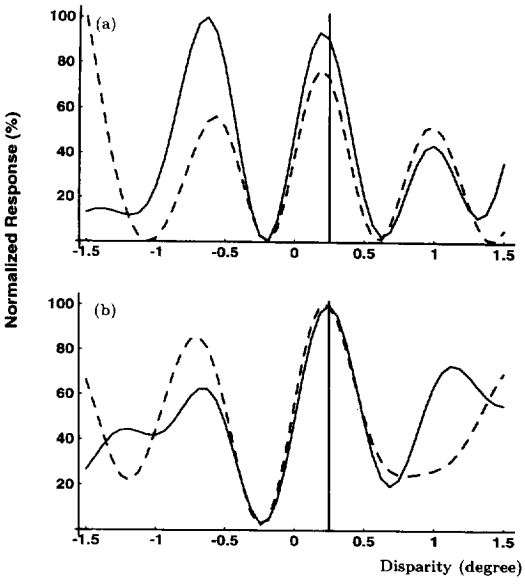

FIGURE 6. Disparity tuning curves to random dot stimuli of the model complex cells (a) without spatial pooling; and (b) with spatial pooling. For each model cell, two sets of independently generated random dot stimuli covering the same disparity range were used to obtain two different tuning curves. For the cell without spatial pooling the peak locations of the tuning curves to the two sets of random dots may often be very different, as is the case in (a). For the cell with spatial pooling the main peak locations of the two tuning curves are always very similar. The expected location of the main peak according to Eq. (7) is indicated by the vertical lines. The parameters used in this simulation were identical to those used in Fig. 5.

which are different for different pairs because they are centered on somewhat different parts of the stimulus. Therefore, the pooling step simply averages over the θ dependent cosine term in Eq. (27), and makes it approximately constant. The approximation in Eq. (28) is thus also valid for random dot type of stimuli after the pooling. We therefore expect that the pooling should significantly improve the reliability of disparity tuning to those patterns whose Fourier phases are not smooth functions of the frequency.

We have confirmed the above analysis through computer simulations. Two model complex cells are considered in our simulations, one with the spatial pooling and the other without. We first examined the sensitivity of these cells to the Fourier phases of line stimuli. For this purpose, we computed, for each complex cell, two disparity tuning curves using two sets of line stimulus covering the same disparity range but with different lateral locations. The results are shown in Fig. 5. As we expected, the pooling does not make much difference in this case: even without the pooling the peak locations of the disparity tuning curves are about the same for the different lateral positions (or equivalently, the Fourier phases) of the line stimuli. We next examined the sensitivity of the same two complex cells to the Fourier phases of random dot patterns. We first generated two independent random dot patterns and then used each of them to create a set of binocular stimuli of various uniform disparities. We then measured the disparity tuning curves of the two model complex cells to these two independent sets of random dot stimuli which contain the same set of disparity values but different Fourier phases. The results are shown in Fig. 6. It is clear that in this case, the pooling greatly improved the reliability of the disparity tuning by reducing the phase dependence. Indeed, without the pooling, the main peaks of the tuning curves are sometimes far away from the expected locations given by Eq. (7), as is the case in Fig. 6(a).

Based on the above results, we have modified our previous procedure for computing disparity maps shown in Fig. 1 to the one in Fig. 7. The second step of the new procedure computes complex cell responses by averaging over *several* quadrature pair responses. Mathematically, this step can be broken down into the two steps shown to the right in Fig. 7, the first of which computes responses of single quadrature pairs (just like step 2 of the old procedure), and the second applies spatial pooling. The final smoothing step in the old procedure has been removed in the new method because it is no longer necessary (see below). Therefore, both the new and old procedures contain four steps in them, and the only difference between them is that the order of the last two steps has been switched.

We have performed computer simulations with the new procedure and an example for the stereogram in Fig. 4(a) is shown in Fig. 8(a). For comparison, the disparity map computed from the same stereogram with our

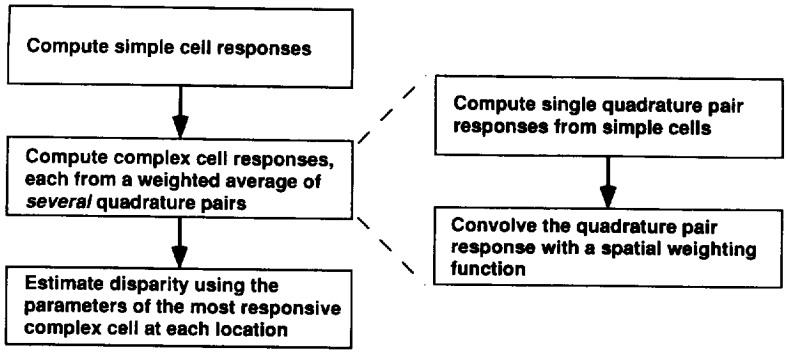

FIGURE 7. The modified algorithm from computing disparity maps from stereograms. The second step can be viewed as being composed of the two steps shown to the right so that there is also a total of four steps in the new algorithm. The only difference between this procedure and the old one shown in Fig. 1 is that the two final steps have been switched.

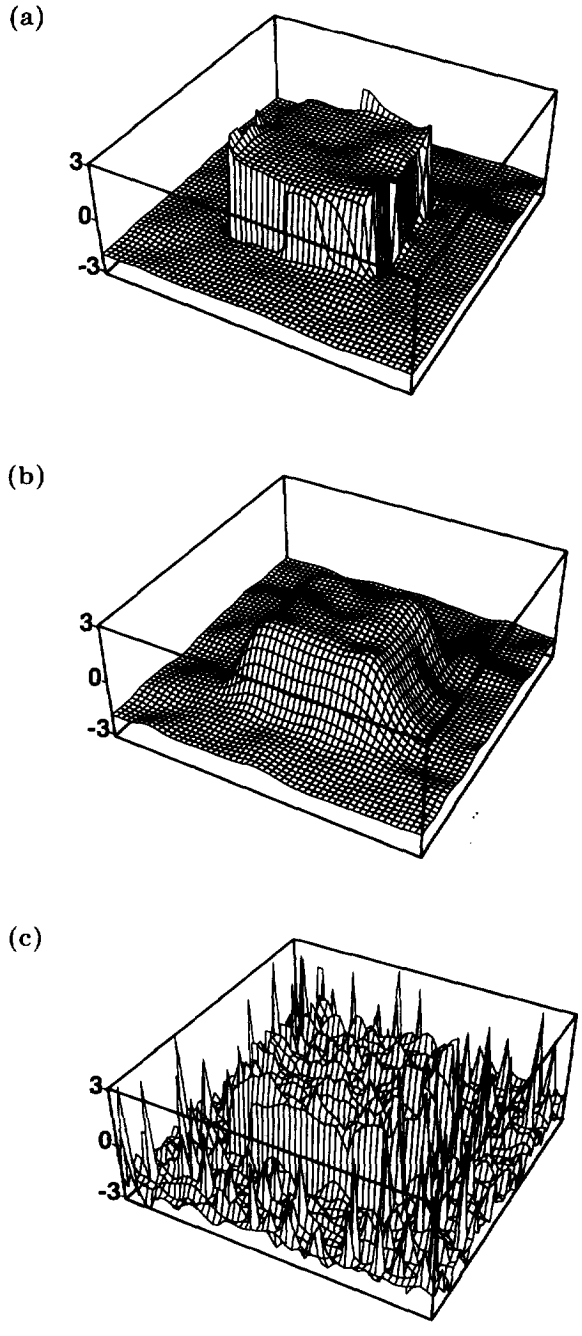

FIGURE 8. Disparity maps of the random dot stereogram in Fig. 4(a) computed with (a) the new algorithm shown in Fig. 7; (b) the old algorithm shown in Fig. 1; and (c) the old algorithm with the final smoothing step omitted. Disparity boundaries computed with the new algorithm are much sharper than those with the old algorithm. Eight complex cells were used at each spatial location. The plot in (b) is copied from Fig. 4(b) and is shown here for comparison. The receptive field parameters used in computing the three disparity maps were identical. The σ_w of the spatial weighting function was 4 pixels. The distance between two adjacent sampling lines in these plots represents a distance of 2 pixels in the stereogram.

previous algorithm is also shown in Fig. 8(b). The disparity map obtained with the new method is significantly better than that with the old method, especially around the disparity transition boundaries: while the transition occurs gradually over a distance of about 15 pixels in the old map, it takes only about 4 pixels in the new map. To our knowledge, Fig. 8(a) is the first demonstration that sharp disparity transition boundaries can be obtained with a physiologically realistic mechanism.

It should be noted that the slow transition with the old method is mainly caused by the final smoothing step (see Fig. 1) which has to be used in order to remove large noisy fluctuations in the disparity maps obtained in the previous step. To see this more clearly, we show in Fig. 8(c) the result from the old method with the final smoothing step omitted. Although the transition boundaries appear sharp, the map is too noisy to be useful. With the new method the final smoothing step is no longer necessary due to the improved reliability of the disparity tuning of the model complex cells. We conclude that the spatial pooling for computing complex cell responses in the new method does not directly "sharpen" the disparity transition boundaries; rather, it helps eliminate the final smoothing step in the old method which destroys the sharp boundaries.

To compare the three disparity maps in Fig. 8 more quantitatively, we plot in Fig. 9 the error distributions for these maps. The errors were obtained by subtracting an idealized disparity map from the computed maps. The

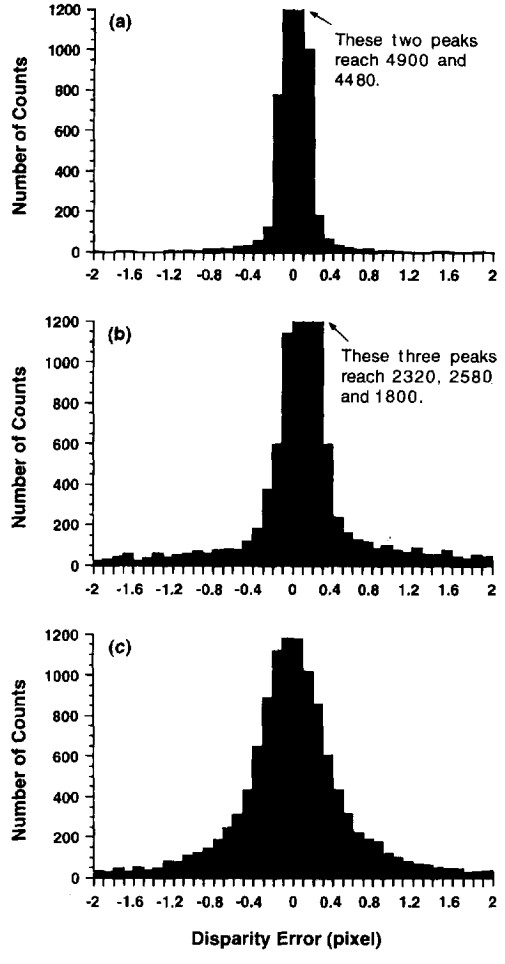

FIGURE 9. The error distributions for the three disparity maps shown in Fig. 8. The errors were obtained by subtracting an idealized disparity map from the computed maps (see text). The error distribution for the new method (a) is more closely centered around 0 than those for the old method with or without the final smoothing step (b and c).

idealized map has disparities of 2 and −2 pixels for the central square region and the surround, respectively, and the transition across the disparity boundaries occurs over 1 pixel.* Fig. 9 indicates that the error distribution for the new method [Fig. 9(a)] is more closely centered around zero than those for the old method with or without the final smoothing step [Fig. 9(b and c)]. The proportions of points with an absolute error less than 0.1 pixel are 78%, 40% and 20% for the three distributions, respectively, and the mean absolute errors† are 0.16, 0.35 and 0.59 pixel, respectively. Although the final smoothing step in the old method also greatly reduces error, it is not as effective as the pooling step at the complex cell level in the new method, and it is not as physiologically justified.

A key parameter in the new method is the width of the Gaussian weighting function (σ_w) for computing complex cell responses through spatial pooling. We noted in a previous publication (Zhu & Qian, 1996) that any $\sigma_w > 1$ can greatly improve the reliability of the complex cells' disparity tuning curves. To see how σ_w affects the performance of the algorithm we plot in Fig. 10 the mean absolute error of the computed disparity map as a function of σ_w. The maps in Fig. 8(c) and Fig. 8(a) correspond to σ_w equal to 0 (no pooling) and 4 pixels in Fig. 10, respectively. The solid, dashed, and dotted curves are the results for all points, points near disparity boundaries, and interior points away from the disparity boundaries in the disparity maps, respectively. It is clear from Fig. 10 that the errors from the boundary regions are much larger than those from the rest of the maps, that the spatial pooling significantly reduces errors in all three curves, and that the effect of the spatial pooling is not very sensitive to σ_w so long as it is larger than 1 pixel. The exact form of the weighting function for spatial pooling is also not important (Zhu & Qian, 1996). Indeed we found that very similar results can be obtained by using a rectangular weighting function covering a line of five consecutive vertical positions. This indicates that it is sufficient for a complex cell to contain about five quadrature-pair subunits to achieve reliable disparity tuning. The spatial pooling step improves the interior points most, with an over 10-fold error reduction. The resulting error for these points is as small as 0.05 pixel. If we identify the widths of the model simple cells (about $2\sigma = 8$ pixels) used in our simulation with the monkey foveal receptive field sizes [0.1–0.2 deg; see Dow *et al.* (1981)] then a 0.05 pixel resolution is equivalent to 2.3–4.5 sec of visual angle, comparable to the human

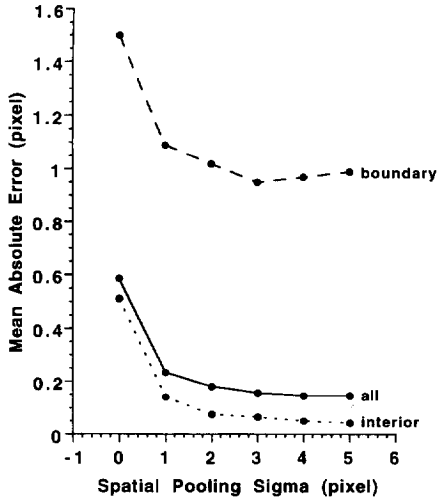

FIGURE 10. The mean absolute error of the computed disparity map is plotted as a function of the width of the Gaussian weighting function (σ_w) used in the spatial pooling step of the new method. The maps in Fig. 8(c and 8a) correspond to σ_w equal to 0 (no pooling) and 4 pixels in Fig. 10, respectively. The solid, dashed, and dotted curves are the results for all points, subset of points within 5 pixels around disparity boundaries, and subset of interior points more than 10 pixels away from the disparity boundaries in the maps, respectively.

stereoacuity (Ogle, 1952; Blackmore, 1970; Westheimer, 1979; Schumer & Julesz, 1984).

We showed in a previous section that stimulus disparity can be computed with only two complex cells at each location. Interestingly, the spatial pooling step for computing complex cell responses does not help improve the two-cell algorithm. The result (not shown) from the new two-cell algorithm is essentially the same as that obtained with the old method shown in Fig. 4(c), with slow transition at disparity boundaries. This is probably due to the fact that the two-cell algorithm depends on the response *magnitudes* while with more cells only the *peak location* of the responses among the cell population is important. The response magnitudes are more likely to be affected by the presence of two different disparities at the transition boundaries than the response peak location.

MULTIPLE SPATIAL SCALES

The results reported so far are all based on a set of front-end filters (binocular receptive fields) at a single spatial scale (i.e., a single set of values for the Gaussian width σ and preferred frequency ω_0 in Eqs (1) and (2)). Since the cells in the visual cortex cover a wide range of σ and ω_0 (DeValois *et al.*, 1982; Shapley & Lennie, 1985) and since the visual system are known to analyze stimuli through multiple frequency channels (Campbell & Robson, 1968; Graham & Nachmias, 1971), it is interesting to compare disparity maps computed by cells at different spatial scales and to consider how these maps may be combined into a unitary percept. Figure 11(a–c) shows the disparity maps of a random dot stereogram computed with filters at three different spatial scales. The parameters for computing Fig. 11(b) are identical to those used in Fig. 4(b). The parameters for Figs 4(a) and 4(c) are scaled down and up by a factor of 1.5 in the spatial

*Note that the actual human perception on a random dot stereogram may not be as perfect as the idealized disparity map. In particular there are two 4-pixel-wide stripes on each side of the central square region along the x-axis whose disparities are undefined because the dots in these stripes do not correspond between the left and right images. The calculated errors are thus somewhat exaggerated around the disparity boundaries.

†We did not use the more standard root-mean-square error in this paper because it tends to over-represent the outliers in the error distributions that mainly come from the disparity boundary regions where the errors are somewhat exaggerated (see the previous footnote).

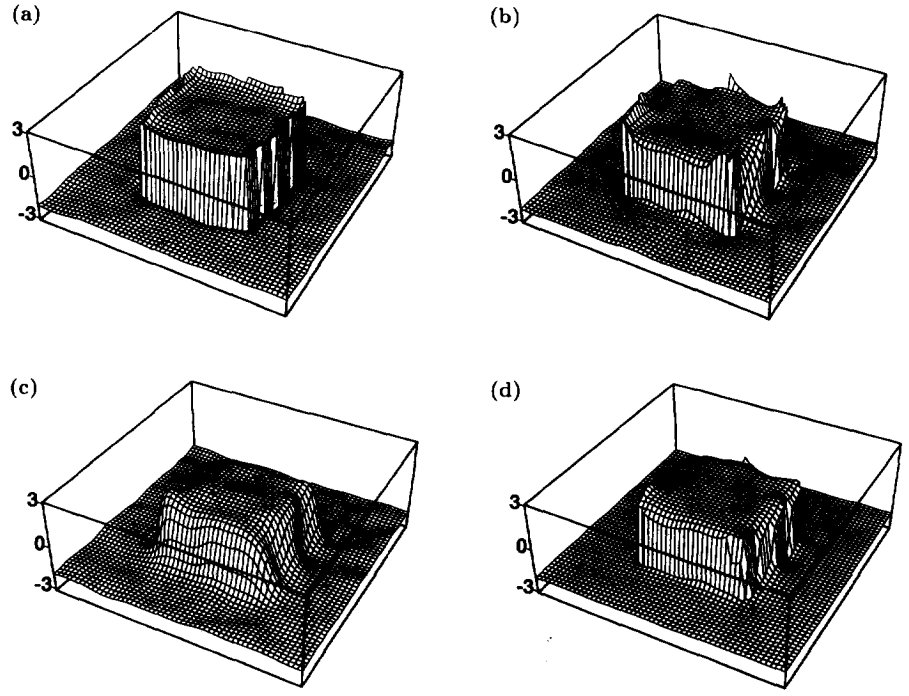

FIGURE 11. The disparity maps of a random dot stereogram (not shown) computed with cells at three different spatial scales (a–c) and the average across the scales (d). The receptive field parameters for (b) are identical to those used in Fig. 4(b). The parameters for (a) and (c) are scaled down and up by a factor of 1.5 in their spatial dimension (or equivalently, scaled up and down in the frequency domain), respectively. The frequency bandwidths of the filters in all three scales are equal to 1.14 octaves.

dimension (or equivalently, scaled up and down in the frequency domain), respectively. The frequency bandwidths of all filters in the three scales are equal to 1.14 octaves. It can be seen from Fig. 11(a–c) that cells at each scale can compute the disparity map independently. As the spatial scale increases, the sharpness of transition at disparity boundaries gradually deteriorates [the transition distances are about 2, 4 and 8 pixels for Fig. 11(a–c), respectively]. The mean absolute errors for the three maps are 0.16, 0.15 and 0.24, respectively. However, larger scales have the advantage of being able to compute a wider range of disparities (see "The Model").

Psychophysical evidence indicates that disparity signals from different frequency channels interact with each other (Wilson *et al.*, 1991; Rohaly & Wilson, 1993, 1994; Smallman, 1995; Mallot *et al.*, 1996). Computational studies have also suggested possible ways of pooling across different scales (Marr & Poggio, 1979; Sanger, 1988; Grzywacz & Yuille, 1990; Fleet *et al.*, 1996). The exact mechanism used by the brain for combining scales, however, remains unknown. The simplest method is to average across the disparity maps computed by different scales (Sanger, 1988). Such an average for Fig. 11(a–c) is shown in Fig. 11(d). The mean absolute error of the whole map is 0.12 pixel, better than those of the individual maps. The transition over disparity boundaries occurs over a distance of about 4 pixels. Obviously, the sharpness of disparity boundaries in the averaged map depends on how many small and large spatial scales are included in the average. An over-representation of large spatial scales in the average will clearly destroy the sharp boundaries.

It should be noted that we are not assuming that the scale averaging is a step for modeling the responses of primary visual cortical cells. Such an operation would render the cells insensitive to spatial frequency (Zhu & Qian, 1996), contradictory to experimental facts. Instead, the population activity of many families of cells at different scales in the primary visual cortex might directly correspond to an overall percept determined by the averaging process. Alternatively, the averaging could be explicitly performed at a stage beyond the striate cortex, such as area MT (Grzywacz & Yuille, 1990).

POSITION-SHIFT RECEPTIVE FIELD MODEL

The binocular receptive field model proposed by Freeman *et al.* (Ohzawa *et al.*, 1990; Freeman & Ohzawa, 1990; DeAngelis *et al.*, 1991) assumes that the left and right receptive field profiles of a simple cell have the same envelopes (on the corresponding left and right retinal locations) but different phase parameters for the excitatory/inhibitory modulations within the envelopes. An alternative assumption preceding this phase-difference model is that there may be an overall positional shift (for both the envelopes and modulations) between the two profiles (Bishop *et al.*, 1971; Maske *et al.*, 1984; Wagner & Frost, 1993). The third possibility is a hybrid which assumes that the two profiles differ by both an overall positional shift and a phase difference (Jacobson *et al.*, 1993; Zhu & Qian, 1996; Fleet *et al.*, 1996). We have previously investigated the subtle but important differences between these receptive field models and suggested methods for correctly distinguishing them

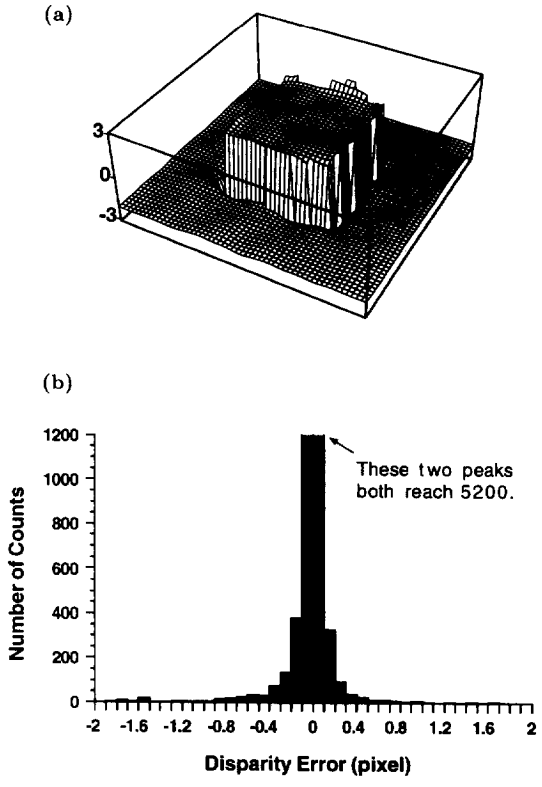

FIGURE 12. (a) Computed disparity map of the random dot stereogram shown in Fig. 4(a) using the position-shift based receptive field models with the new algorithm in Fig. 7. The result is similar to Fig. 8(b) which was computed with the phase-parameter based receptive field model on the same stereogram. Eight complex cells were used at each spatial location. The parameters of the cells were identical to those in Fig. 8(b), except that the phase-parameter differences were replaced by the equivalent positional shift parameters. (b) Error distribution for the map in (a).

experimentally (Qian, 1994b; Zhu & Qian, 1996; see also Fleet *et al.*, 1996).

So far in this paper we have been using the phase-difference based receptive field model in our analyses and simulations. We now demonstrate that our algorithm for disparity computation also works with the position-shift based receptive field models. It is easy to show that if we assume an overall horizontal positional shift Δx between the left and right receptive fields of a simple cell, the complex cell response Eq. (8) should then be replaced by:

$$r_q \approx c^2|\tilde{I}(\omega_0)|^2 \cos^2\left(\frac{\omega_0 \Delta x}{2} - \frac{\omega_0 D}{2}\right). \qquad (29)$$

The preferred disparity of the complex cell is therefore equal to the shift Δx (Zhu & Qian, 1996; Qian & Andersen, 1997). This equation indicates that a population of complex cells with the different position-shift parameter Δx can also form a distributed representation of stimulus disparity, and the same procedure outlined in Fig. 7 can be used to compute disparity maps from stereograms. An example of our computer simulations on the random dot stereogram in Fig. 4(a) is shown in Fig. 12. Both the computed map and the error distribution are very similar to those obtained with the phase-difference receptive field model on the same stereogram [see Fig. 8(a) and Fig. 9(a)]. The proportion of points with an absolute error less than 0.1 pixel is 86%, better than the 78% for the phase-difference algorithm. The mean absolute error, however, is slightly higher at 0.18 pixel (0.16 pixel for the phase algorithm) due to the larger number of outliers in the error distribution.

The results in the previous sections regarding relationship to cross-correlation, two-complex-cell algorithm and spatial pooling also apply to the position-shift based algorithm. However, the position-shift based algorithm does not naturally predict a zero disparity bias because the receptive field shapes of cells tuned to different disparities can all be identical and therefore all have the same DC component. In addition, it does not naturally predict the observed size–disparity correlation (Smallman & MacLeod, 1994; Schor & Wood, 1983) because unlike the phase-difference based algorithm, the preferred disparity of a complex cell is always equal to the shift parameter Δx, regardless of its preferred spatial frequency (ω_0) or the dominant spatial frequency in the stimulus (Ω) (Zhu & Qian, 1996).

SUMMARY AND DISCUSSION

The central question of the stereoscopic depth perception is how the visual system determines which parts on the two retinal images come from the same object in the real world, the so-called correspondence problem. In the case of seeing depth in random dot stereograms the correspondence problem is usually stated as finding explicitly which dot (or other features such as edge of dot) in the left image matches which in the right image. Since all dots in the two images of a random dot stereogram are of identical shape, it is often argued that any two dots, one from each image, could potentially match and that the visual system is faced with an enormously difficult problem of sorting out the right matches from a huge number of false ones. On the other hand, if one considers an implicit version of the correspondence problem by using image patches instead of the fine features for matching, finding disparity in a random dot stereogram becomes a conceptually simple task: it can be solved by computing cross-correlations between the left and right image patches at various relative shifts between them, and then determining which shift produces the largest response.* Since the receptive fields of real visual cortical cells are not point-like, the stereo algorithm used by the brain must also operate on image patches rather than on individual fine features. As we have demonstrated previously (Qian, 1994a) and in this paper, model complex cells with realistic physiological properties can indeed be used to compute disparity maps from random dot stereograms through an operation related to but much more sophisticated than the standard

*When the patch size is reduced to that of a single dot, the implicit version becomes identical to the explicit version of the correspondence problem. At this limit, the cross-correlation response from a correct match and that from a false match are equally strong and that is where the complication of distinguishing correct matches from false ones occurs.

cross-correlation, without facing an explicit correspondence problem. We therefore conclude that random dot stereograms probably do not really pose a computational challenge to the visual system. The explicit version of the correspondence problem may not exist in the brain.

Although many of the existing stereo vision algorithms also avoid the explicit correspondence problem by operating on image patches, most of them cannot be said to be truly physiological because of certain mathematical operations used in them. Our stereo model, on the other hand, is entirely based on known physiological facts. The main goal in this paper is to provide a better and more intuitive understanding of how the model works. We first showed that although we originally derived our algorithm using the Gabor functions as the cells' receptive field profiles, the model works under some very general assumptions about the receptive field properties of binocular cells, and it works for both the phase-difference and the position-shift types of receptive field descriptions. The details of the receptive field profiles are thus not very important in most circumstances. We then showed that if the Gabor functions are used as the receptive field profiles, our stereo algorithm has a small bias towards zero disparity because of the DC components in the Gabor functions. This bias naturally explains the fact that we see zero disparity in horizontally uniform patterns, which by themselves are physically consistent with any disparity values. This result is particularly interesting because there is good evidence indicating that the spatial receptive field profiles of real visual cortical cells can indeed by modeled by the Gabor functions (Jones & Palmer, 1987; Ohzawa *et al.*, 1990).

The DC component of a filter is usually considered as undesirable precisely because of the bias it introduces. Here we have shown that the bias can actually be a useful feature: it allows the visual system to pick the smallest (zero) of the disparity values that are physically consistent with ambiguous stereo stimuli. A similar perceptual bias also exists in motion perception under the name of the "aperture" problem: when we see an oriented pattern moving behind an aperture, we only see the velocity component perpendicular to the orientation of the pattern. Equivalently, one can say that our perception is biased towards seeing the smallest possible speeds that are consistent with ambiguous motion stimuli. It would be interesting to see if the spatiotemporal receptive field profiles of real visual cortical cells could also allow a natural explanation of the motion "aperture" problem just as we showed in this paper for the zero disparity bias in stereo vision.

We also found through mathematical analysis that the complex cell described by Freeman and coworkers essentially sums up two related cross-product terms between the band-pass filtered left and right retinal images. This result is interesting because it provides an intuitive understanding of how the complex cells compute binocular disparity. Indeed, it is difficult to see in the original quadrature pair construction how the complex cells encode disparity. We further compared the complex cells with the standard cross-correlator and pointed out that they avoid several major problems of the latter. In particular we showed that unlike the standard cross-correlators, as few as two complex cells at each spatial location are sufficient for a reasonable estimation of the binocular disparity at that location.

Finally, we showed that our stereo vision algorithm can be significantly improved by considering the additional physiological fact that the receptive field sizes of real complex cells are larger than those of the simple cells. This is incorporated into the model by adding a spatial pooling step for computing complex cells' responses. Owing to the improved reliability of the disparity tuning behavior of the model complex cells, we no longer need the final smoothing step used in our previous algorithm. As a consequence, the disparity maps computed with the new algorithm have sharp transitions at disparity boundaries similar to our perception. In fact, one of the main problems with many existing stereo algorithms is the slow transition at disparity boundaries in the computed disparity maps. Although there are engineering type approaches to fixing the problem, we believe that our algorithm is among the first that solves the problem with a simple and physiologically plausible method. It would also be interesting to experimentally test the idea of the spatial pooling by studying the reliability of complex cells' disparity tuning to line and random dot patterns (cf. Figs 5 and 6).

Psychophysical comparisons

There is a large body of psychophysical literature documenting various aspects of the human stereoscopic depth perception. How our stereo model compares with the existing psychophysical data is the subject of ongoing research. Here we briefly discuss several interesting cases.

It is well known that we can still perceive depth when the contrasts of the two images in a stereo pair are very different so long as they have the same sign (Julesz, 1971). We have shown previously that our algorithm shows the same behavior (Qian, 1994a). Specifically, if the contrast ratio of the lower contrast image to the higher one is γ, the cosine function in Eq. (8) should be multiplied by γ. Our algorithm still works so long as γ is positive (i.e., same contrast sign) but the amplitude of the disparity tuning curves, and consequently the reliability of disparity detection, decreases with decreasing γ (i.e., increasing contrast difference).

Westheimer (1986) found that a few vertical line segments at different disparities, separated laterally along the horizontal fronto-parallel direction, influence each other's perceived depth in the following way: when the lateral distance between the lines is small (less than about 5 min), the lines appear closer in depth as if they are attracting each other. At larger distances, this effect reverses and the lines appear further away from each other (repulsion). When the distance is very large there is no interaction between the lines. We recently analyzed

how the responses of a population of model complex cells centered on one line are influenced by the presence of another line at various distances (Qian & Zhu, 1997). It was found that by averaging across all cell families with different bandwidths and preferred frequencies, the model can naturally explain Westheimer's observation without introducing any *ad hoc* assumptions.

Our model is consistent with the observation that depth in a stereogram can only be observed when there is overlapping spatial frequency content between the two images in a stereogram (Julesz, 1971). This property is shared by all algorithms including ours that perform matching in separate frequency channels. A related observation is that stereopsis is not impaired by the introduction of uncorrelated monocular noise if the noise energy is two octaves or more from that specifying the disparity (Julesz, 1975; Yang & Blake, 1991). To account for this observation, one can simply assume that when averaging results across different frequency channels, the contribution from each channel should be weighted by its disparity signal strength. A number of studies also indicate that strong and sophisticated interactions exist between different frequency channels (Wilson *et al.*, 1991; Rohaly & Wilson, 1993, 1994; Smallman, 1995; Mallot *et al.*, 1996). How to combine outputs from different frequency channels to account for these observations remains an open question. The simple averaging scheme used in Fig. 11(d) is unlikely to be sufficient.

As mentioned in "The Model", when the phase-difference type of receptive field profiles (Ohzawa *et al.*, 1990) are used as the front-end filters, the algorithm predicts a correlation between the perceived disparity range and the dominant spatial frequency in the stimulus (Smallman & MacLeod, 1994; DeAngelis *et al.*, 1995; Zhu & Qian, 1996). Such a correlation has been reported psychophysically (Smallman & MacLeod, 1994; Schor & Wood, 1983). However, the observed disparity range is somewhat larger than that allowed by the algorithm with purely phase-difference types of receptive fields. This discrepancy can be remedied by using a hybrid receptive field model containing contributions from both phase-difference and positional shift (Smallman & MacLeod, 1994; DeAngelis *et al.*, 1995; Zhu & Qian, 1996; Fleet *et al.*, 1996).

The disparity boundaries computed with our algorithm appear to be as sharp as the human perception although we are not aware of any psychophysical studies in this regard to make a quantitative comparison. The error of the computed disparity values at locations away from the disparity boundaries falls in the range of the human stereoacuity (see the section "Improving the Model with Spatial Pooling for Complex Cell Responses"). It is known that the disparity discrimination threshold increases rapidly with the magnitude of the base disparity (Ogle, 1952; Blackmore, 1970; Westheimer, 1979; Schumer & Julesz, 1984). Our model may also be able to explain this observation for the following reason. As we have already mentioned, a family of complex cells with preferred spatial frequency ω_0 can only encode disparity in the range $[-\pi/\omega_0, \pi/\omega_0]$ (Qian, 1994a; Zhu & Qian, 1996). Therefore, for a given stimulus disparity D, only those cell families with preferred spatial frequency ω_0 smaller than π/D can encode the disparity. Consequently, as the base disparity of the stimulus increases, cell families with finer spatial scales will not be able to contribute to the disparity computation, and the variance of the model output will increase. This, in turn, will require a larger disparity increment for reliable discrimination (i.e., a higher discrimination threshold). We are currently investigating this possibility. Our algorithm is also consistent with the observation that depth is perceived in stereograms without localized image features such as zero-crossings (Arndt *et al.*, 1995). This is because the algorithm directly operates on image patches without first extracting image features.

A number of studies (Ramachandran *et al.*, 1973b; Sato & Nishida, 1993; Hess & Wilcox, 1994; Wilcox & Hess, 1995) have suggested the existence of two different stereoscopic mechanisms analogous to the Fourier and non-Fourier systems of motion detection (Ramachandran *et al.*, 1973a; Derrington & Badcock, 1985; Chubb & Sperling, 1988). The Fourier disparity is specified by the relative displacement of luminance profiles (a first-order image property) in the two retinal images, while the non-Fourier disparity is defined by higher-order image properties such as subjective contours, second-order textures, or envelopes of luminance modulations. In a non-Fourier stereogram, the luminance profiles of the two images are either uncorrelated, or correlated but unrelated to the perceived disparity. Our stereo model in its current form can only detect Fourier disparity since it depends on the similarity of luminance profiles in the two retinal images. A second parallel pathway with additional non-linearities has to be added to the model for the detection of the non-Fourier disparities (Wilson *et al.*, 1992). Similarly, our current model cannot explain the perceived depth in stereograms with unmatched monocular elements that simulate occlusions (Shimojo & Nakayama, 1990; Nakayama & Shimojo, 1990; Liu *et al.*, 1994). Finally, the model is limited by only including short-range interactions within the scope of the classical receptive fields of primary visual cortical cells. Long-range connections between these cells and influences outside the classical receptive fields have been documented physiologically (Ts'o *et al.*, 1986; Das & Gilbert, 1995; Allman *et al.*, 1985). In addition, many cells in the extrastriate visual areas, where the receptive fields are much larger, are also disparity selective. How to incorporate these experimental findings into the model to account for perceptual phenomena involving long-range interactions such as "depth capture" (Spillman & Werner, 1996) requires further investigation.

REFERENCES

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 22*, 284–299.

Allman, J., Miezin, F. & McGuinnes, E. (1985). Stimulus-specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience, 8*, 407–430.

Arndt, P. A., Mallot, H. A. & Bulthoff, H. H. (1995). Human stereovision without localized image features. *Biological Cybernetics, 72*, 279–293.

Bishop, P. O., Henry, G. H. & Smith, C. J. (1971). Binocular interaction fields of single units in the cat striate cortex. *Journal of Physiology, 216*, 39–68.

Blackmore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology, 211*, 599–622.

Campbell, F. W. & Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5*, 1986–2006.

Das, A. & Gilbert, C. D. (1995). Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. *Nature, 375*, 780–784.

DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature, 352*, 156–159.

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*, 3–31.

Derrington, A. M. & Badcock, D. R. (1985). Separate detectors for simple and complex grating patterns? *Vision Research, 25*, 1869–1878.

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

Dow, B. M., Snyder, A. Z., Vautin, R. G. & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. *Experimental Brain Research, 44*, 213–228.

Fleet, D. J., Jepson, A. D. & Jenkin, M. (1991). Phase-based disparity measurement. *Computers and Visual Graphics Image Proceedings, 53*, 198–210.

Fleet, D. J., Wagner, H. & Heeger, D. J. (1996). Encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research, 36*, 1839–1858.

Freeman, R. D. & Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vision Research, 30*, 1661–1676.

Graham, N. & Nachmias, J. (1971). Detection of gratings patterns containing two spatial frequencies: a comparison of single-channel and multiple channel models. *Vision Research, 11*, 251–259.

Grzywacz, N. M. & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society London A, 239*, 129–161.

Hannah, M. J. (1974). Computer matching of areas in stereo imagery, PhD thesis, Stanford University, Stanford, CA.

Hess, R. F. & Wilcox, L. M. (1994). Linear and non-linear filtering in stereopsis. *Vision Research, 34*, 2431–2438.

Hubel, D. H. & Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Jacobson, L., Gaska, J. P. & Pollen, D. A. (1993). Phase, displacement and hybrid models for disparity coding. *Investigative Opthalmology and Visual Science Suppl. (ARVO), 34*, 908.

Jones, J. P. & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in the cat striate cortex. *Journal of Neurophysiology, 58*, 1187–1211.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: University of Chicago Press.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 2324*, 34–43.

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

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

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America A, 70*, 1297–1300.

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

Maske, R., Yamane, S. & Bishop, P. O. (1984). Binocular simple cells for local stereopsis: comparison of receptive field organizations for the two eyes. *Vision Research, 24*, 1921–1929.

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

Ogle, K. (1952). Disparity limits of stereopsis. *Archives of Opthalmology, 48*, 50–60.

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.

Panton, D. J. (1978). A flexible approach to digital stereo matching. *Photogrammatic Engineering & Remote Sensing, 44*, 1499–1512.

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

Qian, N. (1994b) Stereo model based on phase parameters can explain characteristic disparity. *Society of Neuroscience Abstracts, 20*, 624.

Qian, N. & Andersen, R. A. (1996). A physiological model for motion-stereo integration and a unified explanation of the Pulfrich-like phenomena. *Vision Research, 37*, 1683–1698.

Qian, N., Andersen, R. A. & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals III: Modeling. *Journal of Neuroscience, 14*, 7381–7392.

Qian, N. & Zhu, Y. (1995). Physiological computation of binocular disparity. *Society of Neuroscience Abstracts, 21*, 1507.

Qian, N. & Zhu, Y. (1997). A physiological Explanation of disparity attraction and repulsion. *Investigative Ophthalmology & Visual Science* (ARVO), *38*, 906.

Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973a) Apparent motion with subjective contours. *Vision Research, 13*, 1399–1401.

Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973b) The role of contours in stereopsis. *Nature, 242*, 412–414.

Rohaly, A. M. & Wilson, H. R. (1993). Nature of coarse-to-fine constraints on binocular fusion. *Journal of the Optical Society of America A, 10*, 2433–2441.

Rohaly, A. M. & Wilson, H. R. (1994). Disparity averaging across spatial scales. *Vision Research, 34*, 1315–1325.

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

Sato, T. & Nishida, S. (1993). Second-order depth perception with texture-defined random-check stereograms. *Investigative Opthalmology & Visual Science Suppl. (ARVO), 34*, 1438.

Schiller, P. H., Finlay, B. L. & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex: I. Spatiotemporal organization of receptive fields. *Journal of Neurophysiology, 39*, 1288–1319.

Schor, C. M. & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research, 23*, 1649–1654.

Schumer, R. & Julesz, B. (1984). Binocular disparity modulation sensitivity to disparities offset from the plane of fixation. *Vision Research, 24*, 533–542.

Shapley, R. & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience, 8*, 547–583.

Shimojo, S. & Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Research, 30*, 69–80.

Smallman, H. S. (1995). Fine-to-coarse scale disambiguation in stereopsis. *Vision Research, 35*, 1047–1060.

Smallman, H. S. & MacLeod, D. I. (1994). Size-disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A, 11*, 2169–2183.

Spillman, L. & Werner, J. S. (1996). Long-range interaction in visual perception. *Trends in Neurosciences, 19*, 428–434.

Stork, D. G. & Wilson, H. R. (1990). Do Gabor functions provide appropriate descriptions of visual cortical receptive fields? *Journal of the Optical Society of America A, 7*, 1362–1373.

Ts'o, D. Y., Gilbert, C. D. & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience, 6*, 1160–1170.

Wagner, H. & Frost, B. (1993). Disparity-sensitive cells in the owl have a characteristic disparity. *Nature, 364*, 796–798.

Watson, A. B. & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 2*, 322–342.

Westheimer, G. (1979). Cooperative neural processes involved in stereoscopic acuity. *Experimental Brain Research, 36*, 585–597.

Westheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. *Journal of Physiology, 370*, 619–629.

Wilcox, L. M. & Hess, R. F. (1995). Dmax for stereopsis depends on size, not spatial frequency content. *Vision Research, 35*, 1061–1069.

Wilson, H. R., Blake, R. & Halpern, D. L. (1991). Coarse spatial scales constrain the range of binocular fusion on fine scales. *Journal of the Optical Society of America A, 8*, 229–236.

Wilson, H. R., Ferrera, V. P. & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9*, 79–97.

Yang, Y. & Blake, R. (1991). Spatial frequency tuning of human stereopsis. *Vision Research, 31*, 1177–1189.

Zhu, Y. & Qian, N. (1996). Binocular receptive fields, disparity tuning, and characteristic disparity. *Neural Computation, 8*, 1647–1677.

Acknowledgements—The work is supported in part by NIH grant MH54125 and a research grant from the McDonnell–Pew Program in Cognitive Neuroscience, both to N. Q.

APPENDIX

Derivation of Eq. (8)

In this section, we derive the complex cell response Eq. (8) under the general assumption that the frequency tuning of the receptive field profiles is much sharper than the frequency spectrum of the input stimulus, and that there is a phase difference $\Delta\phi$ between the left and right receptive field profiles. We will also estimate the error term associated with the approximation method used in the derivation.

The derivation method used here is similar to that used by (Qian & Andersen, 1997). We start by calculating simple cell responses defined in Eq. (3). Applying the Fourier power theorem and using tilde to denote the Fourier transform of a function, Eq. (3) can be written as:

$$r_s = \int_{-\infty}^{+\infty} d\omega[\tilde{f}_l(\omega)\tilde{I}_l^*(\omega) + \tilde{f}_r(\omega)\tilde{I}_r^*(\omega)] \tag{A1}$$

Since $f_l(x)$, $f_r(x)$, $I_l(x)$ and $I_r(x)$ are real functions their Fourier transforms all satisfy the relation: $\tilde{g}(-\omega) = \tilde{g}^*(\omega)$. Equation (A1) can thus be written as

$$r_s = 2\int_0^{\infty} d\omega Re[\tilde{f}_l(\omega)\tilde{I}_l^*(\omega) + \tilde{f}_r(\omega)\tilde{I}_r^*(\omega)] \tag{A2}$$

where *Re* represents the real part of a complex quantity.

Freeman and coworkers (DeAngelis *et al.*, 1991, 1995) proposed based on their quantitative physiological studies that the left and right receptive fields of a binocular simple cell have corresponding retinal locations but different phase parameters for the excitatory/inhibitory modulations within the receptive fields, as represented by Eqs (1) and (2). It is easy to show that, in the Fourier domain, Eqs (1) and (2) differ by $e^{i\text{sign}(\omega)\Delta\phi}$ for well-tuned receptive fields, where $\Delta\phi$ is the phase parameter difference defined in Eq. (6), and the sign function is equal to 1 when its argument is positive, and -1 otherwise.* We can therefore assume that in general the Fourier transforms of the left and right receptive fields are related by

$$\tilde{f}_r(\omega) = \tilde{f}_l(\omega)e^{i\text{sign}(\omega)\Delta\phi}. \tag{A3}$$

Note that the sign function also ensures that upon inverse transform $f_r(x)$ is a real function.

The left and right images of a stimulus patch with constant disparity D can be written as:†

$$I_l(x) = I(x), \tag{A4}$$

$$I_r(x) = I(x + D). \tag{A5}$$

Or equivalently, their Fourier transforms are related by:

$$\tilde{I}_r(\omega) = \tilde{I}_l(\omega)e^{i\omega D} \tag{A6}$$

Substituting Eqs (A3) and (A6) into Eq. (A2) we obtain:

$$r_s = 2Re\int_0^{\infty} d\omega\tilde{f}_l(\omega)\tilde{I}_l^*(\omega)[1 + e^{i\omega\Delta\phi - i\omega D}] \tag{A7}$$

We have dropped the sign function because the integration is carried over the positive frequency only. The terms in the integrand are in general complex, and each can be written as an amplitude multiplied by a complex phase term:

$$\tilde{I}^*(\omega) = |\tilde{I}(\omega)|e^{i\theta_I(\omega)}, \tag{A8}$$

$$\tilde{f}_l(\omega) = |\tilde{f}_l(\omega)|e^{i\theta_f(\omega)}, \tag{A9}$$

$$1 + e^{(i\Delta\phi - \omega D)} = 2|\cos(\frac{\Delta\phi}{2} - \frac{\omega D}{2})|e^{i\theta(\omega)}. \tag{A10}$$

Equation (A7) can then be written as:

$$r_s = 4\int_0^{\infty} d\omega|\tilde{I}(\omega)||\tilde{f}_l(\omega)||\cos\left(\frac{\Delta\phi}{2} - \frac{\omega D}{2}\right)|\cos(\theta_I + \theta_f + \theta). \tag{A11}$$

For simplicity of notation, we did not explicitly write out the ω dependence of the θs in the above equation.

Most primary visual cortical cells are well tuned to spatial frequencies. Assume that the cell in Eq. (A11) is tuned to frequency ω_0 and that its tuning is significantly sharper than that of the other terms in the equation, we can then approximate $\tilde{f}_l(\omega)$ by two delta functions, one peaked at ω_0 and the other at $(-\omega_0)$, and simplify Eq. (A11) into:

$$r_s \approx 4|\tilde{I}(\omega_0)||\cos\left(\frac{\Delta\phi}{2} - \frac{\omega_0 D}{2}\right)|\cos(\theta_I + \theta_f + \theta)\int_0^{\infty} d\omega|\tilde{f}_l(\omega)|. \tag{A12}$$

We now compute complex cell responses using the quadrature pair construction. It is easy to show that the response of the simple cell that forms a quadrature pair with the simple cell in Eq. (A12) is given by:

$$r_s' \approx 4|\tilde{I}(\omega_0)||\cos\left(\frac{\Delta\phi}{2} - \frac{\omega_0 D}{2}\right)|\sin(\theta_I + \theta_f + \theta)\int_0^{\infty} d\omega|\tilde{f}_l(\omega)|, \tag{A13}$$

because the θ_fs of the two simple cells differ by $\pi/2$ while all the other parameters are the same (Adelson & Bergen, 1985; Watson & Ahumada, 1985; Ohzawa *et al.*, 1990; Qian, 1994a). The response of a complex cell constructed from this quadrature pair is then given by:

$$r_q = (r_s)^2 + (r_s')^2 \tag{A14}$$

$$\approx c^2|\tilde{I}(\omega_0)|^2\cos^2\left(\frac{\Delta\phi}{2} - \frac{\omega_0 D}{2}\right), \tag{A15}$$

*Note that under the alternative assumption of an overall horizontal positional shift (Δx) between the left and right receptive fields (Zhu & Qian, 1996; Wagner & Frost, 1993; DeAngelis *et al.*, 1995), the two Fourier transforms will differ by $e^{i\omega\Delta x}$, and a similar derivation can be carried through to obtain Eq (29).

†The disparities of real world stimuli are, or course, not constant. However, this is a good approximation within the spatial windows of the primary visual cortical cells.

where constant c is defined as:

$$c \equiv 4 \int_0^\infty d\omega |\tilde{f}_l(\omega)|. \quad \text{(A16)}$$

This completes the derivation of Eq. (8) in the text.

The above general derivation also allows an easy estimation of the error term associated with Eq. (8). The only approximation we used is treating $|\tilde{f}_l(\omega)|$ in the positive frequency domain as a delta function when obtaining Eq. (A12) from Eq. (A11). To simplify the following notation, let us define:

$$f(\omega) \equiv |\tilde{f}_l(\omega)|, \quad \text{(A17)}$$

$$g(\omega) \equiv 4|\tilde{I}(\omega)|\left|\cos\left(\frac{\Delta\phi}{2} - \frac{\omega D}{2}\right)\right|\cos(\theta_I + \theta_f + \theta). \quad \text{(A18)}$$

With these definitions, Eq. (A11) becomes:

$$r_s = \int_0^\infty d\omega f(\omega) g(\omega). \quad \text{(A19)}$$

We assumed in the above derivation that $f(\omega)$ has a sharp peak at ω_0, while $g(\omega)$ is a relatively slow-varying function of ω, such that:

$$r_s \approx g(\omega_0) \int_0^\infty d\omega f(\omega). \quad \text{(A20)}$$

The error of this approximation is therefore:

$$\Delta r_s = \int_0^\infty d\omega f(\omega)[g(\omega) - g(\omega_0)]$$

$$\approx \int_0^\infty d\omega f(\omega)\left[g'(\omega_0)(\omega - \omega_0) + \frac{g''(\omega_0)}{2}(\omega - \omega_0)^2\right]. \quad \text{(A21)}$$

It is reasonable to assume that ω_0 is the center-of-mass location of $f(\omega)$ in the positive frequency domain:

$$\omega_0 = \frac{\int_0^\infty d\omega f(\omega)\omega}{\int_0^\infty d\omega f(\omega)}. \quad \text{(A22)}$$

It is then easy to show that the first term in Eq. (A21) integrates to zero and the error becomes:

$$\Delta r_s \approx \frac{g''(\omega_0)}{2}\overline{(\Delta\omega)^2}\int_0^\infty d\omega f(\omega) \quad \text{(A23)}$$

where

$$\overline{(\Delta\omega)^2} \equiv \frac{\int_0^\infty d\omega f(\omega)(\omega - \omega_0)^2}{\int_0^\infty d\omega f(\omega)} \quad \text{(A24)}$$

is the variance of $f(\omega)$ around ω_0, and is a measure of its width. The relative error is therefore:

$$\frac{\Delta r_s}{r_s} \approx \frac{g''(\omega_0)\overline{(\Delta\omega)^2}}{2g(\omega_0)} \quad \text{(A25)}$$

We conclude that the relative error is proportional to the width of the simple cell frequency tuning curves.

Derivation of Eq. (20)

We now derive Eq. (20). Following the notations and the approximation methods used in the previous section, we can calculate the filtered left and right images (by a given simple cell) as follows:

$$L_1 = \int_{-\infty}^{+\infty} d\omega \tilde{f}_l(\omega)\,\tilde{I}_l^*(\omega)$$

$$= 2\int_0^\infty d\omega\, Re\,[|\tilde{f}_l(\omega)|e^{i\theta_f(\omega)}|\tilde{I}(\omega)|e^{i\theta_I(\omega)}]$$

$$\approx 2|\tilde{I}(\omega_0)|\cos(\theta_f + \theta_I)\int_0^\infty d\omega|\tilde{f}_l(\omega)|, \quad \text{(A26)}$$

$$R_1 = \int_{-\infty}^{+\infty} d\omega \tilde{f}_r(\omega)\tilde{I}_r^*(\omega)$$

$$= 2\int_0^\infty d\omega Re[|\tilde{f}_l(\omega)|e^{i\theta_f(\omega)}e^{i\Delta\phi}|\tilde{I}(\omega)|e^{i\theta_I(\omega)}e^{-i\omega D}]$$

$$\approx 2|\tilde{I}(\omega_0)|\cos(\Delta\phi + \theta_f + \theta_I - \omega_0 D)\int_0^\infty d\omega|\tilde{f}_l(\omega)|. \quad \text{(A27)}$$

The left and right images filtered by the simple cell that forms a quadrature pair with the cell above are then given by:

$$L_2 \approx 2|\tilde{I}(\omega_0)|\sin(\theta_f + \theta_I)\int_0^\infty d\omega|\tilde{f}_l(\omega)|, \quad \text{(A28)}$$

$$R_2 \approx 2|\tilde{I}(\omega_0)|\sin(\Delta\phi + \theta_f + \theta_I - \omega_0 D)\int_0^\infty d\omega|\tilde{f}_l(\omega)| \quad \text{(A29)}$$

because the two cells have their θ_f differ by $\pi/2$ according to the quadrature pair construction method (Adelson & Bergen, 1985; Watson & Ahumada, 1985; Ohzawa *et al.*, 1990; Qian, 1994a).

It is now easy to verify that

$$L_1^2 + L_2^2 + R_1^2 + R_2^2 \approx \frac{c^2}{2}|\tilde{I}(\omega_0)|^2 \quad \text{(A30)}$$

is approximately a constant, where c is defined in Eq. (9). Similarly, it is easy to see that either $L_1 \times R_1$ or $L_2 \times R_2$ has dependence on the Fourier phases (θ_I) of the stimulus and therefore are not adequate for coding disparity, while their sum:

$$L_1 \times R_1 + L_2 \times R_2 \approx \frac{c^2}{2}|\tilde{I}(\omega_0)|^2\cos(\Delta\phi - \omega_0 D) \quad \text{(A31)}$$

is independent of θ_I. Adding Eqs (A30) and (A31) gives us back the complex cell response expression Eq. (5) in the text.

Derivation of Eq. (22)

We derive Eq. (22) for computing disparity with two complex cells in this section. Assume that the two complex cells are constructed from simple cells with their phase parameter differences equal to $\Delta\phi_1$ and $\Delta\phi_2$, respectively. If the responses of these two cells are r_1 and r_2, then according to Eq. (8) we have:

$$r_1 \approx c^2|\tilde{I}(\omega_0)|^2\cos^2\left(\frac{\Delta\phi_1}{2} - \frac{\omega_0 D}{2}\right) \quad \text{(A32)}$$

$$r_2 \approx c^2|\tilde{I}(\omega_0)|^2\cos^2\left(\frac{\Delta\phi_2}{2} - \frac{\omega_0 D}{2}\right) \quad \text{(A33)}$$

Dividing the above two equations and rearranging, we obtain:

$$a\cos\omega_0 D + b\sin\omega_0 D = r_2 - r_1, \quad \text{(A34)}$$

where a and b are defined in Eqs (23) and (24) in the text. If we further define

$$\tan\delta \equiv \frac{a}{b}, \quad \text{(A35)}$$

we then have

$$\sin\delta = \frac{a}{\sqrt{a^2 + b^2}} \quad \text{(A36)}$$

$$\cos\delta = \frac{b}{\sqrt{a^2 + b^2}} \quad \text{(A37)}$$

and Eq. (A34) becomes

$$\sqrt{a^2 + b^2}\sin(\delta + \omega_0 D) \approx r_2 - r_1. \quad \text{(A38)}$$

Solving for D from this expression we obtain Eq. (22) in the text.