Communicated by Ralph Freeman

Computing Stereo Disparity and Motion with Known Binocular Cell Properties

Ning Qian*
Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology, Cambridge, MA 02139 USA

Many models for stereo disparity computation have been proposed, but few can be said to be truly biological. There is also a rich literature devoted to physiological studies of stereopsis. Cells sensitive to binocular disparity have been found in the visual cortex, but it is not clear whether these cells could be used to compute disparity maps from stereograms. Here we propose a model for biological stereo vision based on known receptive field profiles of binocular cells in the visual cortex and provide the first demonstration that these cells could effectively solve random dot stereograms. Our model also allows a natural integration of stereo vision and motion detection. This may help explain the existence of units tuned to both disparity and motion in the visual cortex.

1 Introduction

It is well known that binocular disparity forms the basis of stereoscopic depth perception. There have been many physiological investigations on the mechanisms of stereopsis (see Freeman and Ohzawa 1990 for a recent review). The best known work is perhaps that of Poggio and his co-workers (Poggio and Fischer 1977; Poggio *et al.* 1985), who found that a large proportion of V1 and V2 cells in alert monkeys are disparity sensitive. They classified their cells into several classes based on tuning behavior. For example, "tuned excitatory" cells respond best to disparities near zero, while "near" or "far" cells prefer a range of crossed or uncrossed disparities. Other investigators argued for a continuous distribution of disparity tuning instead of discrete classes (LeVay and Voigt 1988). More recently, Freeman and his collaborators have carried out quantitative analysis of the receptive field structures of binocular cells in cat primary visual cortex (Freeman and Ohzawa 1990; Ozhawa *et al.* 1990). All these studies indicate that most disparity sensitive cells are broadly tuned. Even the most sharply tuned cells have tuning widths of

*Current address: Center for Neurobiology and Behavior, Columbia University, 722 W. 168th St., New York, NY 10032.

about 0.1°–0.2°, comparable to 2–4 pixels in the stereogram in Fig. 3 of this paper, when viewed at a distance of about 40 cm. It has been shown that the broadly tuned disparity sensitive cells such as those found in the brain can explain some psychophysical results of stereo vision (Lehky and Sejnowski 1990). It remains to be demonstrated if these cells can be used to compute disparity maps from stereograms.

Many models for disparity computation have been proposed over the years (Marr and Poggio 1976, 1979; Quam 1984; Prazdny 1985; Pollard *et al.* 1985; Qian and Sejnowski 1988; Sanger 1988; Yeshurun and Schwartz 1989). Unfortunately, most of them are nonbiological, either because they require sharply disparity-tuned units with preferred disparities covering a wide range of values, or because of certain mathematical operations involved that are unlikely to be physiological or at least have not been demonstrated physiologically. Some models are biologically inspired (Marr and Poggio 1979; Sanger 1988), but they are not solely based on the properties of binocular cells in the brain.

Among the existing algorithms, the one that comes closest to physiology is perhaps the model proposed by Sanger (1988), who used Gabor filters for disparity computation. The model uses the fact that displacement of a function generates a proportional phase shift in its Fourier transformation. The binocular disparity at each location is therefore proportional to the difference of the Fourier phases of the corresponding left and the right image patches. A Gabor function is a product of a gaussian envelope with a sinusoid and can be used to perform an approximate localized Fourier transformation. Sanger thus used sine and cosine Gabor filters to estimate the local Fourier phases of left and right images, and the phase difference at each location was used to find disparity.

Although Sanger's model employs Gabor filters which are known to describe simple cell receptive fields well (Marcelja 1980; Daugman 1985; Jones and Palmer 1987), the filters are used only in a monocular fashion. The binocular interaction in the model occurs only at the final step, when the left and the right phases are compared. Before that, the left and the right images are processed separately. No stage in the model uses binocular receptive fields or disparity tuned cells resembling those found in the visual cortex. Also, the explicit representations of the phases of the left and the right images, and of the phase differences in the model are not physiologically plausible. The simple cells in cortex are phase sensitive, but their responses are not monotonic functions of the image phases as implied by the model.

The work described in this paper is mainly based on the physiological and computational studies by Freeman and Ohzawa (1990), and by Ohzawa *et al.* (1990). These investigators measured the receptive field structures of simple and complex binocular cells in the cat primary visual cortex. They found that simple cells do not reliably signal binocular disparity because they are also sensitive to the contrast and the position of the stimulus. They then showed that a subpopulation of complex cells

is well suited as disparity detectors. Finally, they demonstrated, through computer simulations, that left and right receptive fields of a simple cell can be described by two Gabor functions with a certain phase difference between them and that disparity-selective complex cells can be modeled by combining the outputs of a quadrature pair of simple cells, similar to the procedure for motion energy computation (Adelson and Bergen 1985; Watson and Ahumada 1985). While their work represents a major step toward an understanding of biological stereopsis, it is incomplete from the computational point of view. The work is limited to specific simulations of a few simple and complex cells' responses. They did not explore the explicit relationship between the parameters of the cells and their tuning behavior. More importantly, they did not provide a computational theory (Marr 1982) for computing disparity maps from stereograms.

We propose such a theory in this paper by generalizing and formalizing the model of Ohzawa *et al.* (1990). Specifically, we show through mathematical analysis that disparity tuning curves of model simple cells depend strongly on the Fourier phase of the stimulus used to measure the curve. The expression we derived for simple cell response leads naturally to the quadrature pair method of combining simple cell outputs to form model complex cells that have reliable disparity tuning. A family of such complex cells then constitutes a distributed representation of disparity and the actual disparity values of the stimuli can be easily estimated from such a distribution. We demonstrate the effectiveness of our model by applying it to random dot stereograms. We finally show that our formulation of disparity computation can be combined naturally with the energy models of motion detection (Adelson and Bergen 1985; Watson and Ahumada 1985) into a unified framework.

2 Theory and Simulation

Through single unit recording from cat's primary visual cortex, Freeman and his colleagues (Freeman and Ohzawa 1990; Ohzawa *et al.* 1990; DeAngelis *et al.* 1991) found that a typical binocular simple cell can be described by two Gabor functions, one for each of its receptive fields in the left and the right retinas. A mathematical description of such a pair of receptive fields is given by Normura *et al.* (1990):

$$f_l(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)\cos(\omega x + \phi_l) \tag{2.1}$$

$$f_r(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)\cos(\omega x + \phi_r) \tag{2.2}$$

where σ and ω determine the size and the preferred (angular) spatial frequency of the receptive fields. ϕ_l and ϕ_r are the phase parameters for the left and the right receptive fields, respectively. It has been shown

(Nomura *et al.* 1990) that the different disparity tuning types found in the visual cortex, including those described by Poggio *et al.* (1985), can be generated with appropriate combinations of parameters in equations 2.1 and 2.2.

The response of the simple cell to a stimulus is given by

$$r_s = \int_{-\infty}^{\infty} [f_l(x)I_l(x) + f_r(x)I_r(x)]dx \tag{2.3}$$

where $I_l(x)$ and $I_r(x)$ are the left and the right retinal images of the stimulus. That is, the cell sums the contributions from the two receptive fields linearly (Freeman and Ohzawa 1990; Nomura *et al.* 1990; Ozhawa *et al.* 1990). For a stimulus with a binocular disparity D we can write

$$I_l(x) = I(x), \tag{2.4}$$

$$I_r(x) = I(x + D). \tag{2.5}$$

Under the assumption that the receptive field widths are much larger than the image disparity, the gaussian envelope of the receptive fields can be ignored. It can be shown that the response of a cell with a pair of receptive fields given by equations 2.1 and 2.2, is approximately

$$r_s \approx \rho[\cos(\theta + \phi_l) + cos(\theta + \phi_r - \omega D)] \tag{2.6}$$

or equivalently

$$r_s \approx 2\rho \cos\left(\theta + \frac{\phi_l + \phi_r}{2} - \frac{\omega D}{2}\right) \cos\left(\frac{\phi_l - \phi_r}{2} + \frac{\omega D}{2}\right) \tag{2.7}$$

where ρ and θ are the amplitude and phase of the Fourier transformation of the image $I(x)$ at frequency ω, the preferred spatial frequency of the cell.

Equation 2.7 indicates that the disparity tuning of a binocular simple cell is dependent on the Fourier phases θ of the input stimuli. A special case of this phase dependency has been reported by Ohzawa *et al.* (1990) who found that the disparity tuning of simple cells varies with stimulus position and sign of contrast. Our result is more general because Fourier phase can be affected by other variables besides the position and contrast of the stimulus. Any change to a pattern, other than a constant baseline shift or scaling of brightness, alters the Fourier phase, which in turn affects the disparity tuning. For example, independently generated random dot patterns contain different Fourier phases even when they occupy the same retinal position and have the same overall texture appearance.

Equation 2.7 gives an explicit expression of how disparity tuning of a simple cell depends on the Fourier phase of a stimulus and the cell's parameters. To test this equation, we carried out computer simulation using a model cell with $\sigma = 4$ pixels, $\omega/2\pi = 0.125$ cycles/pixel, and $\phi_l = \phi_r = 0$. Based on the modeling work of Nomura *et al.* (1990), the

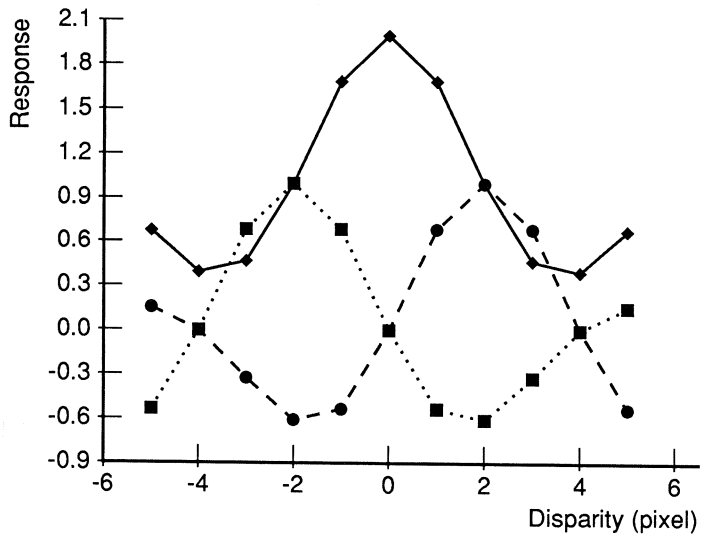

Figure 1: Disparity tuning curves of a simple binocular cell to a vertical line stimulus at three retinal positions. Different position generates different tuning behavior. The solid line is tuned to zero disparity while the dotted and the dashed lines prefer near or far disparities of 2 pixels, respectively. In this simulation, $\sigma = 4$ pixels, $\omega/2\pi = 0.125$ cycles/pixel, and $\phi_l = \phi_r = 0$.

cell should show tuned excitatory behavior. The results of our simulation using a vertical line are shown in Figure 1. When the left image of the line is centered in the left receptive field while the right line position is varied to cover a range of disparity values, the cell is indeed tuned excitatory as shown in Figure 1 (solid line). When the left line position is shifted in one direction or the other and the right line position is varied to cover the same disparity range, however, there is a corresponding shift in the cell's tuning curve and the cell behaves like a near or far cell (Fig. 1, dotted and dashed lines). The amount of shift and the shape of the tuning curves are well predicted by equation 2.7. This is easier to see using the equivalent equation 2.6. The second term in equation 2.6 determines the horizontal shift and the shape of the tuning curves while the first term determines the vertical shift.

For the same model cell we also performed simulations to obtain its disparity tuning curves using independently generated random dot patterns with the same dot density, contrast, and overall position. Each pattern is used to generate a series of random dot stereograms with different disparities for establishing the tuning curve. Three tuning curves

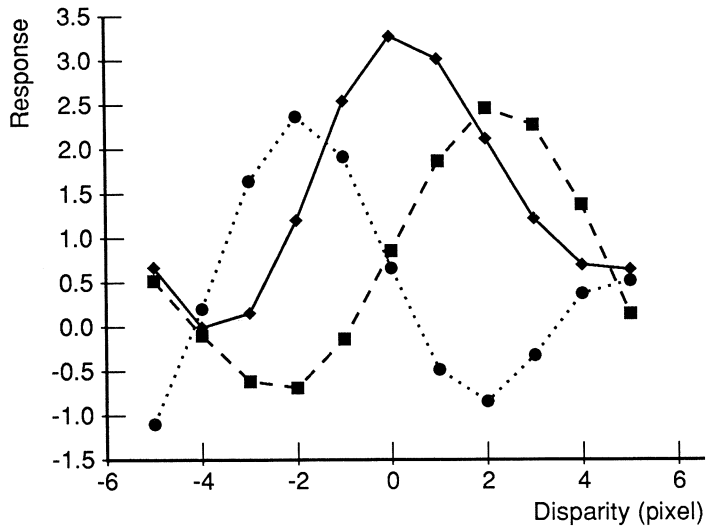

Figure 2: Disparity tuning curves of a simple binocular cell to independently generated random dot patterns. Each pattern is used to generate a series of random dot stereograms for measuring disparity tuning. The model parameters are the same as those in Figure 1.

from the cell are shown in Figure 2. Once again, the cell can be classified as either tuned excitatory, near, or far depending on the patterns used.

The conclusion is that simple cells as described by equations 2.1 and 2.2 do not have reliable disparity tuning and thus cannot be used directly to compute disparity maps from stereograms. We can, however, combine the outputs of the simple cells to form model complex cells that do not depend on the Fourier phase of stimulus. To achieve this, we note that equation 2.7 consists of two cosine terms. The Fourier phase of the stimulus only appears in the first term, which also contains $(\phi_l + \phi_r)/2$. This observation suggests that we can construct a phase-independent complex binocular cell from two simple cells with equal σ, ω, and $(\phi_l - \phi_r)/2$ but with their $(\phi_l + \phi_r)/2$ having a 90° phase difference. If the outputs of these two cells are squared and then summed up, the resulting complex response will be

$$r_c \approx 4\rho^2 \cos^2\left(\frac{\phi_l - \phi_r}{2} + \frac{\omega D}{2}\right) \tag{2.8}$$

which is no longer dependent on the Fourier phases of the input patterns. Since the square of cosine is a periodic function of period π, the range of

$\omega D/2$ in equation 2.8 should be restricted to π, or equivalently, the range of disparity for the cell to detect should be restricted to $2\pi/\omega$, in order to avoid ambiguity. It is easy to show that the preferred disparity of such a complex cell is given by

$$D_{\text{pref}} = \frac{\phi_r - \phi_l}{\omega} \tag{2.9}$$

and its width of tuning (defined at the half peak amplitude) is equal to

$$\Delta D = \frac{\pi}{\omega} \, . \tag{2.10}$$

Note that our construction of the complex cell above is equivalent to using two simple cells with both their ϕ_l's and ϕ_r's in quadrature relationship. We prefer to use the linear combinations $(\phi_l - \phi_r)/2$ and $(\phi_l + \phi_r)/2$ because they are more relevant to disparity computation (see equations 2.8, 2.11, and 2.13).

The method described above is very similar to the quadrature pair method developed for motion energy computation (Adelson and Bergen 1985; Watson and Ahumada 1985). It was first used by Ohzawa *et al.* (1990) to model the disparity selectivity of real complex cells. Their work, however, was limited to computer simulations, and was not based on the theoretical analysis we outlined above. Our equations 2.8, 2.9, and 2.10 are more general and they provide an explicit relationship between the parameters of complex cells and their disparity tuning curves. This relationship forms the basis of a computational theory for solving stereograms to be described next.

Equation 2.8 suggests a simple way of computing stereo disparity using complex binocular cells. If we have a family of complex cells at a spatial location with their $(\phi_l - \phi_r)/2$ covering the range from $-\pi/2$ to $\pi/2$, these cells will then constitute a distributed representation of the stereo disparity present at that location in the input images. Such a representation could be sufficient from a biological point of view. For example, it could be used as input for the control of vergent eye movements. To compute the actual image disparity explicitly, we note that according to equation 2.8 the cell in the family with the strongest response satisfies the condition

$$D = \frac{\phi_r^* - \phi_l^*}{\omega^*} \tag{2.11}$$

where the starred parameters refer to those of the most responsive cell. Thus, by identifying the most responsive cell in the family of complex cells centered at a given location, we can compute the image disparity at that location from the parameters of the cell. Alternatively, we could pick the cell with the highest slope in the distribution instead of the maximum response for better discriminability (Lehky and Sejnowski 1990). In this

case, equation 2.11 should be modified to

$$D = \frac{\phi_r^* - \phi_l^*}{\omega^*} + \frac{\pi}{2\omega^*} \tag{2.12}$$

where the starred parameters are now the parameters of the cell with the highest slope.

It is interesting to note that superficially, equation 2.11 is similar to the one used by Sanger (1988) for disparity computation. The two approaches are actually different. In Sanger's method, ϕ_r^* and ϕ_l^* refer to the Fourier phases of the corresponding left and the right input image patches. The current approach, on the other hand, uses quadrature pairs to deliberately eliminate the image Fourier phase dependence in the simple cell responses. ϕ_r^* and ϕ_l^* in our equation 2.11 are the left and the right phase parameters of the maximally responsive complex cell in the family. They are not the Fourier phases of the input images. Besides, equation 2.11 is not an essential part of our model, and may not correspond to a real step of computation in the brain. It is merely used to demonstrate that an explicit disparity map can be constructed from the distributed representation based on equation 2.8. As we mentioned above, the distributed representation, which is absent in Sanger's model, might be sufficient from the biological point of view.

We have applied our method to random dot stereograms (Julesz 1971). Figure 3 shows an example. The random dot stereogram in Figure 3a has a -2 pixel disparity for the surround, and $+2$ for the center. The dot density is 0.5 and the dot size is 1 pixel. We used 8 complex cells with their $(\phi_l - \phi_r)/2$ distributed evenly in the range $[-\pi/2, \pi/2]$. The σ's and $(\omega/2\pi)$'s are 4 pixels and 0.125 cycle/pixel, respectively, for all cells. The response of the 8 complex cells at each spatial location is computed by first convolving the stereograms with the corresponding 16 simple cells (one quadrature pair of simple cells for each complex cell) and combining the results into complex responses. At each spatial location the highest of the 8 responses is found and equation 2.11 is used to compute the disparity. The disparity map computed this way is then smoothed with a gaussian weighting function with $\sigma = 4$ pixels. The final disparity map is shown in Figure 3b, which agrees well with the correct map. The top and the bottom surfaces have disparities around 2 and -2 pixels, respectively. The error mainly occurs at the transition, which is not as sharp as the perception. To our knowledge, this is the first demonstration that filters with properties similar to real binocular cells in the brain can be used to compute disparity maps from stereograms. Note that the result is obtained with filters of a single spatial scale (i.e., a single set of values for σ and ω). Further improvements could be achieved by combining outputs from several different scales. We could also improve the result by first fitting a smooth curve to the distribution of responses at each location and then identify the maximum.

Our algorithm requires a family of binocular simple cells with various $(\phi_l - \phi_r)/2$. The computation can be made much more efficient if we

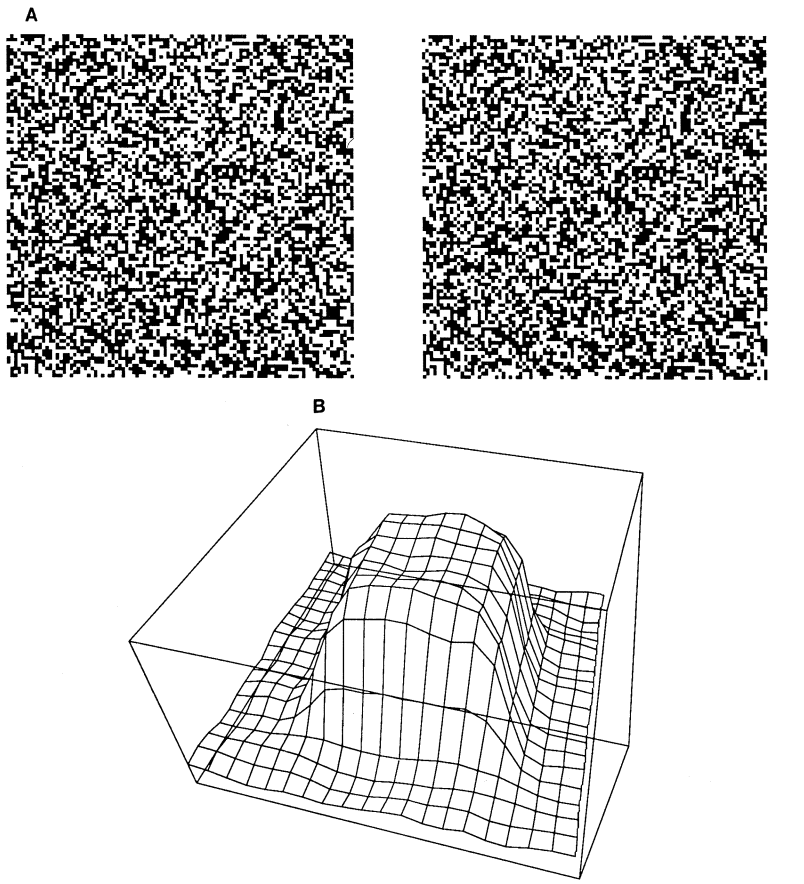

Figure 3: (a) A random dot stereogram with dot density equal to 50% and dot size 1 pixel. The center and the surround have disparities of 2 and −2 pixels respectively. (b) The computed disparity map. See text for the parameters used in the computation.

use the fact that the convolution of a binocular simple cell of arbitrary $(\phi_l - \phi_r)/2$ can be expressed as a linear combination of convolutions at two independent and fixed $(\phi_l - \phi_r)/2$. This is a direct consequence of a trigonometric identity, and is also related to the steerability property discussed by Freeman and Adelson (1991).

It is well known that we can still perceive stereoscopic depth even when the contrast of one image in a stereo pair is different from that of the other (Julesz 1971). The response of our model complex cell under

this condition can be shown to be

$$r_c \approx (1-\gamma)^2\rho^2 + 4\gamma\rho^2\cos^2\left(\frac{\phi_l - \phi_r}{2} + \frac{\omega D}{2}\right) \tag{2.13}$$

where γ is the contrast ratio of the two images in a pair. The expression is different from equation 2.8 by only a dc component and a scale factor, both independent of the disparity D. The actual stimulus disparity can be computed explicitly with the same equation 2.11. We repeated our computer simulation on the stereogram in Figure 3a with the contrast of the right image reduced to half that of the left. The resulting disparity map (not shown) is indistinguishable from Figure 3b.

3 Integrating Motion with Stereo

Our method of combining simple cell outputs for achieving phase independent disparity responses is rather similar to that used in motion energy models (Adelson and Bergen 1985; Watson and Ahumada 1985). This similarity allows a natural integration of motion and stereo vision into a common computational framework. The integration can be achieved by using binocular simple cells with following left and right receptive field structures:

$$f_l(x,y,t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}\right)\cos(\omega_x x + \omega_y y + \omega_t t + \phi_l) \tag{3.1}$$

$$f_r(x,y,t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}\right)\cos(\omega_x x + \omega_y y + \omega_t t + \phi_r) \tag{3.2}$$

where σ's and ω's determine the sizes of the receptive fields and the preferred (angular) frequencies along the spatial and temporal dimensions, and ϕ_l and ϕ_r are again the phase parameters. We assume that these phase parameters are constants independent of x, y, and t. The cell is consequently more sensitive to motion in a constant disparity plane while relatively insensitive to the motion in depth. This is consistent with the physiological finding that few cells in area MT are tuned to motion in depth (Maunsell and Van Essen 1983) and with the psychophysical observation that human subjects are poor at detecting motion in depth based on disparity cues alone (Westheimer 1990).

The filters described by equations 3.1 and 3.2 are nonzero on the negative time axis. They are thus noncausal. This is, however, not a major problem because these filters decay to zero exponentially due to the gaussian envelopes. We could practically make the filters causal by shifting them toward the positive time direction by, for example, $3\sigma_t$. The following results will not be affected by such a shift. A more serious problem with using temporal Gabor filters is that the temporal response of real simple cells is skewed, with its envelope having a longer decay time than

rise time. Also, zero-crossing intervals in the temporal dimension are not equally spaced (DeAngelis *et al.* 1993). We will use temporal Gabor filters in the following analysis for their mathematical simplicity. Similar results can be obtained with more realistic temporal filters and will be presented elsewhere.

Consider a stimulus with a constant disparity D and moving at the speed v_x and v_y along the horizontal and vertical directions, respectively. The left and the right images of such a stimulus is given by

$$I_l(x,y,t) = I(x - v_x t, y - v_y t) \tag{3.3}$$

$$I_r(x,y,t) = I(x - v_x t + D, y - v_y t) \tag{3.4}$$

Let us assume again that the widths of the gaussians in equations 3.1 and 3.2 are large. The response of the simple cell described above to the stimulus is then

$$r_s \approx 2\rho\delta(\omega_t + \omega_x v_x + \omega_y v_y)\cos\left(\theta + \frac{\phi_l + \phi_r}{2} - \frac{\omega_x D}{2}\right) \times \cos\left(\frac{\phi_l - \phi_r}{2} + \frac{\omega_x D}{2}\right) \tag{3.5}$$

where $\delta(\ )$ is the delta function, and ρ and θ are the amplitude and phase of the Fourier transformation of $I(x,y)$. The dependence on the stimulus phase can be removed using the same quadrature pair method discussed before. The resulting complex response is then given by

$$r_c \approx 4\rho^2\delta^2(\omega_t + \omega_x v_x + \omega_y v_y)\cos^2\left(\frac{\phi_l - \phi_r}{2} + \frac{\omega_x D}{2}\right) \tag{3.6}$$

This equation indicates that the cell is indeed sensitive to both motion and stereo disparity. Moreover, the dependencies to motion and stereo are separated into two terms in the product. This is desirable for it allows separate estimation of stereo and motion parameters by using different populations of cells. For disparity computation, we can look at the responses of the family of cells with identical ω_t, ω_x, and ω_y but different $(\phi_l - \phi_r)/2$ as we did in the previous section. Similarly, for velocity field computation we can use a family of cells with constant $(\phi_l - \phi_r)/2$, but different ω_t, ω_x, and ω_y (Heeger 1987; Grzywacz and Yuille 1990). By holding $(\phi_l - \phi_r)/2$ at different values, we could estimate velocity fields at different depth levels.

The unified model presented above could account for several interesting psychophysical observations about the interactions between motion and stereo. For example, it allows representation of more than one velocity vector at the same location if the two vectors have significantly different stereo disparities (Qian *et al.* 1993a). This is because the motion constraint plane determined by the $\delta(\ )$ function in equation 3.6 is weighted not only by the image Fourier power ρ^2 but also by the

disparity-dependent cosine term. When information from a local area is combined to solve the motion aperture problem the cosine term reduces the interference between the motion signals coming from different directions at different disparities. The detailed results of our psychophysical and computational experiments will be presented elsewhere (Qian *et al.* 1993a,b).

4 Discussion

The main purpose of this paper is to show that the binocular receptive field properties of real cortical cells can be used to compute disparity maps from stereograms. As we mentioned in the text, some ingredients of our model have been previously proposed by Ohzawa *et al.* (1990). Our model, however, was *derived* through mathematical analysis of the binocular system, instead of based on examples of computer simulations. This distinction is an important one since our equations, which describe the relationship between the parameters of cells and their disparity tuning, cannot be easily obtained through computer simulations. These equations are the essential part of our model because they constitute a computational theory for solving stereograms.

Our analysis and computer simulations showed that simple cells do not have reliable disparity tuning, because their responses are also dependent on the Fourier phases of the stimuli. A unit that is tuned excitatory under one condition can behave like a near or far cell under other conditions. We suggest that the disparity tuning curves of simple cells are incomplete as they are not uniquely defined for these cells and that the best way to study disparity sensitivity of simple cells is by mapping their binocular receptive field structures (Ohzawa *et al.* 1990).

In order to achieve reliable disparity computation, we combined the outputs of simple cells to eliminate the Fourier phase dependence. This can be realized by squaring and then summing the responses of two simple cells that have the same σ, ω, and $(\phi_l - \phi_r)/2$ but have a 90° phase difference in their $(\phi_l + \phi_r)/2$ (Ohzawa *et al.* 1990; Adelson and Bergen 1985; Watson and Ahumada 1985). An alternative way to eliminate phase dependence, and at the same time preserve disparity tuning, is by averaging over many simple cells with equal σ, ω, and $(\phi_l - \phi_r)/2$ but various $(\phi_l + \phi_r)/2$ (see equation 2.7). This approach is less demanding on the specificity of connections between cells but requires more simple cells for each complex cell. Given the fact that there are large numbers of cells in the visual cortex with a wide range of phase parameters (DeAngelis *et al.* 1993), the second approach may be biologically more plausible. The two methods are equivalent from a computational point of view, however, as they both generate the complex cell response given by equation 2.8. They can thus model real complex cells equally well. We conclude that

the energy method using quadrature pairs is just one way for achieving phase independence and is not an indispensable part of the model.

Our model can explain an observation by Poggio *et al.* (1985), who reported that while both simple and complex cells are disparity tuned to bars, only complex cells show disparity tuning to dynamic random dot stereograms. Each of these stereograms maintains a constant disparity over time but the actual arrangement of dots, and thus the Fourier phase, changes randomly from frame to frame. Since simple cells are sensitive to the Fourier phase as well as to the disparity of the stimulus, they lose their disparity tuning as a result of averaging over the random phases from frame to frame. Complex cells, on the other hand, are sensitive to disparity but are independent of the stimulus Fourier phase. They therefore respond consistently to the fixed disparity in a dynamic random dot stereogram, regardless of the randomly changing Fourier phase.

Due to the similarity of our disparity model and the motion energy models, we are able to combine motion and stereo into a common framework. The resulting model allows a distributed and simultaneous representation of velocity and disparity among a population of cells. This is consistent with the fact that many V1 and MT cells are broadly tuned to direction, speed, and disparity. By looking at different subpopulations of these cells either velocity or disparity information can be extracted. The resulting model is not only physiologically plausible, it also explains several interesting psychophysical observations regarding the interaction between motion and stereo.

Acknowledgments

I would like to thank Professor Richard Andersen for his support and encouragement. I am also grateful to Bard Geesaman, David Bradley, and two anonymous reviewers for their helpful comments. The work was supported by a McDonnell-Pew Postdoctoral Fellowship to the author, and is currently supported by Office of Naval Research Contract N00014-89-J1236 and NIH Grant EY07492, both to Richard Andersen.

References

Adelson, E. H., and Bergen, J. R. 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**(2), 284–299.

Daugman, J. G. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* **2**, 1160–1169.

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

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

Freeman, R. D., and Ohzawa, I. 1990. On the neurophysiological organization of binocular vision. *Vision Res.* **30**, 1661–1676.

Freeman, W. T., and Adelson, E. H. 1991. The design and use of steerable filters. *IEEE Pat. Anal. Mach. Intell.* **13**(9), 891–906.

Grzywacz, N. M., and Yuille, A. L. 1990. A model for the estimate of local image velocity by cells in the visual cortex. *Proc. R. Soc. London A* **239**, 129–161.

Heeger, D. J. 1987. Model for the extraction of image flow. *J. Opt. Soc. Am. A* **4**(8), 1455–1471.

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

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

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

LeVay, S., and Voigt, T. 1988. Ocular dominance and disparity coding in cat visual cortex. *Visual Neurosci.* **1**, 395–414.

Marcelja, S. 1980. Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am. A* **70**, 1297–1300.

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman, San Francisco.

Marr, D., and Poggio, T. 1976. Cooperative computation of stereo disparity. *Science* **194**, 283–287.

Marr, D., and Poggio, T. 1979. A computational theory of human stereo vision. *Proc. R. Soc. London B* **204**, 301–328.

Maunsell, J. H. R., and Van Essen, D. C. 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. ii. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* **49**, 1148–1167.

Nomura, M., Matsumoto, G., and Fujiwara, S. 1990. A binocular model for the simple cell. *Biol. Cybern.* **63**, 237–242.

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

Poggio, G. F., and Fischer, B. 1977. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.* **40**, 1392–1405.

Poggio, G. F., Motter, B. C., Squatrito, S., and Trotter, Y. 1985. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res.* **25**, 397–406.

Pollard, S. B., Mayhew, J. E., and Frisby, J. P. 1985. *Perception* **14**, 449–470.

Prazdny, K. 1985. Detection of binocular disparities. *Biol. Cybern.* **52**, 93–99.

Qian, N., and Sejnowski, T. J. 1988. Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation. *Proceedings of the 1988 Connectionist Models Summer School*, 435–443.

Qian, N., Andersen, R. A., and Adelson, E. H. 1994a. Transparent motion perception as detection of unbalanced motion signals: Psychophysics. Submitted.

Qian, N., Andersen, R. A., and Adelson, E. H. 1994b. Transparent motion perception as detection of unbalanced motion signals: Modeling. Submitted.

Quam, L. H. 1984. Hierarchical warp stereo. *Proceedings of the DARPA Image Understanding Workshop*, 149–155.

Sanger, T. D. 1988. Stereo disparity computation using Gabor filters. *Biol. Cybern.* **59**, 405–418.

Watson, A. B., and Ahumada, A. J. 1985. Model of human visual-motion sensing. *J. Opt. Soc. Am. A* **2**, 322–342.

Westheimer, G. 1990. Detection of disparity motion by the human observer. *Optom. Vision Sci.* **67**, 627–630.

Yeshurun, Y., and Schwartz, E. L. 1989. Cepstral filtering on a columnar image architecture—A fast algorithm for binocular stereo segmentation. *IEEE Pat. Anal. Mach. Intell.* **11**, 759–767.

Received April 29, 1993; accepted August 16, 1993.