Optimal disparity estimation in natural stereo-images
(Supplement)
Johannes Burge & Wilson S. Geisler

Computing the log likelihood with populations of simple and complex cells
Simple cells in visual cortex produce only positive responses, whereas AMA filters produce both positive and negative responses. The linear AMA filter responses (see first term in Eq. 5) and their pairwise sums (see the sum in the third term’s parentheses) can be obtained from a population of simple cells. We define simple cells as units whose outputs result from linear filtering followed by half-wave rectification (i.e. thresholding at zero). Each AMA filter response in the first term can be obtained from two simple cells that are ‘on’ and ‘off’ versions of each AMA filter (Fig. S3a; see Fig. 3). Each pairwise-summed AMA filter response in the third term can be obtained from two simple cells that are ‘on’ and ‘off’ versions of each pairwise-summed filter (Fig. S3a, S4).

The squared AMA filter responses (see the second and third terms in Eq. 5) can be obtained from a population of complex cells. We define complex cells as units whose responses result from linear filtering followed by a squaring non-linearity, as is often the case in cortex. Complex cells can be implemented by summing and then squaring (or squaring and then summing) the outputs of ‘on’ and ‘off’ simple cells (Fig. S3b; the disparity tuning curves of these complex cells are given in Fig. S5).

Finally, neurons whose response rates represent the log likelihood of the AMA filter population responses can be constructed via a weighted sum of simple and complex cell population responses (Eq. 5, Fig. 6, Fig. 7). (Note that the likelihood can be obtained by exponentiating the log-likelihood responses.) In other words, a large collection of log-likelihood (LL) neurons, each with a different preferred disparity, can be constructed from a fixed set of simple and complex cells simply by changing the weights.

Derivation of optimal weights on simple and complex cells
The first term in Eq. 5 (i.e. the linear term) can be conceptualized as a weighted sum of simple cell population responses. The second and third terms (i.e. the quadratic terms) can be conceptualized as weighted sums of complex cell population responses. The constant corresponds to baseline response. The constant and the weights on the linear, squared, and sum-squared filters are given by

\[
\begin{align*}
    w_{i,k} &= C_k^{-1}u_k \\
    w_{i,k} &= -\text{diag}(C^{-1}) + 0.5C^{-1}1 \\
    w_{ij,k} &= -0.5C_{ij}^{-1}, \forall ij, j > i \\
    \text{const}_k &= -0.5u_{k}^{T}C_{k}^{-1}u_{k} + \text{const}_k
\end{align*}
\]

where \( I \) is the identity matrix, \( 1 \) is the ‘ones’ vector, and \( \text{diag}() \) sets a matrix diagonal to a vector. Here, we derive these weight equations based on the multi-dimensional Gaussian approximations to the empirically determined filter response distributions (i.e. conditional response distributions) (see Fig. 4a).
First, we rewrite Eq. 5 in the main text by expanding its third term

$$\ln p(R|\delta_k) = \sum_{i=1}^{n} w_{i,k} R_i + \sum_{i=1}^{n} w_{i,k} R_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij} \left( R_i^2 + 2 R_i R_j + R_j^2 \right) + \text{const}$$  (S2)

Second, we rewrite Eq. 4 in the main text

$$\ln p(R|\delta_k) = -0.5 (R - u_k)^T A_k (R - u_k) + \text{const}$$  (S3)

where $A_k$ equals the inverse covariance matrix $C_k^{-1}$. (Recall that the mean vector $u_k$ and covariance matrix $C_k$ are obtained by projecting a collection of natural stereo-images having disparity $\delta_k$ onto the filters.) From here forward, the disparity index $k$ is dropped for notational simplicity. Third, we set Eqs. S2 and S3 equal to one another, multiply through, and collect terms. The terms of the log likelihood (Eq. S3) are found to have the form

$$-0.5 a_{ii} \left( R_i - u_i \right)^2 = -0.5 \left( a_{ii} R_i^2 - a_{ii} 2 R_i u_i + a_{ii} u_i^2 \right)$$  (S4a)

$$-a_{ij} \left( R_i - u_i \right) \left( R_j - u_j \right) = -\left( a_{ij} R_i R_j - a_{ij} R_i u_j - a_{ij} R_j u_i + a_{ij} u_i u_j \right)$$  (S4b)

where $a_{ii}$ are the on-diagonal elements of $A$, $a_{ij}$ are the off-diagonal elements, and $u_i$ are the mean responses.

The aim is now to express $w_i, w_{ii}$, and $w_{ij}$ in terms of $u_i, a_{ii}$, and $a_{ij}$. First consider $w_i$. In this case, we want to collect terms containing $R_i$: $w_i R_i$ from Eq. S2, and $a_{ii} R_i u_i$ and $a_{ij} R_j u_j$ from Eq. S4a. Setting the $R_i$ terms equal and canceling gives

$$w_i = \sum_{j=1}^{n} a_{ij} u_j$$  (S5a)

Note that the weights on the linear filter responses $w_i$ are zero if the mean linear filter responses $u_i$ are zero. In our case, the filter response distributions means (i.e. the means of the conditional response distributions) are all near zero (see Fig. 4a, Fig. S2). Thus, the weights on the simple cell responses (Fig. 6a) contribute little to the units (LL neurons) that respond according to the log likelihood of the linear filter population response (see Discussion).

Next, consider $w_{ij}, i \neq j$. In this case, we want to collect terms containing $R_i R_j$: $2 w_{ij} R_i R_j$ appears in Eq. S2 and $-a_{ij} R_i R_j$ appears in Eq. S4b. Setting the two $R_i R_j$ terms equal to each other and canceling gives

$$w_{ij} = -0.5 a_{ij}$$  (S5b)
Next, consider \( w_{ij} \). In this case, the total weight on \( R^2_i \) in Eq. S2 must be \(-0.5\sigma_i\). The weight from the third term of Eq. S2 can be found by substituting \( w_{ij} = -0.5\sigma_i \). Subtracting that weight from the total weight specified by Eq. S4 gives \( w_{ij} \):

\[
w_{ij} = -\sigma_i + 0.5 \sum_j \sigma_{ij} \quad \text{(S5c)}
\]

Last, by grouping terms from Eq. S4 that are independent of the filter response, we find the constant in Eq. S2 is given by

\[
\text{const}' = -0.5 \sum_i \left( a_i \sigma_i^2 + \sum_{j \neq i} a_{ij} \sigma_i \sigma_j \right) + \text{const} \quad \text{(S5d)}
\]

Eqs. S1a-d are obtained by rewriting Eqs. S5a-d in matrix notation.

In visual cortex, simple and complex cells have the same maximum firing rates on average (Albrecht & Hamilton, 1982). However, the maximum possible response of the squared filters (see the second term in Eq. 5) is \( R_{\text{max}}^2 \), and the maximum possible response of the sum-squared filters (see the third term in Eq. 5) is given by \( R_{\text{max}}^2 = \left( R_{\text{max}} \sqrt{2 + 2 \mathbf{f}_i \cdot \mathbf{f}_j} \right)^2 \) where \( \mathbf{f}_i \) and \( \mathbf{f}_j \) are the weighting functions defining filters \( i \) and \( j \) (see Fig. 3a). Thus, if these computations were implemented in cortex—that is, if the maximum response of the complex cells equaled the maximum response of the simple cells \( R_{\text{max}} \), rather than \( R_{\text{max}}^2 \) and \( R_{\text{max}}^2 \)—the weights on the complex cell responses must be scaled.

Thus, the optimal weights on complex cells are given by \( w'_i = w_i \), \( w'_{ij} = R_{\text{max}}^{-1} w_{ij} \), and \( w'_{ij} = \frac{R_{\text{max}}^2}{R_{\text{max}}^2} w_{ij} \).

These scale factors are important because in a neurophysiological experiment one would measure \( w' \), not \( w \). Consequently, \( w'_{ij} \) and \( w'_{ij} \) are the normalized weights plotted in Fig. 7b.

**Evaluating the effect of surface slant**

Slant causes disparity to change across the patch. We evaluated the effect of surface slant by determining the optimal filters and estimation performance for a training set with a distribution of slants. If we assume (plausibly) that objects are on average equally likely to be viewed from any given angle, then for a given small angular retinal extent the probability of a given slant about the vertical axis should fall as a half-cosine function. Slant was calculated for the cyclopean eye. To evaluate the effect of surface slant variation, we generated a training set of stereo-images from a set of surfaces with slants that were randomly drawn from a truncated \((\pm 71 \text{ deg})\) cosine probability density (more extreme slants were not practical to include because of the surface size required to produce a projected patch width of 1 deg). The within-patch changes in disparity that were introduced by these slants are comparable to those in natural viewing (Fig. S1g). Results based on this analysis are thus likely to be representative. Then, we determined the optimal binocular filters for this new training set. The filters that were trained on surfaces having different slants (Fig. S1a-c) and the LL neurons that are constructed from them (Fig. S1d,e) are quite similar to those that were trained only on fronto-parallel surfaces. The main (although slight) differences are that the new filters have spatial extents that are slightly smaller and octave bandwidths that are slightly higher.
and that the new LL neuron tuning curves are broader. Disparity estimation performance on slant surfaces is also comparable, but precision is somewhat reduced (Fig. S1f, Fig. 5d). When disparity changes across the patch, the optimal encoding filters must be small enough to prevent the disparity signal from being ‘smeared out’ by within-patch (i.e. within-receptive field) disparity variation. And they must be large enough to prevent a poor signal-to-noise ratio, which would lead to a preponderance of false matches (i.e. inaccurate disparity estimates). The optimal filters appear to strike this balance.

Evaluating the effect of depth variations other than slant

Naturally occurring fine depth structure and occlusions (i.e. within-patch depth variation and depth discontinuities; Fig. S8a-c) were not present in either of our training sets. To determine how the magnitude of these other sources of depth variation compare to that due to slant, we analyzed 40,000 1 deg patches from 30 outdoor range images obtained with a high-precision Riegl VZ400 range scanner. For each patch, we computed the standard deviation of i) the differences between the raw range values and the best-fitting fronto-parallel plane, $\sigma_{\text{non-planar}}$, and ii) the differences between the best-fitting slanted and fronto-parallel planes, $\sigma_{\text{planar}}$. The ratio of these standard deviations $\sigma_{\text{planar}}/\sigma_{\text{non-planar}}$ provides a measure of the relative magnitude of depth variation due to slant and variation due to other factors. If the depth variation is due only to surface slant, then the ratio $\sigma_{\text{planar}}/\sigma_{\text{non-planar}}$ will have a value equal to 1.0. If all of the depth variation is due to factors other than slant, the ratio will have a value of 0.0. Fig. S8d plots the distribution of this ratio for all range patches. In the most common patches, essentially all the variation is captured by slant (i.e., the modal ratio is ~1.0). The median ratio is 0.75, indicating that in the median patch 75% of the depth variation is captured by slant. Given that the optimal binocular filters are robust to variation in surface slant, and given that most of natural depth variation is captured by slant, it is unlikely that depth variation other than slant will qualitatively affect our results.

External variability and neural noise

We examined the relative magnitude of stimulus-induced LL neuron variability and typical neural noise in cortex. Typical cortical neurons have a peak mean response larger than 30 spikes/sec (Geisler & Albrecht, 1997) and intrinsic noise variance proportional to the neuron’s mean response, with a proportionality constant (i.e. Fano factor) less than 1.5 (Tolhurst, Movshon, & Dean, 1983). To err on the side of too much neural noise, we assume that each LL neuron has a maximum average response of 30 spks/s and a Fano factor of 1.5 (b.c. the effect of noise is greater at lower firing rates). We convert the LL neuron responses to spike rates by first obtaining the average response of each LL neuron to its preferred disparity, and then scaling that value to the maximum average response. For an integration time of a typical fixation duration (300 ms), stimulus-induced response variance was ~3x greater than the internal noise. For an integration time equal to a typical stimulus presentation in a psychophysical task (1 sec), stimulus-induced response variance was ~9x greater than the internal noise. Thus, external variability must be considered to understand the influence of neural noise on the behavioral limits of vision under natural conditions. In fact, stimulus-induced variability is likely to be the controlling variability in many natural, single-trial tasks.

Spatial frequency tuning of optimal linear binocular filters

To develop an intuition for why spatial frequencies higher than ~6 cpd carry little information about disparity, we examined the disparity signals resulting from a binocularly-viewed, high-contrast, cosine-windowed edge. The Fourier-decomposition of each eye’s stimulus is a set of
phase-aligned sinewave gratings with the 1/f contrast fall-off typical of natural scenes (Field, 1987). The binocular difference signal (i.e. the difference of the left and right signals) is sinusoidal with contrast amplitude given by

\[ A_b(f|\delta_k) = \sqrt{A_L(f)^2 + A_R(f)^2 - 2A_L(f)A_R(f)\cos(2\pi f\delta_k)} \]  

(S6)

where \( f \) is the spatial frequency, \( \delta_k \) is a particular disparity, \( A_L(f) \) and \( A_R(f) \) are the left and right eye retinal amplitudes, and \( A_b(f|\delta_k) \) is the amplitude of the binocular contrast difference signal. If the two eyes have identical optics, the left- and right-eye retinal amplitudes (which include the effects of the optics and the 1/f falloff of natural image spectra) will be the same for a given disparity. Fig. S7a plots pattern of binocular contrast signals for each of several disparities. Fig. S7b shows the signals after filtering with a bank of log-Gabor filters, each having a 1.5 octave bandwidth. The patterns of binocular contrast barely differ above \(~6\) cpd. A filter population (for disparity estimation) will be most efficient if each filter signals information about a range of disparities.

V1 binocular receptive fields tuned to the highest useful spatial frequency (\(~6\) cpd) would have a spatial extent of \(~8\) arcmin, assuming the typical 1.5 octave bandwidth of neurons in cortex. Given that receptive fields cannot signal disparity variation finer than their own width, this analysis may help explain why humans lose the ability to detect disparity modulations (i.e. sinusoidal modulations in disparity-defined depth) at spatial frequencies higher than \(~4\) cpd (Banks, Gepshtein, & Landy, 2004; Harris, McKee, & Smallman, 1997). Vergence eye movement jitter has previously been proposed as an explanation for why useful binocular information is restricted to low spatial frequencies (Vlaskamp, Yoon, & Banks, 2011): vergence noise can degrade stereo information enough to render high frequency signals unmeasurable (Fig. S7c,d). The analysis presented here suggests that the frequency at which humans lose the ability to detect disparity modulation may simply be explained by the statistics of natural images.

**Optics**

The retinal images used in our analysis were simulated to match the retinal image formation that occurs in the human eye. In humans and non-human primates, accommodative and vergence eye-movement systems are yoked (Fincham & Walton, 1957). Consistent with this fact, we set the focus distance equal to the fixation distance of 40 cm and defocused the images appropriate for each disparity. The image of a surface displaced from the fixation is defocused by

\[ \Delta D \equiv \delta/IPD \]  

(S7)

where \( \delta \) is the disparity of the surface expressed in radians, \( IPD \) is the inter-pupillary distance expressed in meters, and \( \Delta D \) is the defocus expressed in diopters (1/meters). To account for the effect of chromatic aberration we created a different polychromatic point-spread function (PSF) for each disparity (i.e. defocus, Eq. S7). Single-wavelength PSFs were computed every 10 nm between 400 and 700 nm. The wavelength-dependent change in refractive power of the human eye was taken from the literature (Thibos, Ye, Zhang, & Bradley, 1992). Polychromatic PSFs were obtained by weighting the single-wavelength PSFs by the photopic sensitivity function (Stockman & Sharpe, 2000) and by the D65 daylight illumination spectrum, and then summing:
\[ psf_{\text{photopic}}(x, \Delta D) = \frac{1}{K} \sum_\lambda psf(x, \lambda, \Delta D) s_c(\lambda) D65(\lambda) \]  \hspace{1cm} (S8)

where \( s_c(\lambda) \) is the human photopic sensitivity function, and \( K \) is a normalizing constant that sets the volume of \( psf_{\text{photopic}} \) to 1.0. The modulation transfer function (MTF) is the amplitude of the Fourier transform of the PSF.

**Contrast normalization**

Contrast normalization occurs early in visual processing. The standard model of cortical neuron response assumes that the input is the contrast signal, \( c(x) = (r(x) - \bar{r}) / \sigma \), where \( r(x) \) are the noisy sensor responses to the luminance image falling on the retina over a local area and \( \bar{r} \) is the local mean. This contrast signal is then normalized by the local contrast:

\[ c_{\text{norm}}(x) = c(x) / \sqrt{\lVert c(x) \rVert^2 + nc_{50}^2}, \]

where \( n \) is the dimensionality of the vector, \( \lVert c(x) \rVert^2 \) is the contrast energy (or equivalently, \( n \) times the squared RMS contrast), and \( c_{50} \) is the half-saturation constant (Albrecht & Geisler, 1991; Albrecht & Hamilton, 1982; Heeger, 1992). (We are agnostic about the mechanism by which the contrast normalization is achieved.) Finally, the normalized contrast signal is weighted by the cortical neuron’s receptive field and passed through a static non-linearity. For example, in the case of a binocular simple cell, the response of the neuron is obtained by taking the dot product of the contrast-normalized signal with the receptive field and half-wave rectifying,

\[ R_i = R_{\text{max}} \left| c_{\text{norm}}(x) \cdot f(x) \right|. \]
References
Figure S1. Optimal filters and disparity estimation performance for estimating disparity for surfaces with a cosine distribution of slants. (a) Optimal linear binocular filters for estimating the disparity (of the center pixel) of slanted surfaces. The filters were learned on a training set having surfaces with a cosine distribution of slants. The optimal filter shapes are robust to the presence of non-zero surface slants in the training set. However, these filters are of somewhat smaller spatial extent and slightly higher octave bandwidth (compare filters F3-F5 to those in Fig 3). (b) Spatial frequency tuning and bandwidth of filters. (c) Two-dimensional analogs of filters in (a). The two-dimensional filters respond virtually identically to two-dimensional images, as do the one-dimensional filters to one-dimensional signals. (d) Tuning curves of LL neurons trained and tested on surfaces with varying surface slant. (e) Normalized weights on model complex cell responses for constructing five LL neurons with the indicated preferred disparities. (f) Disparity estimation performance with surfaces having a cosine distribution of slants (black line). Also shown are disparity estimates derived from a local cross-correlator (gray line). (g) Disparity gradients from viewing natural scenes (Hibbard, 2008) vs. disparity gradients in our training set. The relationship between slant and disparity gradient is given by $g = {IPD(\tan \theta) / d}$ where $\theta$ is the surface slant and $d$ is the distance to the surface. Hibbard (2008) estimated disparity gradients in natural viewing for distances nearer than 3.5 m (dashed curve) and 0.5 m (dotted curve). At the 0.4m viewing distance assumed throughout the paper, our manipulation of surface slant introduced distribution of disparity gradients (black line) that is comparable to the gradients introduced in natural viewing.
Figure S2. Filter responses, LL neuron responses, and performance with random-dot stereograms and anti-correlated random-dot stereograms. (a) Joint response distributions of filters F1 and F2 when stimulated with random-dot stereograms. Compare to natural stimuli in Fig. 4a. (b) Joint responses of filters F3 and F4 when stimulated with random-dot stereograms. (c) LL neuron responses to random dot stereograms. (d) LL neuron tuning vs bandwidth when stimulated with natural stimuli. Note that bandwidths are elevated compared to when the LL neurons are stimulated with natural stimuli (see Fig. 7c). (e) Disparity estimates and estimate precision (inset) obtained with random-dot stereograms. (fg,h,i) Same as a,b,c,e except that the stimuli were anti-correlated random dot stereograms. Filter response distributions still segregate as a function of disparity (especially with F3 and F4). However, the anti-correlated filter response distributions do not coincide with the 'expected' pattern of response for particular disparities. The disparities are decoded improperly with a decoder trained on natural stimuli. LL neuron tuning curves with anti-correlated stereograms are neither selective nor invariant. Inaccurate, highly variable estimates result.
**Figure S3.** Obtaining AMA filter and model complex cell responses with established neural operations (a) Simple cell responses are obtained by projecting the left- and right-eye signals onto each binocular filter and half-wave rectifying. Subtracting ‘on’ and ‘off’ simple cell responses give the AMA filter responses (see Fig. 3a). (b) Model (and neurophysiological) complex cell responses are obtained by summing and squaring the responses of simple cells. (Model complex cell responses can also be obtained by summing the responses of simple cells that have both a half-wave rectifying and a squaring nonlinearity.) Note that the model complex cells differ from canonical disparity energy binocular complex cells. We label them ‘complex’ because they exhibit temporal frequency doubling and have a fundamental to DC response ratio of less than one. The model LL neurons (see Fig. 6 and Fig. 7) are more similar to disparity energy complex cells although, as noted in the text, there are also important differences.
Figure S4. The eight AMA filters from Fig. 3 (on the diagonal) and their pairwise sums (see Fig. S3a). Solid and dashed curves indicate the left and right eye filter components, respectively. Filter identity is indicated by the subscripts $F_{ij}$. Some filters have shapes similar to canonical V1-like binocular receptive fields. These filters yield model complex cell tuning curves that vary with disparity (Fig. S5). Other filters are irregularly shaped (e.g. $F_{36}$). Irregularly shaped filters yield model complex cell tuning curves that do not vary significantly with disparity (Fig. S5). The inset shows the largest absolute normalized weight (see Supplement, Eqs. S1a-d) given to each filter in constructing the set of LL neurons (Fig. 7a). Lighter colors indicate smaller weights. Unsurprisingly, the irregularly shaped filters (e.g. $F_{36}$) play a generally smaller role in shaping the response properties of log-likelihood (LL) neurons.
**Figure S5.** Disparity tuning curves of the model complex cells implied by the AMA filters (on-diagonal) and their normalized pairwise sums (off-diagonal). Solid curves show the complex cell tuning curves to natural stimuli. The complex response is given by filtering the contrast-normalized stimulus and then squaring: $R_{max} \left( f_e c_{norm} \right)^2$ (see Fig. S4).

Each point on the curve represents the average response to a large number of natural stimuli having the same disparity. Gray area shows response variability due to variation in natural image content. Dashed curves show the complex cell tuning curve to 5% random-dot stereograms (RDSs). In most cases, RDS tuning curves are similar to the tuning curves for natural stimuli, although response magnitude is generally somewhat attenuated. These individual model complex cells have some selectivity for disparity. However, they exhibit little invariance to variation in natural image content. Thus, the pattern of the joint population response must be used to accurately estimate disparity. Neurons that signal the log-likelihood of disparity (LL neurons) can be constructed by appropriately combining the joint responses of model complex cells. LL neurons are highly selective for disparity, and are also strongly response invariant to irrelevant.
Figure S6. Optimal disparity estimation vs. estimation via local cross-correlation. Performance measures were obtained by testing each method on identical sets of left- and right-eye signals, like those shown in Fig. 1d. (Note that this differs somewhat from the most common test of the cross-correlation model, which gives the algorithm windowed access to many patches in each eye. We chose the current method so that each algorithm has access to the same image information, thereby making the comparison meaningful.) Black curves indicate estimation performance via local cross-correlation. Gray curves indicate performance obtained with the eight AMA filters shown in the main text (data replotted from Fig. 5). (a) Median MAP estimates for frontoparallel surfaces only. Error bars indicate 68% confidence intervals. (b) Estimate precision as indicated by 68% confidence intervals. Error bars from (a) are replotted on a linear axis. (Confidence intervals were plotted on a semi-log plot in the main text.) (c) Sign identification performance with frontoparallel surfaces. (d,e). Same as b,c except that the performance is for surfaces with a cosine distribution of slants. Estimation performance with the cross-correlator is (slightly) more biased and is significantly less precise at disparities larger than 7.5 arcmin.
Figure S7. Information for disparity estimation is concentrated in low to mid spatial frequencies. (a) Inter-ocular retinal contrast amplitude of a cosine-windowed high-contrast edge as a function of spatial frequency, for different disparities (Eq. S6). The optics in the two eyes were identical; i.e. $A_L(f) = A_R(f) \approx MTF(f) f$ where $MTF(f)$ is the modulation transfer function associated with a particular disparity (see Methods). (b) Inter-ocular contrast signals that have been passed through a bank of 1.5 octave bandwidth log-Gabor filters. The signals differ little above ~5 cpd. (c) The effect of vergence jitter on inter-ocular contrast signals. Signals were corrupted by 2 arcmin of vergence jitter, a typical value in humans. Vergence noise was simulated by convolving each disparity signal with a Gaussian having a standard deviation of 2 arcmin. Vergence jitter further attenuates high frequency inter-ocular contrast signals. Natural image structure and vergence jitter both contribute to the lack of disparity information at high luminance spatial frequencies. (d) The signals in (c) after having been passed through a bank of 1.5 octave bandwidth log-Gabor filters. (e) The effect of a four diopter difference in refractive power between the two eyes drastically reduces the inter-ocular contrast signal. A four diopter (or larger) difference between the left and right optics, is a primary risk factor for the development of amblyopia (Levi et al, 2011). Note that these signals were simulated for eyes with 2mm pupils. With larger pupil diameters, the same difference in refractive power would further reduce available disparity signals. For example, with a 2mm pupil an 8 diopter difference between the eyes produces the same difference in defocus blur as would a 4mm pupil with a 4 diopters difference between the eyes. In that case, there are no measurable signals above 1 cpd. (f) The signals in (e) after having been passed through a bank of 1.5 octave bandwidth log-Gabor filters. (g) Same as (e) but with a 4mm pupil. (h) Same as (f) but with a 4mm pupil.
**Figure S8.** Depth structure not attributable to slant in natural scenes. (a) A representative natural range image. (b) Overhead plan view of the range along the white scan-line in (a). (c) Range values within a 1 deg patch, marked by the white box in (a) and the black box in (b). The dots represent the raw range values. The plane is the best fitting plane to the range data. (d) Depth variation in 1 deg patches of natural scenes can largely be captured by slanted planes (see supplementary text). Values near one indicate that all range variation can be captured by (planar) surface slant. Values near zero indicate that none of the range variation can be captured by planar surface slant. In the most common patch, nearly all depth variation is attributable to slant. In the median patch, 75% of depth variation is captured by slant.
Figure S9. Contrast normalization and the Gaussian form of the filter response distributions. The standard model of contrast normalization in retinal and cortical neurons is given by $c_{norm}(x) = c(x)/\sqrt{[x]} + nc_{50}^2$ (Albrecht & Geisler, 1991; Albrecht & Hamilton, 1982; Heeger, 1992). Throughout the paper, we assumed that $c_{50}$ had a value of zero. To examine the effect of different values of $c_{50}$ on the form of the filter response distributions, we first contrast-normalized the test stimuli using different values of $c_{50}$. Then, we projected the contrast-normalized the test stimuli from each disparity level onto the filters to obtain the conditional filter response distributions. Finally, we fit a multi-dimensional generalized Gaussian (one-dimension for each filter) to each conditional response distribution, via maximum likelihood estimation. (a) When $c_{50} = 0.0$, the conditional response distributions have tails that tend to be somewhat lighter than Gaussians (i.e. lower kurtosis than a Gaussian). When $c_{50} = 0.1$, the conditional response distributions are most Gaussian on average. For $c_{50} \gg 0.1$, the distributions have heavier tails than Gaussians (i.e. higher kurtosis than a Gaussian). (b) Powers of the multi-dimensional generalized Gaussians that best fit a subset of the two-dimensional conditional filter response distributions. A power of two (dashed line) indicates that the best fitting distribution was Gaussian. Powers greater than or less than two indicate that the responses are better fit by distributions having lighter or heavier tails than a Gaussian, respectively. The best-fitting powers are shown for three different values of $c_{50}$: 0.00, 0.10, and 0.25. The average powers, across the conditional response distributions, for the three tested values of $c_{50}$ were 2.79, 2.07, and 1.17 respectively. The same qualitative results hold for the other two-dimensional response distributions. The pattern also holds for higher-dimensional filter response distributions. In higher dimensions, however, results are noisier because stable estimates of kurtosis require significantly more data.