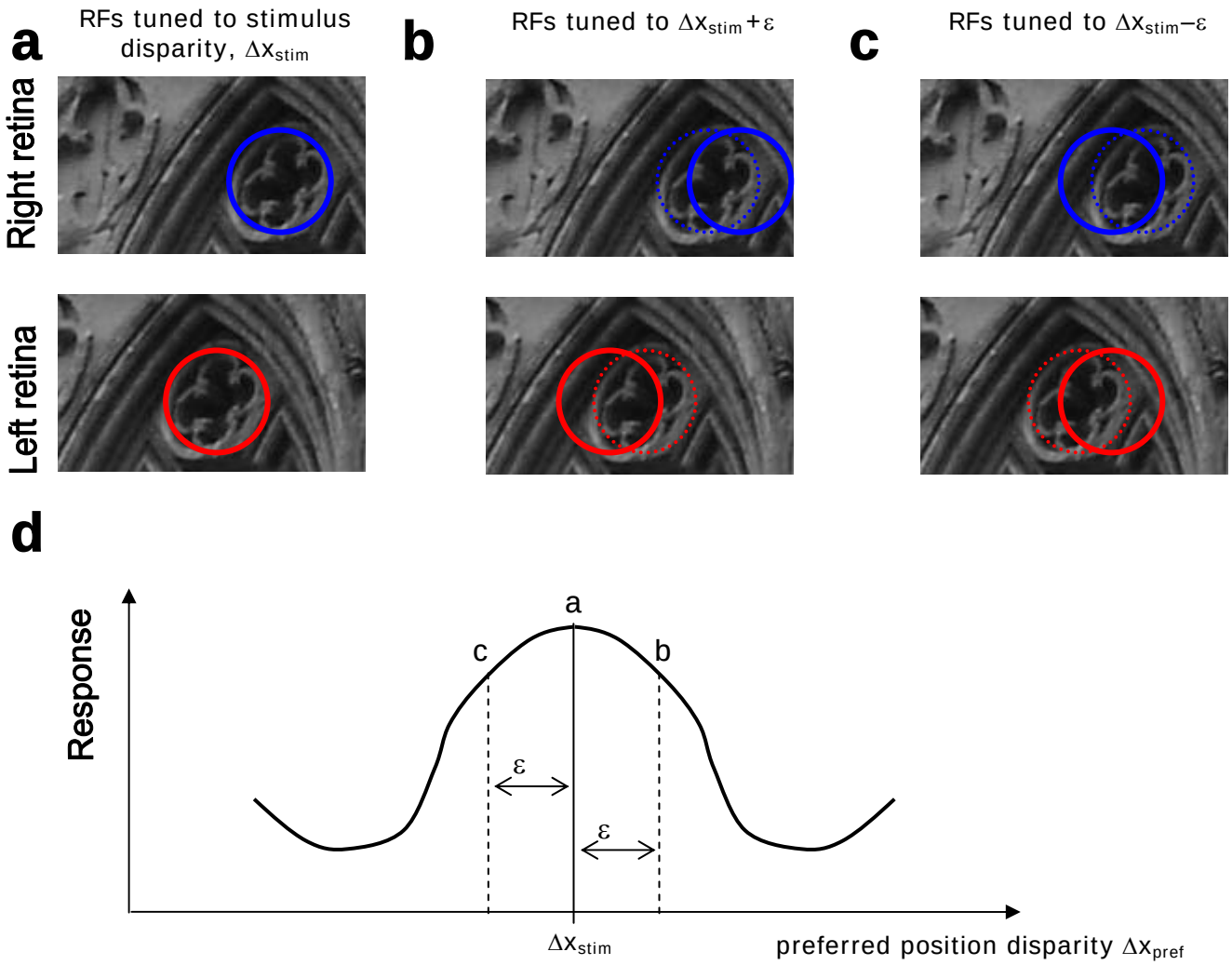
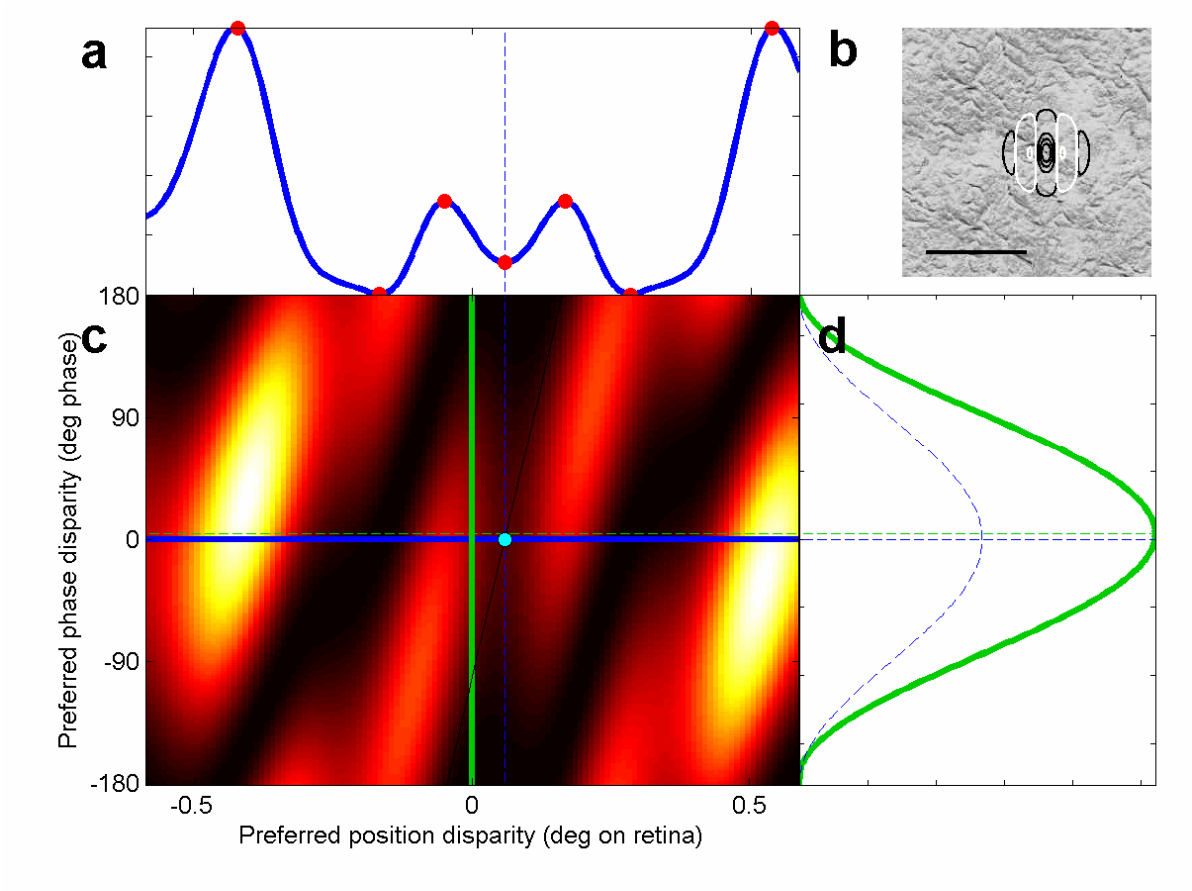
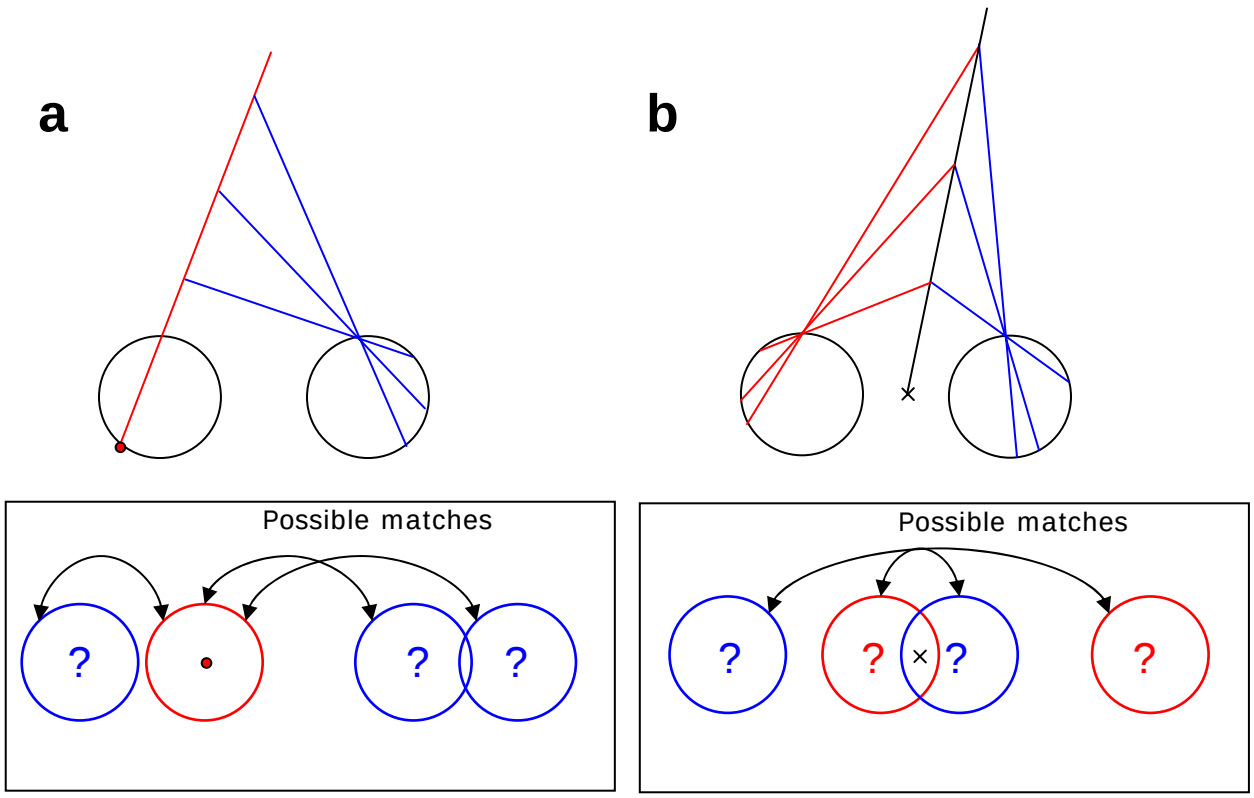


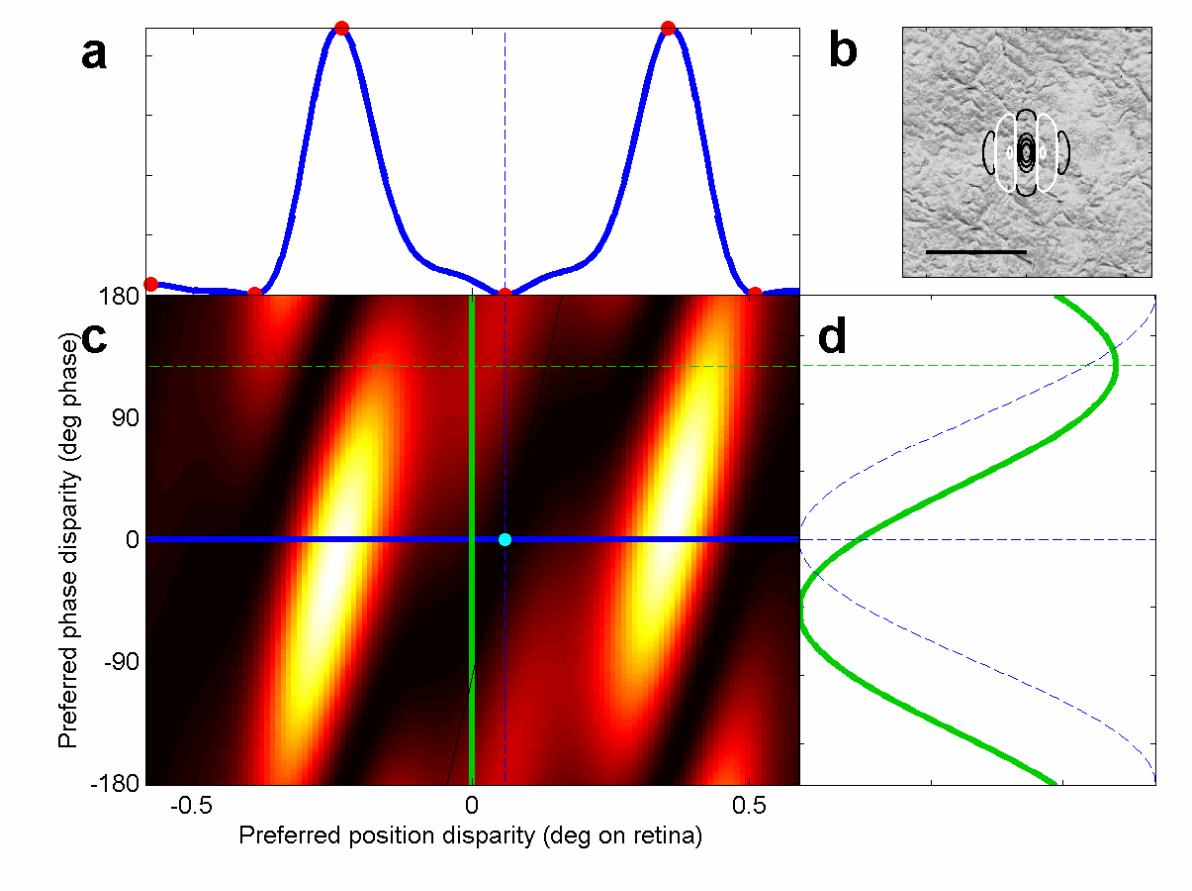
SuppFig 1. Intuitive argument for why TE neurons must have a local extremum at the stimulus disparity. **a-c**: left and right receptive fields (RFs) for neurons tuned to different position disparities. **a**: RFs tuned to stimulus disparity Δx_{stim}; **b,c**: RFs tuned to stimulus disparity $\pm \varepsilon$ (RFs in **a** are shown with dotted lines for comparison). **d**: population response of tuned-excitatory neurons tuned to different position disparities, showing that this response is symmetric about the stimulus disparity.



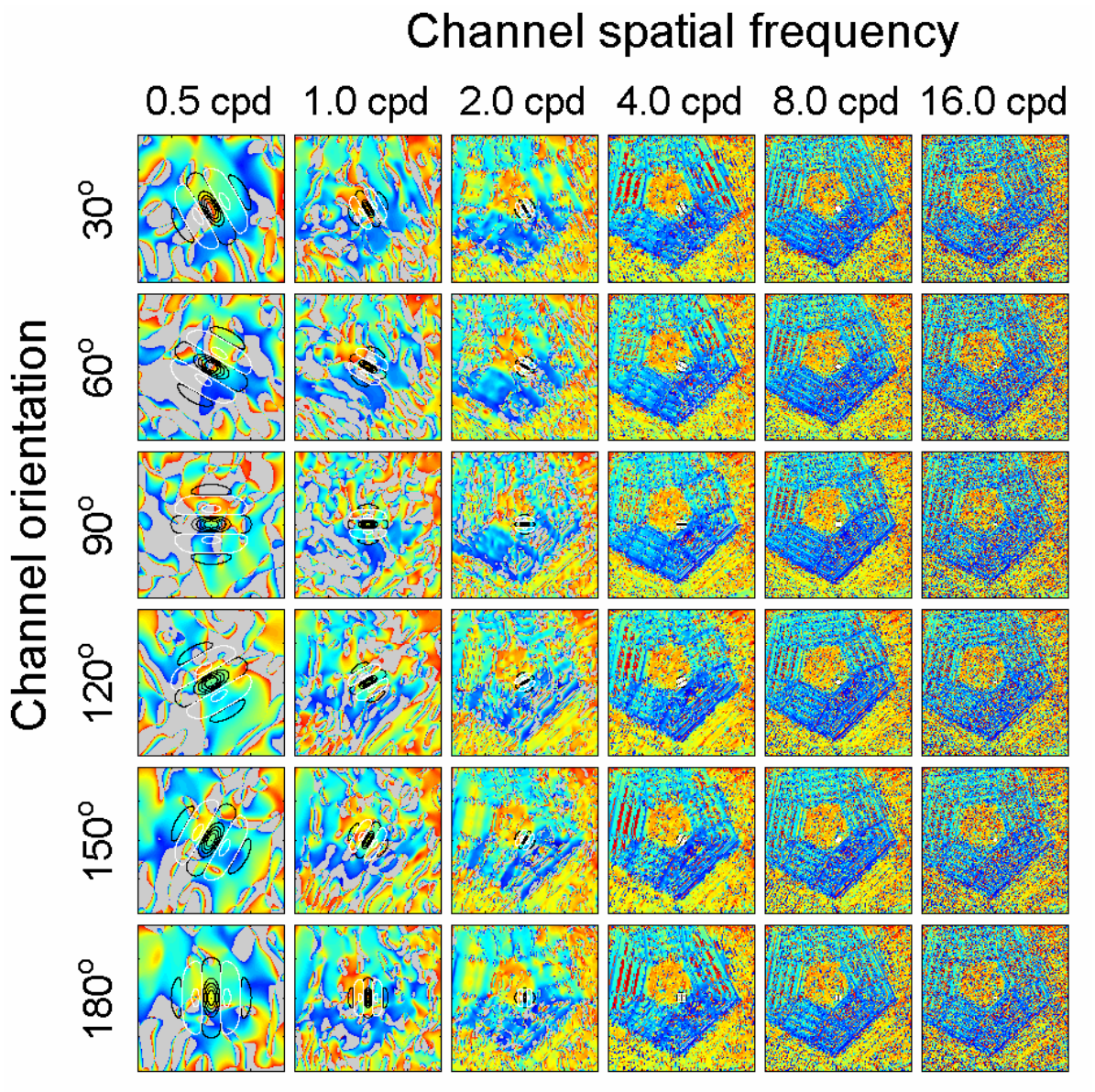
SuppFig 2. As Fig. 2, except at a slightly different point in the scene. Now, the stimulus disparity falls at a local *minimum* with respect to the cells' preferred position disparity, although still, as always, at a local maximum wrt their phase disparity. Once again, the true stimulus disparity (cyan) can be read off from the population response in **c**, since it is the only point on the blue line which is both an extremum wrt position disparity and a maximum wrt phase disparity.



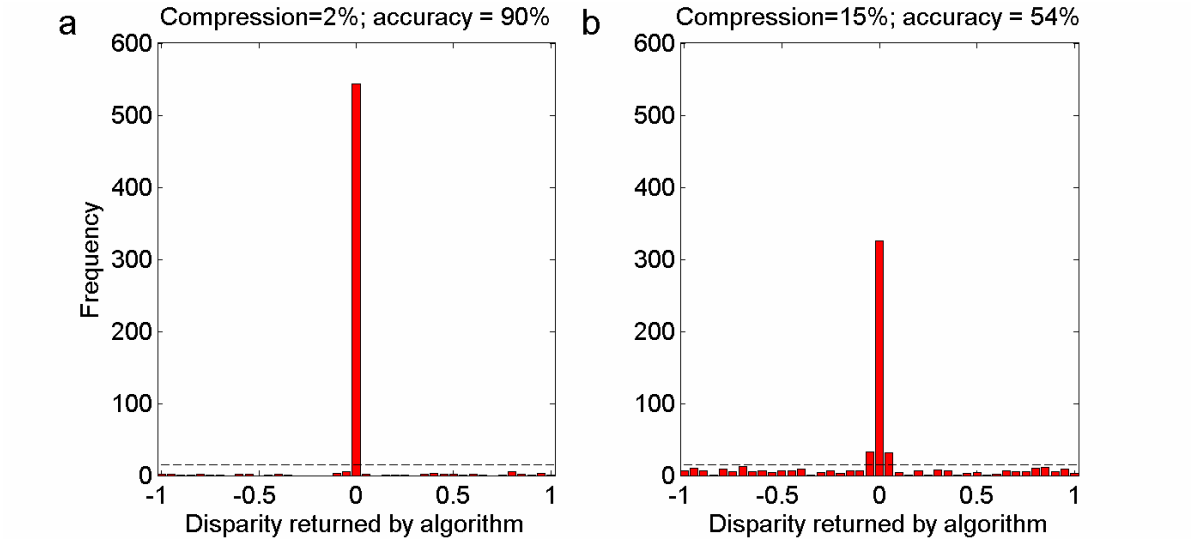
SuppFig 3: Two different approaches to framing the stereo correspondence problem. **a**: pick a feature in one eye (red dot), seek its match in the other eye (blue lines represent possible matches). **b**: pick a visual direction (black), and seek the stimulus disparity at this point (red, blue represent possible left- and right-eye positions). The top row represents the two approaches in terms of position in space (eyes are viewed from above); the bottom row represents them in terms of matching receptive fields on the two retinae. The black arrows in the lower panels represent different possible matches between left (red) and right (blue) receptive fields.



SuppFig 4. Population response to an anti-correlated image. As Fig. 2, except that the contrast of the right image has been inverted, equivalent to a stimulus phase disparity of π.



SuppFig 5. Results for the Pentagon stereogram. Disparity maps returned by single channels tuned to different spatial frequencies (rows) and orientations (columns). Colorscale shows estimated disparity for each point in the image; pseudocolor as in Fig. 5; gray = no estimate returned. Contours show example RF for each channel. Contours are drawn at ± 0.1%, 20%, 40%, 60% and 80% envelope maximum amplitude; black lines show positive values (ON regions) and white show negative (OFF).



SuppFig 6. Results returned by the algorithm for a slanting surface. The stimuli were noise images with a horizontal compression, representing a surface slanting about a vertical axis. A horizontal size ratio H corresponds to a slant of S relative to gaze-normal, where $H = [2R - I \tan S] / [2R + I \tan S]$, I is the interocular distance and R the distance of the surface. **(a)** H = 98%, representing a slant of 6° at 30cm, 9° at 50cm or 18° at 1m; (**b)** H = 85%, representing a slant of 38° at a distance of 30cm, and even more extreme slants at further distances. 100 such stereograms were generated; in each, the disparity at the centre of the receptive field (really 0°) was estimated by 6 channels (2 spatial frequencies 1 octave apart, and 3 orientations 60° apart). The frequency histogram sums individual results for all channels. Errors were counted as those where the algorithm returned a disparity of > 0.025° in magnitude (i.e. outside the central bin). In this simulation, the algorithm was constrained to return an answer between ± 1° (or none). The dashed line shows where the bars would lie if the algorithm were returning random numbers in between this range (chance accuracy = 2.5%).

Why build sensors for impossible stimuli?
Stereo correspondence with position and phase disparity

Jenny C. A. Read & Bruce G. Cumming

Supplementary Information

MATHEMATICAL PROOFS

In the paper, we claimed that, for uniform-disparity images, the true match is distinguished by (A) being at zero phase disparity, (B) being at a local extremum (maximum or minimum) with respect to position disparity, and (C) being at a local maximum with respect to phase disparity. We shall first justify this with an intuitive argument, and secondly with a formal mathematical proof.

We shall prove these results for an energy-model linear simple cell, i.e. a unit with only a single receptive field in each eye, whose response is $E = (L + R)^2$. Since the sum of two functions which each have a maximum at x also has a maximum at x, the same results also hold for an energy-model complex cell, which is the sum of 2 such linear simple cells. Claim (A) above is simply a property of the stimulus (if the stimulus has zero phase disparity, then the true stimulus match is in cells tuned to zero phase disparity). This leaves us with two claims to prove, which we express formally as follows.

We consider the output of a population of binocular energy-model linear simple cells to be a function of two variables: the preferred position disparity Δx_{pref} and the preferred phase disparity $\Delta\phi_{pref}$ of each cell. Then we are seeking to prove two claims concerning the behaviour of this function at the stimulus disparity (i.e. when $\Delta x_{pref} = \Delta x_{stim}$ and $\Delta\phi_{pref} = 0$).

Theorem B: At the stimulus disparity, the function has a local extremum (maximum or minimum) with respect to preferred position disparity.

Theorem C: At the stimulus disparity, the function has a local maximum with respect to preferred phase disparity.

Intuitive argument for Theorems B and C

Formally stated, these theorems may sound complicated, but they are in fact straightforward to the point of being obvious. We therefore begin not with a formal proof, but with an intuitive justification. By hypothesis, the left image is a shifted version of the right, with only position disparity (no phase disparity). This is shown in SuppFig 1: in each panel the top row shows the right eye's image, while the bottom row shows the left eye's image, which is identical apart from being shifted to the left. Consider a TE neuron, with (by definition) no phase disparity, and with the position disparity that matches the stimulus. The image patches seen by its receptive fields are identical (receptive fields marked in blue (right eye) and red (left eye) in SuppFig 1a). Now consider increasing the preferred disparity of the receptive fields by some amount ε while keeping the same cyclopean position. This means displacing the receptive fields by equal amounts in opposite directions (SuppFig 1b). The images seen by the receptive fields are now different. However, compare the effect of *decreasing* the preferred disparity by ε (SuppFig 1c). Now, the image patch seen by the right eye's RF

in SuppFig 1c is the same as the image patch seen by the left eye's RF in SuppFig 1b. However, for a TE neuron the left and right RF profiles are identical, by definition. Thus, the fact that left and right RFs are swapped over in SuppFig 1c compared to SuppFig 1b makes no difference to the response of the neuron. This means that the response at a preferred disparity of (Δx_{stim} + ε) is the same as that at (Δx_{stim} – ε), i.e. the population response, as a function of preferred position disparity, is symmetric about the stimulus disparity (SuppFig 1d). Unless the response is completely flat, this means that there is an extremum – either a maximum or a minimum – at the stimulus disparity (as well, of course, as a large number of other extrema which are *not* at the stimulus disparity!).

At the correct stimulus disparity (a), the left and right receptive fields are viewing identical image-patches. As the neuron's disparity moves away from this (b,c), the binocular correlation reduces. This in turn usually reduces the response of the energy-model unit, meaning that the response is usually a maximum at (a). However, the response of binocular energy-model cells does not depend solely on binocular correlation: contrast energy also affects their response. If the mismatched image-patches viewed in (b,c) have significantly more contrast energy than the matching patches, it is possible for the response at (a) to be a minimum.

By a similar line of reasoning, it is clear that the response as a function of preferred phase disparity must also be symmetric about the stimulus phase disparity (which, for normal images, is zero). In fact, it can be proved that the response is always a maximum at the stimulus phase disparity (see below).

Thus, if we consider a neuronal population varying in preferred position and phase disparity but not in cyclopean position, the stimulus disparity lies at the centre of two axes of symmetry. In theory, either or both of these could be used to identify the stimulus disparity. However, identifying the centre of symmetry along the position-disparity axis is likely to be challenging, given a finite range of position disparities in the population, since the centre of symmetry may lie out towards one end of this range, and may be at either a maximum or a minimum. The symmetry along the phase-disparity axis is much more straightforward to use, since for natural images the stimulus disparity always lies at the centre of the population of phase-disparity sensors, and is a maximum. We suggest that the brain constructs phase-disparity sensors in order to take advantage of this.

Mathematical proof of Theorems B and C

We now prove Theorems B and C formally. These theorems hold for a wide range of receptive fields, more general than the Gabor functions used in our computer simulations. In our proof, we allow the receptive field envelope to be an arbitrary function $V(x,y)$, not necessarily a Gaussian, and we allow both vertical and horizontal position disparity. As before, we assume that the left- and right-eye receptive fields differ only in their position and phase.

We can therefore write the receptive fields as:

$$\rho_L(x,y) = V\left(x + \frac{\Delta x_{pref}}{2}, y + \frac{\Delta y_{pref}}{2}\right)\cos\left[k_x\left(x + \frac{\Delta x_{pref}}{2}\right) + k_y\left(y + \frac{\Delta y_{pref}}{2}\right) - \phi - \frac{\Delta\phi_{pref}}{2}\right]$$

$$\rho_R(x,y) = V\left(x - \frac{\Delta x_{pref}}{2}, y - \frac{\Delta y_{pref}}{2}\right)\cos\left[k_x\left(x - \frac{\Delta x_{pref}}{2}\right) + k_y\left(y - \frac{\Delta y_{pref}}{2}\right) - \phi + \frac{\Delta\phi_{pref}}{2}\right]$$

Equation 1

The output from each eye is the convolution of the retinal image I(x,y) with the receptive field ρ(x,y):

$$L = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy I_L(x,y) V\left(x + \frac{\Delta x_{pref}}{2}, y + \frac{\Delta y_{pref}}{2} \right) \cos\left(k_x x + k_y y - \phi + k_x \frac{\Delta x_{pref}}{2} - \frac{\Delta \phi_{pref}}{2} \right)$$

$$R = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy I_R(x,y) V\left(x - \frac{\Delta x_{pref}}{2}, y + \frac{\Delta y_{pref}}{2} \right) \cos\left(k_x x + k_y y - \phi - k_x \frac{\Delta x_{pref}}{2} + \frac{\Delta \phi_{pref}}{2} \right)$$

Equation 2

For a stimulus with uniform position disparity and no phase disparity, the left and right images may be regarded as shifted versions of a "cyclopean" image *I*. Allowing both horizontal and vertical stimulus disparity, we write:

$$I_L(x,y) = I\left(x + \Delta x_{stim}/2, y + \Delta y_{stim}/2\right), \qquad I_R(x,y) = I\left(x - \Delta x_{stim}/2, y - \Delta y_{stim}/2\right)$$

Equation 3

Then, we can rewrite Equation 2 as

$$L = F\left(\Delta x_{pref} - \Delta x_{stim}, \Delta y_{pref} - \Delta y_{stim}, \Delta \phi_{pref}\right)$$

$$R = F\left(-\Delta x_{pref} + \Delta x_{stim}, -\Delta y_{pref} + \Delta y_{stim}, -\Delta \phi_{pref}\right)$$

where the function F is

$$F\left(\Delta x_{pref} - \Delta x_{stim}, \Delta y_{pref} - \Delta y_{stim}, \Delta \phi_{pref}\right) = \iint dx dy V\left(x + \frac{\Delta x_{pref} - \Delta x_{stim}}{2}, y + \frac{\Delta y_{pref} - \Delta y_{stim}}{2} \right)$$

$$\times \cos\left(k_x \left(x + \frac{\Delta x_{pref} - \Delta x_{stim}}{2} \right) + k_y \left(y + \frac{\Delta y_{pref} - \Delta y_{stim}}{2} \right) - \phi - \frac{\Delta \phi_{pref}}{2} \right) I_L(x,y)$$

Equation 4

In the energy model, the output of a disparity-tuned simple cell is given by the squared sum of left and right inputs:

$$E = (L + R)^2 = L^2 + R^2 + 2LR\,.$$

Now consider differentiating L^2 with respect to Δx_{pref}, holding all other variables constant. We shall denote the result as

$$\frac{\partial L^2}{\partial \Delta x_{pref}} = 2F\left(\Delta x_{pref} - \Delta x_{stim}, \Delta y_{pref} - \Delta y_{stim}, \Delta \phi_{pref}\right) F_1'\left(\Delta x_{pref} - \Delta x_{stim}, \Delta y_{pref} - \Delta y_{stim}, \Delta \phi_{pref}\right).$$

where F_i', F_i'' represents the first, second partial derivative of F with respect to its i^{th} argument, with the other arguments held constant. For R^2, there is an additional minus sign:

$$\frac{\partial R^2}{\partial \Delta x_{pref}} = -2F\left(-\Delta x_{pref} + \Delta x_{stim}, -\Delta y_{pref} + \Delta y_{stim}, -\Delta \phi_{pref}\right) F_1'\left(-\Delta x_{pref} + \Delta x_{stim}, -\Delta y_{pref} + \Delta y_{stim}, -\Delta \phi_{pref}\right).$$

Now evaluate the these derivatives for the case where the cell's disparity tuning matches the stimulus: $\Delta x_{pref} = \Delta x_{stim}$, $\Delta y_{pref} = \Delta y_{stim}$ and $\Delta \phi_{pref} = 0$. Here, the first derivative of L^2 is $2F(0,0,0)F_1'(0,0,0)$, while the first derivative of R^2 is the same but with the opposite sign. Thus, the first derivative of $\left(L^2 + R^2\right)$ with respect to horizontal position disparity is zero. Exactly the same argument holds for the derivatives with

respect to vertical position disparity Δy_{pref} and phase disparity $\Delta\phi_{pref}$. Thus, the sum of these terms has a local extremum (not necessarily a maximum) at the correct stimulus disparity.

We now consider the remaining term 2LR:

$$LR = F\left(\Delta x_{pref} - \Delta x_{stim}, \Delta y_{pref} - \Delta y_{stim}, \Delta\phi_{pref}\right)F\left(-\Delta x_{pref} + \Delta x_{stim}, -\Delta y_{pref} + \Delta y_{stim}, -\Delta\phi_{pref}\right)$$

The first derivative of this with respect to preferred horizontal position disparity, evaluated at the correct stimulus disparity, is

$$\frac{\partial LR}{\partial \Delta x_{pref}} = F_1'\left(\Delta x_{pref} - \Delta x_{stim}, \Delta y_{pref} - \Delta y_{stim}, \Delta\phi_{pref}\right)F\left(-\Delta x_{pref} + \Delta x_{stim}, -\Delta y_{pref} + \Delta y_{stim}, -\Delta\phi_{pref}\right)$$
$$- F\left(\Delta x_{pref} - \Delta x_{stim}, \Delta y_{pref} - \Delta y_{stim}, \Delta\phi_{pref}\right)F_1'\left(-\Delta x_{pref} + \Delta x_{stim}, -\Delta y_{pref} + \Delta y_{stim}, -\Delta\phi_{pref}\right)$$

When these are evaluated at the stimulus disparity, $\Delta x_{pref} = \Delta x_{stim}$, $\Delta y_{pref} = \Delta y_{stim}$ and $\Delta\phi_{pref} = 0$, these terms cancel out, meaning that the binocular term LR also has an extremum at the stimulus disparity. Again, this argument works for differentiation with respect to Δy_{pref} and $\Delta\phi_{pref}$ as well. Thus, the total energy model response, $E = L^2 + R^2 + 2LR$, has a local extremum at the correct stimulus disparity, along all three dimensions of preferred disparity (horizontal position, vertical position and phase).

To examine the nature of this extremum, we need to calculate the second derivatives, again evaluated at the stimulus disparity. Doing this for the individual terms L^2, R^2 and 2LR, and then summing all three, we obtain the second derivative of the energy-model response E:

$$\left.\frac{\partial^2 E}{\partial \Delta x_{pref}{}^2}\right|_{pref=stim} = 8F(0,0,0)F_1''(0,0,0); \qquad \left.\frac{\partial^2 E}{\partial \Delta y_{pref}{}^2}\right|_{pref=stim} = 8F(0,0,0)F_2''(0,0,0);$$
$$\left.\frac{\partial^2 E}{\partial \Delta \phi_{pref}{}^2}\right|_{pref=stim} = 8F(0,0,0)F_3''(0,0,0)$$

Equation 5

Differentiating *F* (Equation 4) with respect to preferred phase disparity, we find that

$$F_3''(0,0,0) = -\frac{1}{4}\iint dxdy\, I_L(x,y)V(x,y)\cos\left(k_x x + k_y y - \phi\right) = -\frac{1}{4}F(0,0,0)$$

So, $\left.\frac{\partial^2 E}{\partial \Delta \phi_{pref}{}^2}\right|_{pref=stim} = -2\left[F(0,0,0)\right]^2$, which is always negative. This means that, as we vary preferred phase disparity for a fixed position disparity preference, the energy model response *E* has a local maximum at a phase disparity of zero (the phase disparity of the stimulus). However, no such neat result follows for the second derivatives with respect to position disparity. These can be either positive or negative. Therefore, the true stimulus disparity may be at a maximum *or a minimum* with respect to preferred position disparity.

SuppFig 2 shows a situation where the stimulus disparity is at a local minimum wrt preferred position disparity.

Note on how to search for stereo matches

Note that in order for theorems B and C to hold, it is essential that all cells be tuned to the same cyclopean location in the visual field. This was specified in Equation 1 by

shifting the receptive fields symmetrically in opposite directions: thus, the mean position of the two eyes' receptive fields is the same for all neurons in the population (SuppFig 3b). Previous workers have often varied disparity by fixing the receptive field in one eye and varying it in the other, so that cyclopean location varies along with disparity tuning (SuppFig 3a). This is the reason, for example, why Fig. 1 of Chen & Qian[22] at first sight appears to contradict the results proved here: in their plot of population response as a function of position disparity tuning, there is usually no extremum at the stimulus disparity. This led them to conclude that position-disparity encoding is intrinsically less accurate than phase disparity, a conclusion we reject (see Discussion). Of course, both approaches sketched in SuppFig 3 are in principle equally valid as a way of framing the correspondence problem. The brain contains neurons with receptive fields at all possible cyclopean positions and disparities (up to some maximum disparity), so it is merely a question of which neurons are considered as "belonging" together for the purposes of reading out the population. Chen & Qian chose to group neurons with a given left-eye receptive field, so that the choice of right-eye receptive field then determined both the neuron's disparity tuning and its cyclopean location (SuppFig 3a). This is responsible for the "noise" in their position disparity detectors. In contrast, we have chosen to look at a group of neurons with one particular cyclopean location, i.e. with monocular receptive fields on either side of this fixed central location (SuppFig 3b). This choice has significant practical advantages. With this grouping, pure position disparity detectors are *guaranteed* to have an extremum at the stimulus disparity. For pure phase disparity units, this is true only when the stimulus disparity is zero; as Chen & Qian found, for realistically broad-band receptive fields, pure phase disparity units become less and less reliable as stimulus disparity increases. Thus when the neuronal population is read out as we propose, position disparity units are actually slightly *more* reliable than phase disparity units, as demonstrated in Fig. 3 (green vs purple histograms). It might be argued that the simulation of Fig. 3 is not a fair test for phase-disparity units, since the stimulus disparity exceeds half a cycle. Broad-band phase-disparity units become less and less accurate – even counting as correct any estimate that is an integral number of cycles away from the stimulus disparity – the further the stimulus disparity is from zero. Chen & Qian envisaged phase-disparity detectors being used for fine disparity judgments once the stimulus disparity had been located to the nearest cycle. To address this, we repeated the simulation of Fig. 3 using a stimulus disparity of 0.03 cycles, and limiting the range of both position- and phase-disparity detectors to $\pm$ 0.5 cycles (not shown). A maximum-energy algorithm using pure phase-disparity detectors is now 44% accurate, while a maximum-energy algorithm using pure position-disparity detectors is 84% accurate.

Theorem D: Stereo energy is a sine function of phase disparity, with period 2π

In the paper, we also claimed that the response of a population of phase disparity detectors all tuned to the same position disparity is a sinusoidal function of their preferred phase disparity. Here, we give a formal proof of this. As before, the proof is for simple cells. Since the sum of two sinewaves with period 2π is also a sinewave with period 2π, the same result holds for complex cells.

Our starting-point is the expressions for L and R given in Equation 2. Notice that these are valid for *any* image-pairs I_L and I_R; the two eyes' images could be completely uncorrelated and these equations would still hold. We now consider the response of an energy-model linear simple cell, $E = L^2 + R^2 + 2LR$. Squaring Equation 2 and shifting the range of integration for convenience, and then using a trigonometric identity to replace the product of cosines with a sum of cosines, we obtain

$$L^2 = \frac{1}{2}\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} dy' I_L\left(x - \frac{\Delta x_{pref}}{2}, y - \frac{\Delta y_{pref}}{2}\right) I_L\left(x' - \frac{\Delta x_{pref}}{2}, y' - \frac{\Delta y_{pref}}{2}\right)$$
$$\times V(x,y)V(x',y')\left\{\cos\left(k_x(x+x') + k_y(y+y') - 2\phi\right) + \cos\left(k_x(x-x') + k_y(y-y') + \Delta\phi_{pref}\right)\right\}$$

Equation 6

The expression for R^2 is similar, except that I_L is replaced with I_R, and the sign of every disparity (Δx_{pref}, Δy_{pref} and $\Delta\phi_{pref}$) is inverted. Similarly, the binocular term $2LR$ is:

$$2LR = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} dy' I_L\left(x - \frac{\Delta x_{pref}}{2}, y - \frac{\Delta y_{pref}}{2}\right) I_R\left(x' + \frac{\Delta x_{pref}}{2}, y' + \frac{\Delta y_{pref}}{2}\right)$$
$$\times V(x,y)V(x',y')\left\{\cos\left(k_x(x'+x) + k_y(y'+y) - 2\phi\right) + \cos\left(k_x(x'-x) + k_y(y'-y) + \Delta\phi_{pref}\right)\right\}$$

Equation 7

Notice that, after the integration has been carried out, each of the expressions for L^2, R^2 and $2LR$ will contain one term which is independent of the cell's preferred phase disparity, and one term which varies as a sinusoidal function of phase disparity, with period 2π. The total energy model response, $E = L^2 + R^2 + 2LR$, is therefore also a sinusoidal function of phase disparity. While the response of a population of energy-model units as a function of position disparity for any fixed phase disparity is "noisy", full of arbitrary peaks and troughs which reflect the particular properties of the images, the response as a function of phase disparity for any fixed position disparity is smooth and predictable, completely specified by the baseline, amplitude and phase of the sinusoid. This is evident in the cross-sections shown in Fig. 2c, SuppFig 2c and SuppFig 4c.

Anti-correlated images

In the paper we briefly discussed the population response to anti-correlated images, when the contrast polarity in one eye has been inverted, equivalent to a stimulus phase disparity of 180°. SuppFig 4 provides a concrete example. All details are as in Fig. 2, except that the polarity of the right image has been inverted. Now, the response of cells tuned to the stimulus position disparity, blue dashed line in SuppFig 4c, is *enhanced* by phase disparity (since the stimulus itself has phase disparity). This means that the maximum occurs at a neuronal phase disparity of 180°. There are no local extrema along the line of zero phase disparity (blue line in SuppFig 4c) which are also local maxima with respect to phase disparity. For example, the extremum at the correct stimulus position disparity (cyan dot in SuppFig 4c) now falls at a local *minimum* with respect to phase disparity (SuppFig 4d).

OBTAINING DISPARITY MAPS

We tested our algorithm by calculating disparity maps, showing estimated disparity across the visual field for an example image. In this section, we record in detail how this was implemented. Rather than include a range of cells tuned to different overall phases ϕ, we calculated the population activity for complex cells, representing the sum of two binocular simple cells whose phases differ by 90° [1]. Thus, the complex cell's response was

$$C = (L_1 + R_1)^2 + (L_2 + R_2)^2 ,$$

where L_1,R_1 are given by Equation 2 with $\phi = 0$ and L_2,R_2 is given by Equation 2 with $\phi = \pi/2$. Similar results were obtained with simple cells. We also shift both receptive fields to different cyclopean positions ($x_{\text{pref}},y_{\text{pref}}$) in order to obtain maps of disparity across the whole image. Thus the left- and right-eye receptive fields are given by

$$\rho_L(x,y) = \rho\left(x - x_{pref} + \Delta x_{pref}/2,\ y - y_{pref};\ \phi + \Delta\phi_{pref}/2\right)$$
$$\rho_R(x,y) = \rho\left(x - x_{pref} - \Delta x_{pref}/2,\ y - y_{pref};\ \phi - \Delta\phi_{pref}/2\right)$$

where the function $\rho(x,y)$ is given in Equation 2 of the main paper.

We then apply our algorithm to obtain an estimate of the stimulus disparity at that point in space, from that particular channel: $\Delta x_{\text{est}}(x_{\text{pref}},y_{\text{pref}};f,\theta)$. For this, it is not necessary to simulate the full hybrid population shown in Fig. 2; we only calculated cross-sections through it, as illustrated in Fig. 4. We first evaluate the complex cell response (Equation 8) for a subpopulation of pure position disparity units, all with preferred phase disparity 0 and with preferred position disparity ranging between ± 0.2° (blue horizontal cross-section in Fig. 2ac and Fig. 4). We locate any extrema, Δx_{TP}, along this cross-section, both maxima and minima. If there are no extrema between ± 0.2°, we say that this channel returns no estimate (gray regions in Fig. 5d-i and SuppFig 5). Obviously, this is most likely to happen for low spatial frequencies, where the response varies slowly as a function of disparity tuning. For high-frequency channels, there will be many extrema within the range ± 0.2° (e.g. for the highest frequency, 16cpd, the allowed disparity range used for the Pentagon test image is more than 6 periods). The challenge is to identify which represents the true stimulus disparity.

To achieve this, for each extremum, we examine a subpopulation of hybrid position/phase disparity detectors, all with position disparity Δx_{TP} but with varying phase disparity. This corresponds to a vertical cross-section (green curve in Fig. 2cd and Fig. 4). We locate the phase disparity, $\Delta\phi_{\max}(\Delta x_{\text{TP}})$, of the maximally-responding cell in the subpopulation with position disparity Δx_{TP}. For a uniform-disparity stimulus, this $\Delta\phi_{\max}$ would be zero when Δx_{TP} is equal to the stimulus disparity. So, we estimate the stimulus disparity as being the value of Δx_{TP} for which $\Delta\phi_{\max}$ is closest to zero.

In this way, we obtained the single-channel estimates shown in SuppFig 5. The colorscale represents the estimated stimulus disparity; gray shows where no estimate was returned by this channel because no extremum was found within the allowed range. Note that the disparity maps are rather insensitive to the orientation of the receptive fields: essentially the same maps are produced for each of the 6 orientations, including 90° (horizontal). This refutes the idea, sometimes encountered, that horizontal disparity can be encoded only by vertically-oriented sensors; see section 3.3 of Read (2005), *Progress in Biophysics and Molecular Biology* **87**, 77-108 for a discussion.

Numerical details

Complex cell response was initially calculated for preferred phase disparity = 0 and for preferred position-disparities Δx_{pref} between fixed limits at intervals of either $\lambda/5$ or 0.1°, whichever is smallest, where λ is the spatial period of the channel. For the Pentagon image, the disparity range was ± 0.2°; for Sawtooth, ± 18 pixels, and for Venus and Tsukuba, –18 to +3 pixels.

These initial calculations provide coarse sampling of the horizontal cross-section shown in blue in Fig. 2. If these responses were monotonic as a function of Δx_{pref}, this channel returned no estimate of stimulus disparity (gray regions in SuppFig 5). Otherwise, the maxima and minima estimated from these coarsely-sampled responses were used as the starting-point for a search algorithm (Matlab's FMINSEARCH), which located the extrema with high accuracy. For each extremum Δx_{TP}, a second optimization then located the value of $\Delta\phi_{\text{pref}}$ which maximises response of cells tuned to that position disparity:

$$\Delta\phi_{\max}\left(\Delta x_{TP}\right) = \arg\max_{\Delta\phi_{pref}}\left(C\left(x_{pref}, y_{pref}, \Delta x_{TP}, \Delta\phi_{pref}, f, \theta\right)\right).$$

The estimated stimulus disparity is then taken as that of the extremum where $\Delta\phi_{\max}$ was closest to zero:

$$\Delta x_{est}\left(x_{pref}, y_{pref}; f, \theta\right) = \arg\min\left(\left|\Delta\phi_{\max}\left(\Delta x_{TP}\right)\right|\right).$$

This algorithm requires calculating the convolutions L_i,R_i many, many times, for different values of position and phase disparity Δx_{pref}, $\Delta\phi_{\text{pref}}$. We employed various short-cuts to speed this up. First, according to Parseval's theorem the same results are obtained by evaluating the convolutions in Fourier space as in physical space. We therefore employed whichever could be made fastest for the channel in question. For high-frequency channels, the receptive fields were small (Fig. 5, SuppFig 5), so we cut down the image to a rectangular strip centred on the particular cyclopean position under consideration. This strip extended 5 RF standard deviations σ vertically, and ($0.4° + 5\sigma$) horizontally to allow for the different horizontal position-disparities. Thus, for each RF that would need to be evaluated, the cut-down image extended at least 5σ from the RF center, ensuring that we included all points where the receptive field envelope exceed 4×10^{-6} of its maximum value. This small image was then passed to the routine that evaluated $\Delta x_{\text{est}}(x_{\text{pref}}, y_{\text{pref}}; f, \theta)$ for that cyclopean position and channel, and used to evaluate the convolutions in physical space.

For low-frequency channels, the receptive fields are large, so most of the original image falls within 5σ of an RF centre. However, its Fourier power spectrum is weighted very heavily towards the lowest frequencies present. Therefore, we can evaluate these convolutions most quickly in Fourier space, cutting down the Fourier spectrum to include only frequencies where the power was > 4×10^{-6} of maximum. This cut-down power spectrum was passed to the algorithm along with the Fourier spectrum of the images, similarly cut-down. The phase of the RF spectrum was then adjusted internally to produce Fourier spectra corresponding to RFs with different retinal locations and phases, and used to evaluate the convolutions in Fourier space.

SLANTED SURFACES

Our algorithm is guaranteed to work only where the disparity is uniform across the receptive field. To investigate how it performs when this condition is violated, we first tested it with noise images portraying slanted surfaces. SuppFig 6 shows the distribution of disparities returned by our algorithm for 100 different white-noise images portraying slanted surfaces. In each case, the right image had uniform power at all spatial frequencies up to some maximum; the Fourier spectrum of the left image was generated by horizontally compressing the Fourier spectrum of the right image by 2% (i.e. horizontal frequencies are multiplied by 0.98, SuppFig 6a) or 15% (SuppFig

6b). Each pair of noise images was fed into our algorithm to obtain an estimate of the disparity at the centre of the surface (true disparity = 0°). For each stimulus, we obtained an estimate from 6 channels (2 spatial frequencies and 3 orientations). The distributions in SuppFig 6 show all these single-channel estimates pooled together. Where the compression was 2%, corresponding to a surface slant of 9° viewed at 50cm, the single-channel estimate was correct 90% of the time. Where the compression was extreme, 15%, the single-channel estimate was correct only around half the time.

Errors made by different channels were largely uncorrelated. After removing correct answers (i.e. those inside the central bin), the pairwise correlation between the erroneous answers returned by different channels was on average 0.04. Thus, considerable improvements in accuracy are possible by combining the outputs of several channels. If we construct a robust average from the 6 channels in this simulation, by removing the 3 most outlying estimates for each stimulus, then the correct disparity is obtained in every single image with a compression factor of 98%, and in 86% of images with a compression factor of 85%.