PLOS | ONE

RESEARCH ARTICLE

Luminance, Colour, Viewpoint and Border Enhanced Disparity Energy Model

Jaime A. Martins[1]*, João M. F. Rodrigues[2], Hans du Buf[1]

1 Vision Laboratory (FCT), ISR-LARSyS, University of the Algarve, Faro, Portugal, **2** Vision Laboratory (ISE), ISR-LARSyS, University of the Algarve, Faro, Portugal

* jamartins@ualg.pt

Abstract

The visual cortex is able to extract disparity information through the use of binocular cells. This process is reflected by the Disparity Energy Model, which describes the role and functioning of simple and complex binocular neuron populations, and how they are able to extract disparity. This model uses explicit cell parameters to mathematically determine preferred cell disparities, like spatial frequencies, orientations, binocular phases and receptive field positions. However, the brain cannot access such explicit cell parameters; it must rely on cell responses. In this article, we implemented a trained binocular neuronal population, which encodes disparity information implicitly. This allows the population to learn how to decode disparities, in a similar way to how our visual system could have developed this ability during evolution. At the same time, responses of monocular simple and complex cells can also encode line and edge information, which is useful for refining disparities at object borders. The brain should then be able, starting from a low-level disparity draft, to integrate all information, including colour and viewpoint perspective, in order to propagate better estimates to higher cortical areas.

Citation: Martins JA, Rodrigues JMF, du Buf H (2015) Luminance, Colour, Viewpoint and Border Enhanced Disparity Energy Model. PLoS ONE 10(6): e0129908. doi:10.1371/journal.pone.0129908

Academic Editor: Zhong-Lin Lu, The Ohio State University, Center for Cognitive and Brain Sciences, Center for Cognitive and Behavioral Brain Imaging, UNITED STATES

Received: January 18, 2015

Accepted: May 14, 2015

Published: June 24, 2015

Data Availability Statement: All relevant data is available in the Middlebury Stereo Vision Page at http://vision.middlebury.edu/stereo/. The full datasets are freely available for download at http://vision.middlebury.edu/stereo/data/ and the evaluation results are shown at http://vision.middlebury.edu/stereo/eval/.

Funding: This work was partly supported by the Portuguese Foundation for Science and Technology (FCT), LARSyS FCT (UID/EEA/50009/2013), EU project NeuroDynamics (FP7-ICT-2009-6, PN: 270247), FCT project SparseCoding (EXPL/EEI-SII/

1. Introduction

Disparity plays an important role in our perception of the environment, giving us precious information for survival. Our brain extracts it from the information that reaches the hypercolumns of V1 via the Lateral Geniculate Nucleus (LGN), which relays information of the left and right retinae. At this early stage, disparity is already key for broad and precise motor control (e.g., walking/running while avoiding obstacles, eye-hand coordination while picking up a pencil), low- and high-level Focus-of-Attention (FoA), object and background segregation, as well as recognition, even with partial occlusions [1].

Computer vision research has significantly advanced the state-of-the-art in disparity estimation models, with many different approaches and applications [2]. However, there is a significant lack of biologically motivated models that computationally implement the Disparity Energy Model (DEM), which integrates key biological evidence from research on the cat's visual cortex and pathways by [3], and more recently from the rhesus monkey's visual cortex [4]. Alternative models also exist for building and combining disparity energy neurons [5]. The

1982/2013) and FCT PhD grant to author JAM (SFRH-BD-44941-2008). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

DEM allowed to explain how neurons tuned to horizontal disparities can have the implicit ability to discriminate vertical disparities [6]. This ability is an emerging property from a neuronal system tuned to horizontal disparities, by decoding vertical ones as a deviation from the expected neuronal responses. This ability also illustrates how the neuronal system can encode much richer information than would be expected and, at the same time, concentrate neuronal resources on the most common cases while keeping the possibility of encoding rare ones.

Most DEM computational implementations found in the literature were unable to give good results on real-world images. Therefore, we first focused on building upon a state-of-the-art theoretical DEM implementation by [6] until we could reliably extract disparity estimations from real-world data. This was documented in [7]. It is still the only DEM-based method ranked on the Middlebury Stereo Evaluation Website [8], against 153 other disparity methods.

Some authors have proposed alternative biological models which are not based on the DEM, e. g., [9] combining geometric information and local edge features, [10] using multiscale lines and edges to retrieve a disparity wireframe model of the scene—the Line and Edge Disparity Model (LEDM) which is further explored in this paper in §5.1—and also du Buf et al. [11], employing the phase differences of simple cell responses to the left and right views. The latter model is often applied to real-world problems, although it has been shown to be very imprecise in terms of localisation of depth transitions.

Most DEM research has considered theoretical or synthetic data, while biological models applied to real-world scenes appeared only recently [7, 9, 10, 12]. This is mainly due to the fact that computational DEM implementations are usually focused on evaluating theoretical results using very specific stimuli, like bar/grating patterns or random-dot stereograms [6], or in psychophysical experiments [4].

In this paper, we propose a disparity map composed of different cell maps built on top of each other, each refining the previously extracted disparity. We also propose that the first, rough disparity (disparity *gist*) is provided by the DEM model [7], after which refinements based on colour, perspective correction (viewpoint) and border information are integrated to achieve the final disparity map. Although the model is still feed-forward or bottom-up, in the future it can be supplemented by feedback loops from higher visual areas V2 and V4 in order to further improve results [1].

In our improved DEM implementation we use two neuronal populations for obtaining disparities:

1. An *encoding population* which uses a set of binocular neurons with a diverse range of cell parameters, e. g., horizontal disparities, spatial frequencies and orientations. This population is trained on random-dot stereograms in order to learn activity codes for many different disparities. The method is similar to that of [6], which is based on the DEM model of [3], with proper normalisation to yield local correlations with neighbourhood weighting [13–15]. Finally, the population is applied to real stereograms in order to obtain local activity codes. This is further explained in §3.1.

2. A higher-level *decoding population* which compares a local activity code, at each image position, with all learned (trained) activity codes, for estimating local disparity. This is further explained in §3.2. Basically, this second population implements a template-matching process similar to those of [16] and Read [6]. This initial DEM model (disparity gist) is then integrated with colour and different viewpoints (§4), and finally with object border information retrieved from the multi-scale line and edge disparity model (LEDM) [10] and low-level processes from object salience research [17] (§5).

Our main contributions in this paper are: (a) Improving previous DEM results in real-world images. (b) The integration of the DEM model with luminance, colour information and viewpoint perspective correction. (c) The integration of two disparity models DEM and LEDM, to improve object boundary precision of the DEM. (d) The integration of different layers of disparity cell maps, with each layer improving the results from layer to layer. (e) The quantitative evaluation of results with real-world scenes, showing that the model can compete with state-of-the-art computer vision algorithms.

2. Disparity-sensitive cells

The primary visual cortex (V1) is composed mostly of simple, complex and end-stopped (hypercomplex) cells arranged into ocular dominance hypercolumns. Computationally, the receptive fields (RFs) of monocular simple cells can be modelled by Gabor wavelets [7, 18, 19], with parameters to specify orientation θ, spatial frequency f (or the wavelength $\lambda = 1/f$), receptive field size σ and spatial phase ϕ, which will be discussed below. We can then model *binocular* simple cells using pairs of *monocular* simple cells with either a position- or phase-shift between RFs (or a combination of both), signalling disparity when both RFs of the binocular cell are fully excited. However, binocular simple cells are also sensitive to stimulus contrast and pattern position within their RFs [3, 18], which makes them unsuitable as disparity detectors.

In contrast, binocular *complex* cells can solve these problems, as there are no separate excitatory and inhibitory subregions within their RFs, making them only sensitive to position, orientation and stimulus size [20]. They also show other desirable properties like sensitivity to fine disparities, immunity to anti-correlated stimuli [3] and they respond accurately to dynamic random-dot stereograms [21]. Two binocular simple cells S_1 and S_2 can be combined into a phase-independent binocular complex cell, provided that their phase difference $|\phi_{S_1} - \phi_{S_2}|$ equals $\pi/2$. Therefore, the response of a binocular complex cell can be obtained by summing the responses of two binocular simple cells with phases in quadrature.

Mathematically, two monocular RFs can be used to model a binocular *simple* cell, with the same size, orientation and spatial frequency, but with different phases ϕ and/or RF positions on the retina $(\Delta x, \Delta y)$ [22]. The left (ρ_L) and right (ρ_R) RFs of binocular simple cells are then defined by

$$\rho_{L,R}(x, y; \theta, \sigma, f, \phi, \Delta\phi) = \exp\left(-\frac{\dot{x}_{L,R}^2 + \dot{y}_{L,R}^2}{2\sigma^2}\right) \cos\left(2\pi f \dot{x}_{L,R} + \phi \pm \frac{\Delta\phi}{2}\right). \tag{1}$$

Since we will use phases in quadrature $\phi \in \{0, -\pi/2\}$ and both ρ_L and ρ_R actually consist of two RFs: the sine and cosine components. In Eq. 1, $\dot{x}$ and $\dot{y}$ are the coordinates relative to the binocular cell's centre, which is (0, 0) at the fovea, and rotated according the cell's preferred orientation θ:

$$\dot{x}_{L,R} = x_{L,R} \cos\theta + y_{L,R} \sin\theta \tag{2a}$$

$$\dot{y}_{L,R} = -x_{L,R} \sin\theta + y_{L,R} \cos\theta. \tag{2b}$$

The left disparity viewpoint is used as reference, requiring the use of binocular cells with left predominance. The main reason for using the left view is that it is often used for defining the ground-truth of real scenes, thus allowing for a quantitative analysis of experimental results. Mathematically, the offset coordinates Δx and Δy, which correspond to the cell's preferred horizontal and vertical disparities, are defined as follows: when the activity code is trained (learned) with random-dot stereograms, the left RF is centred at $(0, \Delta y)$ and the right one at $(-\Delta x, \Delta y)$. When the cells are applied at all input stereogram positions, then $(x_L, y_L) = (x, y + \Delta y)$ and

$(x_R, y_R) = (x - \Delta x, y + \Delta y)$. We note that $\Delta y = 0$ is taken for all cells, as vertical disparity in the fovea is zero [22]. For a detailed mathematical transformation from monocular to binocular simple cells see [18].

3. Luminance Disparity-Energy Model

In this section, we describe the lL-DEM or L-DEM, and show how disparity maps can be extracted by exploiting binocular cell responses and comparing them with previously learned stimuli, using cells sensitive only to luminance variations. The L-DEM was first presented in Martins et al. [7] and is adapted partly for this section, serving here to provide a performance baseline. Understanding this model is also fundamental for understanding all further improvements described in this paper.

For the L-DEM implementation, we use two neuronal populations: (1) an encoding population and (2) a higher-level decoding population. As explained above, for presenting our stereo results we use by default the reference viewpoint (image) of the left eye.

3.1 Disparity encoding population

For the encoding population's binocular simple cells defined in Eq. (1), we selected RF parameters based on [6]:

a. Orientations $\theta_i \in (i \times \pi)/N_\theta$, with the number of orientations $N_\theta = 8$. Our empirical tests showed that using more orientations yielded slightly better disparity estimates, but increases the total cell population. Using eight orientations is a good compromise.

b. Receptive field sizes (scales) $\sigma \in \left\{2\sqrt{2},\ 2,\ \sqrt{2}\right\}$. These are scaled by a factor of $\sqrt{2}$, as is the spatial frequency. Empirical results showed that bigger sizes increase the blur at objects' border regions and smaller sizes lead to errors in disparity estimates.

c. Spatial frequencies $f \in \left\{\sqrt{2}/8,\ 1/4,\ \sqrt{2}/4\right\}$ cycles per pixel. These values are proportional to RF size by $\omega\sigma = \pi$ or $f = 1/2\sigma$. The frequency bandwidth for the three scales was 1.14 octaves.

d. RF phases $\phi \in \{0, -\pi/2\}$, since only two values are needed to build a phase-invariant binocular complex cell from two binocular simple cells [3].

e. RF horizontal position disparity $\Delta x \in \{0, \ldots, 59\}$ in steps of 1 pixel.

f. RF phase disparity $\Delta\phi = 0$, implying no extra phase difference between the left and right RFs of each simple cell (equal phases ϕ for both). It is to be expected that in naturally occurring images, the maximum response of a phase-shift disparity neuron is elicited when there is a different pattern of the same stimulus in the left and right RFs, something that never occurs in the real world [4, 5]. Our empirical tests also showed that the use of phase differences—odd-symmetric disparity tuning curves—did not add significant information and sometimes even degraded the quality of disparity estimates. Other alternative roles for neurons tuned to phase disparities are explained further in [23].

In total, the above selection yields a population of $8_\theta \times 3_{\sigma,f} \times 2_\phi \times 60_{\Delta x} \times 1_{\Delta\phi} = 2880$ binocular simple cells as inputs for 1440 binocular complex cells; see below. The values were chosen to replicate physiological parameters of real cells, for yielding precise disparity estimates in real-world images. The disparity encoding population is then built and trained as follows, based on Read [6]:

Stereo energy coding. Responses of the left and right RFs of binocular simple cells (v_L and v_R) are obtained by convolving (*) the RFs with the corresponding left and right grayscale images $I_{L,R}(x, y)$:

$$v_{L,R}(x, y) = I_{L,R}(x, y) * \rho_{L,R}(x, y). \tag{3}$$

To simplify notation, below we skip (x, y). $I_{L,R}$ are obtained from sampling an RGB colour stereogram using physiologically perceived weights from the luminance Y channel of the CIE XYZ colour-space, which closely resembles human colour perception: $I_{L,R} = 0.2989 \cdot R_{L,R} + 0.5870 \cdot G_{L,R} + 0.1140 \cdot B_{L,R}$.

At each image position, the response S of a binocular simple cell combines the squared responses of the left and right RF components [3, 18]:

$$S = (v_L + v_R)^2 = v_L^2 + v_R^2 + 2\, v_L v_R. \tag{4}$$

S can be split into the monocular term $M = v_L^2 + v_R^2$ and the binocular term $B = 2\, v_L\, v_R$. Biologically, this can be realised by combining the outputs of two energy neurons with phase disparities π apart. If such neurons are identical except for their phase disparities, then the first one computes $(M + B)$ and the second $(M - B)$. Both M and B are then available from the sum and difference of the two responses, i. e., $2M$ and $2B$ [6].

For obtaining the local stereo energy E of a binocular complex cell which is invariant to the phases of local patterns in the input, one can either sum the responses of (a) many binocular simple cells with scattered phases ϕ in $[0, 2\pi]$, or (b) only two cells with phases in quadrature. We could therefore apply the second case with $\phi \in \{0, -\pi/2\}$: $E = \Sigma_{\phi = \{0, -\pi/2\}} S_\phi$. This stereo energy E, for each frequency, orientation and disparity, can be related to the cross-correlation between filtered and windowed images [15]. However, the local stereo energy E cannot be used directly to estimate disparities, as it also reflects monocular energy (stimulus contrast inside each RF) along with binocular energy (stimulus disparity between RFs). This shortcoming is addressed below by using *spatial pooling* and *effective binocular correlation*.

Spatial pooling. Complex cells are normally modelled by taking the square root of the sum of the squared responses of the sine and cosine components of the simple cells. This implies that the RF size of such complex cells is equal to that of the simple cells: the same Gaussian. However, RFs of real binocular complex cells are larger than those of simple cells [18]. Therefore we apply this property by averaging M and B, using grouping cells with a Gaussian RF: $G^{sp}(x, y) = k \exp(-(x^2 + y^2)/2\sigma^2)$. The normalisation factor $k = 1/(2\pi\sigma^2)$ and σ equals the RF size of the corresponding simple cells: $\sigma \in \{2\sqrt{2},\ 2,\ \sqrt{2}\}$. This yields, for the two phases, $M_\phi^{sp} = G^{sp} * M_\phi$ and $B_\phi^{sp} = G^{sp} * B_\phi$. This pooling operation involves using simple grouping cells with a dendritic field size defined by σ and it is crucial to stabilise results in case of real-world images with noise and non-uniform disparity ranges.

Effective binocular correlation. In order to differentiate monocular energy from binocular energy, it is necessary to use normalised binocular correlation detectors [6, 13–15]. These detectors respond maximally (+1) when the left and right RF views are identical, and minimally (−1) when one RF view is an inverted-contrast version of the other. They are implemented by dividing the pooled binocular term by the pooled monocular term, after which the result is pooled once more for increasing robustness:

$$\psi^{sp} = G^{sp} * \left(\frac{\sum_{\phi=\{0,-\pi/2\}} B_\phi^{sp}}{\sum_{\phi=\{0,-\pi/2\}} M_\phi^{sp}} \right). \tag{5}$$

The value of ψ^{sp} relates to the correlation between local, filtered regions of the left and right views [23]. The population of binocular correlation detectors ψ^{sp} is used for encoding disparity in the model. Disparities estimated by using the effective binocular correlation instead of the local stereo energy E are immune to the detrimental effect of monocular contrast, allowing the extraction of disparity from peaks in the population's activity code. ψ^{sp} has also the useful property that it exactly equals 1 when the actual disparity matches a cell's preferred disparity [6]. Please recall that ψ^{sp} is the short notation for $\psi^{\text{sp}}_{f,\theta,\Delta x}(x, y)$, i. e., there are three scales, eight orientations and 60 horizontal position disparities, hence 1440 binocular correlation cells which are later applied at all image positions.

Learning the population code. We trained the energy model to discriminate horizontal stimulus disparities (Δx_{stim}) ranging from 0 to 59 pixels with a stepsize of 1 pixel. Population activity codes were gathered from cell responses to stimuli with known disparities: random-dot stereograms with an uniform disparity, sampled randomly from a Gaussian distribution with zero mean and unit standard deviation, with a Δx_{stim} horizontal offset between the left and right images. Offset gaps were also filled with randomly sampled pixels; see Fig 1. For each Δx_{stim} step we generated 1000 random-dot pairs. Hence, training involved 60,000 stereograms; for details see Martins et al. [7]. For each stereogram, with $I_{L,R}$ the left and right views, we applied Eq (3) and Eq (4), but only at the centre of the left and right images of each stereogram. The values of ψ were computed without spatial pooling, i. e.,

$$\psi = \frac{\sum_{\phi=\{0,-\pi/2\}} B_\phi}{\sum_{\phi=\{0,-\pi/2\}} M_\phi}, \tag{6}$$

because the results are pooled over 1000 random-dot stereograms for each disparity.

During training, and later when applying the population to real images, the effective binocular correlations ψ and ψ^{sp} are encoded as a *mean spike count*,

$$\Psi = (1 + \psi)\, u, \tag{7}$$

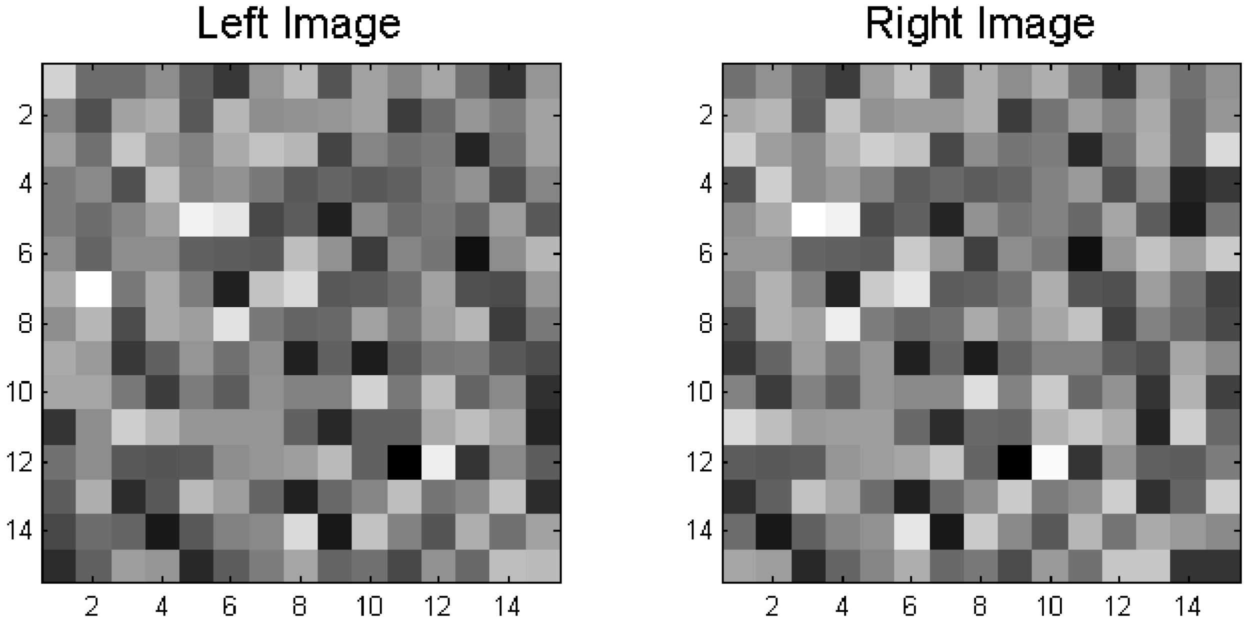


Fig 1. Example of a 15 × 15 random-dot stereogram used in training, with a uniform 2-pixel shift and thus horizontal disparity Δx_{stim} of 2.

doi:10.1371/journal.pone.0129908.g001

where $u = 8$ is the average number of spikes elicited by a binocularly uncorrelated stimulus within the temporal discrimination window. We used parameters similar as [6], with typical values of u around 8 spikes, assuming a firing rate for the optimal disparity of 100 Hz and a temporal window of 160 ms. This yields values of Ψ in the range $[0,2u]$, where $2u$ represents the mean number of spikes that neurons tuned to a specific disparity will fire in the presence of a perfect binocular stimulus of that disparity (maximum correlation).

Finally, Ψ was averaged $[\mathrm{A}(\cdot)]$ over the 1000 different stereograms for each Δx_{stim}, which serves to eliminate random stimulus-dependent noise. This yields an activity code for each trained horizontal disparity Δx_{stim}:

$$W_{f,\theta,\Delta x}^{\Delta x_{\mathrm{stim}}} = \mathrm{A}(\Psi_{f,\theta,\Delta x}^{\Delta x_{\mathrm{stim}}}). \tag{8}$$

In summary, W represents the number of spikes produced by neurons tuned to frequencies f, orientations θ and horizontal disparities Δx, averaged over all 1000 stimuli with the same uniform disparity Δx_{stim}. The population code thus consists of 1440 binocular correlation cell responses (3 scales, 8 orientations and 60 horizontal position disparities) for *each* of the 60 different horizontal stimulus disparities Δx_{stim} of the random-dot stereograms. The adaptation and learning of the encoding cell population to discriminate disparities can be thought of as kin to visual learning in early childhood, assuming that basic neural circuitry is the result of evolution, or, at least, needs adequate training to reach its full potential.

3.2 Disparity decoding population

As mentioned before, learning is done only once and in the centre of the random-dot stereograms. After training, the encoding population can then be applied at all pixel positions (neighbourhoods) of real world input stereograms, excluding the border region. The disparity at each position is estimated by comparing the activity code there with all learned codes. This is done by a second, higher-level *decoding* population. The disparity assigned to each pixel position is the disparity of the best-matching code. Local disparity estimation is a simple matching process [16]: the input code of 1440 responses is matched or correlated with the 60 sets of 1440 trained codes. The final output is selected by the decoding population by a winner-takes-all strategy. Biologically, this probably involves associative memory, which can also be based on a training process [24].

The matching process uses 60 correlation cells ("Corr") which compare $\Psi_{f,\theta,\Delta x}^{\mathrm{sp}}(x,y)$ with $W_{f,\theta,\Delta x}^{\Delta x_{\mathrm{stim}}}$, i. e., the 1440 spike counts at each image position with all previously learned 60 sets of 1440 spike counts:

$$r_{\Delta x_{\mathrm{stim}}}(x,y) = \left[\mathrm{Corr}(\Psi_{f,\theta,\Delta x}^{\mathrm{sp}}(x,y), W_{f,\theta,\Delta x}^{\Delta x_{\mathrm{stim}}})\right]^{+}, \tag{9}$$

where $[\cdot]^{+}$ is half-wave rectification. This avoids the problem of disparity in anti-correlated stereograms by setting any negative correlations to zero [25]. Note that $r_{\Delta x_{\mathrm{stim}}}$ is a vector of 60 correlation values, each related to a specific Δx_{stim} disparity that the population was trained to recognise, from 0 to 59. The maximum correlation yields the luminance-disparity map $D^{\mathrm{L}}(x,y) = \underset{\Delta x_{\mathrm{stim}}}{\operatorname{argmax}} \left[r_{\Delta x_{\mathrm{stim}}}(x,y)\right]$. Biologically, this corresponds to the activation of a single disparity cell at each position, inhibiting the other 59 cells. Mathematically, the implemented matching process (Corr) is the Pearson product-moment correlation coefficient with $\mathrm{A}[\cdot]$ the average, σ_Ψ and σ_W the standard deviations of all 1440 responses:

$$r_{\Delta x_{\mathrm{stim}}}(x,y) = \left[\frac{\mathrm{A}[\Psi^{\mathrm{sp}}(x,y)W] - \mathrm{A}[\Psi^{\mathrm{sp}}(x,y)]\mathrm{A}(W)}{\sigma_{\Psi^{\mathrm{sp}}(x,y)}\sigma_W}\right]^+. \tag{10}$$

3.3 Experimental results

The obtained results for this method were first published in Martins et al. [7], where we tested the Luminance Disparity Energy Model (L-DEM) on various reference stereograms from the Middlebury stereo evaluation set. These are: *tsukuba*, *venus*, *teddy* and *cones* [26, 27], *aloe* and *cloth3* of the 2006 dataset, and *dolls*, *moebius* and *reindeer* of the 2005 dataset [28].

For reference, Fig 2(a)-2(e) shows the L-DEM results for the *tsukuba* stereo pair [7, 27]. This algorithm was able to obtain good results for the Middlebury evaluation test (ranked there as "BioDEM") [7], which are detailed in §6. We will compare further disparity improvements using these results as baseline.

4. Luminance, Colour and Viewpoint DEM

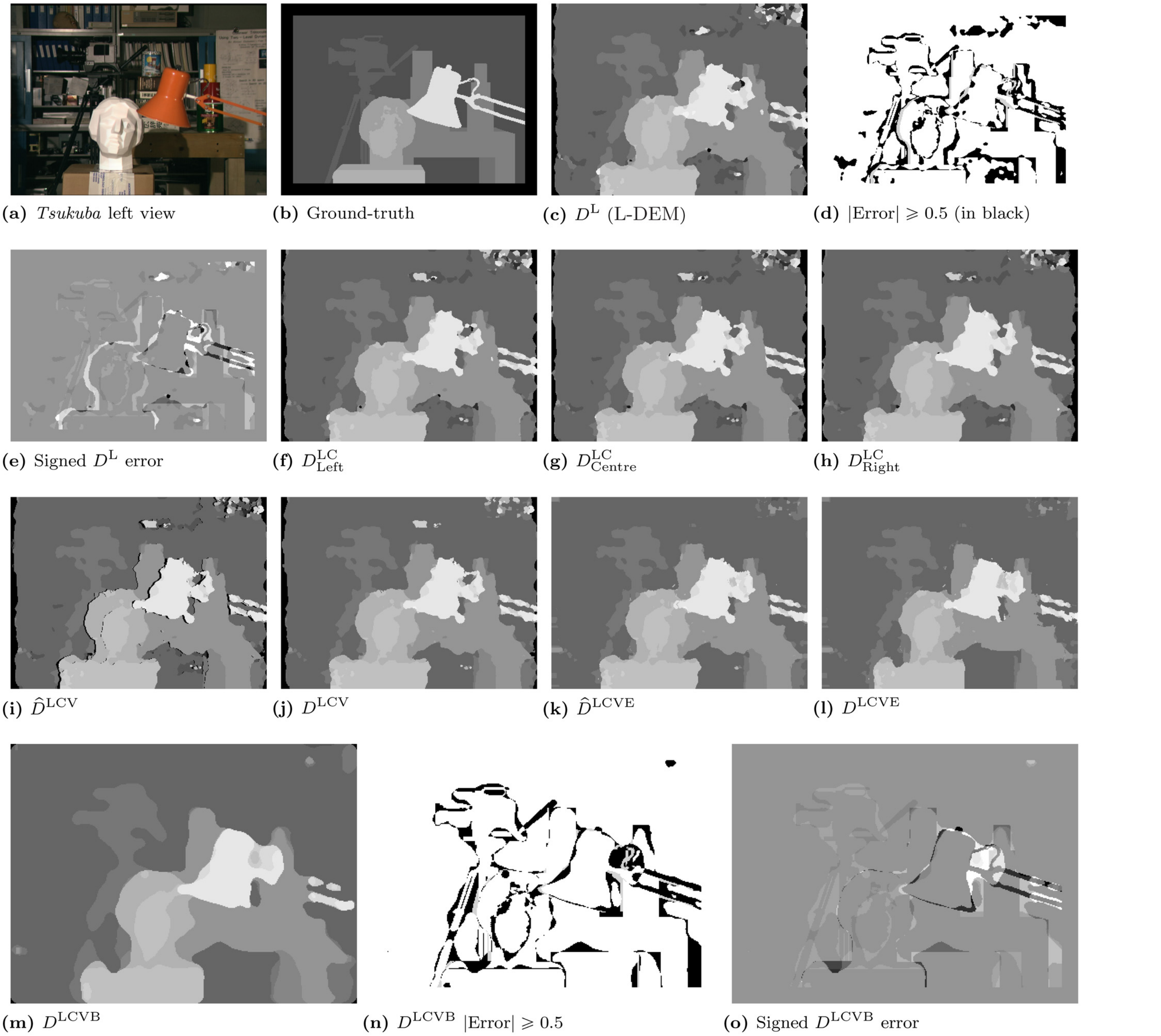
This section addresses an improved disparity model, the Luminance, Colour and Viewpoint Disparity Energy Model(LCV-DEM), which integrates colour and viewpoint (perspective) information to increase accuracy of the L-DEM.

Research involving the chromatic representation in area V1 has shown that cone responses from the retina turn into three relatively independent spatio-chromatic colour channels after the LGN [29], which are then transformed in several neural pathways, mixing colour responses with those of other cells [30]. The majority of neurons in V1 seem to respond to pure isoluminant stimuli (i. e., they are colour sensitive even in the absence of luminance changes), and around 50% of all neurons are sensitive to both luminance and isoluminant stimuli. They are classified as either "colour-luminance" or "luminance-preferring" cells with a varying degree of cone opponency [31]. There is also evidence that chromatic features are useful for binocular correspondence in complex images, suggesting the possibility of independent contributions from both luminance and colour channels [32, 33]. In addition, it has been reported that there exist V2 neurons of macaques that are sensitive to both colour and disparity, supporting the notion that the primate visual system combines disparity and colour as early as in area V2 [34].

For the LCV-DEM implementation we initially chose the LMS colour space, which mimics the trichromatic neuronal encoding of cone responses after the LGN [30]. However, the results obtained with the LMS colour space were not significantly better than those with a simple variation of RGB (each channel codes both luminance and colour). This is not surprising. Since neuronal cells have so many different combinations of luminance or colour predominance, the system is able to be independent of the colour method used, as long as there is enough variety of weight predominance between the different colour channels. We did, however, get better results when using physiologically perceived colour weights for encoding luminance (the Y channel of the XYZ colour space), suggesting that not only disparity is heavily luminance based, but also that it depends on luminance being perceptually representative of the scene being observed.

4.1 Disparity encoding population

The extended model uses the same population parameters as L-DEM, defined in §3.1, with, in addition to points (a) to (f), point



(a) *Tsukuba* left view (b) Ground-truth (c) D^{L} (L-DEM) (d) $|\mathrm{Error}| \geqslant 0.5$ (in black)

(e) Signed D^{L} error (f) $D^{\mathrm{LC}}_{\mathrm{Left}}$ (g) $D^{\mathrm{LC}}_{\mathrm{Centre}}$ (h) $D^{\mathrm{LC}}_{\mathrm{Right}}$

(i) $\hat{D}^{\mathrm{LCV}}$ (j) D^{LCV} (k) $\hat{D}^{\mathrm{LCVE}}$ (l) D^{LCVE}

(m) D^{LCVB} (n) D^{LCVB} $|\mathrm{Error}| \geqslant 0.5$ (o) Signed D^{LCVB} error

Fig 2. Disparity results in the cell map layers for the *tsukuba* [27] stereo pair. (a): *Tsukuba*'s left view. **(b)**: Ground-truth. **(c)**: D^{L} (L-DEM) result. **(d)**: Bad pixels (black) with an absolute disparity error ≥ 0.5 and **(e)**: signed disparity error returned by the Middlebury evaluation test [8]. **(f)**: Left–viewpoint $D^{\mathrm{LC}}_{\mathrm{Left}}$. **(g)**: Centre–viewpoint $D^{\mathrm{LC}}_{\mathrm{Centre}}$. **(h)**: Right–viewpoint $D^{\mathrm{LC}}_{\mathrm{Right}}$. **(i)**: Left-viewpoint corrected $\hat{D}^{\mathrm{LCV}}$. **(j)**: Background and occlusion corrected D^{LCV}. **(k)**: Line and edge region enhanced $\hat{D}^{\mathrm{LCVE}}$. **(l)**: Object border enhanced D^{LCVE}. **(m)**: The final disparity map D^{LCVB}, after median smoothing. **(n)**: Bad pixels (black) with absolute disparity error ≥ 0.5. **(o)**: Signed disparity error of D^{LCVB}. Images (a) and (b) are reprinted from [26] under a CC BY license, with permission from Daniel Scharstein, original copyright 2002.

doi:10.1371/journal.pone.0129908.g002

g. RF dominance $\mu \in$ {Left, Centre, Right}: three values of binocular RF dominance, representing three possible configurations for RF organisation around a centre point, as further explained below.

We can improve disparity estimates by using two more RF dominances. As previously mentioned, the binocular simple cell RFs are defined by $\rho_{L,R}$ in Eq (1), where $(\dot{x}, \dot{y})$ are offset coordinates relative to the centre (0,0) and rotated to the cell's preferred orientation according to Eq (2). For μ = Left we use $(x_{L,\,R}, y_{L,\,R})$ as shown in §2, representing both RFs centered around $-\Delta x/2$. For μ = Centre the RFs are equidistant from (0,0) and their coordinates are $(x_L, y_L) = (x + \Delta x\,/2, y + \Delta y/2)$ and $(x_R, y_R) = (x–\Delta x/2, y - \Delta y/2)$. For μ = Right the RFs are shifted to the right and centered at $\Delta x/2$, resulting in coordinates $(x_L, y_L) = (x + \Delta x, y + \Delta y)$ and $(x_R, y_R) = (x, y)$.

Stereo energy coding. The LCV-DEM model also employs pairs of binocular simple cells in quadrature in order to construct phase-invariant complex cells. The responses of simple cells are obtained similarly to Eq (3), but now with the previous DEM luminance-only channel (l) complemented by three new luminance/colour channels: $c \in \{l, r, g, b\}$ with $r = R + G/4+B/4$, $g = R/4+G + B/4$, $b = R/4+G/4+B$ and l as in L-DEM (see §3). This represents luminance-colour sensitive cells with different RGB component predominance, with the l channel representing luminance-predominant cells using physiologically perceived colour weights (corresponding to the Y channel of the XYZ colour space). Responses of the left and right RFs of binocular simple cells ($v_L^{\mu,c}$ and $v_R^{\mu,c}$) are obtained by convolving (*) the RFs with the corresponding left and right images $I_{L,R}^c(x, y)$:

$$v_{L,R}^{\mu,c}(x, y) = I_{L,R}^c(x, y) * \rho_{L,R}^{\mu,c}(x, y). \qquad (11)$$

The augmented parameter set results in an encoding population of $8_\theta \times 3_{\sigma,\,f} \times 2_\phi \times 60_{\Delta x} \times 1_{\Delta\phi} \times 3_\mu \times 4_c$ = 34, 560 binocular simple cells (17,280 complex cells), twelve times larger than L-DEM due to the three different viewpoints μ and four luminance/colour channels c.

4.2 Disparity decoding population

The implementation uses the same decoding method as L-DEM, as specified in §3.2. However we are processing each of the four colour channels c independently—this allows us to show the benefits of colour without having to train the population again.

For each (x, y), the correlation (Corr) coefficient is now calculated between $\Psi_{\mu,c}^{\text{sp}}$ and $W_{f,\theta,\Delta x}^{\Delta x_{\text{stim}}}$. The correlation vector $r_{\Delta x_{\text{stim}},\,\mu,\,c}$ now holds 60 × 3 × 4 cell responses, 60 for each μ and c combination. At this step, three viewpoint–based D_μ^{LC} disparity maps are built independently (examples are shown in Fig 2f-2h). The disparities assigned to each position (x, y) will be the values d_μ of the maximum correlations, where $d_\mu = \text{argmax}_{\Delta x_{\text{stim}}} \left(r_{\Delta x_{\text{stim}},c} \right)_\mu$, over all Δx_{stim} and c values, for each μ. This yields three different disparity maps $D_\mu^{\text{LC}}(x, y) = d_\mu$. Biologically this corresponds to an activation of a single disparity cell per pixel and per viewpoint μ.

Viewpoint correction layer. Outputs from cell layers D_μ^{LC} are combined in a viewpoint correction layer, where the information from the three viewpoint disparity maps is used to select the most accurate information. This can be seen as a fusion of the disparity maps relative to the perspective of an observer with a left-side viewpoint. It is done by shifting the maps to the right accordingly (each pixel's shift distance depends on its disparity value) and by computing the median M[·, ·, ·] of the three maps:

$$\begin{aligned}\hat{D}^{\mathrm{LCV}}(x,y) \quad = \quad & \mathrm{M}\Big[D^{\mathrm{LC}}_{\mathrm{Left}}(x,y), \\ & D^{\mathrm{LC}}_{\mathrm{Centre}}(x+0.5\,D^{\mathrm{LC}}_{\mathrm{Centre}}(x,y),y), \\ & D^{\mathrm{LC}}_{\mathrm{Right}}(x+D^{\mathrm{LC}}_{\mathrm{Right}}(x,y),y)\Big].\end{aligned} \tag{12}$$

The resulting map can be seen in Fig 2(i). Combining viewpoints effectively increases the accuracy of disparity estimates at the left and right borders of objects, which are usually inaccurate due to viewpoint occlusion (i. e., each eye will see some information that the other does not). This leads to a correspondence problem, which is greater when the distance between left and right images of the pair is larger. For illustration purposes, Fig 3 shows a better example of the benefits of combining viewpoints, for the *cones* stereo pair. Here, the left and right images are more separate, with a maximum disparity of 59 pixels vs. only 15 pixels for *tsukuba*. *Cones*' disparity maps highlight the greater differences between viewpoints. The fusion of all three maps is shown in image (i), where black pixels represent uncertain disparity regions, which we address below.

Background and occlusion correction layer. The map $\hat{D}^{\mathrm{LCV}}(x,y)$ needs to be corrected in order to eliminate uncertain/unknown disparities due to incorrect disparity assignments in background regions or from occluded regions where disparities were shifted. To remove these, we use a two-step approach: First, we determine which disparity is the probable background and assign it as the farthest disparity in $\hat{D}^{\mathrm{LCV}}(x,y)$. Computationally, this process is done in four steps:

(a) Count how many cells (N_d) are activated per disparity value $d \in \{0, \ldots, 59\}$;

(b) normalise the counts by dividing each value by the square of the respective disparity: $\hat{N}_d = N_d/d^2$ (this gives less priority to the nearest/highest disparities, since it is expected that the background should be farthest);

(c) the background disparity is chosen as $d_{\mathrm{bck}} = \mathrm{argmax}_d\left(\hat{N}_d\right)$; and

(d) to every disparity value $\hat{D}^{\mathrm{LCV}}(x,y) < d_{\mathrm{bck}}$ is assigned the value d_{bck}.

Afterwards, remaining inactive disparity cells receive the minimum median cell value of the closest active disparity cells in the epipolar plane, yielding $D^{\mathrm{LCV}}(x,y)$. Results for *tsukuba* are shown in Fig 2(j). The result is much better when compared to Fig 2(i).

4.3 Experimental results

We tested the LCV-DEM on the same Middlebury stereograms used in L-DEM [26–28]. Fig 2(f) to 2(h) illustrate disparity maps for *tsukuba*—the $D^{\mathrm{LC}}_{\mathrm{Left}}$, $D^{\mathrm{LC}}_{\mathrm{Centre}}$ and $D^{\mathrm{LC}}_{\mathrm{Right}}$ images show results after luminance/colour grouping with three different viewpoints: Left, Centre and Right. The $\hat{D}^{\mathrm{LCV}}$ shows the integration of the three viewpoint maps into a single Left viewpoint (Fig 2i), and D^{LCV} shows the final LCV-DEM map after the background and occlusion correction layer (Fig 2j).

The quantitative results from the Middlebury stereo evaluation are discussed in §6, comparing L-DEM with LCV-DEM. We can visually verify (see Fig 2c and 2j) that there are several improvements from D^{L} to D^{LCV}, nevertheless, the edges and regions around objects still lack a precise boundary definition. In the next section we will explain a complementary stereo model

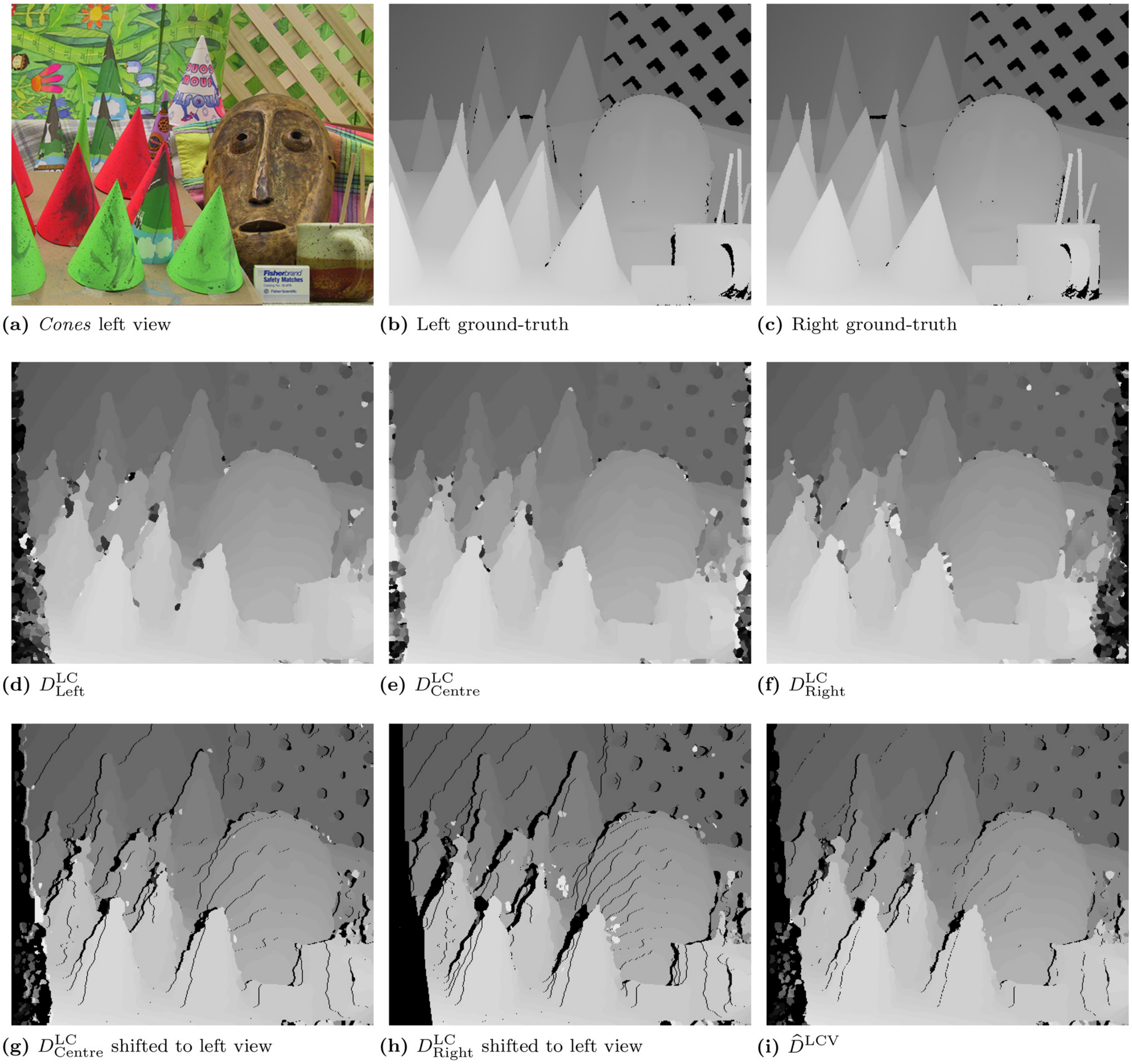


Fig 3. Example of viewpoint correction results for the *cones* [27] stereo pair. (a): *Cones* left view of the pair. **(b)**: Left viewpoint ground-truth. **(c)**: Right viewpoint ground-truth. **(d)**: Left–viewpoint $D^{\mathrm{LC}}_{\mathrm{Left}}$. **(e)**: Centre–viewpoint $D^{\mathrm{LC}}_{\mathrm{Centre}}$. **(f)**: Right–viewpoint $D^{\mathrm{LC}}_{\mathrm{Right}}$. **(g)**: Centre to Left viewpoint disparity shift. **(h)**: Right to Left viewpoint disparity shift. **(i)**: Fusion of the three Left (shifted) maps into $\hat{D}^{\mathrm{LCV}}$. Images (a), (b) and (c) are reprinted from Scharstein and Szeliski [26] under a CC BY license, with permission from Daniel Scharstein, original copyright 2003.

doi:10.1371/journal.pone.0129908.g003

to assign disparity to line and edge features, and show how the integration of both disparity maps can be achieved.

5. Boundary enhanced LCVB-DEM

Another role for monocular simple and complex cells in V1 is the ability to extract multiscale lines and edges that are significant for object categorisation and recognition [19]. If lines and edges are extracted in V1, where left and right retinal projections are close together, one might even assume that depth is attributed to them. In other words, a "3D wire-frame representation" could be built in V1 for handling 3D objects and scenes. Although this idea is speculative, many V1 cells have been found to be tuned to different combinations of frequency (scale), orientation, colour and disparity. If not coded explicitly, disparity could be coded implicitly. This allows us to develop an alternative disparity model, where we assume that lines, edges and disparity are coded explicitly—the Line and Edge Disparity Model (LEDM).

Since disparity along object borders is the biggest problem for the presented DEM models, we also integrate at this step a low-level object salience model [17] that complements line and edge information from LEDM. This allows us to combine edge conspicuity with line/edge disparity information readily available in V1/V2. Using both on top of the LCV-DEM allows us to correct disparity values astride object borders. This yields our final model, the Luminance, Colour, Viewpoint and Boundary enhanced Disparity Energy Model (LCVB-DEM).

5.1 Line and Edge Disparity Model

Line and edge detection is based on responses of even and odd monocular simple cells, corresponding to the real and imaginary parts of a Gabor filter [19]. These responses are denoted by $R_E^{s,i}(x,y)$ and $R_O^{s,i}(x,y)$, with scale s given by λ and orientation i according to θ. We used the same 8 orientations as for the binocular cells in the previous models, and scales s corresponding to $4 \leq \lambda \leq 24$ with a step size $\Delta\lambda = 2$. Positive/negative *lines* are detected where R_E has a local maximum/minimum and R_O has a zero crossing. For *edges*, the even and odd responses are swapped. In total, there are four possibilities for positive and negative Line/Edge features (L/E). An improved scheme [19] consists of combining responses of monocular simple and complex cells, i. e., simple cells serve to detect positions and L/E types, whereas complex cells are used to increase confidence. Monocular complex cell responses are modelled by the modulus $C^{s,i}(x,y) = \left[\{R_E^{s,i}(x,y)\}^2 + \{R_O^{s,i}(x,y)\}^2\right]^{\frac{1}{2}}$. Spurious cell responses beyond line and edge terminations are suppressed by lateral and cross-orientation inhibition, and assemblies of grouping cells serve to improve L/E continuity in the case of curved L/Es. We denote the line and edge cell map by $\mathrm{LE}^s(x, y)$. Fig 4 shows in (a) and (b) the multiscale line and edge coding for the *tsukuba* stereogram, at fine ($\lambda = 4$) and coarse ($\lambda = 24$) scales. Different grey levels, from white to black, show detected L/Es: positive/negative lines and positive/negative edges, respectively. We can see that many small L/Es are detected at fine scales, whereas coarse scales highlight global structures.

Keypoint maps are also exploited in the LEDM model, as these code line and edge crossings, singularities and points with large curvature. They are built from two types of end-stopped cells, single and double, which are modelled by the first and second derivatives of $C^{s,\,i}$. End-stopped responses are refined by tangential and radial inhibition to obtain precise keypoint cell maps $\mathrm{KP}^s(x, y)$ [35]. Fig 4(c) shows the *tsukuba* keypoint map at a coarse scale ($\lambda = 24$).

The disparity assigned to each L/E is based on a left–right correspondence over scales:

1. First, we suppress L/Es which may be due to noise: at each scale s of the left and right maps $\mathrm{LE}_{L,R}^s(x,y)$, we compute the maximum response of the monocular complex cells $C^{s,\,i}$ where L/Es have been detected. Any L/Es with a small amplitude ($C^{s,\,i}$ below 5% of the maximum response) are inhibited, yielding $\widehat{\mathrm{LE}}_{L,R}^s(x,y)$. The 5% threshold is necessary to eliminate detected L/Es at small gradients that do not represent region transitions. This value depends on the noise sensitivity of the Gabor responses and it was empirically determined. We found 5% to be consistently stable across many cases.

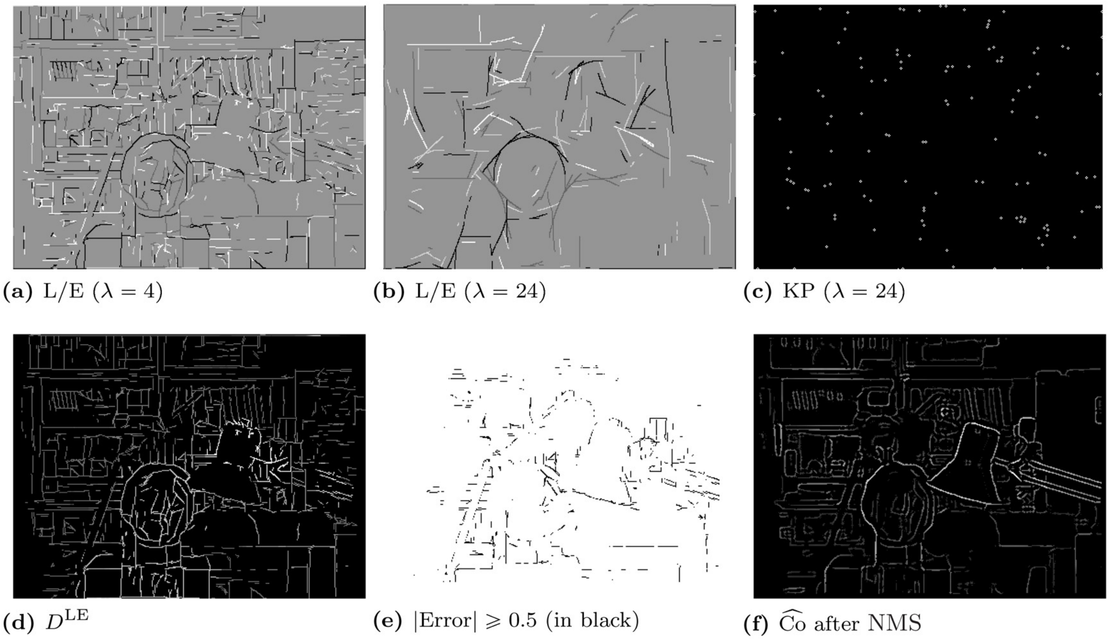


Fig 4. Line and edge disparity, and conspicuity results. (a, b): Multiscale line and edge coding at $\lambda = 4$ and $\lambda = 24$. **(c)**: Keypoint map at $\lambda = 24$. **(d)**: D^{LE} map with brightness-coded disparities. **(e)**: Bad pixels (black) with an absolute disparity error ≥ 0.5 [8]. **(f)**: Conspicuity map $\widehat{\mathrm{Co}}$ after applying INMS.

doi:10.1371/journal.pone.0129908.g004

2. In the left map, at each L/E position (x_L, y_L) and at the finest scale ($s = 1$), $\widehat{\mathrm{LE}}_L^1$ is used to define regions of interest which are centred on each L/E position (x_L, y_L). These regions are formed by grouping cells with circular RFs. At the *same* position (x_L, y_L), other grouping cells are activated at all other scales, still in the left cell map, with RF sizes depending on the scale: $2\,\lambda_s$. This *scale space* of the left cell maps (or hierarchical set of grouping cells with RFs at all L/E positions from $s = 1$) is used to accumulate displacement evidence of similar L/Es at similar scales, but with relative (shifted) positions in the right cell maps, $\widehat{\mathrm{LE}}_R^s$; see below.

Basically, the RFs serve to compare L/Es in the left and right cell maps as a function of the shift along the epipolar plane. This is done at all individual scales, after which the scales are combined. The right scale space $\widehat{\mathrm{LE}}_R^s$ can shift Δx (epipolar line) with a step size $\delta x = 1$, such that $0 \leq \Delta x \leq 59$, for a total of 60 shifts, at which the L/Es in both scale spaces are binocularly compared, according to specific rules: the Δx with the maximum L/E *correspondence* (defined below) is then assigned to the disparity map $D^{\mathrm{LE}}(x, y)$, where (x, y) still corresponds to L/E positions (x_L, y_L) of $\widehat{\mathrm{LE}}_L^1$ (Left viewpoint).

Computationally, at each scale and within each RF, four *correspondence* measures are combined with different weight factors:

(M_1) Counting all line/edge L/Es with the same position, the same type (L or E) and the same polarity (+ or –);

(M_2) As in M_1 but only counting matching L/Es irrespective of type and polarity;

(M_3) Counting the number of complex cells with similar amplitudes at all L/E positions, i. e., $|\ C_L^{s,i} - C_R^{s,i}\ | \leq 2$;

(M_4) Counting the number of keypoints with about the same coordinates in $\mathrm{KP}_{L,R}^s$, i. e., in small cell clusters of size 3×3.

Before combining the four measures, they are first normalised: M_1, M_2 and M_3 are divided by the number of L/Es in $\widehat{\mathrm{LE}}^s_L$, whereas M_4 is divided by the number of keypoints in KP^s_L, each number being computed within each respective RF. The normalised numbers are denoted by $\overline{M}_i$ and the final *correspondence* is determined by combining the weighted and normalised measures over all scales:

$$\hat{C}_{\Delta x} = \sum_s (4\,\overline{M}_1 + \overline{M}_2 + \overline{M}_3 + 4\,\overline{M}_4). \tag{13}$$

The weights for each factor were empirically determined after several trials. Finally, the horizontal disparity Δx belonging to the maximum $\hat{C}$ value is stored in the depth map $D^{\mathrm{LE}}(x, y)$. For more implementation details see Rodrigues et al. [10].

LEDM was applied to the Middlebury stereo pairs, exemplified with *tsukuba* in Fig 4. The results were very good, with disparities correctly assigned to object borders in image (d). The disparity error image (e) displays the incorrect values as black pixels, showing that almost all lines and edges have a correctly assigned disparity (80.7% at a 0.5 max error and 90.6% at a 1.0 max error).

5.2 Line and Edge region enhancement

To enhance disparity accuracy in line and edge regions and to remove small gaps we combine LCV-DEM with LEDM into an intermediate representation $\hat{D}^{\mathrm{LCVE}}$, similar to Rodrigues et al. [10]. For each L/E pixel in the D^{LE} map we define a small cluster at the L/E position plus its N_4 neighbourhood (left, right, top and bottom neighbours) and compare its median to the median of a similar cluster in D^{LCV}, at the same position. If the clusters have similar median values (less than a threshold t), the D^{LCV} cluster response at the L/E position is propagated into $\hat{D}^{\mathrm{LCVE}}$ as detailed below. Mathematically, $\forall(x, y)$,

$$\left|\mathrm{med}(\mathrm{N}_4[D^{\mathrm{LE}}(x, y)]) - \mathrm{med}(\mathrm{N}_4[D^{\mathrm{LCV}}(x, y)])\right| \leq t\,, \tag{14}$$

where $t \in \{1,\ldots,5\}$ is an integer value that represents the maximum allowed difference and med(·) the median. If Eq (14) is *false*, then the D^{LCV} cluster response is assumed to be wrong, and its region is filled in $\hat{D}^{\mathrm{LCVE}}$ using the value of $D^{\mathrm{LE}}(x, y)$. This way, we correct the LCV-DEM results using the LEDM responses. This process starts with $t = 1$ and it is applied in several cell layers, recursively, on top of the newly created $\hat{D}^{\mathrm{LCVE}}$ map, i. e., if it is not possible to fill it any more, but there are still gaps, we increment t by 1 and repeat the same procedure. In our experiments 5 was the maximum value. Biologically, this could correspond to 5 layers of $\hat{D}^{\mathrm{LCVE}}$ that activate neighbouring "idle" cells. The result can be seen in Fig 2(k), where many small regions have been corrected.

5.3 Object Boundary enhancement

Despite the above process to correct ambiguous regions, some boundaries can still be improved. In real scenes, disparity borders are mostly found at the contours of real objects, so we use a disparity sharpening process based on local contrast of disparity values, conspicuity information and line/edge boundaries to reach the final stage of this whole process, yielding D^{LCVB}—Luminance, Colour, Viewpoint and Boundary enhanced Disparity Energy Model (LCVB-DEM). This process requires three steps:

Edge conspicuity. In general, object borders are perceptually salient in a scene. In order to detect them, we first define *edge conspicuity* $\widetilde{\mathrm{Co}}(x, y)$ as a low-level V1 process. Mathematically, it is the maximum difference between colours in $I_L^c(x, y)$, with $c \in \{l, r, g, b\}$, at four pairs of symmetric positions with pixel distance $\|\vec{\delta}_i\| = 1$ from point (x, y), i. e., on horizontal, vertical and two diagonal lines [17]. Conspicuity $\widetilde{\text{Co}}(\xi)$ is the maximum Euclidean distance of all four colour pairs,

$$\widetilde{\text{Co}}(\xi) = \max_{i=1}^{4} \sqrt{\sum\nolimits_c [I_L^c((x, y) - \vec{\delta}_i) - I_L^c((x, y) + \vec{\delta}_i)]^2}. \tag{15}$$

In order to remove low responses due to small colour gradients that do not represent edges, responses lower than 10% of $\max(\widetilde{\text{Co}})$ are inhibited. A value of 10% for this threshold was found to be a consistently good choice for many cases. This value is linked to the perceptual nature of differentiating colours and is an empirically determined constant. We can think of this inhibition process as following the Weber-Fechner law (just-noticeable differences) in psychophysics, with this threshold being Weber's constant. The remaining active cells are selected by Non-Maximum Suppression (NMS), which yields conspicuity edge positions $\widehat{\text{Co}}$. Fig 4(f) shows the *tsukuba* $\widehat{\text{Co}}$ map after NMS.

Border Detection. We use a specific binary border-detection cell layer B_d that combines cell responses from $\widehat{\text{Co}}$, D^{LE} and D^{LCV}. $B_d(x, y)$ cells are only active when the following condition is *true*: $\forall(x, y) : \widehat{\text{Co}}(x, y) > 0 \vee [D^{\text{LE}}(x, y) > 0 \wedge | D^{\text{LE}}(x, y) - D^{\text{LCV}}(x, y) |> 0]$, i. e., at conspicuous borders and at lines/edges when they correspond to object borders and not to homogeneous disparity regions. Then, we devise two approaches to detect and correct bad disparity estimations by analysing regions that are *far* or *near* B_d active cells:

- The *far* case will cover regions where there are no active B_d cells nearby, i. e., regions that should have a homogeneous disparity value. Here we analyse relationships between small disparity *peaks* or *bumps* and their surrounding areas. For *peaks*, if the inside median disparity of a small cell cluster (10px radius) M_{in} is different from that of its border (outside perimeter) M_{out}, and if $M_{\text{in}} > M_{\text{out}}$ then the cell cluster disparity is reduced to the value M_{out}, eliminating the disparity peak. For *bumps*, if $M_{\text{in}} < m_{\text{out}}$, with m_{out} the minimum value of the border region (perimeter), then the cell cluster disparity is increased to m_{out}, slightly bumping the disparity depression to a coherent region background value (using M_{out} here could lead to wrong results near regions with objects, as bumps could wrongly be raised to their disparity instead).
- For regions *near* active B_d cells, i. e., near object borders, every active border in B_d activates a filling in process. We assume that the entire disparity map D^{LCV} is covered by overlapping Φ cells with RFs of 3×3 pixels and one pixel distance between their centres, which compute the median disparity in their RF. On each side and orthogonal to a B_d edge, a cluster of three orthogonal neighbouring Φ cells starts close to the edge and moves until a maximum distance of 25 pixels. If the three neighbouring cells are denoted by Φ_1 (closest to border), Φ_2 (middle) and Φ_3 (farthest from border), then disparity Φ_2 is propagated to the border at the first position where $|\Phi_2 - \Phi_3| \geq 2$ and $\Phi_2 = \Phi_1$. Hence, a stable disparity value (before the first significant disparity transition) is propagated until a B_d edge. In this process we apply median disparities in order to skip disparity changes which do not likely correspond to true object borders.

The completion of both approaches returns an enhanced disparity map D^{LCVE}; see Fig 2(l).

Median smoothing. Finally, the last step serves to correct all locally inconsistent disparities by assigning to each (x, y) position the most probable disparity within a small RF. This process is similar to a median smoothing filter and is achieved by applying circular cell clusters to

$D^{\mathrm{LCVE}}(x, y)$ (6px radius; slightly bigger or smaller sizes do not affect the global ranking in the Middlebury test, despite slightly improving/degrading individual images). This yields the final disparity map LCVB-DEM denoted by D^{LCVB}, shown in Fig 2(m).

5.4 LCVB-DEM Experimental Results

Fig 2 shows the D^{LCVE} map in (l) and the final disparity map D^{LCVB} in (m). By subjecting the last result to the Middlebury evaluation test we obtain the *"Bad pixels absolute disparity error $\geq$ 0.5"* and *"Signed disparity error"* of D^{LCVB}, respectively shown in (n) and (o). When comparing (m) with the results obtained from L-DEM in (c) we can observe significant improvements. Nevertheless, the number of pixels with wrong disparity estimates, although reduced, is still significant (see Fig 2n, at regions near depth discontinuities) and the biggest errors are located at the border of the desk-lamp and its support (Fig 2o).

Fig 5 details schematically all intermediate disparity maps needed to create the LCVB-DEM model, divided into three big layers. Our first DEM implementation from Martins et al. [7] (detailed in §3 as the L-DEM) is highlighted in grey. In the next section we will show results for other images and discuss the different disparity models qualitatively and quantitatively.

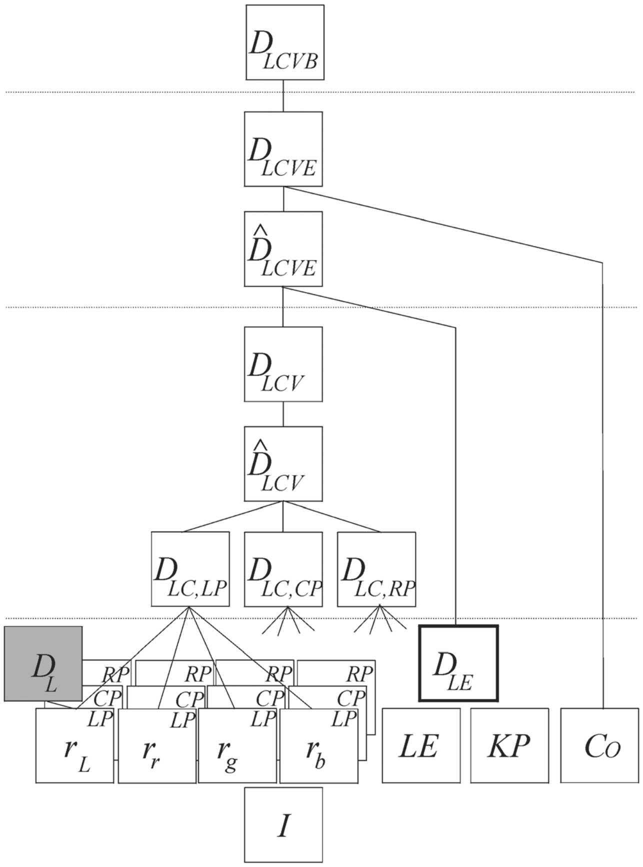


Fig 5. Summary of the different features and disparity maps leading to the LCVB-DEM.

doi:10.1371/journal.pone.0129908.g005

6. Results

As mentioned in §3.3, we also tested the model at the different implementation steps on various stereograms, including *tsukuba*, *venus*, *teddy*, *cones*, *aloe*, *cloth3*, *dolls*, *moebius* and *reindeer* [26–28]. Fig 6 shows the left view of the pair along with the left groundtruth and our final LCVB-DEM result.

Fig 7 shows the evolution of error results, using the Middlebury evaluation page [8], for *tsukuba*, *venus*, *teddy* and *cones*, at different steps of the model which correspond to the final disparity maps in each layer (see Fig 5). In the bottom table of Fig 7, top to bottom are shown: (1) L-DEM; (2) Luminance and Colour Disparity Energy Model (LC-DEM) which includes colour but not viewpoint correction; (3) LCV-DEM with colour and viewpoint; (4) LCVE-DEM with colour, viewpoint and LEDM; and finally (5) LCVB-DEM which integrates all above steps with object border enhancement. The results show improvements in all layers of the model, with the number of error pixels mostly decreasing consistently.

We can see that our model performs best in non-occluded regions but it is not as good near depth discontinuities. This was expected, because L-DEM and LCV-DEM struggle at border transitions, which is why the LEDM model is used to improve the LCV-DEM; it improves results but without yet achieving outstanding results—still, the error for regions near depth discontinuities decreases more than a factor of two in the *venus* case. The *all regions* columns refer to entire images, even regions which are half-occluded. *Avg % bad pixels* gives a general indication of how well the methods perform, as it shows the average percentage of bad pixels (wrong estimates) over all twelve columns. In all cases, the bad pixels were counted by applying the smallest error criterion possible: a disparity difference with the ground-truth greater than 0.5; for details see Scharstein and Szeliski [8].

Overall, best results were obtained for images without many small details. This is related to the size of the RFs in the cell population; smaller RFs are required to resolve the smallest details, but unfortunately they also increase binocular correspondence errors. Fig 8 shows our result when compared to the ranked results of other methods, which can include more sophisticated post-processing and top-down methodologies, like image segmentation, for yielding massively improved pixel-to-pixel correspondences. This table was replicated from the Middlebury online evaluation webpage, applying the smallest available error threshold (≤ 0.5) to emphasise that a biologically-inspired algorithm can achieve competitive results.

We can also see that the LCVB-DEM method improves the results achieved with the L-DEM (*BioDEM*) method. Overall, we achieved a good position in the average ranking table: rank 95.6 between 5.4 (best) and 159.5 (worst), on a total of 162 evaluated methods. With LCVB-DEM we significantly rise 31 positions, from position 126.6 to 95.6 (table retrieved on *13th January 2015*) relative to *BioDEM*. If we average the rankings of individual results in the columns devoted to non-occluded regions, our method would rise 20.6 positions, to rank 75. This confirms that the biggest improvement can be achieved by even more accurate estimates near depth discontinuities. Finally, to the best of our knowledge, our method is ranked highest when compared with other biologically inspired methods [12–36].

7. Discussion

We presented a hierarchical model of four disparity estimation methods, based on the biological lDEM. It can achieve good results if compared with computer vision methods [8] and it advances the state-of-the-art of biologically inspired methods [12, 36]. The advantage of the proposed DEM approach is that it does not rely on extrinsic knowledge of cell parameters to estimate disparities, requiring only trained cell populations. All used DEM-like models rely on two neuronal populations: (1) an *encoding* population that learns to discriminate disparities

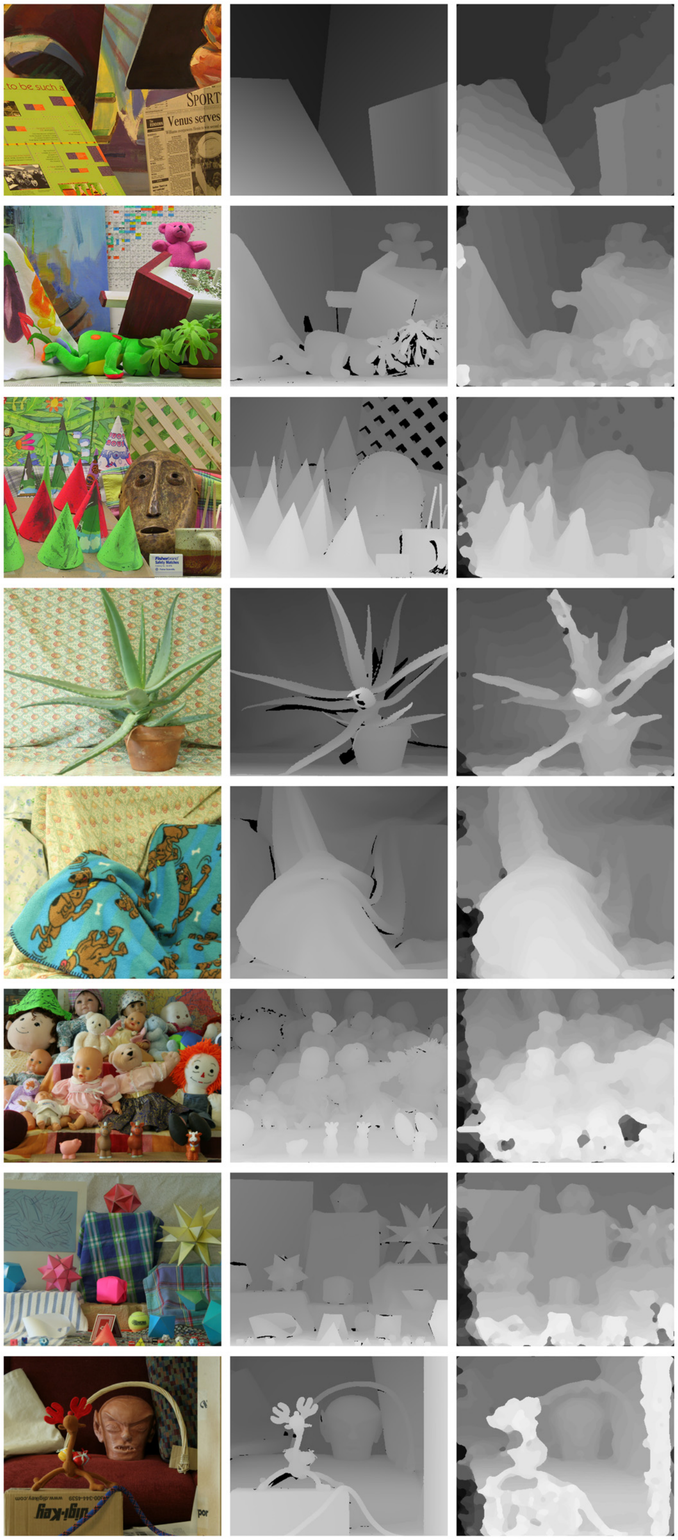
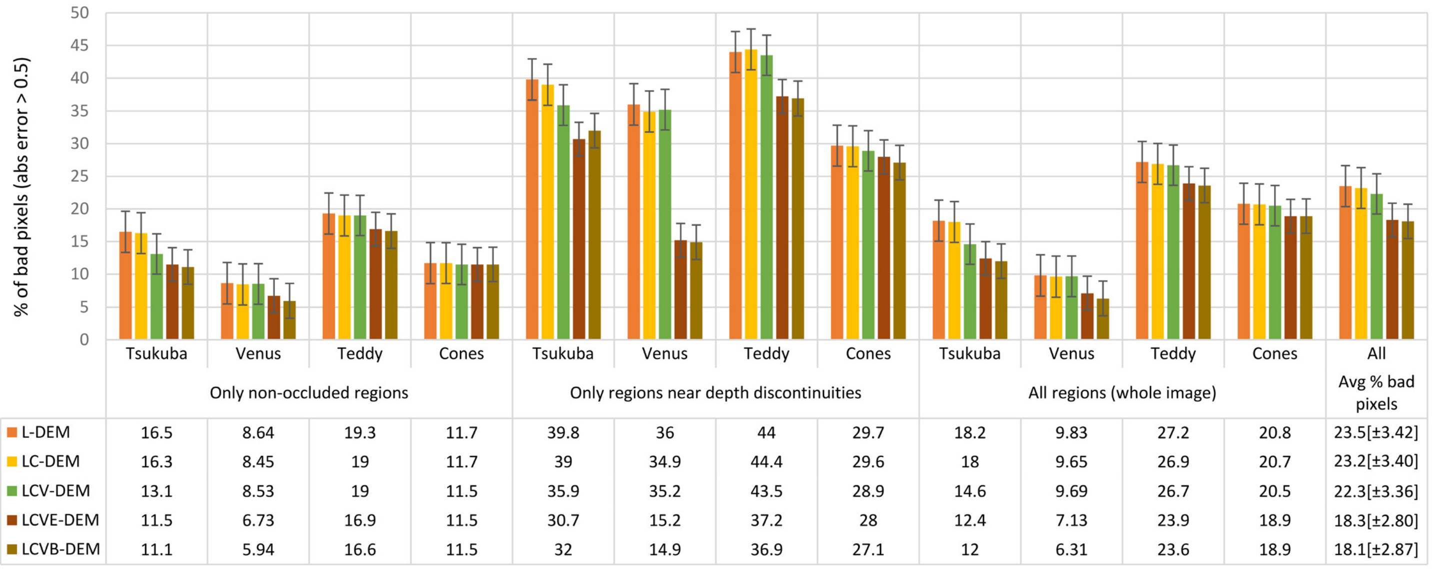


Fig 6. LCVB-DEM Middlebury dataset results. Each row shows the left view of each stereogram, its ground-truth, and final disparity result. Top-to-bottom:*venus*, *teddy*, *cones*, *aloe*, *cloth3*, *dolls*, *moebius* and *reindeer* stereograms. Images in the first and second columns are reprinted from Scharstein and Szeliski [26, 27], Scharstein and Pal [28] under a CC BY license, with permission from Daniel Scharstein, original copyright 2002–2006.

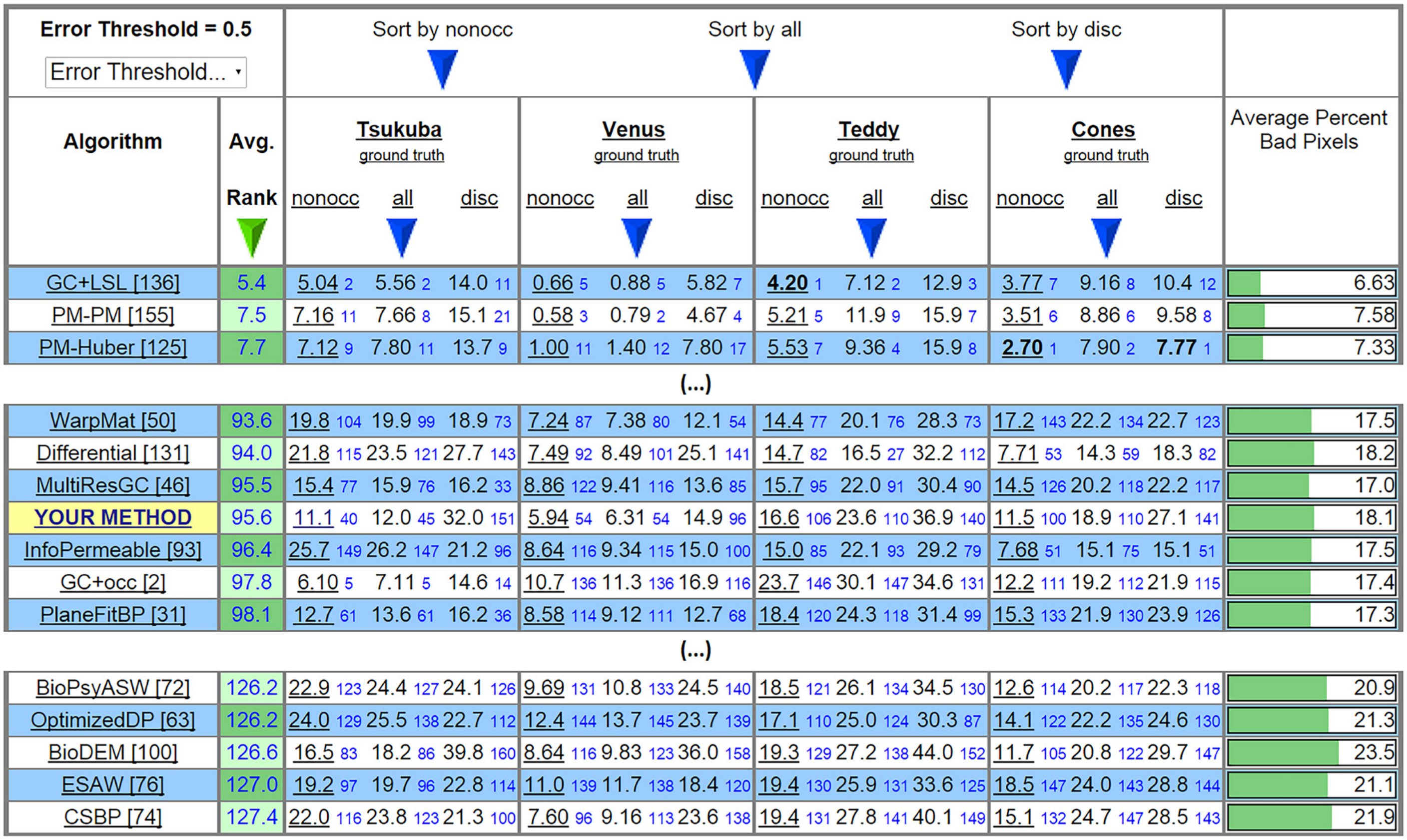
doi:10.1371/journal.pone.0129908.g006



	Tsukuba	Venus	Teddy	Cones	Tsukuba	Venus	Teddy	Cones	Tsukuba	Venus	Teddy	Cones	All
	Only non-occluded regions				Only regions near depth discontinuities				All regions (whole image)				Avg % bad pixels
🟧 L-DEM	16.5	8.64	19.3	11.7	39.8	36	44	29.7	18.2	9.83	27.2	20.8	23.5[±3.42]
🟨 LC-DEM	16.3	8.45	19	11.7	39	34.9	44.4	29.6	18	9.65	26.9	20.7	23.2[±3.40]
🟩 LCV-DEM	13.1	8.53	19	11.5	35.9	35.2	43.5	28.9	14.6	9.69	26.7	20.5	22.3[±3.36]
🟫 LCVE-DEM	11.5	6.73	16.9	11.5	30.7	15.2	37.2	28	12.4	7.13	23.9	18.9	18.3[±2.80]
🟫 LCVB-DEM	11.1	5.94	16.6	11.5	32	14.9	36.9	27.1	12	6.31	23.6	18.9	18.1[±2.87]

Fig 7. Evaluation of disparity errors, for the different model layers. Error bars represent the standard error of the mean for each model layer and are specified between brackets in the last column.

doi:10.1371/journal.pone.0129908.g007



Error Threshold = 0.5

Error Threshold... ▾

Sort by nonocc ▼ Sort by all ▼ Sort by disc ▼

Algorithm	Avg. Rank ▼	Tsukuba ground truth			Venus ground truth			Teddy ground truth			Cones ground truth			Average Percent Bad Pixels
		nonocc	all ▼	disc	nonocc	all ▼	disc	nonocc	all ▼	disc	nonocc	all ▼	disc	
GC+LSL [136]	5.4	5.04 2	5.56 2	14.0 11	0.66 5	0.88 5	5.82 7	**4.20** 1	7.12 2	12.9 3	3.77 7	9.16 8	10.4 12	6.63
PM-PM [155]	7.5	7.16 11	7.66 8	15.1 21	0.58 3	0.79 2	4.67 4	5.21 5	11.9 9	15.9 7	3.51 6	8.86 6	9.58 8	7.58
PM-Huber [125]	7.7	7.12 9	7.80 11	13.7 9	1.00 11	1.40 12	7.80 17	5.53 7	9.36 4	15.9 8	**2.70** 1	7.90 2	**7.77** 1	7.33
								(...)						
WarpMat [50]	93.6	19.8 104	19.9 99	18.9 73	7.24 87	7.38 80	12.1 54	14.4 77	20.1 76	28.3 73	17.2 143	22.2 134	22.7 123	17.5
Differential [131]	94.0	21.8 115	23.5 121	27.7 143	7.49 92	8.49 101	25.1 141	14.7 82	16.5 27	32.2 112	7.71 53	14.3 59	18.3 82	18.2
MultiResGC [46]	95.5	15.4 77	15.9 76	16.2 33	8.86 122	9.41 116	13.6 85	15.7 95	22.0 91	30.4 90	14.5 126	20.2 118	22.2 117	17.0
YOUR METHOD	95.6	11.1 40	12.0 45	32.0 151	5.94 54	6.31 54	14.9 96	16.6 106	23.6 110	36.9 140	11.5 100	18.9 110	27.1 141	18.1
InfoPermeable [93]	96.4	25.7 149	26.2 147	21.2 96	8.64 116	9.34 115	15.0 100	15.0 85	22.1 93	29.2 79	7.68 51	15.1 75	15.1 51	17.5
GC+occ [2]	97.8	6.10 5	7.11 5	14.6 14	10.7 136	11.3 136	16.9 116	23.7 146	30.1 147	34.6 131	12.2 111	19.2 112	21.9 115	17.4
PlaneFitBP [31]	98.1	12.7 61	13.6 61	16.2 36	8.58 114	9.12 111	12.7 68	18.4 120	24.3 118	31.4 99	15.3 133	21.9 130	23.9 126	17.3
								(...)						
BioPsyASW [72]	126.2	22.9 123	24.4 127	24.1 126	9.69 131	10.8 133	24.5 140	18.5 121	26.1 134	34.5 130	12.6 114	20.2 117	22.3 118	20.9
OptimizedDP [63]	126.2	24.0 129	25.5 138	22.7 112	12.4 144	13.7 145	23.7 139	17.1 110	25.0 124	30.3 87	14.1 122	22.2 135	24.6 130	21.3
BioDEM [100]	126.6	16.5 83	18.2 86	39.8 160	8.64 116	9.83 123	36.0 158	19.3 129	27.2 138	44.0 152	11.7 105	20.8 122	29.7 147	23.5
ESAW [76]	127.0	19.2 97	19.7 96	22.8 114	11.0 139	11.7 138	18.4 120	19.4 130	25.9 131	33.6 125	18.5 147	24.0 143	28.8 144	21.1
CSBP [74]	127.4	22.0 116	23.8 123	21.3 100	7.60 96	9.16 113	23.6 138	19.4 131	27.8 141	40.1 149	15.1 132	24.7 147	28.5 143	21.9

Fig 8. Middlebury Stereo Evaluation table [8], with methods ordered by total Average Rank using the strictest error threshold (0.5). LCVB-DEM is listed as "YOUR METHOD" and the original L-DEM method is listed as "BioDEM". The small blue numbers beside each column indicate the Average Rank of each individual result against all other methods. The table was retrieved on *13th January 2015*.

doi:10.1371/journal.pone.0129908.g008

from repeated presentations of random and binocularly uniform stimuli, resulting in a population activity code (i. e., a mean spike count) for each stimulus disparity; and (2) a *decoding* population that associates each code to a specific disparity value, using synaptic weights that store the mean activity of the population [1]. After foveal training, the populations are ready to evaluate disparities at any retinotopic (image) position, each local activity code being decoded into a single disparity value. Although not explored here, we also expect the decoding population to have some degree of neural plasticity and context-awareness, dynamically adapting itself to correlate the decoding weights to local image content.

All proposed models use a large number of cells: the L-DEM model starts with 2880 binocular simple cells which are combined into 1440 complex cells, at each retinotopic (image) position; LCVB-DEM increases that number to 17,280 complex cells. Nevertheless, these are trivial numbers when compared to total V1 size, estimated at about 190 million cells [37], but that number could well be near 243 million (average volume of V1 of 5,405 $\mathrm{mm}^3 \times 45{,}000$ cells/mm^3).

The role of colour in biological disparity models is still rather speculative [32], with little research into biological disparity models that employ colour, even in view of already existing evidence that disparity-sensitive neurons can also be isoluminant-sensitive [33, 34]. Meanwhile, our empirical evidence suggests that mixing colour weights may definitively play a significant role in improving the *luminance* discrimination of cells, which can significantly improve disparity estimations. Empirically, using different Y-channel luminance formulas in the XYZ colour space significantly affected the accuracy of the disparity maps, suggesting that the brain's luminance pathway (where L- and M-cone responses are combined) plays a key role in the stereo matching process by maximising the differences between regions of a scene. This is expected evolutionarily, since the brain needed to develop a robust disparity system that worked well for various survival-related tasks, especially in the dark, when scotopic colour perception is unreliable. Nevertheless, colour can still play an important role in defining disparity transitions by highlighting conspicuous object borders [17].

The role of *perspective correction*, to shift the viewpoint of disparity maps in order to yield better estimates, is also biologically plausible: even uV1 cells display the ability to shift their RFs [38]. Basically, this process increases the robustness of binocular correspondence (i. e., stereo-matching) by combining the responses of three binocular RF perspectives, instead of just one, at each image position. This is especially useful for scenes with many occlusions or periodic textures. The method chosen for perspective shifting, shown in Eq (12), could also be particularly useful for combining many different perspectives in multi-view stereo. In this paper we considered the left view, but this was because of a practical reason. In biological vision models this should be the central view in order to mimic cyclopean vision and minimise object border occlusions between left/right perspectives.

A big advantage of the models is that they exploit cell types that are already available in the cortex: monocular simple cells can be paired to construct binocular cells. They are also useful for coding lines and edges, as in the lLEDM, or even for object segregation or brightness perception [19]. Also, as shown by Pugeault et al. [9], different spatial structures can be linked both in 2D and 3D by using constraints like good continuity. These structures can be complemented with other features, like optical flow, colour and texture, to help in object recognition. The LEDM exploits the structural organisation of V1 hypercolumns, with very close left and right retinal projections, associating depth to detected lines and edges at a low level, i. e., a sort of "wireframe" representation [1]. This is useful for post-processing of DEM estimates in occluded regions, where some detail is visible in one projection but not in the other. This allows the LCVB-DEM to use LEDM and conspicuity edges to steer and correct disparity estimations on both edge sides, while smoothing disparities in regions without edges. The role of phase tuning in sharpening edge disparities is also yet to be explored [11].

Finally, we propose and illustrate that the classical DEM (L-DEM) and LEDM can be used to create a disparity "gist" map, i. e., they are robust enough to quickly draft the environment, either from binocular energy complex cells or from object contours (the bottom layer of Fig 5). Such maps are sufficient for person or robot navigation, as they are based on quickly extracted visual features in a very low-level layer. In a second layer, the DEM is combined with colour and perspective correction, giving a more accurate disparity map, but still lacking well-defined borders around objects. In the third layer, information about edges is integrated into the LCVE-DEM disparity map. The fourth and final layer sharpens object borders using saliency data on top of LCVE-DEM, yielding LCVB-DEM. In summary, we have two disparity gist-like maps, one with localised edge information (LEDM) and one with spatially inaccurate, but precise region information (L-DEM), which are later combined with colour and viewpoint to form a more robust map (LCVB-DEM).

For further research, it makes sense to explore some alternative and promising combinations of binocular cells that proved to yield more biologically accurate disparity tuning curves in rhesus monkeys [4, 5]. The role of phase-tuned cells is also an interesting topic [19, 23], as their use can be seamlessly integrated into our model, signalling false disparity matches that can be immediately corrected at a low-level.

Supporting Information

S1 Fig. Tsukuba LCVB-DEM result.
(PNG)

S2 Fig. Venus LCVB-DEM result.
(PNG)

S3 Fig. Teddy LCVB-DEM result.
(PNG)

S4 Fig. Cones LCVB-DEM result.
(PNG)

S5 Fig. Aloe LCVB-DEM result.
(PNG)

S6 Fig. Cloth3 LCVB-DEM result.
(PNG)

S7 Fig. Dolls LCVB-DEM result.
(PNG)

S8 Fig. Moebius LCVB-DEM result.
(PNG)

S9 Fig. Reindeer LCVB-DEM result.
(PNG)

Author Contributions

Conceived and designed the experiments: JAM JMFR HDB. Performed the experiments: JAM JMFR HDB. Analyzed the data: JAM JMFR HDB. Contributed reagents/materials/analysis tools: JAM JMFR HDB. Wrote the paper: JAM JMFR HDB.

References

1. Lee TS, Mumford D, Romero R, Lamme VA. The role of the primary visual cortex in higher level vision. Vision Res. 1998; 38(15-16):2429–2454. doi: 10.1016/S0042-6989(97)00464-1 PMID: 9798008
2. Szeliski R. Stereo correspondence. Comput Vis. 2011;p. 467–503.
3. Ohzawa I, DeAngelis GC, Freeman RD. Encoding of binocular disparity by complex cells in the cat's visual cortex. J Neurophysiol. 1997 Jun; 77(6):2879–909. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9212245. PMID: 9212245
4. Tanabe S, Cumming BG. Mechanisms Underlying the Transformation of Disparity Signals from V1 to V2 in the Macaque. J Neurosci. 2008 Oct; 28(44):11304–14. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2601577&tool = pmcentrez&rendertype = abstract. doi: 10.1523/JNEUROSCI.3477-08.2008 PMID: 18971472
5. Haefner RM, Cumming BG. Adaptation to Natural Binocular Disparities in Primate V1 explained by a Generalized Energy Model. Neuron. 2008 Jan; 57(1):147–58. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2344156&tool = pmcentrez&rendertype = abstract. doi: 10.1016/j.neuron.2007.10.042 PMID: 18184571
6. Read JCA. Vertical Binocular Disparity is encoded implicitly within a Model Neuronal Population tuned to Horizontal Disparity and Orientation. PLoS Comp Bio. 2010 Apr; 6(4):e1000754. Available from: http://dx.plos.org/10.1371/journal.pcbi.1000754. doi: 10.1371/journal.pcbi.1000754
7. Martins JA, Rodrigues JMF, du Buf JMH. Disparity Energy Model using a Trained Neuronal Population. In: Proc. IEEE Int. Symp. Signal Proc. Inf. Technol. Bilbao, Spain; 2011. p. 287–292. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6151575http://hdl.handle.net/10400.1/2078.
8. Scharstein D, Szeliski R. Middlebury Stereo Vision—Evaluation Webpage, http://vision.middlebury.edu/stereo/eval; 2012. Available from: http://vision.middlebury.edu/stereo/eval.
9. Pugeault N, Woergoetter F, Krueger N. Disambiguating multi-modal scene representations using perceptual grouping constraints. PLoS One. 2010; 5(6):e10663. doi: 10.1371/journal.pone.0010663 PMID: 20544006
10. Rodrigues JMF, Martins JA, Lam R, du Buf JMH. Cortical multiscale line-edge disparity model. In: Proc. Int. Conf. Image Anal. Recognit. Aveiro, Portugal: Springer LNCS 7324; 2012. Available from: http://w3.ualg.pt/~jrodrig/papers_pdf/iciar2012.pdf.
11. du Buf JMH, Terzic K, Rodrigues JMF. Phase-differencing in stereo vision: solving the localisation problem. In: Proc. 6th Int. Conf. Bio-inspired Syst. Signal Process. Barcelona, Spain; 2013. p. 254–263. Available from: http://w3.ualg.pt/~jrodrig/papers_pdf/biosignals2013.pdf.
12. Mutti F, Gini G. Bio-inspired disparity estimation system from energy neurons. Proc IEEE Int Conf Appl Bionics Biomech. 2010; 1(1):1–6.
13. Read JCA, Cumming BG. Does Depth Perception require Vertical-Disparity Detectors? J Vis. 2006 Jan; 6(12):1323–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17209738. doi: 10.1167/6.12.1 PMID: 17209738
14. Banks MS, Gepshtein S, Landy MS. Why is Spatial Stereoresolution so low? J Neurosci. 2004 Mar; 24(9):2077–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14999059. doi: 10.1523/JNEUROSCI.3852-02.2004 PMID: 14999059
15. Filippini H, Banks M. Limits of Stereopsis explained by Local Cross-Correlation. J Vis. 2009; 9(8):1–18. Available from: http://www.journalofvision.orgwww.journalofvision.org/content/9/1/8.short.
16. Tsai JJ, Victor JD. Reading a Population Code: a Multi-Scale Neural Model for representing Binocular Disparity. Vis Res. 2003 Feb; 43(4):445–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12536001. doi: 10.1016/S0042-6989(02)00510-2 PMID: 12536001
17. Martins JA, Rodrigues JMF, du Buf JMH. Local object gist: meaningful shapes and spatial layout at a very early stage of visual processing. Gestalt Theory. 2012; 34(3/4):349–380. Available from: http://gth.krammerbuch.at/content/vol34-issueheft3-4.
18. Chen Y, Qian N. A Coarse-to-Fine Disparity Energy Model with both Phase-Shift and Position-Shift Receptive Field Mechanisms. Neural Comput. 2004; 16(8):1545–1577. Available from: Object Categorization and Recognition, and Consciousness. BioSystems. 2009;95:206–226. Available from: http://www.sciencedirect.com/science/article/pii/S0303264708002372. doi: 10.1162/089976604774201596 PMID: 15228745
19. Rodrigues JMF, du Buf J. Multi-scale Lines and Edges in V1 and beyond: Brightness, Object Categorization and Recognition, and Consciousness. BioSystems. 2009; 95:206-226. Available from: http://www.sciencedirect.com/science/article/pii/S0303264708002372.

20. Skottun BC, Freeman RD. Stimulus specificity of binocular cells in the cat's visual cortex: ocular dominance and the matching of left and right eyes. Exp Brain Res. 1984; 56:206–216. doi: 10.1007/BF00236275 PMID: 6479258

21. Poggio GF, Motter BC, Squatrito STY. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. Vis Res. 1985; 25:397–406. doi: 10.1016/0042-6989(85)90065-3 PMID: 4024459

22. Read JCA, Phillipson GP, Glennerster A. Latitude and Longitude Vertical Disparities. J Vis. 2009 Jan; 9(13):11.1–37. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2837276&tool = pmcentrez&rendertype = abstract. doi: 10.1167/9.13.11

23. Read JCA, Cumming BG. Sensors for Impossible Stimuli may solve the Stereo Correspondence Problem. Nat Rev Neurosci. 2007 Oct; 10(10):1322–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2075086&tool = pmcentrez&rendertype = abstract. doi: 10.1038/nn1951

24. Yang S, Yao X. Population-Based Incremental Learning With Associative Memory for Dynamic Environments. IEEE Trans Evol Comp. 2008 Oct; 12(5):542–561. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4454713. doi: 10.1109/TEVC.2007.913070

25. Read JC, Eagle RA. Reversed Stereo Depth and Motion Direction with Anti-correlated Stimuli. Vis Res. 2000 Jan; 40(24):3345–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11058733. doi: 10.1016/S0042-6989(00)00182-6 PMID: 11058733

26. Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int J Comput Vis. 2002; 47:7–42. doi: 10.1023/A:1014573219977

27. Scharstein D, Szeliski R. High-accuracy stereo depth maps using structured light. Proc IEEE Comp Soc Conf Comp Vis Pattern Recogn. 2003;p. 195–202.

28. Scharstein D, Pal C. Learning Conditional Random Fields for Stereo. Proc IEEE Conf Comp Vis Pattern Recogn. 2007;p. 1–8.

29. Field GD, Chichilnisky EJ. Information processing in the primate retina: circuitry and coding. Annu Rev Neurosci. 2007; 30:1–30. doi: 10.1146/annurev.neuro.30.051606.094252 PMID: 17335403

30. Wade A, Augath M, Logothetis N. fMRI measurements of color in macaque and human. J Vis. 2008; 8(10):1–19. Available from: http://w.journalofvision.org/content/8/10/6.short. doi: 10.1167/8.10.6

31. Johnson EN, Hawken MJ, Shapley R. Cone inputs in macaque primary visual cortex. J Neurophysiol. 2004; 91(6):2501–2514. doi: 10.1152/jn.01043.2003 PMID: 14749310

32. den Ouden HEM, van Ee R, de Haan EHF. Colour helps to solve the binocular matching problem. J Physiol. 2005; 567(2):665–671. Available from: http://jp.physoc.org/content/567/2/665.abstract. doi: 10.1113/jphysiol.2005.089516 PMID: 15975983

33. Krauskopf J, Forte JD. Influence of chromaticity on vernier and stereo acuity. J Vis. 2002; 2(9):6. Available from: http://ww.journalofvision.org/content/2/9/6.short. doi: 10.1167/2.9.6

34. Ts'o DY, Roe AW, Gilbert CD. A hierarchy of the functional organization for color, form and disparity in primate visual area V2. Vision Res. 2001 May; 41(10-11):1333–1349. Available from: http://dx.doi.org/10.1016/S0042-6989(01)00076-1. doi: 10.1016/S0042-6989(01)00076-1 PMID: 11322978

35. Rodrigues JMF, du Buf J. Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. BioSystems. 2006; 2:75–90. Available from: http://sapientia.ualg.pt/handle/10400.1/181. doi: 10.1016/j.biosystems.2006.02.019

36. Nalpantidis L, Gasteratos A. Biologically and psychophysically inspired adaptive support weights algorithm for stereo correspondence. Rob Auton Syst. 2010; 58(5):457–464. Available from: http://www.sciencedirect.com/science/article/pii/S0921889010000394. doi: 10.1016/j.robot.2010.02.002

37. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? Neuron. 2012 Feb; 73(3):415–34. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3306444&tool = pmcentrez&rendertype = abstract. doi: 10.1016/j.neuron.2012.01.010 PMID: 22325196

38. Fu YX, Shen Y, Gao H, Dan Y. Asymmetry in Visual Cortical Circuits Underlying Motion-induced Perceptual Mislocalization. J Neurosci. 2004; 24(9):2165–2171. Available from: http://citeseer.uark.edu:8080/citeseerx/viewdoc/summary;jsessionid=D810CAB6AEFC6DE425B25A76353637E6?doi=10.1.1.134.4746. doi: 10.1523/JNEUROSCI.5145-03.2004 PMID: 14999067