Cortical computations that support stereo vision

Jenny C. A. Read and Bruce G. Cumming

Abstract
Our visual system fuses the images from left and right eyes into a single percept, while using the disparities between them to extract information about depth. Disparity-tuned neurons in primary visual cortex are believed to carry out the initial processing which ultimately makes this possible. A relatively simple model (the energy model) provides the most successful account of the mechanism by which these neurons signal disparity. Even this model fails to capture several properties of real neurons. We describe how a simple modification of the energy model enables us to account for three puzzling experimental observations.

1000 word summary
Because our eyes are displaced from each other horizontally, there are small differences – disparities – between the views from the two eyes. These disparities are sufficient for depth perception even in the absence of other cues. The initial stages in the processing of binocular disparity are believed to occur in primary visual cortex (V1), the first place in the visual system where information from the two eyes is combined. Many cells in V1 modulate their firing rate according to the disparity of the visual stimulus.

At this early stage in the visual system, the cells' response seems to be a relatively simple function of the retinal input. Many aspects of the behaviour of these cells have been explained using an "energy model". The first stage in this model is a linear spatial summation of image luminance over the receptive field of the cell, weighted by a receptive field function which describes the cell's response to luminance at each point. Thus the cell computes the convolution v of the retinal image $I(x,y)$ with its receptive field function $\rho(x,y)$:

$$v = \int_{-\infty}^{+\infty} I(x,y)\rho(x,y)dxdy .$$

Disparity-sensitive neurons have (at least) two receptive fields, one in each retina. Thus they compute both v_L and v_R. The energy model postulates that the cell's response is given by the square of the sum of left- and right-eye inner-products:

$$C = (v_L + v_R)^2.$$

However, as more data has been gathered on the properties of disparity-tuned neurons, it has become clear that this very simple model fails to explain several properties of real neurons.

- Respnse to anticorrelated stimuli.

Anti-correlated stimuli are those in which one eye's image has been replaced by its photographic negative. As the disparity of these images is varied, the energy model predicts that neurons will modulate their firing as strongly as for the corresponding correlated stimuli. In fact, the amplitude of modulation is significantly reduced in V1 neurons.

- Shape of the disparity tuning curve

Given the receptive field profiles in the two eyes, the energy model enables us to predict how the cell modulates its firing rate according to the disparity of the stimulus. Real cells show systematic deviations from this prediction. In particular, neurons are often observed to be highly selective for the spatial frequency of sinusoidal luminance gratings. This indicates that their receptive field functions must include both ON regions (where bright stimuli tend to increase the cell's firing, and dark stimuli to reduce it) and OFF region (where the opposite holds). The energy model then predicts that, although some favoured disparities will increase the cell's firing rate above that obtained with uncorrelated images, other disparities should reduce its firing rate below that background level. In fact, most cells show much less inhibition of this type than predicted.

- Response to monocular stimulation

The energy model predicts that the response to visual stimuli presented to one eye only will be the square of the convolution of the image with the receptive field in that eye:

$$C = (v_L)^2 .$$

Thus, it predicts that cells which are sensitive to disparity must show some response to monocular stimulation in either eye. In fact, cells are frequently observed which show no response to monocular stimulation in one eye (so would traditionally be classified as receiving input from only one eye), but which nevertheless are clearly sensitive to binocular disparity (which shows that they must be receiving input from both eyes).

For each of these problems, there are a number of modifications which could be made to the energy model to account for the experimental observations. However, we concentrate on one simple modification which we believe can account for all three. We propose that the convolutions from each eye are separately passed through a thresholding non-linearity before being combined. Thus, our modified model computes

$$C = (\Theta(v_L) + \Theta(v_R))^2 \quad ,$$

where Θ is a threshold function characterized by some threshold $\theta > 0$: $\Theta(x) = x - \theta$ if $x > \theta$, $\Theta(x) = 0$ otherwise.

This is physiologically plausible, since firing rates cannot be negative: neurons naturally implement threshold non-linearities at their output. Thus, our model can be viewed as proposing that the convolutions from the two retinae are passed first through monocular neurons, before being combined at a binocular neuron.

We examine how this new model can help us account for the three observations highlighted above.

- Response to monocular stimulation

If we further postulate that, in some cases, the monocular neurons make inhibitory synapses onto the binocular neuron, it is immediately clear how cells which appear to be monocular when stimulated via one eye at a time can show disparity selectivity when driven binocularly. Fig. 1 illustrates the situation where the right eye provides solely inhibitory input onto the binocular cell. Clearly the binocular cell will never fire in response to right-eye monocular stimulation. However, input to the right eye will still modulate the response to input in the left-eye, allowing disparity tuning to emerge. This

is not possible in the energy model, since that postulates that both eyes provide both excitatory and inhibitory input.

- Response to anticorrelated stimuli.

We can also see qualitatively how this modification to the energy model can allow a reduced response to anti-correlated stimulation. The energy model may be expanded as

$$C = (v_L)^2 + (v_R)^2 + 2v_L v_R .$$

It is only the last term which describes the disparity-specific response, since this is the only term in which information from both eyes is combined. Anti-correlation means that one image is replaced with its negative, so the response of the energy model changes to

$$C = (v_L)^2 + (v_R)^2 - 2v_L v_R .$$

The disparity-modulated term changes sign, but its amplitude is unaffected.

Once we have applied a threshold to each of the convolutions, this argument no longer applies, since $\Theta(-x) \neq -\Theta(x)$. Simulations with this new model demonstrate that it shows reduced amplitude to anti-correlated stimuli, in agreement with experiment.

- Shape of the disparity tuning curve

The threshold tends to reduce the effect of the OFF regions of the receptive field, meaning that the "inhibitory" disparities become less powerful in suppressing the cell's response. Simulations with our new model produce disparity tuning curves which more closely resemble those found in experiments.

Thus we have found a simple and physiologically plausible modification to an existing successful model, which enables us to retain the successes of the original model while extending it to account simultaneously for three puzzling experimental observations.
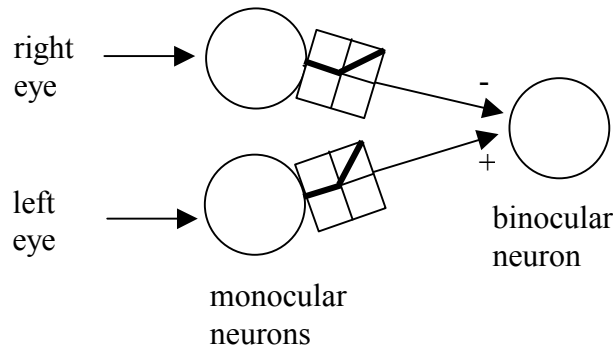

Fig. 1