

Sensors for impossible stimuli may solve the stereo correspondence problem

Jenny C A Read¹ & Bruce G Cumming²

One of the fundamental challenges of binocular vision is that objects project to different positions on the two retinas (binocular disparity). Neurons in visual cortex show two distinct types of tuning to disparity, position and phase disparity, which are the results of differences in receptive field location and profile, respectively. Here, we point out that phase disparity does not occur in natural images. Why, then, should the brain encode it? We propose that phase-disparity detectors help to work out which feature in the left eye corresponds to a given feature in the right. This correspondence problem is plagued by false matches: regions of the image that look similar, but do not correspond to the same object. We show that phase-disparity neurons tend to be more strongly activated by false matches. Thus, they may act as ‘lie detectors’, enabling the true correspondence to be deduced by a process of elimination.

Over the past 35 years, neurophysiologists have mapped the response properties of binocular neurons in primary visual cortex and elsewhere in considerable detail. A mathematical model, the stereo energy model¹, has been developed that successfully describes many of their properties. In this model, binocular neurons can encode disparity in two basic ways: phase disparity (Fig. 1a), in which receptive fields differ in the arrangement of their ON and OFF regions, but not in retinal position, and position disparity (Fig. 1b), in which left and right-eye receptive fields differ in their position on the retina, but not in their profile². Several recent studies in V1^{3–6} have concluded that most disparity-selective cells are hybrid⁷, showing both position and phase disparity. These neurophysiological data present a challenge to computational models of stereopsis. Why does the brain devote computational resources to encoding disparity twice over, once through position and once through phase?

Phase disparity presents a particular puzzle, as it does not correspond to anything that is experienced in natural viewing. For a surface such as a wall in front of the observer, where disparity is locally uniform, the two eyes' images of a given patch on the surface are related by a simple position shift on the retina (Fig. 1b). For an inclined surface, with a linear disparity gradient, the two image patches are also compressed and/or rotated with respect to one another; that is, they differ in spatial frequency and/or orientation (Fig. 1c). Higher-order changes in disparity, such as those produced by curved surfaces, produce images whose spatial frequency and orientation differences vary across the retina (Fig. 1d)⁸. Disparity discontinuities, which occur at object boundaries, produce different disparities in different regions of the retina⁹. However, phase-disparity neurons do not appear to be constructed to detect any of these possible situations. They respond optimally to stimuli in which the left and right eye's image are related by a constant shift in Fourier phase; that is, each Fourier component is displaced by an amount proportional

to its spatial period. Physically, such a stimulus would correspond to a set of transparent luminance gratings whose distance from the observer is a function of their spatial frequency (Fig. 1a), a situation that never occurs naturally, and that we therefore characterize as ‘impossible’, even though it can be simulated in the laboratory.

Recent physiological experiments support this conclusion. When presented with various possible disparity patterns, V1 neurons prefer stimuli with uniform disparity¹⁰, in contrast to higher visual areas, where neurons are found that respond optimally to depth discontinuities (V2^{9,11}), disparity gradients (V4¹², MT¹³, IT¹⁴, IP¹⁵) and disparity curvature (IT¹⁴). Psychophysical data supports the interpretation that disparity is initially encoded as a set of piece-wise fronto-parallel patches, which are then combined in higher brain areas to generate tuning for more complicated surfaces^{10,16,17}. However, when V1 neurons are probed with impossible stimuli that are designed to be optimal for phase-disparity detectors, many of them respond better to these than to any naturally occurring pattern of disparity¹⁸. Together, these results suggest that apparent tuning to phase disparity is not an artifact of a preference for a physically possible, but nonuniform, disparity; rather, the phase-disparity detectors found in V1 are genuinely tuned to impossible stimuli.

This raises the conundrum of why the brain has apparently built detectors for stimuli that are never encountered. We present one possible answer, by demonstrating that phase disparity detectors could potentially make a unique contribution to solving the stereo correspondence problem. Here, the major challenge is identifying the correct stereo correspondence amid a multitude of false matches. Matching regions of a real image contain no phase disparity. They will therefore preferentially activate pure position-disparity sensors. However, false matches are under no such constraints; they will have neither pure position disparity nor pure phase disparity, and

¹Institute of Neuroscience, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK. ²Laboratory of Sensorimotor Research, National Eye Institute, 49 Convent Drive, US National Institutes of Health, Bethesda, Maryland 20892-4435, USA. Correspondence should be addressed to J.C.A.R. (j.c.a.read@ncl.ac.uk).

Received 21 June; accepted 23 July; published online 9 September 2007; doi:10.1038/nn1951

may be best approximated by a mixture of both. Thus, it is quite possible for them to preferentially activate hybrid neurons with a mixture of position and phase disparity. Consequently, the preferential activation of these neurons is a signature of a false match. Neurons with phase disparity, precisely because they are tuned to impossible stimuli, enable false matches to be detected and rejected.

We prove mathematically that, for uniform-disparity stimuli, this method is guaranteed to find the correct disparity, even within a single spatial-frequency and orientation channel, and even when the disparity is larger than the period of the channel. This is a notable success, given that previous neuronal correspondence algorithms have had to compare information across channels to overcome the false-match problem in this situation, even in uniform-disparity stereograms^{7,19–22}. Of course, real images contain multiple disparities, so the proof no longer holds. In such images, our method may give conflicting results in different spatial frequency and orientation channels. However, we find that a simple robust average of the results from different channels produces good disparity maps, with no need for any interaction between channels. Thus, this method is also useful in complex natural images where its success is not mathematically certain. We suggest that the brain may similarly use phase-disparity neurons to eliminate false matches in its initial piecewise frontoparallel encoding of depth structure. This provides, for the first time, a clear computational rationale for the existence of both phase-disparity and position-disparity coding in early visual cortex.

RESULTS

Neither position nor phase detectors can recover disparity

Figure 2 illustrates the responses of model neurons in a simple situation where an observer stands in front of a wall with stucco texture (as shown in **Fig. 2b**). This visual scene, a single frontoparallel surface, has no depth structure at all (the stucco texture is flat). Yet even in this simplest of cases, identifying the wall's disparity from a population of neurons that have solely position disparity is by no means straightforward. The receptive fields of these cells have identical profiles in the two eyes, and differ only in their position on the retina (pure position disparity, that is, zero phase disparity, as in **Fig. 1b**). We plotted the response of a simulated population of these neurons (**Fig. 2a**). Each neuron's firing rate is calculated according to the stereo energy model¹ (Methods, equation 1) and plotted as a function of its position disparity. Using the modern version^{6,23–25} of the terminology introduced previously²⁶, we shall refer to these pure position-disparity cells as tuned-excitatory cells (**Supplementary Fig. 1** online). In this modern usage, these terms describe only the shape of the disparity tuning curve, not the preferred disparity. Therefore, a tuned-excitatory cell is defined for us by having a symmetrical tuning curve with a central peak, irrespective of whether this preferred disparity is crossed, uncrossed or zero.

How can we deduce the stimulus disparity from the response of this population? Perhaps the simplest approach is to find the preferred disparity of the maximally responding tuned-excitatory cell and to take this as an estimate of the stimulus disparity^{7,27}. However, this maximum-energy algorithm is not guaranteed to give the right answer. Its problem is that the stereo energy computed by these model cells depends not just on the correlation between the left and right images, but also on the contrast within each receptive field. Thus, mismatched image patches that have high contrast can easily have more stereo energy than corresponding patches that happen to have low contrast. In experiments, the response of the population is often averaged over many images with the same disparity structure and with purely random contrast structure, such as random-dot patterns²⁸. This means that the effect of contrast will average out, and that the cell with the maximum

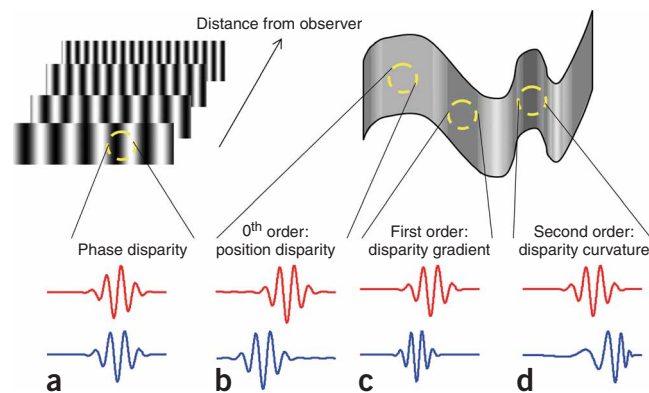
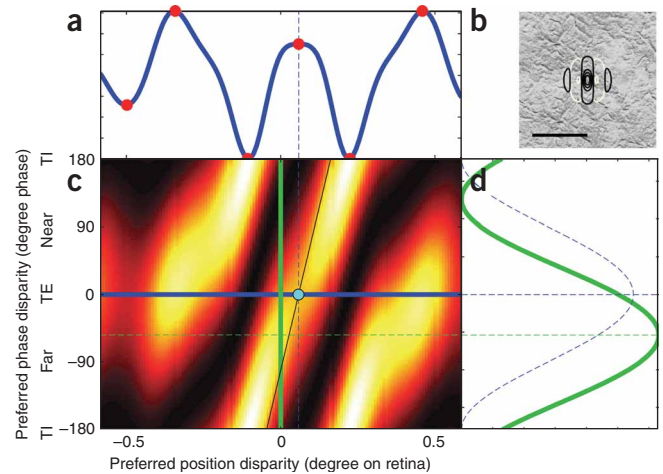


Figure 1 Different types of disparity. **(a)** Disparity stimulus that optimally stimulates a neuron with phase disparity (example receptive fields shown below, odd-symmetric receptive field in left eye and even-symmetric in the right eye). This is a set of luminance gratings (shown as opaque, but actually transparent), at varying distances that depend on their spatial frequency. **(b–d)** Varying-disparity surface, showing different types of physically possible disparity, and the relationship between the two images of a little patch of narrow-band contrast on the surface (red, left eye; blue, right eye). These can also be thought of as the two receptive fields of a binocular neuron that is optimally tuned to the disparity of the surface at each position. 0th-order disparity, regions of uniform disparity, is shown in **b**; optimal detectors have pure position disparity. First-order disparity, regions where disparity is varying linearly, is shown in **c**; optimal detectors have both a position shift and a spatial frequency difference between the eyes (one receptive field compressed relative to the other). Second-order disparity, disparity curvature, is shown in **d**. Receptive fields are related by a position offset and compression that varies across the receptive field. The other major situation that occurs in natural viewing is disparity discontinuities at object boundaries. This is not shown here because it cannot be detected by a single pair of receptive fields.

average response will indeed be that tuned to the stimulus disparity. However, in real life the brain does not have this luxury. It has to make a judgment about single stereo images, and here a maximum-energy algorithm is likely to fail^{7,27}. For example, the false matches at -0.35° and $+0.46^\circ$ both elicited larger responses than the true match at 0.06° (**Fig. 2a**). Indeed, we point out here that the stimulus disparity may not be at even a local maximum. If the contrast at the cyclopean location happens to be particularly low, then the stimulus disparity can actually fall at a local minimum of the population response (**Supplementary Fig. 2** online). This occurs if the decrease in binocular correlation produced by moving the receptive fields onto noncorresponding regions of the image is outweighed by a fortuitous increase in contrast energy. Thus, all we can deduce is that the stimulus disparity must be one of the values where the response of tuned-excitatory cells has a local turning point (**Fig. 2a**, marked in red). This still left us with six possible matches even in this restricted range (3 cycles of the cell's spatial period). Ignoring these problems and simply picking the tuned-excitatory cell with the largest response led to the maximum-energy algorithm finding the correct answer only 29% of the time (**Fig. 3**).

Physiologically inspired stereo correspondence algorithms have more commonly used pure phase-disparity detectors^{21,27,29–31}. These algorithms consider a population of model neurons whose receptive field envelopes are centered on the same position in both retinas, but that differ in the pattern of their ON and OFF regions (**Fig. 1a**). The response of a simulated population of pure phase-disparity detectors as a function of their preferred phase disparity is a sinusoid (**Fig. 2d**). This is a general property of a population of phase-disparity detectors tuned to a given position disparity, here zero (**Supplementary Note** online,

Figure 2 Response of a neuronal population to a broadband image with uniform disparity of 0.06° . (a–d) All neurons are tuned to vertical orientations and a spatial frequency $f = 4.8$ cycles per degree; the bandwidth is ~ 1.5 octaves. Example receptive field profile is shown in **b** (black contour lines, ON regions; white, OFF regions; scale bar is 0.5°). Receptive fields differ in their phase and in their position on the retina, but all cells in this population have the same cyclopean location, that is, the same mean position of their binocular receptive fields (**Supplementary Fig. 3**). The curves in **a** and **d** represent cross-sections through the response surface (**c**); that is, activity of two neuronal subpopulations, one with pure position disparity (**a**), and one with pure phase disparity (**d**). Each pixel in **c** represents one neuron. The pixel's horizontal and vertical location indicates the neuron's preferred position and phase disparity, respectively. The pixel color indicates the neuron's firing rate in response to this image. This was calculated according to the stereo energy model¹, equation (1). Neurons with zero phase disparity (blue line) are called tuned-excitatory (TE)-type, neurons with $\pm 180^\circ$ of phase disparity are tuned-inhibitory (TI), and neurons with $+90^\circ$ phase disparity are 'near' and those with -90° phase disparity are 'far'. The cyan dot in **c** marks the stimulus disparity. The dashed blue curve in **d** shows the response of the subpopulation where all neurons are tuned to the position disparity of the stimulus, but have varying phase disparity. The sloping black line in **c** shows the linear relationship between position and phase disparity for sine functions: $\Delta x = \Delta\phi/(2\pi f)$. For sufficiently narrow-band neurons, the yellow diagonal stripes of high neuronal response would all be parallel to the black line, and the stimulus disparity, modulo one period, could be read off from the maximally responding pure phase-disparity neuron. However, reading along this line on the present plot, we see that the stimulus disparity of 0.06° (cyan dot in **c**) corresponds to a phase disparity of -101° (0.28 cycles), yet the maximally responding pure phase-disparity neuron is tuned to just -54° (0.15 cycles, green dashed line).



Theorem D). For sufficiently narrow-band cells, the stimulus disparity, modulo the preferred spatial period of the cell, can be read off from the peak of this sinusoid. If the sinusoid peaks for cells tuned to a phase disparity of $\Delta\phi_{\text{pref}}$, then the stimulus disparity is $\lambda\Delta\phi_{\text{pref}}/2\pi \pm n\lambda$, where λ is the period of the spatial-frequency channel under consideration, and n is any integer. Thus, a narrow-band population can correctly identify stimulus disparities up to a half-cycle limit²⁷. Combining information from different spatial-frequency channels, in a coarse-to-fine scheme^{22,32}, could expand this range, at least in uniform-disparity stimuli where the disparity sampled by the large, low-frequency detectors is the same as that experienced by the smaller, high-frequency detectors. However, psychophysical evidence has yielded limited support for the idea of a coarse-to-fine hierarchy^{33,34}. Pure phase-disparity detectors cannot explain how humans are able to perceive disparity in narrow-band stimuli well above the half-cycle limit^{35,36}.

Furthermore, pure phase-disparity detectors perform reliably only if they are sufficiently narrow band. Previous theoretical work has concentrated on this mathematically tractable case. For realistic V1 bandwidths, pure phase-disparity detectors failed even for uniform-disparity stimuli within the half-cycle limit (**Fig. 2**). The model neurons in our simulations have a spatial frequency bandwidth of 1.5 octaves, which was the mean value for a sample of 180 V1 neurons in our previous physiological experiments^{23,37}, in agreement with previous estimates³⁸. The stimulus has a uniform disparity of 0.06° , which is 0.28 of a cycle (101° phase). Yet the most active phase-disparity detector was tuned to only 0.15 cycles (54° phase, green curve in **Fig. 2d**). Thus, the pure phase-disparity population would give the wrong answer in this case. This failure is common (purple histogram, **Fig. 3**). Here, the maximum-energy algorithm was tested on pure phase-disparity cells in a uniform-disparity stimulus whose disparity (0.42°) lay outside the half-cycle limit ($\pm 0.25^\circ$). Obviously, phase-disparity detectors cannot signal the true disparity in this case. However, they also failed to correctly detect the correct disparity even modulo their spatial period. In this case, the stimulus disparity minus one spatial period was -0.08° (short black arrow). The maximum-energy algorithm found this value on less than one-quarter of images. Thus, with a realistic spatial frequency bandwidth, being even 1 cycle away from the stimulus disparity has a catastrophic effect on the accuracy of pure phase-disparity detectors.

Position and phase detectors together can recover disparity

We have seen, then, that neither pure position-disparity (**Fig. 2a**) nor pure phase-disparity detectors (**Fig. 2d**) can reliably signal the correct disparity, even in the simplest possible case where the stimulus contains only one disparity. However, the brain also contains hybrid position/phase-disparity sensors. We measured the responses of hybrid energy-model neurons with all possible combinations of position and phase disparity (up to 0.6° position disparity) (**Fig. 2c**). Simply extending the maximum-energy algorithm to this full population resulted in little improvement (blue histogram, **Fig. 3**). However, there is a guaranteed means of identifying the correct match in uniform-disparity stimuli from the response of this population.

Real stimuli do not contain phase disparity (**Fig. 1**); that is, if a detector is already tuned to the correct position disparity, then its response can only be reduced by any tuning to nonzero phase disparity (dashed blue curve, **Fig. 2d**). Thus, for the subpopulation of hybrid sensors whose position disparity matches the stimulus, the maximum response will be in the sensor with zero phase disparity. However, for a subpopulation whose position disparity corresponds to a false match, the response is far more likely to be maximum at a nonzero phase disparity (for example, green curve in **Fig. 2d**). Expressed formally, this means that the true match is distinguished by (i) being at zero phase disparity, (ii) by being at a local extremum (maximum or minimum) with respect to position disparity (Theorem B in the **Supplementary Note**), and (iii) by being at a local maximum with respect to phase disparity (Theorem C in the **Supplementary Note**). The first condition says that the true match will lie on the zero phase disparity curve (blue line, **Fig. 2a,c** and **Fig. 4**). The second condition states that the true match will be one of the extrema on this curve (dots in **Fig. 2a** and **Fig. 4**). The final condition states that the true match will be the extremum marked with the cyan dot, as this is the only one located at a local maximum with respect to phase disparity. False matches that satisfy the first and second conditions may occur, but are very unlikely to satisfy the third condition as well. In other words, phase disparity sensors, precisely because they are tuned to impossible patterns of disparity, can identify false matches. These three conditions are guaranteed to hold in any stereo image where the disparity is uniform across the visual field (see **Supplementary Note** for mathematical

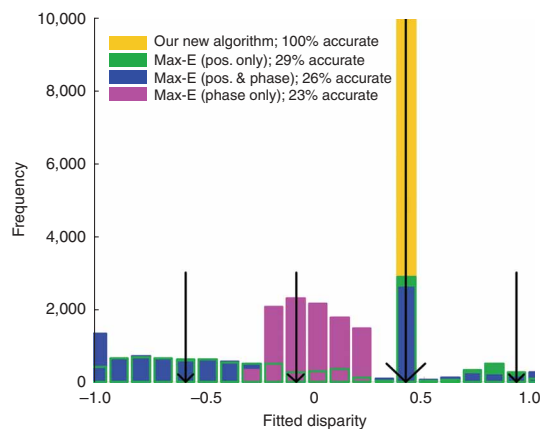


Figure 3 Comparison of our algorithm with four possible implementations of a maximum-energy algorithm, tested on a uniform-disparity noise stimulus. The histogram summarizes results for 10,000 noise images with a disparity of 0.42° , marked by the large arrow. All results are for the same channel (spatial frequency, $f = 2$ cycles per degree, orientation is vertical, bandwidth is 1.5 octaves). The shorter arrows mark disparities that are integer multiples of the spatial period away from the correct disparity. Orange, our algorithm, using both position- and phase-disparity detectors, always returned the correct disparity. Blue, evaluating the response of the full population, including cells with both position and phase disparity, and taking the stimulus disparity to be the preferred disparity of the maximally-responding cell (that is, $\Delta x_{\text{pref}} - \Delta \phi_{\text{pref}} / (2\pi f)$), gave the correct answer only 26% of the time. Green, as for blue, except considering only pure position-disparity detectors (TE cells, $\Delta \phi_{\text{pref}} = 0$); performance is similar. Purple, as for blue, except considering only pure phase-disparity detectors ($\Delta x_{\text{pref}} = 0$). This can only return answers within half a cycle of zero (that is, $\pm 0.25^\circ$). '% accurate' refers to the proportion of results that lie within the bin centered on the true disparity of 0.42° , except for the phase-disparity case (purple), where bins differing from this by integer multiples of the period 0.5° (marked with small arrows) were also considered to be correct.

proofs). This enables the correct disparity to be uniquely identified, even in a single spatial-frequency channel where the stimulus disparity is many cycles of the channel frequency.

Anticorrelated images

Anticorrelated images are those for which the contrast in one eye's image is inverted. Dense anticorrelated stimuli provide a multitude of false matches, but no overall depth percept; they have an unsettling, shimmering appearance^{39–42}. Most existing stereo algorithms pick one false match from each channel. The fact that no depth is perceived in these stimuli is then explained as being due to cross-channel conflict, because a different false match is returned from each channel¹⁹. Our findings on the use of phase disparity suggest an additional possibility. Anticorrelation corresponds to a phase disparity of 180° . Thus, out of the subpopulation of detectors tuned to the stimulus position disparity, the ones responding most will be tuned to a phase disparity of 180° (Supplementary Fig. 3 online). An algorithm that is looking for a subpopulation where the peak falls at 0° will therefore ignore this subpopulation. Indeed, there will, in general, be no point on the zero phase-disparity line that is both a local extremum with respect to position disparity, and a local maximum with respect to phase disparity. Thus, even in a single channel, an anticorrelated stimulus visibly fails to conform to the expected pattern of population activity for a uniform-disparity stimulus. Note that this is only true when the whole population of hybrid position- and phase-disparity sensors is considered; if we consider only a subpopulation of pure position- or pure phase-disparity detectors, there is nothing in any one channel to

indicate that the stimulus is unusual. We suggest that this violation of the expected pattern across the full population, as well as cross-channel conflict, may contribute to the lack of a depth percept and the distinctive appearance of this stimulus.

Performance on more complex depth structures

We implemented the above ideas in a computer algorithm (Fig. 4). This algorithm is guaranteed to return the correct disparity in a uniform-disparity stimulus (Fig. 3). However, real visual scenes seldom have the convenient property of containing only a single disparity. How will our algorithm fare when this requirement is not met? As an example, we tested the algorithm on slanted surfaces (Supplementary Fig. 4 online). The disparity at the center of the receptive field was 0° , but 1° to the left or to the right it was $\pm 0.02^\circ$ or $\pm 0.16^\circ$. This latter case is a very extreme disparity gradient. When viewed up close at 30 cm, it corresponds to a surface slanted at nearly 40° to the frontoparallel, and at larger viewing distances the slant becomes even more extreme. Nevertheless, even here the algorithm performs well. On any single image, a single channel has about a 50% chance of returning the correct disparity, and a 50% chance of returning essentially a random value. Notably, the errors made by different channels are virtually uncorrelated. Any given channel may be misled by a particular feature of the image and pick the wrong disparity. But other channels do not see that feature and are not misled. This means that by combining the outputs from a few different channels, the true disparity can be reliably recovered. For example, carrying out a robust average over just six channels increases the accuracy from 54% to 85%. With less extreme slants, even a single channel is reliable. For instance, when the surfaces are slanted 18° away from frontoparallel, viewed at a distance of 1 m, the algorithm performed with 90% accuracy even in a single channel, and with 100% accuracy when the outputs from six channels are robustly averaged.

We measured our algorithm on a test stereogram that is widely used in the computer vision literature (Fig. 5). Good (though noisy) disparity maps are produced from the output of a single spatial-frequency and orientation channel (Fig. 5d–h; all channels are shown in Supplementary Fig. 5 online). When the outputs of several spatial-frequency and orientation channels are averaged, even better results are

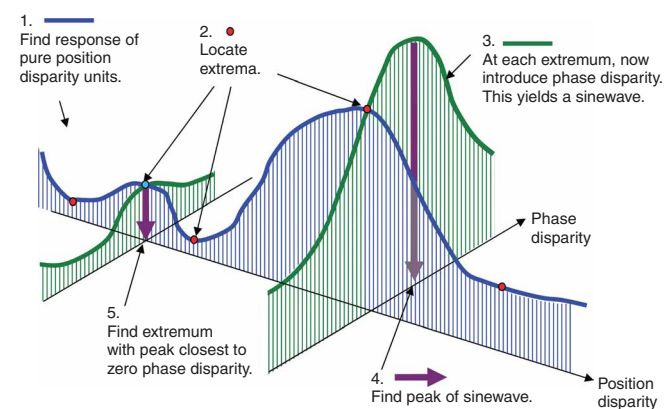


Figure 4 Sketch of the algorithm used to estimate stimulus disparity in a single channel. The blue and green curves represent horizontal and vertical cross-sections through the population response in Figure 2. Thus, the blue curve represents the response of a population of tuned-excitatory cells (with pure position disparity, no phase disparity), whereas the green curves represent the response of a population of hybrid cells, with varying phase disparity, but all having the same position disparity. Note that 'position disparity' and 'phase disparity' in this figure refer to the tuning preferences of the cells, not to properties of the stimulus.

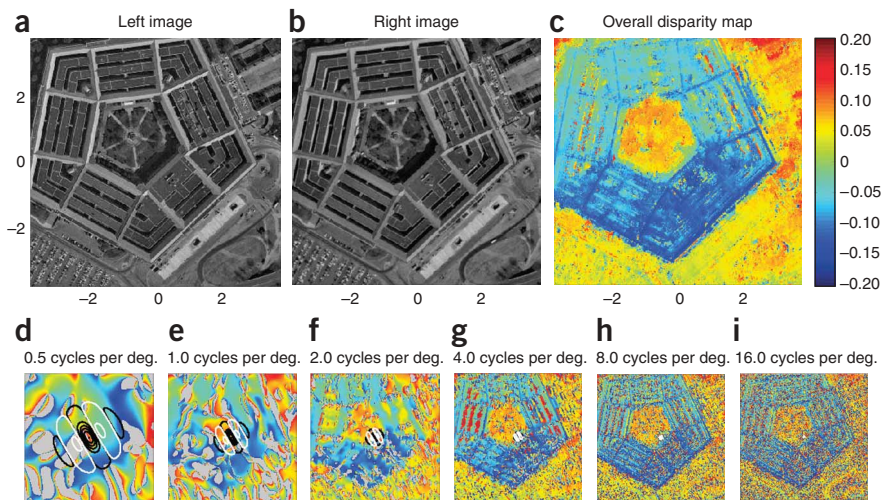


Figure 5 Applying our principle to a classic test stereogram. **(a,b)** Left and right images of the Pentagon. **(d–i)** Disparity maps obtained from single channels, with spatial frequencies indicated. Because orientation tuning made little difference to the disparity maps obtained, we show results for only one preferred orientation, 30° (**Supplementary Fig. 5** shows results for all six orientations). Contours show example receptive fields for each channel. Color scale for all panels is as given in **c**. Gray regions indicate that no candidate matches were encountered in the range examined. A robust average of disparity maps obtained by all 36 channels (6 spatial frequencies, 6 orientations) is shown in **c**.

obtained. We robustly averaged the outputs of 36 channels (6 orientation and 6 spatial frequency; **Fig. 5c**). Not only was the broad outline of the Pentagon itself reproduced, but also the details of its roof and even the trees in the courtyard.

This method allows good disparity maps to be obtained in this complex image, even from a single channel. Of course, the very lowest frequency channel (0.5 cycles per degree, **Fig. 5d**) is unable to report accurate values, as its large receptive fields span regions of different disparity. Errors introduced by this channel are kept to a minimum because our algorithm is capable of reporting failure; when no extrema were encountered in the disparity range examined, no disparity estimate was produced (gray pixels), and the channel did not contribute to the overall disparity estimate (**Fig. 5c**) for this cyclopean position. However, when the channel does return an estimate, it is usually at least roughly correct, so the vague outline of the pentagon emerges even here. For the highest frequency channels, the problem is

the exact opposite. Here, the receptive fields are small enough that they usually sample a disparity that remains at least roughly uniform. However, they are exposed to more false matches. One way of overcoming this well-known problem is to use the outputs of lower-frequency channels, in a coarse-to-fine hierarchy^{22,32}. In contrast, we show that by combining position- and phase-disparity sensors as we propose, quite reliable disparity maps can be extracted from a single channel, even where the disparity exceeds the half-cycle limit (**Fig. 5**).

We also tested our algorithm on three images from the Middlebury stereo repository⁴³ (<http://cat.middlebury.edu/stereo>). Because this database records the correct disparity for each image pair, this enables a quantitative evaluation of our algorithm. The stimulus disparity range was similar for each image, about 10 pixels. The algorithm searched over a much wider range, 30 pixels, greatly increasing the chance of false matches. Yet in every image, the algorithm succeeded in recovering good disparity maps. We used three quantitative measures of fit quality, R, B and M. R, the root-mean-square error, and B, the percentage of pixels where the disparity error is > 1 pixel, have been employed previously⁴³. M is the median absolute error, measured in pixels; in every case, this is less than half a pixel. Our algorithm considerably underperforms the best machine-vision stereo algorithms, for which B is typically an order of magnitude less. It is clear that the high percentage of ‘bad’ pixels arises predominantly not from high-frequency noise scattered over the disparity map, but from a blurring of the depth discontinuities. Edges that are straight and abrupt in the stimulus are reconstructed as wavy and gradual, whereas disparity over broad regions is generally correct. This is compatible with a notable feature of human stereopsis: poor spatial resolution for stereoscopic structure^{10,16}. Thus, although other machine algorithms produce more veridical depth maps, they probably outperform human stereopsis also.

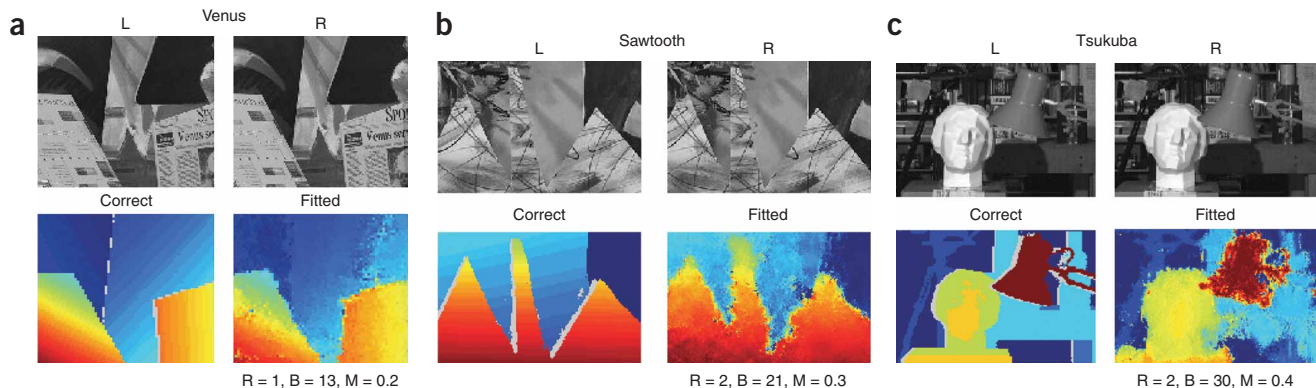


Figure 6 **(a–c)** Three test stereo-pairs from the Middlebury repository. In each panel, the top row shows left and right images over the region where disparity was evaluated (for speed, we did not evaluate disparity at every pixel in the image, although the full-size images were used as input to the algorithm). Bottom left, pseudocolor plot shows the ‘ground truth’ disparity given in the repository, with occluded regions grayed out. Bottom right, pseudocolor plot shows, on the same color scale, the disparity fitted by our algorithm (robust average over 6 spatial frequency \times 6 orientation channels). Below the plots are three quantitative measures of fit quality, R (RMS error in pixels), B (percentage of pixels with error > 1 pixel) and M (median absolute error in pixels). All three measures are evaluated over the entire fitted region, including occlusions.

DISCUSSION

Stereopsis is a perceptual module whose neuronal mechanisms are understood in considerable detail. The process begins in primary visual cortex, which contains sensors tuned for both position and phase disparity, and for combinations of both. It is currently unclear as to why the brain should encode disparity twice over in this way. Phase disparity is particularly puzzling, as it can encode disparities up to only half a cycle, with optimal detection at a quarter-cycle³⁵, whereas position disparity, and human perception, is subject to no such limit^{35,36}. Furthermore, we point out here that phase-disparity detectors are tuned to patterns of luminance that are not found in real stimuli. Of course, such luminance patterns may occasionally occur, for example, in regions of the image that are uncorrelated as a result of occlusions. The point is that, unlike position-disparity detectors, phase-disparity detectors do not reliably signal a particular physical disparity.

Why build detectors for phase disparity?

Why then does the brain contain both sorts of detector? One early suggestion⁴⁴ was that phase disparity might compensate for unwanted position disparities that are introduced by the finite precision of retinotopy. However, subsequent studies have failed to show the inverse correlation between position- and phase-disparity tuning that this would imply^{4,6}.

Only one previous study²² has suggested a theoretical reason as to why the brain might need both types of detector. This study²² suggested, on the basis of the simulated response to many different bar and random-dot patterns, that populations of pure phase-disparity detectors are more reliable than their pure position-disparity counterparts. The study found that for any given image the maximally responding phase-disparity unit was usually the one that was tuned to the stimulus disparity ($\Delta\phi_{\text{pref}} = 2\pi f \Delta x_{\text{stim}}$), whereas the position-disparity units showed much wider scatter in the location of the peak. Thus, it suggested that position disparity was used to overcome the half-cycle limitation, shifting disparity into the right range, whereas phase disparity was used to make a precise judgment. However, the study's observation that phase detectors give more reliable signals depends critically on the particular set of detectors that it happened to consider: a group of neurons that all had the same left-eye receptive field (see **Supplementary Note** and **Supplementary Fig. 6** online). Thus, in that population, cyclopean location covaried with position disparity, introducing additional noise into the disparity signal. When cyclopean location and position disparity are varied independently, as we did here (**Supplementary Fig. 6**), position-disparity detectors are actually more reliable than phase-disparity detectors over the same range (compare green and purple histograms in **Fig. 3**). Thus, nothing in the existing literature explains why the brain should deliberately construct detectors with phase disparity.

One possibility is that such detectors are not, in fact, constructed deliberately. It may simply be too hard to build neurons that are tuned strictly to real-world disparities, and so phase disparities represent a form of noise that the visual system copes with. Phase-disparity detectors do respond to the disparity in real images, even though they are not ideally suited to these, which is why stereo correspondence algorithms can be constructed out of phase-disparity detectors^{27,30,31,45,46}. Perhaps, therefore, there is not enough pressure to iron out accidental phase disparities that arise during development.

Here, we point out another possibility. Precisely because phase disparity does not occur in real-world images, neurons that have both position- and phase-disparity signal false matches, and so help to solve the correspondence problem. If stimulus disparity is uniform across the receptive field, then the neuron that is optimally tuned to the stimulus will be a pure position-disparity neuron (equivalently, a tuned-excitatory cell^{2,23,26}), with position disparity equal to that of

the stimulus and with zero phase disparity. Introducing phase disparity into this cell's receptive fields would make it less well tuned to the image and reduce its response. Thus, the tuned-excitatory cell will respond more strongly than its nontuned-excitatory neighbors with the same position disparity. This is the signature of the true match. False matches may elicit a strong response in particular tuned-excitatory cells, but the response will be even stronger when an appropriate phase disparity is added in. Thus, these hybrid position/phase-disparity sensors act as lie detectors. Their activity unmasks false matches.

Implementation in the brain

The algorithm used for our simulations (**Fig. 4** and **Supplementary Note**) is not physiologically realistic; it was designed for speed in a serial processor, not as a model of a massively parallel system like the brain. Nonetheless, the principle underlying our simulations could be implemented in visual cortex by a cooperative network in which lie detectors with hybrid position/phase disparity suppress activity in the corresponding tuned-excitatory cells. A winner-take-all competition between neurons with different phase disparity could ensure that the only surviving activity in tuned-excitatory cells is where zero phase disparity produced the best response. This would silence almost all tuned-excitatory cells except those located at the true match. Irrespective of the precise neuronal circuitry, our model predicts that activity in neurons with nonzero phase disparity should tend to reduce either the actual activity in the corresponding tuned-excitatory cells, or the downstream effect of that activity. One way in which this might be tested experimentally is to measure the dynamics of responses to stimuli that produce local peaks in a population of tuned-excitatory cells, but which are vetoed by phase-disparity detectors (as happens, according to our model, in anti-correlated stereograms). The effect of this veto should appear later in the response, so our model predicts that dynamic measures of responses to anti-correlated stereograms would show a characteristic evolution over time.

Conclusion

We report here a previously unknown way of identifying the stimulus disparity from a population of hybrid position- and phase-disparity neurons modeled on primary visual cortex. Where the stimulus disparity is uniform, this method gives the correct answer with 100% reliability, even in a single spatial-frequency/orientation channel, and even if the stimulus disparity is many cycles of the channel's spatial period. This success is notable given that most existing algorithms have to compare information from several channels to overcome aliasing, even for a uniform-disparity stimulus. For realistic stimuli with varying disparity, the method is not guaranteed, but simulations suggest that it is still very successful. This is the first theory to explain the existence of large numbers of visual cortical neurons that respond best to stimuli that never occur in natural viewing¹⁸. Our proposal also provides a new insight into why the modulation of V1 neurons for anti-correlated stimuli does not result in a depth percept. The theory was inspired by the observed properties of visual cortex, and can be implemented in a physiologically plausible manner. The idea of lie detectors specifically tuned to false matches may also turn out to be useful in machine stereo algorithms.

METHODS

The model neurons in this paper are constructed according to the stereo energy model described previously¹. We begin with binocular simple cells. These are characterized by a receptive field in each eye: $\rho_L(x,y)$ and $\rho_R(x,y)$. The output from each eye is given by the convolution of the retinal image with the receptive field

$$L = \int_{-\infty}^{+\infty} dx dy I_L(x,y) \rho_L(x,y),$$

and similarly for the right eye. The retinal images $I_L(x,y)$ and $I_R(x,y)$ are expressed relative to the mean luminance, so that positive values of I represent bright features and negative values represent dark features. The output of an energy-model binocular simple cell is

$$E = (L + R)^2 \quad (1)$$

(strictly, a physiological simple cell would be $E = |L + R|^2$; the unrectified expression, equation (1), would represent the sum of two physiological simple cells¹).

All simulations used Gabor receptive fields with an isotropic Gaussian envelope. The basic cyclopean receptive field profile is

$$\rho(x, y; \phi) = \cos(2\pi f x' - \phi) \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right), \quad (2)$$

where $x' = x \cos \theta + y \sin \theta$; $y' = y \cos \theta - x \sin \theta$ where f is the cell's preferred spatial frequency, θ is its preferred orientation and ϕ is its phase. The s.d., σ , is

$$\sigma = \frac{\sqrt{\ln 2}}{2\pi f} \times \frac{2^{1.5} + 1}{2^{1.5} - 1},$$

resulting in a full-width, half-maximum bandwidth of around 1.5 octaves¹⁹.

The left- and right-eye receptive fields are shifted in position and phase depending on the tuning of the neuron. For a neuron tuned to position disparity Δx_{pref} and phase disparity $\Delta \phi_{\text{pref}}$ the left- and right-eye receptive fields are

$$\begin{aligned} \rho_L(x, y) &= \rho(x + \Delta x_{\text{pref}}/2, y; \phi + \Delta \phi_{\text{pref}}/2); \\ \rho_R(x, y) &= \rho(x - \Delta x_{\text{pref}}/2, y; \phi - \Delta \phi_{\text{pref}}/2). \end{aligned}$$

Tuned-excitatory cells have $\Delta \phi_{\text{pref}} = 0$ by definition; tuned-inhibitory cells have $\Delta \phi_{\text{pref}} = \pi$.

Figure 2 and Supplementary Figures 2 and 3 show the responses of simple cells (equation (1)), tuned to a particular phase ϕ . To avoid dependence on any one phase, the results in Figures 5 and 6, and Supplementary Figures 4 and 5, were obtained using the response of phase-independent complex cells, given by the sum of two simple cells in quadrature¹. The mathematical results underlying our algorithm hold for both simple and complex cells (Supplementary Note).

We adopted a simple robust-averaging heuristic for combining the disparity maps produced by different spatial-frequency and orientation channels (Supplementary Fig. 5) into a single best-estimate disparity map (Fig. 5c). At each cyclopean position ($x_{\text{pref}}, y_{\text{pref}}$), we calculated the average disparity from all channels that returned an estimate of that position, $\langle \Delta x_{\text{est}}(x_{\text{pref}}, y_{\text{pref}}, f, \theta) \rangle_{f, \theta}$. Then we removed the channel whose estimate was furthest from the mean and calculated the mean of the remaining channels. We repeated this procedure until only half of the channels remained.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

J.C.A.R. performed all analyses and simulations and wrote the manuscript. B.G.C. supervised the project.

Published online at <http://www.nature.com/natureneuroscience>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- 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).
- DeAngelis, G.C., Ohzawa, I. & Freeman, R.D. Depth is encoded in the visual cortex by a specialised receptive field structure. *Nature* **352**, 156–159 (1991).
- Anzai, A., Ohzawa, I. & Freeman, R.D. Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proc. Natl. Acad. Sci. USA* **94**, 5438–5443 (1997).
- Anzai, A., Ohzawa, I. & Freeman, R.D. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J. Neurophysiol.* **82**, 874–890 (1999).
- Livingstone, M.S. & Tsao, D.Y. Receptive fields of disparity-selective neurons in macaque striate cortex. *Nat. Neurosci.* **2**, 825–832 (1999).
- Prince, S.J., Cumming, B.G. & Parker, A.J. Range and mechanism of encoding of horizontal disparity in macaque V1. *J. Neurophysiol.* **87**, 209–221 (2002).

- Fleet, D.J., Wagner, H. & Heeger, D.J. Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res.* **36**, 1839–1857 (1996).
- Orban, G.A., Janssen, P. & Vogels, R. Extracting 3D structure from disparity. *Trends Neurosci.* **29**, 466–473 (2006).
- Bredfeldt, C.E. & Cumming, B.G. A simple account of cyclopean edge responses in macaque V2. *J. Neurosci.* **26**, 7581–7596 (2006).
- Nienborg, H., Bridge, H., Parker, A.J. & Cumming, B.G. Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *J. Neurosci.* **24**, 2065–2076 (2004).
- von der Heydt, R., Zhou, H. & Friedman, H.S. Representation of stereoscopic edges in monkey visual cortex. *Vision Res.* **40**, 1955–1967 (2000).
- Hinkle, D.A. & Connor, C.E. Three-dimensional orientation tuning in macaque area V4. *Nat. Neurosci.* **5**, 665–670 (2002).
- Nguyenkim, J.D. & DeAngelis, G.C. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J. Neurosci.* **23**, 7117–7128 (2003).
- Janssen, P., Vogels, R. & Orban, G.A. Three-dimensional shape coding in inferior temporal cortex. *Neuron* **27**, 385–397 (2000).
- Taira, M., Tsutsui, K.I., Jiang, M., Yara, K. & Sakata, H. Parietal neurons represent surface orientation from the gradient of binocular disparity. *J. Neurophysiol.* **83**, 3140–3146 (2000).
- Tyler, C.W. Depth perception in disparity gratings. *Nature* **251**, 140–142 (1974).
- Prince, S.J., Eagle, R.A. & Rogers, B.J. Contrast masking reveals spatial-frequency channels in stereopsis. *Perception* **27**, 1345–1355 (1998).
- Haefner, R. & Cumming, B.G. Spatial nonlinearities in V1 disparity-selective neurons. *Soc. Neurosci. Abstr.* **583.9** (2005).
- Read, J.C.A. A Bayesian model of stereopsis depth and motion direction discrimination. *Biol. Cybern.* **86**, 117–136 (2002).
- Read, J.C.A. A Bayesian approach to the stereo correspondence problem. *Neural Comput.* **14**, 1371–1392 (2002).
- Tsai, J.J. & Victor, J.D. Reading a population code: a multi-scale neural model for representing binocular disparity. *Vision Res.* **43**, 445–466 (2003).
- Chen, Y. & Qian, N. A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural Comput.* **16**, 1545–1577 (2004).
- Read, J.C.A. & Cumming, B.G. Ocular dominance predicts neither strength nor class of disparity selectivity with random-dot stimuli in primate V1. *J. Neurophysiol.* **91**, 1271–1281 (2004).
- Cumming, B.G. & DeAngelis, G.C. The physiology of stereopsis. *Annu. Rev. Neurosci.* **24**, 203–238 (2001).
- Freeman, R.D. & Ohzawa, I. On the neurophysiological organisation of binocular vision. *Vision Res.* **30**, 1661–1676 (1990).
- Poggio, G.F. & Fischer, B. Binocular interaction and depth sensitivity of striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.* **40**, 1392–1405 (1977).
- Qian, N. Computing stereo disparity and motion with known binocular cell properties. *Neural Comput.* **6**, 390–404 (1994).
- Poggio, G.F., Motter, B.C., Squatrito, S. & Trotter, Y. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res.* **25**, 397–406 (1985).
- Qian, N. & Zhu, Y. Physiological computation of binocular disparity. *Vision Res.* **37**, 1811–1827 (1997).
- Qian, N. Binocular disparity and the perception of depth. *Neuron* **18**, 359–368 (1997).
- Sanger, T. Stereo disparity computation using Gabor filters. *Biol. Cybern.* **59**, 405–418 (1988).
- Marr, D. & Poggio, T. A computational theory of human stereo vision. *Proc. R. Soc. Lond. B* **204**, 301–328 (1979).
- Smallman, H.S. & MacLeod, D.I. Spatial scale interactions in stereo sensitivity and the neural representation of binocular disparity. *Perception* **26**, 977–994 (1997).
- Rohaly, A.M. & Wilson, H.R. Nature of coarse-to-fine constraints on binocular fusion. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **10**, 2433–2441 (1993).
- Prince, S.J. & Eagle, R.A. Size-disparity correlation in human binocular depth perception. *Proc. Biol. Soc.* **266**, 1361–1365 (1999).
- Prince, S.J.P. & Eagle, R.E. Weighted directional energy model of human stereo correspondence. *Vision Res.* **40**, 1143–1155 (2000).
- Read, J.C.A. & Cumming, B.G. Testing quantitative models of binocular disparity selectivity in primary visual cortex. *J. Neurophysiol.* **90**, 2795–2817 (2003).
- DeValois, R.L., Albrecht, D.G. & Thorell, L.G. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559 (1982).
- Cumming, B.G., Shapiro, S.E. & Parker, A.J. Disparity detection in anticorrelated stereograms. *Perception* **27**, 1367–1377 (1998).
- Cogan, A.I., Lomakin, A.J. & Rossi, A.F. Depth in anticorrelated stereograms: effects of spatial density and interocular delay. *Vision Res.* **33**, 1959–1975 (1993).
- Julesz, B. *Foundations of Cyclopean Perception* (University of Chicago Press, Chicago, 1971).
- Rogers, B. & Anstis, S. Reversed depth from positive and negative stereograms. *Perception* **4**, 193–201 (1975).
- Scharstein, D. & Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **47**, 7–42 (2002).
- Parker, A. Misaligned viewpoints. *Nature* **352**, 109 (1991).
- Jepson, A.D. & Jenkin, M.R.M. The fast computation of disparity from phase differences. in *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer Society Conference on.*, 389–403 (San Diego, California, USA, 1989).
- Fleet, D.J. Disparity from local weighted phase-correlation. in *Systems, Man, and Cybernetics, 1994. 'Humans, Information and Technology', 1994 IEEE International Conference on.* **1**, 48–54 (San Antonio, Texas, USA, 1994).