Computing Stereoscopic Disparity with Binocular Cortical Simple and Complex Cells

Florentin Wörgötter* and Alex Cozzi

Institut für Physiologie, Ruhr-Universität Bochum, D-44780 Bochum, Germany,
email:worgott@neurop.ruhr-uni-bochum.de

The analysis of the depth coordinates of objects in a visual scene is of vital importance for animals as well as in technological applications like autonomous robot navigation or product quality control. In the two-dimensional camera projections of a three-dimensional visual scene depth information is initially lost but can be recovered when using two cameras in a stereoscopic setup. The projection of an object located at a finite distance from the camera will be laterally displaced in the left camera image as compared to the right camera image. This displacement, called disparity, can then be used to retrieve the depth coordinate of the object. In principle this can be done by locally cross-correlating the left and the right camera images. In 1988 Sanger [1] proposed as an alternative that instead also the phase relation between two spatial band-pass filter responses (Gabor filters) could be used to measure the local disparity. Simple cells in the visual cortex have receptive fields which can be described as Gabor filters (spatial band-pass filters) [2, 3, 4]. Most of them are driven binocularly and tuned to respond most strongly to stimuli at a certain preferred distance from the fixation plane [5, 6, 7]. Cortical complex cells receive input from simple cells. As a consequence also complex cells implicitly encode visual depth [8]. Here we formalize the computational procedure which could underlie the extraction of depth information from complex cell responses and solve it analytically for two different stimulus situations. The theory predicts that a strong discrepancy should exist between the actual and the perceived depth of sine-wave luminance modulated ("grating") stimuli: If the spatial frequency of the grating is increased it should appear to move closer to the observer.

In general there are several strategies of how to retrieve depth information from a sequence of images, like depth from motion (flow-field analysis), depth from shading and depth from stereopsis, on which we concentrate in this article. In a stereoscopic approach usually two cameras are mounted with a horizontal distance between them. As a consequence objects displaced in depth from the fixation point are projected onto image regions which are horizontally shifted with respect to the image center. This shift is called *disparity* and it can be used to determine the depth of the object. Due to the geometry of the optic system it is thereby sufficient to restrict disparity analysis to the projection of corresponding linear segments (lines) in the left and right eye (epipolar line constraint). It is therefore not necessary to extend the problem to two dimensions, which raises computational complexity. It can, however, improve the results.

In the most straightforward approaches that address the problem of depth from stereo, the disparity is computed by searching the maximum of the cross-correlation between image windows along the epipolar lines of the left and right image. More recently spatially localized band-pass filters have been utilized to compute the disparity. This method computes the convolution between Gabor kernels (Eq. 1) and the left and right image parts. The alternating subfield arrangement in simple cells has been successfully described by such Gabor functions [3, 4]. A Gabor function is a sine-wave multiplied and, thus, damped by a Gaus-

sian envelope [2]. Thus, these cells represent localized spatial band-pass filters which are tuned to the resonance frequency k of the sine-wave and located at x_0 in the visual field where the Gaussian envelope has its center. In complex notation:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-x_0)^2}{2\sigma^2}} e^{i(kx-\phi)} \quad (1)$$

where σ is related to the width of the receptive field. The phase parameter ϕ represents the fact that most cells in the visual cortex have a receptive field which is mixed from a pure cosine- and a pure sine-type. We will set x_0 and ϕ to zero because they do not affect our results except by adding unnecessary mathematical complexity. Thus, in Eq. 1 the real component represents the cosine- and the imaginary component the sine-shaped receptive fields in Fig. 1. These are the archetypes of receptive fields that exist in monocularly driven simple cells. The linear part of the response of such a cell is given by the convolution of the receptive field with the stimulus f(x):

$$M_{l,r}(x) = G(x) * f_{l,r}(x) \quad (2)$$
$$= \int_{-\infty}^{+\infty} G(x - x')f_{l,r}(x')dx'$$

In the case of stereo-vision the responses of receptive fields in the two eyes need to be considered. Binocularly driven simple cells exhibit distinctive disparity tuning curves [5, 6, 7] and Nomura et al. [12] provided evidence that those can be obtained by summing the responses of two corresponding monocular simple cells from both eyes for which DeAngelis et al. found experimental support [13]. In order to explain deviations from a linear behavior furthermore Heeger suggested that simple cells exhibit a squared response characteristic [14]. Accordingly, we define for binocular disparity tuned simple cells (Fig. 1):

$$S_1^e(x) = [Re(M_l) + Re(M_r)]^2 \quad (3)$$
$$S_2^e(x) = [Re(M_l) - Re(M_r)]^2$$
$$S_3^e(x) = [Im(M_l) + Im(M_r)]^2$$
$$S_4^e(x) = [Im(M_l) - Im(M_r)]^2$$

$$S_1^o(x) = [Re(M_l) + Im(M_r)]^2$$
$$S_2^o(x) = [Re(M_l) - Im(M_r)]^2$$

$$S_3^o(x) = [Im(M_l) + Re(M_r)]^2$$
$$S_4^o(x) = [Im(M_l) - Re(M_r)]^2$$

where Re is the real and Im the imaginary part of a complex number.

The responses of these binocular simple cell receptive fields are contrast dependent because they are not normalized. A sensible normalization must take the energy of the stimuli into account [15] that can be computed with the well consolidated, so called "quadrature pair cell model" [16, 17, 18, 19]. For binocular input this model combines the cell responses in the following way arriving at a binocular complex cell C^e:

$$4C^e(x) = S_1^e - S_2^e + S_3^e - S_4^e \quad (4)$$

$$= 4[Re(M_l)Re(M_r) + Im(M_l)Im(M_r)]$$

The terms show that two times two binocular simple cells are combined in a push-pull arrangement ([20], see [21] for a review on push-pull models). The constant factor 4 can be disregarded in the following. This model is strongly supported by the findings of Ohzawa et al. [8] who provided evidence that a class of complex cells exists which seems to compute the positive parts (i.e., $S_1 + S_3$) of Eq. 4.

C^o is created in the same way as outline in Eq. 4 from the simple cells S^o. It should be noted that the complicated structure of Eq. 4 can be reduced in complex notation using Eq. 2 to:

$$C(x) = M_l M_r^* = C^e(x) + iC^o(x) \quad (5)$$

where the asterisk denotes the complex conjugate. The energy of a signal Z is given by $\| Z \| = \sqrt{ZZ^*}$. Thus, the last equation immediately shows that $C(x)$ represents something like the "cross-energy" of the responses from the two eyes. It implicitly encodes the disparity of the stimulus but it is still dependent on the location x, which means that different read-outs will be obtained from different locations in the receptive field. This unwanted dependency can be eliminated by normalizing the "cross-energy" signal with the left- and right-eye energy signals $C_l = \| M_l \|$, $C_r = \| M_r \|$, which can be obtained from the responses of other complex cells C_l, C_r which get only monocular inputs [18]. We define:

$$D = \frac{C}{C_l C_r} = \frac{M_l M_r^*}{\sqrt{M_l M_l^* \; M_r M_r^*}} \quad (6)$$

Figure 1: Schematic wiring diagram for disparity estimation with cortical cells. In order to get C^o interchange the two cells along the arrow and change the summation at C from $++--$ to $+-+-$.

For monocular responses a related normalization process has been theoretically proposed by Heeger [22] for which experimental support became available later [23].

We will consider two simple stimulus situations: Small spot stimuli and, alternatively, sine-wave stimuli. We assume that a disparity of d exists between the left and the right input. Spot stimuli are commonly described by delta-functions. Thus, in the first case we have: $f_l(x) = \delta(x)$ and $f_r(x) = \delta(x - d)$.

The result is obtained from the convolution of the δ-function stimuli with the receptive field archetypes (Eq. 2) as $\int_{-\infty}^{+\infty} G(x - x')\delta(u_{l,r})dx'$, where $u_l = x'$ and $u_r = x' - d$, respectively. This yields:

$$M_l(x) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-x^2}{2\sigma^2}}}_{=A(x)} e^{ikx} \quad (7)$$

and
$$M_r(x) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-d)^2}{2\sigma^2}}}_{=A(x-d)} e^{ik(x-d)}$$

where A is the Gaussian attenuation factor. Accordingly we get:

$$M_l M_r^* = A(x)A(x-d)e^{ikx}e^{-ik(x-d)} \quad (8)$$
$$= A(x)A(x-d)e^{ikd}$$

Furthermore we find that

$$\sqrt{M_l M_l^* \ M_r M_r^*} = A(x)A(x-d) \quad (9)$$

and obtain the result of Eq. 10 after normalization as:

$$D = e^{ikd} = \cos[kd] + i\sin[kd] \quad (10)$$

The disparity d can now be computed by inverting the trigonometric function of either the real or the imaginary part of the last equation. Neurons can achieve this by

applying an appropriate (sigmoidal-like) activation function. Finally the result must be divided through the constant k, which is the resonance frequency of the initial simple cell Gabor-type receptive fields.

$$d = \frac{1}{k} \arccos[Re(D)] \qquad (11)$$

$$\text{or} \quad d = \frac{1}{k} \arcsin[Im(D)]$$

Since k is known, d can be exactly computed when two spot stimuli are presented. The intuitive reason for this result is that the power spectrum of a dot-stimulus (δ-function) contains all frequencies. Thus, the cortical band pass filters (Gabor type simple cells) can resonate at their own reference frequency k and use it for all subsequent computations.

The situation changes when we use two sine-wave grating stimuli with stimulus frequency ω and disparity d between them. Thus, $f_l(x) = sin[\omega x]$ and $f_r(x) = sin[\omega(x - d)]$. In this case the calculation of the result is very tedious and shall not be presented here. As final result we get:

$$D = e^{i\omega d} = \cos[\omega d] + i \sin[\omega d] \qquad (12)$$

While this looks very similar to Eq. 10, the central difference is that the result now depends on the stimulus frequency ω which is unknown to the system. The disparity can now only be approximately computed by using the (known) resonance frequency k and we get:

$$\hat{d} = \frac{\omega}{k} d \qquad (13)$$

Thus, $\hat{d}$ is only equal to d if the stimulus frequency is identical to the resonance frequency of the simple cell Gabor receptive fields: $\omega = k$, which is not the general case.

At this point it is important to realize that this effect is already introduced by the band-pass characteristic of cortical simple cells regardless of the subsequently following computational steps. Unlike for the dot-stimulus, the grating stimulus contains only the frequency ω, which will be retrieved by the cortical simple cells and the system is forced to compare it with its own intrinsic resonance frequency k. In technical systems this error is avoided by actually measuring the local frequency ω. This is achieved by the calculation of the derivatives of the left and right filter

output phase-components [9]. Derivatives are very sensitive to noise and it is rather complicated to compute a derivative with a neuronal network. Therefore, it seems unlikely that the cortical network implements this particular a correction mechanism.

Without any correction, however, the theory predicts that for a high spatial frequency $\hat{d}$ is much larger than d, thus a narrow grating should appear more in front than a wide grating presented at the same true disparity d.

The theory suggests that neuronal operations in simple and complex cells - many of which have already been observed experimentally - can in a very direct way lead to disparity estimates of the objects in a visual scene. Thus, it seems that the computation of visual disparities, which is a central component for the perception of depth, is already to a large degree solved by the cells in the primary visual cortex.

References

[1] Sanger, T.D. Stereo disparity computation using gabor filters. Biol. Cybern. 59, 405-418 (1988).

[2] Gabor, D. Theory of communication. J. IEEE Lond. 93, 429-457 (1946).

[3] Daugman, J.G. Two-dimensional spectral analysis of cortical receptive field profiles. Vision Res. 20, 847-856 (1980).

[4] Jones, J.P. & Palmer, L.A. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. J. Neurophysiol. 58, 1187-1211 (1987).

[5] Barlow, H.B., Blakemore, C. & Pettigrew, J.D. The neural mechanism of binocular depth discrimination. J. Physiol.. 193, 327-342 (1967).

[6] Pettigrew, J.D., Nikara, T. & Bishop, P.O. Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. Exp. Brain Res. 6, 391-410 (1968).

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

[8] Ohzawa, I., DeAngelis, G.C. & Freeman, R.D. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science 249, 1037-1041 (1990).

[9] Fleet, D., Jepson, A. & Jenkin, M. Phase-based disparity measurement. Computer Vision, Graphic and Image Process. 53, 198-210 (1991).

[10] Cozzi, A., Crespi, B., Valentinotti, F. & Wörgötter, F. Performance of phase-based algorithms for disparity estimation. Machine Vis. Appl. 9, 334-340 (1997).

[11] Hubel, D.H. & Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160, 106-154 (1962).

[12] Nomura, M., Matsumoto, G. & Fujiwara, S. A binocular model for the simple cell. Biol. Cybern. 63, 237-242 (1990).

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

[14] Heeger, D.J. Half-squaring in responses of cat striate cells. Vis. Neurosci. 9, 427-443 (1992).

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

[16] Pollen, D.A. & Ronner, S.E. Phase relationships between adjacent simple cells in the visual cortex. Science 212, 1409-1411 (1981).

[17] Pollen, D.A. & Ronner, S.F. Spatial computation performed by simple and complex cells in the visual cortex of the cat. Vision Res. 22, 101-118 (1982).

[18] Pollen, D.A. & Ronner, S.F. Visual cortical neurons as localized spatial frequency filters. IEEE Trans. SMC 13, 907-915 (1983).

[19] Liu, Z., Gaska, J.P., Jacobson, L.D. & Pollen, D.A. Interneuronal interaction between members of quadrature phase and anti-phase pairs in the cat's visual cortex. Vision Res. 32, 1193-1198 (1992).

[20] Wörgötter, F., Nelle, E., Li, B., Wang, L. & Diao, Y. A possible basic cortical microcircuit called "cascaded inhibition" Results from cortical network models and recording experiments from striate simple cells. Exp. Brain Res. 122, 318-332 (1998).

[21] Palmer, L.A., Jones, J.P. & Stepnowski, R.A. Striate receptive fields as linear filters: characterization in two dimensions of space. In: Cronly-Dillon J, Leventhal AG (eds). The neural basis of visual function. (Vision and visual dysfunction, vol 4) Macmillan, London (1991).

[22] Heeger, D.J. Normalization of cell responses in cat striate cortex. Vis. Neurosci. 9, 181-197 (1992).

[23] Carandini, M. & Heeger, D.J. Summation and Division by neurons in primate visual cortex. Science 264, 1333-1336 (1994).

[24] Blakemore, C. The range and scope of binocular depth discrimination in man. J. Physiol. 211, 599-622 (1970).

[25] Badcock, D.R. & Schor, C.M. Depth-increment detection function for individual spatial channels. J. opt. Soc. Am. A. 2, 1211-1216 (1985).

[26] Wörgötter, F., Suder, K., Zhao, Y., Kerscher, N., Eysel, U.T. & Funke, K. State-dependent receptive-field restructuring in the visual cortex. Nature 396, 165-168 (1998).

[27] Lehky, S.R. & Sejnowski, T.J. Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. J. Neurosci. 10, 2281-2299 (1990).