VISION RESEARCH

Encoding and estimation of first- and second-order binocular disparity in natural images

CrossMark

Paul B. Hibbard [a,b,*], Ross Goutcher [c], David W. Hunter [b]

[a] Department of Psychology, University of Essex, Colchester CO4 3SQ, UK
[b] School of Psychology and Neuroscience, University of St Andrews, St Mary's Quad, South Street, St Andrews, KY16 9JP Scotland, UK
[c] Psychology, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK

ARTICLE INFO

ABSTRACT

The first stage of processing of binocular information in the visual cortex is performed by mechanisms that are bandpass-tuned for spatial frequency and orientation. Psychophysical and physiological evidence have also demonstrated the existence of second-order mechanisms in binocular processing, which can encode disparities that are not directly accessible to first-order mechanisms. We compared the responses of first- and second-order binocular filters to natural images. We found that the responses of the second-order mechanisms are to some extent correlated with the responses of the first-order mechanisms, and that they can contribute to increasing both the accuracy, and depth range, of binocular stereopsis.
© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. The physiology of binocular vision

The binocular disparities provided by our front-facing eyes are an important cue to depth and distance. These disparities are encoded by binocular neurons in the visual cortex. These neurons, whose responses are affected by the images formed in both eyes, are found in many cortical areas, including V1 (Barlow, Blakemore, & Pettigrew, 1967; Hubel & Livingstone, 1987; Poggio & Poggio, 1984), V2 (Bredfeldt & Cumming, 2006; Thomas, Cumming, & Parker, 2002; Zhou, Friedman, & von der Heydt, 2000), V3/V3A (Anzai, Chowdhury, & De Angelis, 2011; Cottereau, McKee, & Norcia, 2012) V4 (Shiozaki, Tanabe, Doi, & Fujita, 2012; Umeda, Tanabe, & Fujita, 2007) and V5/hMT+ (Krug & Parker, 2011). This wide spread of disparity sensitive areas across the cortex allows disparity processing to be specialised for distinct functional roles (Krug & Parker, 2011; Parker, 2007; Roe, Parker, Born, & DeAngelis, 2007).

The binocular energy model (DeAngelis, Ohzawa, & Freeman, 1991; Fleet, Wagner, & Heeger, 1996; Ohzawa, DeAngelis, & Freeman, 1990) provides a good approximation to the responses

of the initial stages of binocular processing in the primary visual cortex, although it does have some notable limitations (Faria, Batista, & Araújo, 2013; Haefner & Cumming, 2008; Read, Parker, & Cumming, 2002; Tanabe & Cumming, 2008). The basic structure of this model is outlined in Fig. 1. The first stage of the model consists of Gabor filters, in quadrature pairs, applied to each eye's image. Binocular responses are created by first summing over corresponding pairs of filters in the two eyes, before squaring and summing again across the two components of the quadrature pair. For example, if the filters are in even-symmetric and odd-symmetric phase, the responses of the even filters to left and right eyes' images are summed, as are the responses of the odd filters to the two images. The energy response is then created by squaring, then adding, these results.

Because the binocular energy model contains receptive fields in each eye, the overall response is affected by binocular disparities. This means that energy neurons can be created that are tuned for particular binocular disparities. In the case that the left and right eyes' filters are identical, the model neuron will tend to respond best when the disparity is zero. Tuning to non-zero disparities can be achieved by introducing inter-ocular differences in the location or phase of the Gabor filters. Examples of these are shown in figure 1b and c, and both types of tuning are found in cortical neurons; typically, individual neurons will be tuned to non-zero disparities in both position and phase (Prince, Cumming, & Parker, 2002).

* Corresponding author at: Department of Psychology, University of Essex, Colchester CO4 3SQ, UK.
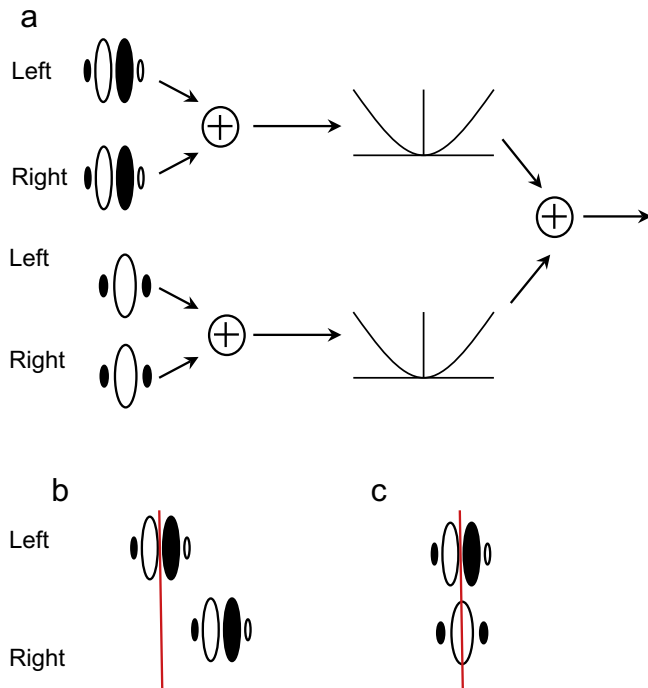   *E-mail address:* phibbard@essex.ac.uk (P.B. Hibbard).

Fig. 1. Outline of the binocular energy model. (a) Images are filtered with quadrature pairs of Gabor receptive fields. The white areas represent excitatory regions of the receptive fields, and the dark areas inhibitory regions. Responses are summed across the two eyes for corresponding filters, then squared. Finally, these squared outputs are summed across the two halves of the quadrature pair. (b) Positional disparity tuning is achieved if the two eyes' receptive fields are in different locations. Here, the vertical red line shows the centre of the left eye's receptive field; the right eye's receptive field is identical in shape but shifted to the right. (c) Phase disparity tuning is achieved if the two eyes' receptive fields have a different shape. Here, the two receptive fields are in the same location, but the left eye's is odd-symmetric while the right eye's is even symmetric.

The binocular energy model is related to the concept of the use of cross-correlation as a way to match binocular images and estimate disparity (Allenmark & Read, 2011; Banks, Gepshtein, & Landy, 2004; Fleet et al., 1996). In this model, the correlation between samples taken from the left and right images is calculated, at a range of disparities. High values of correlation can then be used as an indication of the similarity between the two samples, as a function of disparity, and thus used to estimate binocular disparity. This approach has been used to account for the spatial resolution of binocular depth (Banks et al., 2004) and the maximum gradient of disparity that supports binocular depth perception (Filippini & Banks, 2009). In order to approximate a cross-correlation on the basis of energy outputs, it is necessary to pool information across orientation, spatial frequency and space (Allenmark, 2011; Fleet et al., 1996) and to normalise responses to take account of differences in contrast between the samples from the two eyes (Hibbard, 2008). Thus, although the spatial limitations imposed by the initial binocular sampling in V1 can account for a number of limitations of depth perception (Banks et al., 2004; Filippini, 2009), the actual calculation of cross-correlation is likely to involve a more extended network of processing.

1.2. Depth perception from second-order binocular disparity

The first-stage filters of the binocular energy model are tuned to the orientation and frequency of the luminance variation in the monocular images. The disparity tuning of these model neurons means that they are sensitive to differences in the location of these luminance variations, which are referred to as first-order cues. In addition, depth can also be perceived on the basis of disparities in contrast-variations in this first-order information, even when there are no disparities in the first-order cues themselves (Edwards, Pope, & Schor, 1999, 2000; Hess & Wilcox, 1994; Langley, Fleet, & Hibbard, 1999a, 1999b; McKee, Verghese, & Farell, 2004; Schor, Edwards, & Pope, 1998; Wilcox & Hess, 1995, 1996, 1997). An example of these second-order, contrast-envelope cues is shown in Fig. 2. Here, the contrast of a Gaussian random noise sample (the first-order carrier) is modulated by a vertical sinewave (the second-order envelope). The noise carrier is identical for the left and right images, and presented at zero disparity. However, there is a non-zero (crossed) disparity in the modulation envelope. The fact that observers are able to perceive depth from contrast envelope cues in these stimuli, and others such as contrast modulated sinusoids, and Gabor patches in which the carrier and envelope disparity can be manipulated independently, has been used to argue for the existence of an additional, second-order stereoscopic channel (Hess & Wilcox, 1994; Langley et al., 1999a, 1999b; Wilcox & Hess, 1996).

1.3. Second-order mechanisms in the visual cortex

The perception of depth from disparities in contrast envelopes is an example of a more general phenomenon of second-order vision. This refers to situations in which perceived structure in the image cannot be derived from the outputs of linear filters at single points in the image (Graham & Sutter, 1998). In the case of the vertically-oriented contrast envelope in Fig. 2, for example, a linear filter tuned to the orientation and frequency of the envelope will not respond strongly and selectively to this stimulus.

The mechanisms that allow for the perception of this structure have variously been described as second-order, non-Fourier or complex (Graham, 2011; Graham & Sutter, 1998). Typically, they are modelled as two linear filtering stages, in which the responses of the first-stage filters are first transformed by a rectifying non-linearity, and secondly "paste[d] together" by the second-stage filters with larger receptive fields than those of the first-stage filters. (Graham & Sutter, 1998, p232).

Neurons in the prestriate cortex of both cats and monkeys have been found that are consistent with this filter-rectify-filter (FRF) model, and are therefore likely candidates as the physiological substrate of second-order channels. (Li et al., 2014; Mareschal & Baker, 1998a, 1998b, 1999; Rosenberg, Husson, & Issa, 2010; Rosenberg & Issa, 2011; Tanaka & Ohzawa, 2006; Zhou & Baker, 1993, 1994, 1996). These studies have identified neurons that respond to both first-order luminance-defined stimuli, and second-order, contrast-defined stimuli. These neurons have very similar tuning for orientation and spatial frequency in both luminance gratings, and the envelope of contrast-modulated gratings. In the cat, neurons have tended to be tuned to slightly higher luminance frequencies than envelope frequencies (Mareschal & Baker, 1998, 1999), while in the macaque monkey, this ratio is reversed (Li et al., 2014). For contrast modulated stimuli, these neurons are typically tuned for both the orientation and spatial frequency of both the carrier and envelope of the stimulus. In the FRF model, this tuning to the carrier reflects the properties of the early filter, and the tuning to the envelope the properties of the later filter. For contrast-modulated stimuli, the preferred carrier frequency reported was typically considerably higher than that of the contrast envelope. While there is a lot of variability across individual neurons, average estimates of the optimal spatial frequency ratio of 10 (Mareschal & Baker, 1999), 11.0 (Tanaka & Ohzawa, 2006) and 8.2 (Li et al., 2014) have been reported. There is also no clear relationship between the orientation tuning for the carrier and envelope (Mareschal & Baker 1998a, 1999), although Li et al. (2014) found a greater preponder-