NOTE **Communicated by Terence Sanger**

Relationship Between Phase and Energy Methods for Disparity Computation

Ning Qian
Center for Neurobiology and Behavior, Columbia University, New York, NY 10032, U.S.A.

Samuel Mikaelian
Center for Neural Science, New York University, New York, NY 10003, U.S.A.

The phase and energy methods for computing binocular disparity maps from stereograms are motivated differently, have different physiological relevances, and involve different computational steps. Nevertheless, we demonstrate that at the final stages where disparity values are made explicit, the simplest versions of the two methods are exactly equivalent. The equivalence also holds when the quadrature-pair construction in the energy method is replaced with a more physiologically plausible phase-averaging step. The equivalence fails, however, when the phase-difference receptive field model is replaced by the position-shift model. Additionally, intermediate results from the two methods are always quite distinct. In particular, the energy method generates a distributed disparity representation similar to that found in the visual cortex, while the phase method does not. Finally, more elaborate versions of the two methods are in general not equivalent. We also briefly compare these two methods with some other stereo models in the literature.

1 Introduction

Binocular disparity is defined as the positional difference between the two retinal projections of a given point in space. It is well known that the horizontal component of disparity provides the sensory cue for stereoscopic depth perception. Many computational models for disparity estimation have been proposed. Here we compare two computational models that appear to have the strongest biological relevance: the phase method and the energy method.

The phase method (Sanger, 1988; Fleet, Jepson and Jenkin, 1991) for disparity computation is based on the mathematical result that the displacement of a function generates a proportional phase shift in its complex Fourier transform. The binocular disparity at each location is therefore proportional to the difference of the Fourier phases of the corresponding left and right

Neural Computation **12, 279–292** (2000)

image patches. Sanger (1988) used sine and cosine Gabor filters to estimate the local Fourier phases of both left and right images, and then calculated the Fourier phase difference at each location to find disparity. The algorithm has been tested on both random dot stereograms and natural images (Sanger, 1988).

The energy method for disparity computation (Qian, 1994, 1997; Zhu & Qian, 1996; Qian & Zhu, 1997; Qian & Andersen, 1997; Fleet, Wagner, & Heeger, 1996) was derived from the energy model of binocular cell responses in the cat's primary visual cortex (Ohzawa, DeAngelis, & Freeman, 1990, 1996, 1997; Freeman & Ohzawa, 1990; DeAngelis, Ohzawa, & Freeman, 1991). Through quantitative physiological experiments, Freeman and coworkers found that a typical binocular simple cell can be described by two Gabor functions—one for its left and the other for its right receptive fields. There is a relative phase difference (or positional shift, or both) between the two receptive fields that generates disparity sensitivity. The activities of such simple cells were determined by first convolving the left and right images with the left and right receptive fields, respectively, and then summing the two contributions. Freeman et al. further found that responses of binocular complex cells in cat's primary visual cortex can be well modeled by summing the squared outputs of a quadrature pair of simple cells. We analyzed this so-called energy model for binocular complex cells and found that a population of such complex cells can be used to extract stimulus disparity (see Qian, 1997, for a review). The resulting energy method for disparity computation has been tested on random dot stereograms (Qian, 1994; Qian & Zhu, 1997).

These two methods are different in several ways. First, although both methods use Gabor filters at the front end, the phase method applies two *monocular* Gabor filters separately and delays binocular comparison until the last step, when the Fourier phases of the left and right image patches are subtracted, while the energy method uses a set of *binocular* filters that combines contributions from the two eyes at the first step. In addition, the phase method explicitly computes and represents the complex Fourier phases of the left and right image patches, while the energy method uses the quadrature-pair construction (or the equivalent phase averaging; see section 2) at the complex cell stage to remove the simple cells' dependence on image Fourier phases. Furthermore, complex cell responses in the energy method form a distributed representation of disparity; such a representation is absent in the phase method. Despite these differences, we show that the simplest versions of the two methods are exactly equivalent at the final steps where disparity values are made explicit. The intermediate steps, however, differ in the two methods and have different physiological implications. In addition, the two methods can be elaborated in different ways, resulting in nonequivalent algorithms.

2 Results

Before we start, a few potential confusions in terminology should be clarified. First, the phase method should not be confused with the phase-difference (also called phase-parameter or phase-shift) receptive field model. The former specifies a method for disparity estimation from stereograms, while the latter is a model for describing receptive field profiles of binocular simple cells often used in the energy method. Second, the phase parameters in the phase-difference receptive field model (ϕ_l and ϕ_r in equations 2.1 and 2.2) should not be confused with the image Fourier phases. The former determine the positions of the excitatory-inhibitory bands relative to the receptive field center, while the latter are the phases of the complex Fourier transforms of retinal images. Finally, mathematically complex quantities, with real and imaginary components, should not be confused with complex cell properties.

We now demonstrate the exact equivalence of the simplest versions of the phase and the energy methods. Elaborations of the methods such as frequency pooling (for both methods), confidence measure (for the phase method), and spatial pooling (for the energy method) are ignored in this section (but see section 3). We start by reformulating the energy method (Qian, 1994) and convert it into a form identical to that for the phase method (Sanger, 1988).

Consider a binocular simple cell centered at $x = 0$, with the left and right receptive fields given by the following Gabor functions (Marcelja, 1980; Daugman, 1985; McLean & Palmer, 1989; Ohzawa et al., 1990):

$$f_l(x) = g(x)\cos(\omega x + \phi_l) \tag{2.1}$$

$$f_r(x) = g(x)\cos(\omega x + \phi_r) \tag{2.2}$$

with

$$g(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{2.3}$$

where ω is the preferred (angular) horizontal spatial frequency of the cell, σ is the gaussian width determining the receptive field size, and ϕ_l and ϕ_r are the left and right phase parameters. Its response to a stereo image pair $I_l(x)$ and $I_r(x)$ is given by:

$$r_s = \int_{-\infty}^{\infty} \left[f_l(x)I_l(x) + f_r(x)I_r(x)\right] dx. \tag{2.4}$$

(For a layer of topographically arranged cells of the same type, the convolution operation between images and receptive fields should be used instead.) The simple cell with a quadrature phase relationship to the above cell has receptive fields of the same form but with the cosine functions replaced by sines (Pollen, 1981; Adelson & Bergen, 1985; Watson & Ahumada, 1985; Ohzawa et al., 1990; Qian, 1994). It is easy to see that the responses of these two simple cells are the real and imaginary parts of

$$R_+ \equiv \int_{-\infty}^{\infty} g(x) e^{i\omega x} \left[e^{i\phi_l} I_l(x) + e^{i\phi_r} I_r(x) \right] dx. \tag{2.5}$$

According to the quadrature-pair construction, the complex cell response is given by the sum of the squared responses of these two simple cells, (or equivalently, four half-wave rectified simple cells (Ohzawa et al., 1990)), and can thus be written as

$$r_q = |R_+|^2 = \left| \int_{-\infty}^{\infty} g(x) e^{i\omega x} \left[I_l(x) + e^{i\phi_-} I_r(x) \right] dx \right|^2, \tag{2.6}$$

where

$$\phi_- \equiv \phi_r - \phi_l \tag{2.7}$$

is the phase parameter difference. Therefore, complex cell responses depend on only ϕ_- instead of on ϕ_r and ϕ_l individually. For a given binocular stimulus, complex cells with different ϕ_- will give different responses. These cells at a given spatial location form a distributed representation of the stimulus disparity at that location. According to the energy method (Qian, 1994), the stimulus disparity can be explicitly estimated from the distribution as:

$$D = \frac{\hat{\phi}_-}{\omega}, \tag{2.8}$$

where $\hat{\phi}_-$ is the phase difference ϕ_- that maximizes the complex cell response r_q. Let

$$\int_{-\infty}^{\infty} g(x) I_l(x) e^{i\omega x} dx \equiv \chi_l(\omega), \tag{2.9}$$

$$\int_{-\infty}^{\infty} g(x) I_r(x) e^{i\omega x} dx \equiv \chi_r(\omega), \tag{2.10}$$

so that $\chi_l(\omega)$ and $\chi_r(\omega)$ are the Fourier transforms of the left and right image patches under the gaussian envelope $g(x)$, respectively. Equation 2.6 then becomes:

$$r_q = \left| \chi_l(\omega) + e^{i\phi_-} \chi_r(\omega) \right|^2. \tag{2.11}$$

To find $\hat{\phi}_-$, note that the maximum of r_q is attained when the two terms inside the norm of the above expression are along the same direction in the complex plane, that is,

$$\hat{\phi}_- = \arg(\chi_l) - \arg(\chi_r) \tag{2.12}$$

$$= \arctan\left[\frac{\int_{-\infty}^{\infty} dx\, I_l(x)g(x)\sin\omega x}{\int_{-\infty}^{\infty} dx\, I_l(x)g(x)\cos\omega x}\right] - \arctan\left[\frac{\int_{-\infty}^{\infty} dx\, I_r(x)g(x)\sin\omega x}{\int_{-\infty}^{\infty} dx\, I_r(x)g(x)\cos\omega x}\right]. \tag{2.13}$$

Equation 2.13 combined with equation 2.8 is precisely the expression for the phase method proposed by Sanger (1988). The two terms in equation 2.13 are the local complex Fourier phases of the left and right image patches, respectively, estimated through the sine and cosine Gabor filters. This establishes the equivalence between the two methods.

It is interesting to note that ϕ_- simply compensates for the different Fourier phases of the right and left image patches in equation 2.11. Scaling the relative contrast of the image pair while preserving the contrast polarity will not affect the disparity estimates because only the relative amplitudes (but not phases) of χ_l and χ_r will be changed by this manipulation (Qian, 1994). When the contrast polarity of one image in the pair is reversed, however, ϕ_- has to be shifted by π to compensate, resulting in an inverted complex cell turning curve. This prediction is also contained in equation 2.13 of Qian (1994) and has been verified experimentally (Ohzawa et al., 1990; Cumming & Parker, 1997; Masson, Busettini, & Miles, 1997).

We have previously discussed the similarities and differences between the energy method and the cross-correlator (Qian & Zhu, 1997). With the current formulation, it is easy to see that finding $\hat{\phi}_-$, which maximizes the quadrature-pair response, is equivalent to determining the Fourier phase of the cross-correlation between the left and right image patches under the gaussian envelope $g(x)$. This is because the Fourier transform of the cross-correlation,

$$C_{lr}(x) = \int_{-\infty}^{\infty} g(x')I_l(x')\, g(x+x')I_r(x+x')\, dx', \tag{2.14}$$

is simply $\chi_l(\omega)\chi_r^*(\omega)$ where $*$ denotes complex conjugation. Therefore, the phase of this cross-correlator, or its normalized version, is just the expression for $\hat{\phi}_-$ in equation 2.12. Correlation methods usually use the peak location of $C_{lr}(x)$ to estimate disparity. If $C_{lr}(x)$ is sharply peaked at x_0, its Fourier phase is approximately ωx_0. It is in deviations from this approximation that the cross-correlator and the energy method can yield different estimates.

We used the quadrature-pair construction for obtaining complex cell responses above. As we pointed out previously (Qian & Zhu, 1997), a less demanding and physiologically more plausible alternative is to integrate squared responses of many simple cells all with the same ϕ_- but with

$$\phi_+ \equiv \phi_r + \phi_l \tag{2.15}$$

uniformly distributed in the entire 2π range (see equation 2.19). The energy method with this uniform phase averaging approach is still equivalent to the phase method because the phase averaging gives exactly the same complex cell expression as the quadrature-pair construction. To demonstrate, define:

$$R_- \equiv R_+^* = \int_{-\infty}^{\infty} g(x) e^{-i\omega x} [e^{-i\phi_l} I_l(x) + e^{-i\phi_r} I_r(x)] dx, \tag{2.16}$$

and rewrite the simple cell response as:

$$r_s = \frac{1}{2}(R_+ + R_-). \tag{2.17}$$

Since r_s is real, we have:

$$r_s^2 = |r_s|^2 = \frac{1}{4}\left(|R_+|^2 + |R_-|^2 + 2\mathrm{Re}R_+ R_-^*\right) = \frac{1}{2}\left(|R_+|^2 + \mathrm{Re}R_+^2\right). \tag{2.18}$$

According to the phase averaging approach, the complex cell response is given by the integration:

$$r_a = \frac{1}{2\pi}\int_0^{2\pi} r_s^2 d\phi_+ = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{2}\left(|R_+|^2 + \mathrm{Re}R_+^2\right) d\phi_+, \tag{2.19}$$

while ϕ_- is kept constant. Since $|R_+|^2$ is not a function of ϕ_+ (see equation 2.6), the averaging leaves the first term unchanged. The second term integrates to zero because:

$$\mathrm{Re}R_+^2 = \mathrm{Re}\left[\left(\int_{-\infty}^{\infty} dx g(x) e^{i\omega x}\left[e^{-i\phi_-/2} I_l(x) + e^{i\phi_-/2} I_r(x)\right]\right)^2 e^{i\phi_+}\right] \tag{2.20}$$

and

$$\int_0^{2\pi} e^{i\phi_+} d\phi_+ = 0. \tag{2.21}$$

Therefore, the phase averaging approach generates a complex cell response proportional to that of the quadrature pair in equation 2.6,

$$r_a = \frac{1}{2}|R_+|^2 = \frac{1}{2} r_q, \tag{2.22}$$

where the proportionality constant is immaterial for disparity computation. Intuitively, one can imagine dividing the 2π range for ϕ_+ into many small intervals in equation 2.19. Then each pair of intervals differing by $\pi/2$ can be considered a quadrature pair. Therefore, the above phase averaging can be viewed as averaging together a continuum of quadrature-pair responses, all with the same ϕ_-.

Two types of receptive field models for binocular simple cells have been proposed: the position-shift model (Bishop, Henry, & Smith, 1971) and the phase-differences model (Ohzawa et al., 1990; DeAngelis et al., 1991). The latter was used in the above derivations. Real cortical cells may use a combination of both mechanisms for coding disparity (Jacobson, Gaska, & Pollen, 1993; Zhu & Qian, 1996; Fleet et al., 1996; Anzai, Ohzawa, & Freeman, 1997). We showed previously that although there are important differences between them, under reasonable assumptions, both models (or their hybrid) can be used with the energy method for computing disparity maps from stereograms (Zhu & Qian, 1996; Qian & Zhu, 1997). However, the exact equivalence between the phase and the energy methods does not hold when the position-shift receptive field model is used in the energy method. To see this, note that according to the position-shift receptive field model, equation 2.2 should be replaced by

$$f_r(x) = f_l(x+d) = g(x+d)\cos[\omega(x+d)+\phi_l], \tag{2.23}$$

where d represents the amount of positional shift between the left and right receptive field centers. The complex cell responses in equation 2.6 should then be written as:

$$r_q = \left| \int_{-\infty}^{\infty} e^{i\omega x} \left[g(x)I_l(x) + e^{i\omega d} g(x+d)I_r(x) \right] dx \right|^2, \tag{2.24}$$

and the stimulus disparity is given by (Zhu & Qian, 1996):

$$D = \hat{d}, \tag{2.25}$$

where $\hat{d}$ is the positional shift that maximizes complex cell response. Equation 2.24 can be rewritten as

$$r_q = \left| \chi_l + e^{i\omega d} \chi_r(d) \right|^2, \tag{2.26}$$

with the definitions:

$$\int_{-\infty}^{\infty} g(x)I_l(x)e^{i\omega x}dx \equiv \chi_l\,, \tag{2.27}$$

$$\int_{-\infty}^{\infty} g(x+d)I_r(x)e^{i\omega x}dx \equiv \chi_r(d)\,. \tag{2.28}$$

It is obvious that a closed-form expression similar to equation 2.13 cannot be obtained for $\hat{d}$ because χ_r is also a function of d, and the maximum of

r_q as a function of d may not occur when the two terms inside the norm of equation 2.26 have the same direction on the complex plane. Therefore, the energy method with the position-shift receptive field model is not equivalent to the phase method. However, under the oft-invoked assumption that the disparity is small compared to the receptive field size, the d-dependence of χ_r in equation 2.28 is negligible and ωd in equation 2.26 plays the same role as ϕ_- in equation 2.11. Clearly, in this approximation the energy method with the positional-shift model corresponds to the phase method in the same way as the energy method with the phase-difference model does.

3 Discussion

We have demonstrated that the simplest versions of the energy and the phase methods for disparity computation are exactly equivalent at the final stages where the disparity values are made explicit. The equivalence still holds when the quadrature-pair construction in the energy method is replaced by a less demanding phase-averaging procedure. However, when the position-shift type of simple cell receptive field model is used to replace the phase-difference type in the energy method, the two methods are no longer exactly equivalent. This result is consistent with the fact that both the phase method and the energy method combined with the phase-difference receptive field model predict a relationship between the spatial scale ω and the computed disparity range (the so-called size-disparity correlation), while the methods with the position-shift receptive field model do not make such a prediction (Sanger, 1988; Smallman & MacLeod, 1994; Qian, 1994; Zhu & Qian, 1996; Fleet et al., 1996).

Regardless of the receptive field models used, the intermediate results in the two methods are always very different. Indeed, the energy method contains simple and complex cell stages that simulate binocular interactions observed in simple and complex cortical cells, while none of the steps in the phase method resembles actual binocular cells. More importantly, the complex cells in the energy method form a distributed representation of binocular disparity at each location; such a representation is absent in the phase method. The distributed representation is useful because it could directly guide motor behavior such as vergence eye movement (Masson et al., 1997) without first generating an explicit disparity map. In fact, the final explicit disparity extraction in both methods does not seem to happen in the brain since "grandmother cells" for disparity coding have never been recorded from the visual cortex. While an explicit, grandmother-cell type of representation is more convenient for us to comprehend, the brain appears to rely more on implicit, distributed representations for perception and control.

The simplest versions of the phase and the energy methods can also be elaborated in many ways. For example, both methods can be extended to combine results from different spatial scales (i.e., filters with different ωs).

Obviously, if the same extension is made after the final stages in both methods, they remain equivalent. Any elaborations at intermediate steps or those that can only be applied to one method, however, will in general render the two methods nonequivalent. One example is the spatial pooling procedure in the energy method, which greatly improves the algorithm by making complex cell responses more independent of image Fourier phases (Zhu & Qian, 1996; Qian & Zhu, 1997). This procedure cannot be easily incorporated into the phase method because image Fourier phases are precisely the useful information in the phase method. Similarly, the confidence measure in the phase method for appropriately weighting different spatial scales (Sanger, 1988) cannot be readily extended to the energy method since the measure is derived from the Fourier amplitudes of the left and right image patches; these monocular amplitudes are not computed in the energy method that starts with binocular filters. Of course, a different spatial pooling step could be incorporated into the phase method. For example, one could average the Fourier phases estimated from several nearby spatial locations. Similarly, a different confidence measure could be added to the energy method. One possibility would be using the normalized range of the complex cell responses,

$$\frac{r_q(\hat{\phi}_-) - r_q(\hat{\phi}_- + \pi)}{r_q(\hat{\phi}_-) + r_q(\hat{\phi}_- + \pi)} = \frac{2\,\mathrm{Re}(e^{-i\hat{\phi}_-}\chi_l\chi_r^*)}{|\chi_l|^2 + |\chi_r|^2} = \frac{2|\chi_i||\chi_{\hat{r}}|}{|\chi_l|^2 + |\chi_r|^2}, \tag{3.1}$$

as the confidence measure because a larger modulation would allow a more reliable estimation of $\hat{\phi}_-$.

In conclusion, the equivalence between the simplest versions of the phase and energy methods demonstrated in this article indicates that the two methods can be viewed as different implementations of the same underlying mathematical principle. The differences in implementation, however, allow the methods to be elaborated, interpreted, and used in different ways by subsequent processes. Our results are reminiscent of the relationship between various motion models that have been documented in the literature (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Borst & Egelhaaf, 1989; Simoncelli & Adelson, 1991; Heeger & Simoncelli, 1992).

3.1 Other Methods. One of the reviewers pointed out that disparity can also be computed by calculating the magnitude of the bandpass-filtered sum of the left and right image patches. This magnitude is equivalent to setting ϕ_- to zero in equation 2.6 (or setting d to zero in equation 2.24) and dropping the immaterial squaring outside the norm. For a filter well-tuned to horizontal frequency ω and for left and right image patches given by $I_l = I(x)$ and $I_r = I(x + D)$, this magnitude is approximately proportional to $|\cos(\omega D/2)||\tilde{I}(\omega)|$. If the image spectrum $|\tilde{I}(\omega)|$ does not vary much with ω, the disparity D can be estimated by using two or more filters with different

ω. We will refer to this algorithm as the frequency method because it applies identical filters to the left and right images and consequently has to rely on more than one frequency band for disparity computation. In essence, the frequency method fixes ϕ_- (or d) to zero and varies ω, while the energy method fixes ω and varies ϕ_- (or d). Obviously the frequency method is not equivalent to the energy (or the phase) method because the former cannot compute disparity within a single frequency band, while the latter can. Existing psychophysical evidence is against the frequency method in its pure form: human subjects do perceive stereoscopic depth in band-limited stereograms (Julesz, 1971). Psychophysical observations also indicate that different frequency channels interact with each other in human stereo vision (Wilson, Blake, & Halpern, 1991; Rohaly & Wilson, 1993, 1994; Smallman, 1995; Mallot, Gillner, & Arndt, 1996). It is thus possible that the brain may combine the frequency and the energy methods by using cells with different ω *and* different ϕ_- (or d) simultanously in disparity computation. Alternatively, the brain could first apply the energy method to obtain one disparity estimate from each frequency channel and then pool the estimates across channels at a later stage (Sanger, 1988; Qian & Zhu, 1997). Further studies are needed for differentiating these two possibilities.

In addition to the phase and the energy methods, many other stereo models have been proposed over the years. Similar to the phase method that first estimates Fourier phases of the left and right image patches separately, most of the other models also start with a monocular detection of some matching primitives. For example, several studies (Marr & Poggio, 1976; Pollard, Mayhew, & Frisby, 1985; Prazdny, 1985; Qian & Sejnowski, 1989; Marshall, Kalarickal, & Graves, 1996) let individual dots in random-dot stereograms be the matching primitives. Marr and Poggio (1979) later used zero crossings at different spatial scales as the primitives. For convenience, we classify these methods as *initially monocular*. In contrast, the energy method is *initially binocular* because it starts with binocular filtering without performing any monocular feature detection. In this aspect it is similar to the cepstral method (Yeshurun & Schwartz, 1989) and the frequency method discussed above. The cepstral and the energy methods are very different in other aspects, however. For instance, the former contains a complex logrithmic operation and complex Fourier transforms, while the latter does not. (Note that the Fourier transform used in this article and elsewhere in connection with the energy method is only for mathematically analyzing the method; it is not a step in the method.) In addition, the cepstral method does not use binocular receptive fields from the physiological experiments.

The initially monocular algorithms generally contain a subsequent step that matches the monocular primitives to solve the correspondence problem. Most methods (Marr & Poggio, 1976, 1979; Pollard et al., 1985; Prazdny, 1985; Qian & Sejnowski, 1989; Marshall et al., 1996) do so by introducing explicit constraints or rules that determine correct matches between the left and right primitives among all possible matches within a certain disparity range. These methods will be referred to as the *explicit methods*. In contrast, the phase method is *implicit* (Sanger, 1988) because it finds disparity by simply subtracting the two monocular phases without performing any explicit matching. (If the monocular Fourier amplitudes are used to calculate a confidence weighting factor [Sanger, 1988], then the matching is slightly more explicit). The initially binocular algorithms defined above are necessarily implicit since there are no monocular primitives for explicit matching in the first place. An advantage of the implicit methods is that they can have subpixel disparity resolution without the burden of sorting through a huge number of potential matches (Sanger, 1988; Qian & Zhu, 1997). As we discussed elsewhere (Qian, 1997), the implicit methods are also more physiologically plausible.

To summarize, most stereo models are initially monocular for feature extraction and are then explicit in binocular matching. The cepstral, the energy, and the frequency methods, on the other hand, are initially binocular and implicit. The phase method is somewhere in between: it is initially monocular and is implicit. Obviously, there are many other ways of classifying stereo models. For example, some models measure and rely on disparity gradients, while others do not; some implement the uniqueness constraint through inhibitory interactions, while others emphasize facilitation and allow multiple matches of monocular primitives.

The current formulation of the energy method assumes that the disparity is constant over the small image patches covered by the receptive fields (Qian, 1994; Qian & Zhu, 1997). Because of this assumption, the performance tends to degrade when there are large disparity gradients in the images such as across disparity boundaries. However, the energy method can be extended to encompass disparity gradients. Specifically, it can be shown that a disparity gradient of $D'(= dD(x)/dx)$ can be best measured by binocular cells whose left and right preferred spatial frequencies (ω_l and ω_r) are related by $\omega_r = (1 + D')\omega_l$. Therefore, the maximum measurable disparity gradient depends on the differences between ω_l and ω_r among real binocular cells. If ω_l and ω_r are equal, then the cells' frequency tuning widths have to be at least $(1 + D')\omega_l$ in order to detect D' (Sanger, 1988). This extension of the energy method can be viewed as an implicit counterpart of the gradient limit rule employed explicitly by Pollard et al. (1985). It is also worth pointing out that the energy method in its current form (Qian, 1994; Qian & Zhu, 1997) already contains an implicit implementation of the uniqueness constraint (used by many explicit models in various forms). Because of the approximately cosine tuning behavior, there is only one maximum of the complex cell response as a function of ϕ_- within the $[-\pi, \pi)$ range even when there are two overlapping stimulus disparities. The stereo transparency problem could still be solved by an interdigitating representation of the two disparities by cells tuned to different spatial locations, similar to a previous motion-transparency model (Qian, Andersen, & Adelson, 1994). This approach is also related to those explicit methods that use the uniqueness constraint to model stereo transparency (Qian & Sejnowski, 1989; Pollard & Frisby, 1990). It thus appears that despite the major differences between the explicit and implicit methods, they could share certain conceptual similarities.

Acknowledgments

This work was supported by a Sloan Research Fellowship and NIH grant MH54125, both to N. Q. We are grateful to the anonymous reviewers for their helpful comments.

References

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

Anzai, A., Ohzawa, I., & Freeman, R. D. (1997). Neural mechanisms underlying binocular fusion and stereopsis: Position vs. phase. *Proc. Nat. Acad. Sci. USA, 94*, 5438–5443.

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

Borst, A., & Egelhaaf, M. (1989). Principles of visual motion detection. *Trends Neurosci., 12*, 297–306.

Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature, 389*, 280–283.

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

Fleet, D. J., Jepson, A. D., & Jenkin, M. (1991). Phase-based disparity measurement. *Comp. Vis. Graphics Image Proc., 53*, 198–210.

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

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

Heeger, D. J., & Simoncelli, E. P. (1992). Model of visual motion sensing. In L. Harris & M. Jenkins (Eds.), *Spatial vision in humans and robots*. Cambridge: Cambridge University Press.

Jacobson, L., Gaska, J. P., & Pollen, D. A. (1993). Phase, displacement and hybrid models for disparity coding. *Invest. Ophthalmol. and Vis. Sci. Suppl. (ARVO), 34*, 908.

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

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

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

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

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proc. R. Soc. Lond. B, 204*, 301–328.

Marshall, J. A., Kalarickal, G. J., & Graves, E. B. (1996). Neural model of visual stereomatching: Slant, transparency, and clouds. *Network: Comput. Neural Sys., 7*, 635–639.

Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature, 389*, 283–286.

McLean, J., & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Res., 29*, 675–679.

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.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1996). Encoding of binocular disparity by simple cells in the cat's visual cortex. *J. Neurophysiol., 75*, 1779–1805.

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

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

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

Pollen, D. A. (1981). Phase relationship between adjacent simple cells in the visual cortex. *Nature, 212*, 1409–1411.

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

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

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

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

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

Qian, N., & Sejnowski, T. J. (1989). Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation. In D. S. Touretzky, G. E. Hinton, & T. J. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 435–443). San Mateo, CA: Morgan Kaufmann.

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

Rohaly, A. M., & Wilson, H. R. (1993). Nature of coarse-to-fine constraints on binocular fusion. *J. Opt. Soc. Am. A*, *10*, 2433–2441.

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

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

Simoncelli, E. P., & Adelson, E. H. (1991). Relationship between gradient, spatio-temporal energy, and regression models for motion perception. *Invest. Ophthalmol. and Vis. Sci. Suppl. (ARVO)*, *32*, 893.

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

Smallman, H. S., & MacLeod, D. I. (1994). Size-disparity correlation in stereopsis at contrast threshold. *J. Opt. Soc. Am. A*, *11*, 2169–2183.

van Santen, J. P. H., & Sperling, G. (1985). Elaborated Reichardt detectors. *J. Opt. Soc. Am. A*, *2*, 300–321.

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

Wilson, H. R., Blake, R., & Halpern, D. L. (1991). Coarse spatial scales constrain the range of binocular fusion on fine scales. *J. Opt. Soc. Am. A*, *8*, 229–236.

Yeshurun, Y., & 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.

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

Received September 30, 1998; accepted February 12, 1999.