Encoding of 3D surface priors in the spatial pattern of interaction among V1 disparity neurons

Xiong Li¹,², Jason Samonds¹, Yuncai Liu² and Tai Sing Lee¹

¹Center for the Neural Basis of Cognition and Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
²Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China.

Our perceptual systems are shaped by our natural environment. Statistical studies on natural scenes have provided important insights to the development of simple cell receptive fields and the connectivity among orientation selective neurons in the primary visual cortex. Here, we extended such study to the neural encoding of statistical regularities of 3D surfaces in natural scenes. Analysis of 3D range data of natural scenes suggests the correlation of binocular disparity signals at two spatial locations is a function of the relative spatial orientation between these locations: stronger correlation for cardinal orientations than for non-cardinal orientations. This cardinal preference likely arises from the prevalence of upright or horizontal objects in natural scenes. We investigated the implications of this statistical regularity in the neuronal connectivity in the primary visual cortex. We showed that this statistical regularity is reflected in the pairwise interaction between disparity neurons in V1, indicative of a pattern of underlying connectivity that can serve as a surface prior to facilitate 3D perceptual inference.

1 Introduction

Bayesian inference has long been proposed as a fundamental computational principle in the brain. The earliest idea can be traced back to Helmholtz’s theory of unconscious inference to reason about perception. Recently, the Bayesian paradigm has gained increasing popularity in psychology and neural theory, with empirical psychological evidence coming from the sensory, decision and motor systems. Central to Bayesian perceptual inference is the encoding of statistical regularities in natural scenes as priors to support the inference process. Significant progress has been made in understanding scene statistics of edges and boundary contours, where studies in scene statistics have demonstrated that the orientation association field characterized in psychophysical studies can be linked to the co-occurrence statistics of edge elements in natural scenes. Neurophysiological and anatomical evidence further suggest that the association field are likely implemented in the primary visual cortex (V1) using horizontal recurrent connec-
tions among orientation tuned neurons \(^?\). On the other hand, relatively little is known about the neural encoding of 3D surface regularities underlying 3D surface perceptual inference and completion. It is possible that co-occurrence statistics of the binocular disparity signals derived from 3D natural scenes could also be encoded in the recurrent connections in the primary visual cortex, building a disparity association field. In this paper, we tested this idea by evaluating the correspondence between the spatial distribution of correlation between disparity signals in natural environment and the spatial distribution of functional connectivity between disparity neurons in the primary visual cortex.

We measured the correlation between disparity signals derived from databases of 3D range data from natural scenes and found the correlation varies as a function of the spatial orientation between two points in the image space: stronger along cardinal spatial orientations (horizontal and vertical) than along the non-cardinal orientations. To understand how correlation in natural scene signals can be linked to neuronal connectivity, we simulated a network of disparity neurons to assess the correlation of the model neurons’ spiking activities in response to these natural scene signals. The stimulus induced correlation in neurons can induce connectivity between neurons over time based on the established Hebbian learning mechanism. We carried out neurophysiological experiments using multi-electrode arrays to study correlated activities of disparity-tuned neurons in response to stimuli devoid of the cardinal preference in correlation and nevertheless found that correlation between simultaneously recorded disparity-tuned neurons in V1 exhibited the same pattern of cardinal spatial orientation preference. This cardinal effect in pairwise interaction remains valid when the neuronal interaction is analyzed in the context of a larger network using Markov random field Ising model approach. A predominantly negative third order interaction revealed by the Markov random field model suggests that the pairwise interactions operate in a network with additional sparsity or normalization constraints.

To assess the possible contribution to this effect by the neural circuit underlying the contour orientation association field between orientation-tuned cells, we also examined the co-occurrence statistics of oriented elements in natural scenes and concluded that orientation association field cannot give rise to this cardinal spatial orientation effect we observed in disparity neurons. We concluded that this effect is a signature of a statistical regularity in natural scenes intrinsic to the disparity domain. This supports the existence of a disparity association field dissociable from the orientation association field. Taken together, the findings reported here provide evidence that statistical regularities in 3D surfaces in natural scenes are encoded in the pairwise connectivity between disparity neurons in the primary visual cortex. These pairwise connections can serve as
surface priors to facilitate 3D perceptual inference, as has been demonstrated in a large class of Markov random field models in computer vision.

2 Results

**Depth statistics in 3D natural scenes** The prevalence of upright and horizontal structures in natural scenes has given rise to the more frequent appearance of vertical and horizontal oriented edges in natural images. This has lead to more neurons tuned to the cardinal orientations, leading to a higher sensitivity to the cardinal orientation as measured in perceptual experiments. To understand the statistical regularities of the surfaces of these same structures in natural scenes provided to our binocular visual experience, we analyzed frequency of occurrence and co-occurrence of binocular disparity signals derived from two range image data sets, the CMU data set and the Brown data set. Using the same method used in Liu et al., we can compute a disparity image from a range data image relative to a fixation depth point. The computed binocular disparity images would vary as the ‘fixation point’ moves from one location to another location in the scene image, assuming the two eyes verge to the depth values of that particular point. As our objective is to evaluate the implication of scene statistics in neural coding, and since our neurophysiological experiments were done on disparity-tuned neurons within 6-degree visual angle eccentricity in the parafoveal regions of V1, we studied only the binocular disparity signals within 12 × 12 degree visual angle window centered at the fovea. Fixation points were sampled at random from each range image. Random sampling approximated human fixation experience, as Liu et al. have shown the distribution of fixation depth roughly matches the distribution of absolute depths in a visual scene.

Fig. 1(a) shows an example window (red square) surrounding one fixation point (red dot) in an image range in a forest scene. The range image within this window was converted to binocular disparity values from absolute depth values based on the position and the depth of the fixation point. This binocular disparity image is shown in Fig. 1(b). To match spatial receptive field locations of the neurons, we select a target location within this image on a circle 4 degree eccentricity away from the center or fovea, frequency distribution of disparity values at this target location, or co-occurrence frequency distribution of disparity values surrounding this location at different spatial orientation configuration were studied. The target location is specified by $(\delta, R)$. The spatial orientation configuration around the target location is specified by $\theta$ in Fig. 1(b). In this study, we quantized the angle $\theta$ to 12 spatial orientations at 15 degree steps to span 180 degrees. Fig. 1(c) illustrate three chosen random target locations on a disparity image, each shown with four possible
spatial orientation configurations.

The distribution of the frequency of occurrence of disparity values (first order measures) at a million (1.27 million) of such target positions follows a Laplacian distribution, as shown in Fig. 1(d), similar to earlier report by Liu et al.\textsuperscript{10} The second order co-occurrence statistics of disparity values at two locations surrounding target points (1 degree apart), measured in terms of Pearson correlation across 106 images, is found to depend on the spatial configuration orientation $\theta$, as shown in Fig. 1(e), where four example spatial orientation configurations are also illustrated as icons. The correlation in disparity values is significantly stronger along the cardinal directions or orientations (i.e. $\theta = 0^\circ/180^\circ$ (horizontal) and 90 degrees (vertical)) than the other noncardinal orientations. We call this cardinal preference in spatial disparity correlation.

This cardinal preference could arise from the prevalence of vertical and horizontal structures in our natural environment due to the pull of gravity: trees and buildings tend to stand upright or lie down horizontally, where depth values are more similar along vertical lines or along the horizontal lines respectively in the images. As we can see on the body of the dinosaur in Fig 1(c), more points along the cardinal orientations belong to the same surfaces and hence share the same depth, whereas in the noncardinal orientations, points are more likely to reside in two distinct surfaces. To test this idea, we computed the distribution of the disparity differences in the cardinal orientation, and that in the $\theta = 45, 135$ diagonal orientation, and compared the standard deviation and the peaks of the two distributions. The standard deviation and the peak of the disparity difference is significantly broader (std: 0.378 vs 0.36, ($p = 2.9 \times 10^{-10}$) and statistically significantly lower ($p = 7.6 \times 10^{-19}$) respectively for the diagonal orientation. This broader distribution in the diagonal orientation supports the idea that noncardinal locations are more likely to be on different surfaces, but it is also consistent with local surfaces being more likely to tilt vertically or horizontally than along the diagonal. For example, the sides of a standing tree trunks are characterized by local surfaces that tilt horizontally, whereas the ground plane are tilted vertically. To examine this possibility, we fitted local surface patches with planes of different slants and tilts, and found that vertically and horizontally tilted planes are indeed more prevalent than the diagonal planes (Fig 1(f)). These two factors, higher likelihood of similar depth along the horizontal and vertical lines, and the higher likelihood of vertical and horizontal tilt surfaces, are complementary. Both are related to the prevalence of vertical and horizontal structures and both likely contribute to the cardinal preference effect in spatial correlation in the depth domain.
Figure 1: (a) A sample of range image from Brown data set. (b) The spatial arrangement of an unbiased stimulated network for disparity statistics. $O$ is the fixation point when computing relative disparity, which is randomly chosen in statistics. $\theta$ and $d$ are respectively the configural orientation and receptive field distance (in pixel) of that pair. (c) An illustration of pairs in different configural orientations. (d) The probability density $p(s)$ of disparity sampled from Brown and CMU data sets. The positive mean (0.0113) indicates that the mean disparity slightly prefers ‘close’. (e) The Pearson’s correlation of disparity pair as a function of the configural orientation. The correlation in cardinal ordination is significantly stronger than that in non-cardinal orientation ($p = 0.026 < 0.05$). (f) Frequency of tilted planes in natural scenes as a function of slants.
Functional connectivities predicted by scene statistics  Hebbian learning, which dictates that neurons fire together will wire together, predicts correlation in disparity signals across space in natural scenes can induce correlation in neuronal spiking activity between appropriately tuned neurons, leading to stronger connectivity between these neurons. To understand how disparity signals correlation can be translated into neuronal connectivity, we simulated a neuronal network composed of hypercolumns of disparity neurons and exposed this network to disparity signals from natural scenes to compute their response correlation at different spatial orientation configuration. The set of disparity-tuned neurons can be predicted from the distribution of disparity signals in natural scenes (Fig. 1d) based on the Infomax principle, following Ganguli and Simoncelli’s works 8. The tuning curves of this set of neurons in each hypercolumn is shown in Fig. 2(a). The density of neurons at each disparity is proportional to scene disparity distribution $p(s)$, where $s$ is scene disparity, and the tuning width is proportional to $p(s)^{-1}$ 8. This set of neurons is designed to optimally encode the disparity signals experienced at each space location relative to the fixation point. We follow Ganguli and Simoncelli’s works and constrain the tuning curves to be Gaussians for simplicity 8. While real neurons have heterogeneous tuning curves to disparity, but they all exhibit individual preference to certain range of disparity and are approximated reasonably well by Gaussian tuning curves.

In the simulation, disparity signals drawn from each of image patches from natural scenes were presented to the set of model disparity-tuned neurons in the the hypercolumns in each spatial orientation configuration, as shown in Fig. 2(b). Each model neuron’s spikes are generated by an independent homogeneous Poisson process. The correlation of the spiking activities of each pair of disparity neurons in the two hypercolumns was computed and quantified using a standard CCH measure, which is the area under the peak in the cross-correlation histogram (CCH) within a +/- 15 msec window. The same CCH measure will be used to measure functional connectivity between neurons in the physiological experiments (see Methods and 7). Since neurons in the simulated network do not have horizontal connections, the CCH response correlation of the neurons reflects solely the stimulus correlations, rather than functional connectivity. Hebbian learning predicts that correlated activities arise from signal correlations are ultimately encoded in the connectivity of the neurons.

Fig. 2(c) shows the averaged CCH correlation measures across all possible pairs of model disparity neurons in the two corresponding hypercolumns in each spatial orientation configuration. This average CCH measure, a summary measure of correlated activities of disparity neurons in the two hypercolumns, shows the same cardinal preference as the spatial Pearson correlation measure.
Figure 2: (a) A population of disparity tuning curves given by the Infomax principle, where the density is proportional to \( p(s) \) and the tuning width is proportional to \( p(s)^{-1} \). (b) Illustration of the stimulated neural network. The neurons in a cortical column share the same spatial receptive field, quantized by \((x, y)\). \( \theta_{12} \) is the configural orientation. (c) The average CCH of stimulated neural pairs as a function of the configural orientation, where the effects of unbiased sampling is removed (see Methods). (d) The correlation between CCH and DSI. The correlation in cardinal configural orientation is greater than that non-cardinal configural orientation (0.58 > 0.49, \( p = 0.019 < 0.05 \)).

on the disparity signals shown in Fig. 1(e). This summary measure furnishes a prediction that we will test in neurophysiology.

One of the sources underlying the cardinal preference in disparity correlation in natural scenes is that surfaces tend to be more continuous and different image points are more likely to share similar depth values along the cardinal orientations than along the noncardinal orientations, as discussed in the last section. This predict that similarly tuned neurons across hypercolumns in the cardinal orientation configurations would exhibit stronger correlation in their activities than their counterparts in the noncardinal orientation configurations. Also, in the cardinal orientation, across two hypercolumns, neurons of similar disparity tunings will have much stronger connections than neurons of dissimilar disparity tunings and have a greater share of overall connectivity. On the other hand, because depth values in the noncardinal orientation configurations tend to exhibit greater variation, neurons with dissimilar disparity tunings might have a greater chance of experiencing correlated input signals in the noncardinal configurations than in the cardinal configurations. Thus, the connectivity is more independent of the neurons’ disparity tunings, and more diffuse across disparity-tuned neurons in the two hypercolumns in the noncardinal orientation. To evaluate this conjecture, we plotted a scatter plot of the CCH of each model neuron pairs against the disparity similarity index (DSI) of the pairs in Fig. 2(d). The data points were divided into
the cardinal group (+/- 22.5 degree from vertical or horizontal) and non-cardinal group (+/- 22.5 degrees from diagonals). We found that there is indeed a correlation between correlated activities (or predicted connectivity) and the degree of similarity of the two neurons’ disparity tuning for both groups, indicating neurons of similar disparity tunings tend to be more strongly correlated in activity or more strongly wired together. The correlation is however stronger for the cardinal orientation than for the noncardinal orientation \((r_{0.58} > r_{0.49})\), and that the linear regression fit has a steeper slope for the cardinal group. The noncardinal group’s CCH is less dependent on disparity tuning similarity and appears to be stronger than the cardinal group for neurons that are dissimilar in disparity tunings. This confirms our conjecture and provides a second quantitative prediction to be tested in neurophysiology.

**Cardinal Preference in the correlated activities of disparity tuned neurons** To evaluate the two predictions on cardinal effects in neuronal connectivity, we recorded from neurons in the primary visual cortex in response to disparity stimuli using multi-electrode recording techniques (See method, as well as \(^5\)). Stimuli tested were 12 Hz dynamic random dot stereograms (DRDS) that was presented for one second each trial while the monkey performed a fixation task. Each DRDS defines a disparity attribute that is uniform inside a 3-5 degree aperture window over the receptive fields of the neurons being recorded. The stimuli were used to obtain disparity tuning of the neurons but were also appropriate to test for the cardinal effect in connectivity. This is because even though the disparity cue is uniform, and thus there is stimulus correlation across the stimulus, this correlation is nevertheless isotropic and uniform across space, invariant to the different spatial orientation configurations.

For each neuron, we measured its spatial receptive field, preferred disparity tuning and orientation tuning curves. For each pair of neurons, we computed five measures: cross-correlation histogram (CCH) (see Methods) \(^5\); disparity tuning similarity (DSI); orientation tuning similarity (OSI); spatial configural orientation \(\theta\), and distance \(d\) between the spatial locations of their receptive fields (Fig. 3(c)). The correlation measure (CCH) is known to depend on the distance \(d\) \(^7\) between the pair. To remove the effect of \(d\) on the variation of CCH, we considered only cell pairs with distance \(d\) within a particular range (middle 25% 75%) and computed the average CCH for neuronal pairs in this set for each spatial configuration orientation \(\theta\). The results for the three monkeys confirmed the two predictions derived from scene statistics. First, the functional connectivity CCH measure is significantly stronger for the cardinal spatial orientations than for the non-cardinal spatial orientations (Fig. 3(d)). Given the relative dependence on spatial orientations among the three monkeys are qualitatively the same, we averaged the results from the three mon-
keys as shown in Fig. 3(e). The average CCH in cardinal orientation is significantly greater than that in non-cardinal orientation for each of the three monkeys (what is the p values here?), as well as their combined average ($p = 2.5 \times 10^{-5}$).

Second, the correlation between functional connectivity (CCH) and disparity tuning similarity (DSI) in this set of disparity tuned pairs is also consistent with the prediction from scene statistics. For the 161 pairs from the three monkeys, we found correlation between CCH and DSI to be stronger, with a more positive regression slope for the cardinal orientation than for the non-cardinal orientation (Fig. 3(f)), exactly as predicted by the neural correlated activities based on the scene statistics (Fig. 2(d)). The overall correlation between CCH and disparity tuning similarity (DSI) to be 0.46. The difference in correlation between CCH and DSI between the cardinal group and the non-cardinal group, i.e. $r_{\text{card}} > r_{\text{non}}$, was shown to be statistically significant as follows: we divide the data points of cardinal/nondinal orientation into several groups and compute $r_{\text{card}}/r_{\text{non}}$ for each group. We perform t-test over the two sets of $r_{\text{card}}$ and $r_{\text{non}}$. The p value of the test was found to be $p = 4.5 \times 10^{-3}$.

How about the distribution or composition of the pairs?

**Cardinal preference in functional connectivity is not predicted by orientation association network** To what extent the cardinal preference in functional connectivity we observed is due to the V1 network among orientation selective neurons, known as the orientation association field ??. In an orientation association field, an orientation-selective neuron would receive facilitation from neurons of the same orientation preference along the longitudinal direction but would receive inhibition from neurons of the same orientation on its two lateral flanks. In contour computation, longitudinal facilitation can help complete contour, and lateral inhibition could suppress alternative contour hypothesis in a neighborhood, called non-maximum suppression. The connection pattern of the orientation association field among vertically oriented neurons is supposed to exhibit strong connectivity in the vertical spatial orientation configuration but weak or negative connectivity in the horizontal spatial orientation configuration. The connection patterns among the horizontally oriented neurons is supposed to exhibit the strong in the horizontal spatial orientation configuration, and weak or negative connectivity in the vertical spatial orientation configuration. Given the orientation specific nature of the association field, the superposition of all the association fields between neurons in two hypercolumns, with mostly weak connections at best, or potential cancellation at worst, would produce a much weaker overall connection strength across the two hypercolumns in each spatial orientation configuration. In contrast, strong connections exist between
Figure 3: (a) Recording neural activities from monkey A and B using tunable electrodes. (b) Multielectrode array used to record monkey C. 96 electrodes in total. (c) The recovered spatial receptive fields of neurons. $\alpha$ is the configural orientation of the neurons recorded by channels 8 and 86. (d) The spatial correlation pattern – the average CCH as a function of the configural orientation – of monkeys A, B and C. CCH in cardinal group is greater than CCH in non-cardinal group ($p = 0.0 < 0.05$). (e) The average spatial correlation pattern of three monkeys. (f) Scatter points over CCH and DSI. The correlation between CCH and DSI depends on the configural orientation. The correlation in cardinal orientation is stronger than that in the non-cardinal orientation ($0.5 > 0.3$, $p = 0.005 < 0.05$). The effect of the receptive fields distance is removed.
all similarly disparity-tuned neuronal pairs in the adjacent hypercolumns. So the overall strength between disparity neurons due to the superposition of all the disparity association fields, across hypercolumns are much stronger. Intuitively, this is because contour association requires connection along a line or a curve, whereas surface completion requires connections over a 2D surface. These intuitions lead us to conjecture a weaker overall connectivity and potentially negligible cardinal preference effect in connectivity among orientation-selective neurons than the connectivity pattern among disparity-tuned neurons.

To evaluate this conjecture, we compute the prediction on the connectivity strength of orientation-tuned neurons based on the statistics of edge signals in natural images, in a manner similar to what we have done for the network of disparity tuned neurons for 3D scenes as shown in Fig. 2. Orientation edge signals were first extracted using Gabor-based complex cell models from natural images \[32\] and then their first order occurrence distribution and second order co-occurrence distribution were computed. The occurrence frequency of the oriented edge elements are found to favor the vertical and horizontal cardinal orientations, as is already known from other scene statistics and psychophysics study \[7,7\]. We used Ganguli and Simoncelli’s paradigm \[8\] to obtain a distribution of orientation tunings (see Fig. 4(a)) and simulated a neural network with hypercolumns populated by orientation selective neurons following such a distribution. The correlated activity of neuronal pairs in response to natural images, measured in averaged CCH (across all orientation-tuned neuronal pairs between two adjacent ‘hypercolumns’) as a function of spatial configural orientation is summarized in Fig. 4(b). As predicted, a weak trend of cardinal preference in connectivity can be observed, but the difference between the cardinal and the non-cardinal averaged CCH is not statistically significant \(t-test, p = 0.00 > 0.05\). Thus, despite there is a cardinal effect in the distribution of edge orientation in scene statistics, the asymmetric co-occurrence statistics and the overall weak co-occurrence statistics among edges of different orientation do not predict a significant cardinal effect in connectivity among orientation selective neurons in the primary visual cortex. This suggests the cardinal effect we observed among the disparity neurons are not due to the orientation association fields.

Next, we evaluated whether the disparity neurons we studied follow the prediction of connection patterns of orientation association field. We first separate the cell pairs into four major spatial orientation configurations. For each spatial configuration, we divided the neuron pairs into three groups based on the two neurons’ orientation tunings: the ‘best group’ contains neuronal pairs with the strongest expected connectivity as the neurons’ orientations are aligned along the longitudinal direction; the ‘worst group’ contains neurons with the weakest expected connectivity.
as neurons of similar orientation preference with their spatially receptive fields arranged laterally (see icons Fig. 4(c)); the medium group contains all other pairs of neurons, which are the majority, in which the neurons in each pair are of different orientation preferences and weak connectivity is also expected based on orientation association field. Fig. 4(c) shows that among the disparity neurons we studied, the average CCH of the best group was not stronger than that of the worst group in each of four major spatial orientation configurations, contradictory to the prediction of the orientation association fields. In fact, the medium group’s connectivity was as strong as the best group’s except for the vertical spatial orientation configuration. Overall, the connectivity patterns among the disparity neurons we studied are insensitive to the specific orientation tunings of the neuronal pairs, suggesting that the functional connectivity reported here is not very orientation specific, and might concern more with associating surface cues, rather than boundary/edge cues. It is important to note that this finding arises because our cell pool are selected based on disparity-tuning. We would expect to observe trends more consistent with orientation association fields when the cell pool are selected based on orientation selectivity, as other studies have suggested (citations). The fact that the co-occurrence statistics of orientation edges in natural scenes do not predict the strong and significant cardinal effect we observed among the disparity-tuned neurons, and the fact that the disparity neurons we studied do not follow the pattern of connectivity suggested by the orientation association fields lead us to conclude that the cardinal preference in connectivity reported here is not due to the orientation association fields known to exist between orientation selective neurons in V1, but are intrinsic to disparity neurons.

Figure 4: (a) A population of orientation tuning curves given by the Infomax principle. (b) The average CCH of stimulated neural pairs as a function of the configural orientation ($p = 0.17 > 0.05$). (c) The average CCH of neuronal pair groups, where neuronal pairs of a configural orientation are divided into three groups according to their orientation tuning.

**Scene surface priors in neuronal ensembles** Neurons could compute and encode information as a population in a larger network. When the activities of an ensemble of neurons are analyzed
together, would their interaction as a whole exhibit the same pattern of pairwise interaction inferred based on the interaction of pairs of neurons? Are there higher order interaction? To address these questions, we fitted a Markov random field of binary variables, or Ising model \(^7,^6\), to the data recorded using the Utah array to study small network behaviors. The spike trains of subsets of 8 neurons were binned in 10 msec (20 msec?) window to form a time series of 8-duple vectors with binary elements (0 or 1). Ising models with first, second and third order interactions were used to fit the data. The parameters of these interactions \((\alpha, \beta, \gamma)\) were learned using a maximum likelihood approach (see Method) to fit the spike patterns observed in the spike trains of the simultaneously recorded neurons in each session (see Fig. 5(c)). The \(\beta\) parameters as a function of spatial configuration orientation was assumed to vary smoothly and regularized. 200 (?) small networks were studied based on data from seven recording sessions. Fig. 5(d)) shows the configuration of a particular 8-neuron subnetwork. Interaction parameters among all the neurons in the 8-neuron subnetwork were estimated simultaneously. Then the \(\beta\) of of all the pairs across all the sampled subnetworks at the required distance for each spatial orientation configurations were averaged to yield Fig. ??, which exhibits the same cardinal preference in spatial interaction between the neurons as revealed earlier in cross-correlation (CCH) analysis, as shown in Fig. 3(e). Furthermore, the scatter plot of second order interaction parameters \(\beta\) against the disparity tuning similarity index (DSI) as shown in Fig ?? is also comparable with Fig. 3(f). Thus, the MRF analysis confirms our observations on pairwise interaction based on the CCH analysis.

We evaluated the role of the the first, second and third order terms in predicting or explaining the observed spike patterns. We used a cross-validation technique by learning the model based on 20 percent of the data and computed the log ratio error, \(\frac{1}{n} \sum_{i=1}^{n} \log_{2} \frac{P_i(\text{empir.})}{P_i(\text{pred.})}\), as in \(^7\), between the empirical measurement and the prediction. This log ratio measures the relative error in terms of information loss in bits. Fig 5(e) shows the performance of the model with only the first term (independent), the first two terms (Ising2) and all three terms (Ising3) in predicting spike patterns observed in networks of different sizes. We can observe an increase in performance of the model with the addition of the second and the third term, particularly in larger networks, suggesting these interactions are meaningful in predicting population spiking patterns. Fig 5(h) shows the distribution of the coefficients of the three order terms respectively. The coefficients of the first term are all negative, indicating the neurons are biased against high firing rates, i.e. they tend to fire sparingly. The coefficients of the second terms are predominantly positive, indicating predominantly pairwise positive interaction. Interestingly, the third order term is predominantly negative, suggesting it is exerting a modulating or normalization effect that discourage neurons in the network to all fire at
the same time due to all the pairiwse interactions. This effect is supported by the observation that, introducing the third order term improve the accuracy in predicting how many neurons will fire together at the same time in a set of 8-neuron networks, as shown in Fig 5(h).

When only 20 percent of the data were used in the cross-validation analysis, the pairwise interaction term deduced for each network tended to be noisy. Thus, we have regularized the pairwise interaction term in the Ising model using scene statistical prior (i.e. the spatial orientation dependency or the cardinal preference) (see Methods for details). Actually I don’t fully understand how this is done – please explain. Essentially, this is achieved by constraining the distribution of the second order interaction terms by the distribution profile of spatial orientation dependent interaction predicted by scene statistics (Fig. 2(c)). Without this regularization, the improvement of the Ising model with pairwise interaction learned from 20 percent of the data over the independent model is negligible. The benefit of this prior is improving the performance of the model in predicting neural spiking patterns in the population is a further indication that the spatial orientation configuration dependency is likely an integral part of the underlying functional connectivity of the V1 cortical networks.

This figure needs to be modified. Figure 5e and 5f to go, Fig 5g becomes Fig 5e, and Figure 5f and g should be the distributions of the three terms’ coefficients. Figure 5h remain Figure 5h.

3 Discussion

Statistical regularities in natural environment can be learned and exploited by our visual systems for perceptual inference. While significant advances have been made in relating scene statistics to the development of receptive fields (citations), and the neural mechanisms of the perceptual organization of contours (citations), relatively little is known about the encoding of 3D surface priors in neural circuits. In earlier works, we have demonstrated that the recurrent interaction within disparity tuned neurons follow roughly the cooperation and competition patterns predicted by Marr and Poggio (1976), and that the correlation between luminance and disparity as elucidated in scene statistics study can be found as correlation in the tuning preferences in disparity and luminance in the neuronal population. In this work, we went further in demonstrating that cardinal preference in pairwise functional connectivity between disparity-tuned neurons in the primary visual cortex that is predicted by statistical analysis of 3D natural scenes and Hebbian learning. That the statistical regularities of 3D natural scenes manifest in the patterns of functional connectivities of
Figure 5: (a) The average second-order interaction parameter $\beta$ of Ising model as a function of configural orientation. (b) Scatter points over DSI and $\beta$. The Pearson correlation of DSI and $\beta$ in cardinal orientation is significantly stronger than that in non-cardinal orientation ($0.0<0.0$, $p = 0.0 < 0.05$). (c) The prediction of spiking pattern or word. Most patterns occurrence with a pretty low probability, i.e. region A. (d) The frequencies of prediction errors for two models. For region A, RIM tends to have smaller error than that of naive Ising model. (e) The prediction error as a function of the size of network. (f) The prediction error as a function of the number of synchronous neurons or the order of the spiking pattern.
the disparity neurons is an evidence in support of the idea that surface priors are likely encoded in pairwise horizontal connections between neurons.

In demonstrating (1) that the association network learned from disparity signals in 3D range data predicts the cardinal preference in functional connectivity observed, while (2) association network learned from orientated edge signals in natural scenes does not, and (3) that the functional connectivity between disparity tuned neurons is relatively independent of the orientation selectivity of the neuronal pairs, we argued for the existence of a disparity association network that is dissociable from the better known orientation association network. Furthermore, our study of the larger network interaction using MRF Ising model shows that this disparity association network is subject to a sparse constraint, as demonstrated by the need of a third order interaction term, with mostly negative coefficients, in predicting complex spike patterns in the neuronal population. This sparsity prior can prevent the positive pairwise connectivity coding surface priors from over-exciting all the neurons in the network simultaneously, and has been hypothesized to be an important principle in neuroscience.

The disparity association field with pairwise connections, as its counterpart Markov random field in computer vision, can be useful for noise removal, surface interpolation and 3D inference by bringing global context to resolve local ambiguity. Markov random fields with pairwise connectivity has been shown to be effective for many classes of early vision computation. Marroquin et al. (PNAS 1988) in particular proposed a V1 circuit based on MRF model with pairwise connectivity for stereo inference based on disparity-tuned neurons. Our finding on scene surface priors being encoded in pairwise connectivity of V1 neurons, also recovered as the second order interaction terms of a MRF Ising model, suggests that Markov random field framework for early vision might indeed be a reasonable theoretical framework to conceptualizing the functional neural circuits in the primary visual cortex, as envisioned in some earlier works. While recent pioneering works in neural data analysis have experimented with Ising models to show that weak pairwise connections can actually have significant large network effect in accounting for complex spiking patterns observed in neuronal population, this paper provides the first demonstration that the connectivity patterns decoded in V1 network can be matched to scene statistics and thus potentially serve a meaningful computational function, as surface priors for Bayesian perceptual inference.
Methods

Recording neural activities We record neural activities from three monkeys A, B and C. Monkeys A and B are recorded using 2-3 electrodes (Fig. 3(a)), while C is recorded using fixed Multi-electrode array with 96 channels (Fig. 3(b)). In each recording session, we record multiple trials. First, we identify the spatial receptive fields of all recorded neuron using grating stimulus (monkeys A and B) or white noise based reverse correlation (monkey C) \[xx\]. For monkeys A and B, we tune the electrodes to find disparity tuned and orientation tuned neurons with receptive fields. For monkey C, about half neurons have clear receptive fields. Second, we record 60-110 trials for each recording session when disparity stimulus presented. For each recording trial, Random Dot Stereogram (RDS) with 11 disparities or conditions are presented. Third, we record 30 trials for each recording session when grating stimulus presented. The grating stimulus are composed of 16 conditions which are uniformly distributed orientations from 0 to 180 degree.

The record date of each neuron is a spike train over time, with a time window of 10ms, (Fig. 3(c)). With the above recordings, for each pair of neurons, we compute the configural orientation, receptive field distance, disparity tuning similarity (DSI), preferred orientation of individual neuron and orientation tuning similarity (OSI).

Ranking disparity tuning neurons and orientation tuning neurons We select disparity tuning neurons from all recorded neurons according to the following criterions: (1) the maximum firing rate over 11 disparities should be greater that 2 spikes per second; (2) the average ranking of d prime ranking, overlap ranking and p-value ranking; (3) the peak value of the tuning curve should be significantly greater than other values. For monkeys A and B, we move the electrodes to find disparity neurons. For monkey C, about 25% percent of recorded neurons are disparity neurons.

The criterions for disparity tuning neuron selection can be applied to the orientation tuning neuron selection. The only difference is that the orientation tuning curves have two peaks.

Cross-Correlation Histogram (CCH) The correlation is controlled by the physical connectivity among neurons, correlation in stimulus, firing rate and lag time. We here use cross-correlation histogram (CCH) \(^5\) to measure the correlation between neural pair. CCH seeks to reduce the effects of firing rate and lag time, and therefore measures the functional correlation reflecting neural inherit correlation and stimulus correlation. The effects brought by the last two factors are considered in CCH.

\[
D(t_1, t_2) = \frac{\langle x(t_1) \rangle \langle x(t_2) \rangle}{\sqrt{C_{xx}(t_1, t_1)C_{yy}(t_2, t_2)}},
\]  
(1)
where

\[ C_{xy}(t_1, t_2) = \langle x(t_1), x(t_2) \rangle - \langle x(t_1) \rangle \langle x(t_2) \rangle, \]  
\( \langle x(t_1) \rangle \langle y(t_2) \rangle = \frac{1}{N} \sum_{n=1}^{N} \frac{\theta_x(n)}{\theta_y} \langle x(t_1) \rangle \frac{\theta_y(n)}{\theta_x} \langle y(t_2) \rangle. \]  

(2)

(3)

where \( n \) is the \( n \)-th trial of \( N \) total trials, \( \theta(n) \) is average firing rate for trial \( n \), and \( \bar{\theta} \) is the average firing rate for all trials.

Then \( D(t_1, t_2) \) is a function over the lag time \( \triangle t = t_2 - t_1 \), and roughly takes the peak value when \( \triangle t = 0 \). We use the area under peak, i.e., the integration of \( D(t_1, t_2) \) over \([ -\triangle t_1, \triangle t_1 \]  

Relative disparity from range image Disparity neurons in visual cortex process disparity information instead of depth information. Disparity is computed from range with respect to a fixation point. We use the method described in ???. Two complementary range image databases, CMU and Brown, are used for scene statistics, where CMU database is collected from inside city while Brown database is collected from forest.

Reduce the effect of biased disparity tuning similarity As the correlation between neurons is mainly affected by four factors: disparity tuning similarity (DSI), orientation tuning similarity (OSI), configural orientation and receptive field distance. When we investigate the relationship between CCH and configural orientation, we should reduce the effects potentially brought by the rest three factors. Here we use a two-steps strategy: (1) select data points with median DSI, OSI and RF distance; (2) use bootstrapping to create a new set of data points which have the same average DSI, OSI and RF distance at all orientations. TODO: figure

Simulated neuron and population of tuning To compute the spatial correlation pattern in natural scene images, we stimulate a neural network composed of several column. Each column contains a population of neurons. We here determine the population using Infomax principle.

The population of tuning curves depend on the prior \( p(s) \) of disparity (stimulus) \( s \) in natural scene. Define \( D(s) = \int_{-\infty}^{\bar{s}} Np(s')ds' \). Then \( s_n = D^{-1}(n) \) is the preferred stimulus of the \( n \)-th tuning curve. The tuning curve can be approximately expressed as,

\[ h_n(s) \approx g(s_n)h(Np(s_n)(s - s_n)) \]

where \( h(\cdot) \) is set to normal Gaussian function in this work.
**Regularized Ising Model (RIM)** The Ising model with pair-wise interaction is:

\[
P(\{\sigma_i\}) = \frac{1}{Z(\alpha, \beta)} \exp \left\{ \sum_i \alpha_i \sigma_i + \sum_{i<j} \sigma_i \sigma_j \beta_{ij} \right\}
\]  

where \(i, j\) index the neurons. To introduce the configural orientation prior over \(\beta\), we equally quantize the configural orientation into \(G\) points. That is, each neural pair \((i, j)\) belongs to one of \(G\) groups according to their configural orientation, denoted as \(g(i, j)\). We assume that, for each group \(g\), \(\beta\) are drawn from a same Gaussian distribution. Then the joint distribution can be expressed as,

\[
P(\{\sigma_i\}, \beta) = \frac{1}{Z} \exp \left\{ \sum_i \alpha_i \sigma_i + \sum_{i<j} \sigma_i \sigma_j \beta_{ij} \right\} \prod_{i<j} N(\beta_{ij}; \theta_{g(i,j)})
\]  

where \(g(i, j)\) is the group index of the pair \((i, j)\). Neural pairs in the same group share the same prior.

This model can be efficiently solved using an iterative strategy. First, we estimate \(\hat{\beta}_{ij}\) using MAP estimation,

\[
\hat{\beta}_{ij} = \max_{\beta_{ij}} \sum_i \log P(\{\sigma_i\}, \beta)
\]

Second, update parameters using the estimated \(\hat{\beta}\),

\[
\mu_g = \frac{1}{N} \sum_{g(i,j)=g} \hat{\beta}_{ij}, \quad \delta^2_g = \frac{1}{N} \sum_{g(i,j)=g} \left(\hat{\beta}_{ij} - \mu_{g(i,j)}\right)^2
\]


Acknowledgements   Put acknowledgements here.

Competing Interests   The authors declare that they have no competing financial interests.

Correspondence   Correspondence and requests for materials should be addressed to Tai Sing Lee (email: tai@cnbc.cmu.edu).