Image correlates of crowding in natural scenes

Schepens Eye Research Institute, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA; School of Psychology, The University of Western Australia, Perth, Australia

Thomas S. A. Wallis

Peter J. Bex

Schepens Eye Research Institute, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Visual crowding is the inability to identify visible features when they are surrounded by other structure in the peripheral field. Since natural environments are replete with structure and most of our visual field is peripheral, crowding represents the primary limit on vision in the real world. However, little is known about the characteristics of crowding under natural conditions. Here we examine where crowding occurs in natural images. Observers were required to identify which of four locations contained a patch of “dead leaves” (synthetic, naturalistic contour structure) embedded into natural images. Threshold size for the dead leaves patch scaled with eccentricity in a manner consistent with crowding. Reverse correlation at multiple scales was used to determine local image statistics that correlated with task performance. Stepwise model selection revealed that local RMS contrast and edge density at the site of the dead leaves patch were of primary importance in predicting the occurrence of crowding once patch size and eccentricity had been considered. The absolute magnitudes of the regression weights for RMS contrast at different spatial scales varied in a manner consistent with receptive field sizes measured in striate cortex of primate brains. Our results are consistent with crowding models that are based on spatial averaging of features in the early stages of the visual system, and allow the prediction of where crowding is likely to occur in natural images.

Keywords: crowding, texture, contours, natural scenes, anisotropy, pooling, feature integration, peripheral vision, reverse correlation, Bayesian Information Criterion, model selection


Introduction

While reading this sentence, notice that text rapidly becomes illegible and details become indistinguishable with distance from the current point of regard. This effect is called crowding (Bouma, 1970). This paucity of detail is not just an effect of reduced acuity or contrast sensitivity—the presence of contours can easily be detected—it is just that they cannot be individuated. It is therefore crowding that represents the primary limit on the functionality of peripheral vision.

Crowding causes deficits across a wide variety of visual tasks, including Vernier acuity (Westheimer & Hauske, 1975), orientation discrimination (Andriessen & Bouma, 1975; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), letter identification (Bouma, 1970; Flom, Heath, & Takahashi, 1963; Toet & Levi, 1992), and face recognition (Louie, Bressler, & Whitney, 2007; Martelli, Majaj, & Pelli, 2005; see Levi, 2008; Pelli & Tillman, 2008; Whitney & Levi, 2011 for recent reviews). The emerging consensus from these studies is that crowding is a consequence of spatially pooling features within receptive fields of increasing size: information is averaged (Balas, Nakano, & Rosenholtz, 2009; Dakin, Cass, Greenwood, & Bex, 2010; J. Freeman & Simoncelli, 2011; Greenwood, Bex, & Dakin, 2009; Parkes et al., 2001; van den Berg, Roerdink, & Cornelissen, 2010; J. Freeman, Chakravarthi, & Pelli, 2012) or not resolved by attention (He, Cavanagh, & Intriligator, 1996; Nandy & Tjan, 2007; Strasburger, 2005) and therefore some is lost.

Most investigations of crowding have used simplified stimuli presented on otherwise featureless backgrounds, a situation quite unlike the typical natural world. While these studies have provided important insights into the essential process of crowding, the extent to which this understanding holds true in natural vision is less clear.
The most robust finding from the crowding literature is that image features are pooled within a region whose size is approximately half the retinal eccentricity (Bouma, 1970; Pelli & Tillman, 2008). Given that natural images are cluttered, this finding implies that most of the time, we should be unable to identify anything outside the fovea because there is nearly always another contour in the “crowding window.” Alternatively, recent work suggests that image grouping processes can minimize the effects of crowding in complex images. For example, crowding can be attenuated when flanks can be grouped together and/or segmented from a central target (Bex, Dakin, & Simmers, 2003; Livne & Sagi, 2007, 2010; Saarela, Sayim, Westheimer, & Herzog, 2009), facial expression can be recognized even though facial features are crowded (Fischer & Whitney, 2011; Martelli et al., 2005), and identification of objects containing internal structure is relatively less affected by crowding than for object silhouettes or for letters (Wallace & Tjan, 2011). Given that natural scenes generally contain meaningful objects, this class of observation suggests that natural images may be relatively resistant to crowding.

A number of authors have examined crowding-related phenomena in natural images viewed by the peripheral visual field. To and colleagues (To, Gilchrist, Troscianko, & Tolhurst, 2011; To, Lovell, Troscianko, & Tolhurst, 2010) showed that the subjective magnitude of hue or orientation differences in peripherally-viewed natural images was less than predicted from contrast sensitivity. Similarly, Kingdom, Field, and Olmos (2007) demonstrated that subjects were highly insensitive to naturally-occurring geometric transformations that produced local luminance changes that otherwise could be easily detected. Bex (2010) showed that spatial distortions could be detected in natural images, but sensitivity decreased as edge density increased in real and random phase images. These findings demonstrate that spatial discrimination of cluttered natural images is impaired in the peripheral visual field, and that the presence of edges plays an important role in these discriminations. Since crowding depends on visual features such as edges, this suggests a critical role for crowding in peripheral vision.

Several authors (J. Freeman & Simoncelli, 2011; Balas et al., 2009; Parkes et al., 2001) have proposed that crowding may be an emergent property of statistical averaging among image features. J. Freeman and Simoncelli (2011) developed a crowding model based on a texture synthesis algorithm (Portilla & Simoncelli, 2000), modified so that spatial structure is synthesized within regions whose size scales with eccentricity. To test the model, this scale factor was varied to produce a set of naturalistic stimuli that were progressively “texturized” with eccentricity. Human observers discriminated between pairs of such texturized images and the slope of the scaling factor for images that were perceptual metamers (i.e., not a discriminable difference) was used to estimate perceptual field sizes for texture discrimination. These size estimates closely followed those estimated for receptive fields in visual area V2 in non-human primates. The authors conclude that V2 is a likely candidate for the mid-level visual mechanism at which crowding occurs.

Averaging models are therefore able to account for crowding effects across a range of stimuli, from oriented gratings (Parkes et al., 2001) and simple objects (Dakin et al., 2010; Greenwood et al., 2009; van den Berg et al., 2010; Balas et al., 2009) to natural images (Balas et al., 2009; J. Freeman & Simoncelli, 2011). While we have good estimates of the statistical distribution of luminance (van Hateren & van der Schaaf, 1998), contrast (Balboa & Grzywacz, 2003) and edges (Bex, 2010; Bex, Solomon, & Dakin, 2009) in natural images, there is no firm understanding of how these image properties determine crowding in natural images. We sought to improve this understanding using a modified reverse correlation paradigm to identify where crowding can be expected to occur in arbitrary natural images.

To assess crowding in natural scenes, stimuli must be embedded within scenes rather than presented on blank backgrounds if they are to capture the influence of the spatial characteristics of natural images. Furthermore, in order to eliminate luminance and contrast cues that could be detected independently of crowding, the stimuli must be locally matched to the image they replace. We considered three methods to meet these criteria. First, randomizing the phase spectrum of an image patch is frequently used to generate naturalistic images (e.g., Oppenheim & Lim, 1981), while maintaining the amplitude spectrum. However, phase scrambled patches lack edges (which arise from correlations in phase across spatial scale), and have Gaussian luminance distributions, so are therefore not useful for discriminations of spatial structure. Second, introducing spatial distortions avoids both these problems, but their visibility in natural images depends on the spatial scale of the distortion (Bex, 2010).

We therefore employed “dead leaves” stimuli (Lee, Mumford, & Huang, 2001; Matheron, 1975; Pitkow, 2010; Ruderman, 1997) that were embedded in a natural image. Opaque elementary shapes (in this case, ellipses of pseudo-random aspect ratio) were laid over a circular image segment with no constraint for overlap (see Figure 1A). The ellipses therefore form naturalistic edges where some objects occlude others, a property that mimics the provenance of local image structure in natural scenes (Lee et al., 2001; Ruderman, 1997; Pitkow, 2010). In our variant of this method, each ellipse is assigned the intensity of the pixel at its centre in the underlying natural image. The dead leaves patches therefore have the same space-averaged luminance and contrast as the image segments they replace.
and differ from the background image only in structure such as texture and contours (see Figure S1).

**Methods**

**Observers**

Three observers participated in the experiment, the authors and one naïve participant. All observers had normal or corrected-to-normal visual acuity. All participants gave informed consent and the methods were approved by the Institutional Review Board of the Schepens Eye Research Institute in accordance with the Declaration of Helsinki.

**Stimuli and procedure**

Stimuli were presented using the PsychToolbox (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007) for MATLAB (The Mathworks) running on a Windows 7 computer. The display was an Apple Studio Display 17” CRT with 1024 × 768 pixels resolution refreshed at 75 Hz. The monitor was gamma-corrected via the graphics card control panel after calibration with a Minolta LS110 luminance meter, and was set to a maximum luminance of 80 cd/m² (mean 40 cd/m²). Observers viewed the screen from 57 cm.

Natural scenes were selected at random from a set of calibrated natural images (van Hateren & van der Schaaf, 1998). A 23.4 × 23.4° (750 × 750 pixel) section of an image in the set was selected at random. The mean intensity of the image was set to the mean of the monitor, the global RMS contrast (\(\sigma(L)/\mu(L)\)) was set to 25%, and the resulting image was clipped to the range [0–255]. Many natural scenes contain large inhomogeneities, such as relatively sparse areas of sky at the top of the image, which may influence crowding and response biases. We therefore randomized the orientation of the underlying image at 0°, 90°, 180° or 270° across trials. Images were presented for 200 ms followed by a mask (a phase scrambled version of the image), which remained on the screen until the observer responded.

Figure 1. “Dead leaves” in natural images. (A) Twelve circular patches of elliptical primitives are laid over a natural image at three eccentricities from the centre on each cardinal axis. The radius of the patch increases with eccentricity. Each ellipse takes on the greyscale value of the image pixel at its centre, matching the dead leaves patches in space-averaged luminance and contrast. On fixating the cross in the centre of the image, you may note that some patches are harder to discriminate than others—thus local image properties modulate crowding. In the experiment, observers reported the location of a single patch of dead leaves relative to fixation (4AFC). Our analysis aimed to determine the local image properties in the underlying natural image that are correlated with discrimination performance. (B) Seven image statistics were computed at four Gaussian-weighted scales (here, the finest scale, \(\sigma = 2\) pixels, is shown). These statistics are used to predict performance on the discrimination task, along with task parameters such as patch size.
responded. A fixation spot subtending 0.25° was presented in the centre of the image.

A patch of dead leaves was presented above, below, left or right of fixation (at random across trials), centered at 2, 4 or 8° of eccentricity (varied across blocks of trials). The observers’ task was to indicate the location of the dead leaves patch relative to fixation (4 alternative forced-choice). The radius of the patch was under the control of an adaptive staircase. The area of the patch was tiled with elliptical primitives, placed within the patch region with an average density of 1 per 64 square pixels, with a minimum of one ellipse. The horizontal and vertical dimensions of each ellipse were independently randomly drawn from a uniform distribution with a lower bound of 4 pixels and an upper bound of 32 pixels. These constraints were selected to ensure that dead leaves patches were matched to the average local slope and edge density of the ensemble of natural images (see Figure S1). The orientation of each ellipse was randomly drawn from a uniform distribution. The centers of the ellipses were constrained to fall within the circular patch area. The intensity of each ellipse was set to the intensity value of the pixel in the natural image at the centre location of the ellipse. These constraints ensured that on average across the set of natural images, the image statistics of the dead leaves patches are very similar to the image segments they replace. This matching can be seen in Figure S1. At a coarse spatial scale, the dead leaves patches are matched in luminance, RMS contrast, orientation, orientation variability and local amplitude spectrum slope to the patch that they replace. They are less-well matched on edge density and phase congruency, which is unsurprising since these properties are closely tied to contours. Thus, the dead leaves patches serve to alter local contour information while remaining closely matched to the image segment they replace in other image statistics.

Two different adaptive staircases were used in the experiment: in approximately 60% of trials the patch size was determined using a 2 up 2 down staircase (Wetherill & Levitt, 1965) based on responses from all patch locations, while the remaining trials were collected using the Psi method (Kontsevich & Tyler, 1999) with separate staircases for each visual field location. Both methods were set to converge on 50% correct performance in order to provide sufficient numbers of hits and misses for efficient reverse correlation. One block consisted of 400 trials. At eccentricities of 2, 4, and 8°, TW completed 5,718, 5,600, and 5,614 trials, PB completed 5,418, 5,700, and 5,800, and N1 completed 5,600, 6,000, and 5,600 trials respectively (total trial numbers not factors of 400 are due to aborted blocks).

**Image analysis**

We aimed to determine the local image properties correlated with dead leaves discrimination in natural images. We concentrated our analysis on seven image statistics. Luminance corresponds to pixel intensity. RMS contrast is the variation in pixel intensity over space. Edge density is the space-averaged binary output of the Sobel edge detector, where higher values denote more “edge” pixels per unit of area. Orientation is calculated using steered filters (W. Freeman & Adelson, 1991), and we test orientation relative both to the image and to the screen. Orientation variance is the variability in orientation over space, bounded [0–1]. Local amplitude spectrum slope is the log-log local slope of the Fourier amplitude spectrum at every point in the image, where more negative slopes correspond to greater power at low spatial frequencies than high, indicating that the image is more blurred. Finally, the maximum moment of phase congruence (Kovesi, 2003) is an additional measure of the presence of edges that is less correlated with image contrast and more scale invariant than methods that depend on intensity gradients. Details for the calculation of these statistics are given in the Appendix. Spatial maps of each local image statistic were computed for each natural image (see for example Figure 1B), then were Gaussian weighted over four spatial scales (σ = 2, 8, 32 or 128 pixels, corresponding to 0.06, 0.25, 1, and 4°). The value of each weighted statistic corresponding to the centre of the dead leaves patch on each trial (a scalar) was entered into further analyses.

**Reverse correlation**

We used logistic regression to estimate the correlation between manipulated parameters (such as patch size), local image statistics, and trial-by-trial performance. Logistic regression describes the change in logged odds (a linear transformation of the expected value of proportion correct) as a function of a linear system of predictors and weights. Fits were performed using the MATLAB (Mathworks) GLMFIT function. We performed a stepwise model selection procedure to determine which predictor variables to include in the model, in which experimentally-manipulated variables and then stochastically-varying image statistic predictors were sequentially added to the model (see Appendix and Tables S1–S5). With a large dataset such as ours, traditional null hypothesis significance testing and associated statistics such as p values, F ratios and $R^2$ may overestimate the evidence against the null hypothesis, overfit the data, and make comparisons between the relative importance of predictor variables inferentially hazardous (Wagenmakers,
Manipulated predictors

Four factors were independently manipulated during the experiment: the size of the dead leaves patch, its location relative to fixation, its eccentricity from fixation, and the orientation of the underlying natural image segment. The relationship between these factors and task performance is shown in Figure 2. Performance improves with increasing patch size, and there is an interaction between patch size and eccentricity such that larger patches are required at larger eccentricities to reach the same level of performance. These findings are consistent with known properties of crowding.

In addition, performance is generally better when the dead leaves patch was presented to the left or right of fixation rather than above or below fixation (compare red to blue data and curves, Figure 2), and to a lesser degree, when the patch is presented below fixation compared to above fixation. This effect is independent of the rotation of the underlying image, suggesting that these visual field effects are relative to the intrinsic perceptual space of the observer rather than to the geotopic orientation of the image (e.g., sky above). Finally, the orientation of the image segment did not improve the model, suggesting that performance is governed by local image properties rather than by the global image orientation (see Figure S2 for thresholds according to image segment orientation).

As noted in the Introduction, a common diagnostic tool for identifying crowding is that the size of the integration region producing crowding scales with eccentricity with a factor of approximately 0.5 (“Bouma’s Law”; Pelli & Tillman, 2008; Levi, 2008; Bouma, 1970), with an empirically observed range from approximately 0.3 to 0.7 (Pelli, Palomares, & Majaj, 2004; Bouma, 1970; Toet & Levi, 1992; Levi & Carney, 2009; Chung, Levi, & Legge, 2001; Strasburger, Harvey, & Rentschler, 1991; Kooi, Toet, Tripathy, & Levi, 1994). To estimate the value of Bouma’s constant for our data, we calculated the patch size that produces 62.5% correct performance (representing the mid point of the theoretical dynamic range of task performance, from 25 to 100% correct) when all other predictors are held constant. This was performed by solving Equation 2 (see Appendix) for patch size at each eccentricity, setting $p$ to be 0.625. The dummy variables for visual field location equal to their marginal frequency and all other predictors to their mean. We reason that the diameter of the threshold patch size corresponds approximately to the spatial region over which dead leaves patches are difficult to discriminate from the surrounding natural image. That is, patches smaller than the integration region are averaged with the surrounding natural structure, whereas patches larger than the region are not averaged with natural structure and are therefore discriminable.

These results are shown in Figure 3. While estimates fall below the rule-of-thumb 0.5 slope, they are well within the range of Bouma’s Law, and well above scaling factors for overlay masking and position uncertainty (Pelli et al., 2004; Michel & Geisler, 2011; White, Levi, & Aitsebaomo, 1992). We believe this provides further evidence to suggest that our task is measuring crowding, rather than masking or position uncertainty alone.

Image predictors

Which local image statistics in the natural image are associated with task performance once patch size and eccentricity have been factored out? We concentrate on seven statistics (see Figure 1B). While many of these
statistics are highly correlated with each other, our model selection attempts to reduce the set of predictors to those that most parsimoniously capture variation in task performance.

All predictors except orientation (either relative to the image or absolute) significantly improved fits over models containing the manipulated experimental parameters alone. However, cumulatively entering these predictors and testing model fit against the BIC in a stepwise fashion (see Supplementary Table S4) determined that the most parsimonious cumulative model contained only RMS contrast and edge density.

The relationship between RMS contrast, edge density and performance across spatially-weighted pooling regions was quantified by calculating these statistics at more (finer-grained) scales than at the four scales used in model-fitting (results of the following analysis for the original four scales are provided in Tables S6 and S7). We picked 12 equally-log-spaced scales from 1 to 256 pixels (0.03 to 8 degrees). Since the units of RMS contrast and edge density are domain-dependent and the distributions observed in the image database are skewed, we needed to calculate a meaningful metric to compare predictors in the central portion of the observed distributions. To this end, we computed the value of the linear predictor (logit; see Equation 2) with patch size set to its threshold value (see Figure 3), and all other predictors set to their mean or marginal values for an eccentricity, except for the predictor at the scale of interest. We computed the logit for two values of the predictor of interest for each scale and eccentricity: the first in which the predictor was set to the 16th percentile of the observed distribution (approximately one standard deviation below the mean

Figure 2. Performance as a function of patch size and target visual field location relative to fixation, for eccentricities of 2°, 4°, and 8° (rows top to bottom), and target visual field location for three observers (columns left to right). Note logarithmic scale of patch size. Curves show fits of the logistic model favored by model selection. Models were fitted to the raw binary response data. For illustration, these data have been binned into six equally-sized bins (data points). Model fits are plotted from the smallest to the largest patch size presented. Each data point represents approximately 230 trials. Error bars show the 95% beta distribution confidence intervals on the bins.
Points have been jittered horizontally to aid visibility.

were 0.44, 0.41, and 0.28 for TW, PB, and N1, respectively.

has a slope of 0.37. Least-squares slopes for individual subjects
function of eccentricity. Solid line shows best fit to all points and
Figure 3. Threshold patch diameter for each observer as a
between positive and negative values (see insets, Figure
eccentricity. The resulting difference scores alternated
of extrastriate cortex.

for a normal distribution) and the second in which the
predictor was set to the 84th percentile. This allows us
to examine the changes in task performance predicted
by the model across the central operating range of the
predictor. The difference score between the two logits
($\Delta_{\text{logit}} = \text{logit}_{84} - \text{logit}_{16}$) can be thought of as a
normalized regression weight for the predictor at that
scale and eccentricity. If the difference is positive, then
as the predictor increases the model predicts that the
observer’s task performance will improve. If the
difference is negative, an increase in the predictor is
correlated with more incorrect responses. The absolute
magnitude of the difference indicates the importance of
that scale for the model, such that a larger magnitude
indicates that changes in the predictor at that scale are
more strongly associated with performance. The value
of $\Delta_{\text{logit}}$ was calculated for each scale at each
eccentricity. The resulting difference scores alternated
between positive and negative values (see insets, Figure
4A–B). This indicates that contrasts between closely-
spaced neighboring scales are important predictors of
performance. More strikingly, we noticed that the
absolute magnitude of these difference scores changed
lawfully with the scale of the spatial averaging (see
Figure 4A–B), even after size and eccentricity had been
factored out in this analysis. Given that these weights
for spatial averaging are correlated with psychophysical
performance, we speculated the area of spatial averag-
ing might be related to receptive field sizes of neurons
supporting task performance.

To quantify this possibility, we fit a three-parameter
Gaussian function (mean, bandwidth and amplitude) to
the absolute predictor weights at each eccentricity.
While the individual signed data are noisy, on average
the absolute values across observers show evidence of
size tuning for both RMS contrast and edge density.
Note that the effect of patch size and eccentricity has
already been accounted for in this analysis by setting
patch size to its threshold value for each eccentricity.
For RMS contrast, the peak locations of the best-fitting
Gaussian functions increase slightly with eccentricity,
consistent with the increase in receptive field sizes with
eccentricity. J. Freeman and Simoncelli (2011) recently
observed that psychophysical discrimination thresholds
for texturized natural scenes scaled with Bouma’s law,
which is consistent with the scaling of receptive field
sizes in V2. While it is highly speculative to attempt to
relate the underlying physiology to psychophysical
performance, we apply similar logic to our results from
Figure 3 and Figure 4A–B by converting the mean
parameter from the Gaussian fits ($\mu$) into a “receptive
field size” by plotting the full width at half maximum of
the corresponding Gaussian scaling region (i.e., $2\sqrt{2}\ln 2\mu$). The resulting values are plotted against
estimates of receptive field sizes of visual areas V1, V2,
and V4 of macaque (J. Freeman & Simoncelli, 2011) in
Figure 4C. Consistent with Bouma’s law, threshold size
for the detection of dead leaves patches is close to the
receptive field sizes of neurons in area V2 of primate
brains. Once patch size is factored out, however, spatial
averaging estimates for RMS contrast are consistent
with the smaller receptive fields of V1 (see Figure 4C). The uncertainty associated with the fits for edge density
does not allow us to draw any strong conclusions, but
these peaks are generally larger than those for RMS
contrast and are closer to the larger receptive field sizes
of extrastriate cortex.

The observation that the coefficients for RMS
contrast are not exclusively positive means that
performance is not simply a linear function of local
image contrast (note that the global contrast of all
images was the same), but is dependent on center and
surround contrasts at multiple scales. To verify that the
importance of contrast is not driven solely by patches
below contrast detection threshold, we repeated our
model selection procedure after excluding trials in
which the local RMS contrast at the target location fell
below 10% (a value that is well above contrast
increment detection threshold in natural scenes at most
scales (Bex et al., 2009). Excluding trials for which local
contrast fell below this value at a coarse scale ($\sigma = 4^\circ$
/128 pixels; 8% of trials) resulted in no substantive
changes to model selection. Excluding trials where local
contrast fell below 10% in a more fine region ($\sigma = .25^\circ$
/
8 pixels; 45% of trials) did change some outcomes of the model selection procedure (these analyses are detailed in the captions of Tables S1–S5), but importantly, edge density and RMS contrast were still included in the final models for two of three observers. The remaining observer’s preferred model included only edge density. These control analyses confirm that the present task examines crowding and not target detection, as our primary findings hold for image patches that were easily detectable.

**Predictive performance of the model**

To assess the predictive performance of our model, we performed bootstrapped cross validation (see Figure 5). For all observers, the area under the ROC curve for the final model falls in the range of .75 to .8, with little variation across samples. This is a good indication that the model is not overfit, since it stably predicts out-of-sample data on a criterion not used for model selection. Figure 5 also shows two reduced models for comparison. The “Task” model includes only manipulated parameters, whereas the “+ RMS” model is the task model with RMS contrast included. It can be seen from these distributions that including RMS contrast significantly improves the predictive performance of the model relative to task parameters alone. The effect of including edge density is smaller but still statistically substantial.

Since our model contains pixel-based predictors, it is possible to evaluate the predictions of the model for an arbitrary image. To the extent that crowding in natural scenes reflects our task, we believe that this type of computation could be used to predict the likelihood of “crowding” in arbitrary natural images. We illustrate model predictions for the natural image from Figure 1 (see Figure 6). Here, we averaged model coefficients across observers then evaluated the model’s predicted proportion correct for every pixel in the image,
assuming central fixation and marginalizing over visual field location. We show predictions for several combinations of image size and Bouma’s constant (the scaling of the integration region as a function of eccentricity). Finally, the importance of local image statistics depends on the region of integration: if the region of integration is larger (Bouma’s constant of 0.5) as a function of eccentricity, the magnitude of the influence of features at very local scales is reduced. Note that we found a constant of approximately 0.4 across observers (Figure 3), meaning that the demonstration images shown here span likely ranges of the integration region for human observers.

Figure 5. Model cross-validation. 70% of the data was randomly selected and fit; these coefficients were used to predict responses in the remaining 30% of data. This process was repeated 4999 times for each observer, and distributions of model fit assessed by the area under the Receiver Operating Characteristic (ROC) curve are shown. A value of .5 would indicate that the model discriminates hits from misses no better than chance, whereas a value of 1 would indicate perfect discrimination performance. Bootstrap distributions are represented as box-and-whisker plots where the horizontal line shows the median, the height of the square shows the interquartile range, the whiskers are twice the interquartile range, and samples lying outside this range are presented as points. Here we show the performance of three models: the “Task” model contains only manipulated parameters (patch size, eccentricity and target location); the “+ RMS” model is the task model with RMS contrast added, and the “Final” model is the final model after model selection (containing RMS contrast and edge density). Including image predictors in the model greatly improves predictive performance compared to considering task parameters alone.

Discussion

Some natural scenes are more cluttered than others, and some areas within a given natural scene are more cluttered than others. Crowding therefore will not occur uniformly across the visual field in the natural world. Consequently, predicting where crowding is likely to occur in natural images is a necessary step in developing methods for assisting those with low vision, and could improve the prediction of eye movements in natural scenes. More fundamentally, understanding the image correlates of crowding in natural scenes will help to constrain models of crowding. We developed a paradigm that allowed us to study crowding in natural scenes independently of simple luminance and contrast detection. Three observers identified the location of modifications of spatial structure (dead leaves) in natural images. We examined the correlations between performance and local image statistics at the location of the dead leaves using linear modeling. Our model allows prediction of where discrimination of local structure will be poor for any arbitrary natural image (Figure 6).

The discrimination of contrast-matched changes in spatial structure in natural images will necessarily involve the combined activity of several mechanisms that have been distinguished by studies using simplified stimuli. For example, performance will likely be influenced by spatial position uncertainty (e.g., Pelli, 1985, 1981), overlay masking (e.g., Legge & Foley, 1980; Pelli & Farell, 1999) and surround suppression (Petrov, Carandini, & McKee, 2005; Polat & Sagi, 1993; Bair, Cavanaugh, & Movshon, 2003), which have been distinguished from “pure crowding” per se (Levi, Harriharan, & Klein, 2002; Pelli et al., 2004; Petrov, Popple, & McKee, 2007). However, there are several aspects of our results that indicate the task under study is primarily a crowding task.
Consistent with one of the most commonly-agreed signatures of crowding, we found that threshold patch sizes increased with eccentricity with a constant of approximately 0.4 (Figure 3). This is within the range of previous estimates of “Bouma’s law” (Pelli & Tillman, 2008; Levi, 2008; Bouma, 1970; Pelli et al., 2004), indicating that our results are likely a consequence of crowding rather than masking, which remains approximately invariant of eccentricity (e.g., Mullen & Losada, 1999), surround suppression (which

Figure 6. Model outputs for the image used in Figure 1, assuming an observer fixates centrally. Here we plot the probability of “crowding” as 1-p, where p is the model’s predicted probability of a correct response. Lighter pixels denote areas in which dead leaves discrimination is more difficult; that is, where contrast-matched contours are more likely to be crowded. In the left images, patch diameter (Bouma’s constant) was set to 0.3 × eccentricity, whereas in the right images patch diameter was 0.5 × eccentricity. In the top images, the image size is set to 24” (i.e., the borders of the image are imagined to be at 12” eccentricity) whereas in the bottom images the retinal image size is doubled (i.e., viewing distance is halved). These images can be used to gain a sense of how the model weighs the size of the integration region and eccentricity against local image statistics.

Downloaded From: https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933492/ on 10/19/2018
increases with eccentricity with a shallower slope) (Petrov & McKee, 2006), or position uncertainty (which also increases with a far shallower slope (Michel & Geisler, 2011; White et al., 1992). Indeed, differences in this eccentricity scaling factor have been characterized as a critical distinction between crowding and “ordinary masking” (Pelli et al., 2004).

In addition, we found evidence of visual field anisotropies such that modifications on the horizontal meridian were more easily detected than modifications on the vertical meridian. While visual field anisotropies consistent with these results have been reported previously for both contrast detection (Najemnik & Geisler, 2005, 2008) and for crowding (Liu, Jiang, Sun, & He, 2009; Toet & Levi, 1992; see Petrov & Meleshkevich, 2011), our task does not allow us to rule out response bias as a contributing factor. That is, visual field anisotropies in this task could be produced by differences in sensitivity across the visual field, biases toward some responses over others, or some combination of the two. We are currently working to disentangle these factors using a variant of Luce’s Choice Model (Lesmes, Wallis, & Bex, 2011; Luce, 1963).

Our analysis does not allow us to investigate another often-reported anisotropy of crowding, the “inward-outward” effect, in which flankers radial but more eccentric than the target produce greater crowding than flankers closer to the fovea (Toet & Levi, 1992; Bex et al., 2003; Petrov & Meleshkevich, 2011). This anisotropy is generally proposed to be a consequence of the two-dimensional geometric projection of M-scaled receptive fields onto an isotropic area of cortex (Motter, 2009), with a possible role for trans-saccadic attentional integration (Nandy & Tjan, 2012). Our approach could be extended to investigate this effect by testing skewed spatial averaging regions as well as symmetric Gaussians: if sensitivity were spatially anisotropic, preferred averaging regions should exhibit the characteristic “teardrop” shape of crowding interference zones (Toet & Levi, 1992).

Importantly, our finding that crowding is greater in regions where local contrast is low does not simply reflect a failure to detect contours under these conditions: contrast and edge density remained important predictors after excluding trials where local contrast was below 10%, a conservative cutoff for detection thresholds in natural images (Bex et al., 2009). While crowding is often characterized as contrast-independent once flankers are detectable (Pelli et al., 2004), the strength of crowding effects have been demonstrated to scale with contrast, depending on other stimulus parameters (Chung et al., 2001). Our finding that local contrast is an important determinant of crowding is in good agreement with this. Similarly, while spatial uncertainty, overlap masking, and surround suppression are necessarily involved in the observers’ perceptual processing of all stimuli, these factors cannot fully explain performance in the task. Every location in our natural image stimuli will suffer from these effects, including the target and the three non-target locations. It is not that subjects confuse the spatial location of the target with non-target locations; the four locations are too remote from one another for such positional confusions. Nor are observers unable to detect the presence of the target because of masking or surround suppression: observers are aware of spatial structure everywhere in the image, just as you are now while reading this text. Performance in our task is poor because the visible spatial structure of the dead leaves patches is not discriminably different from the surrounding natural image—this phenomenon is the hallmark of crowding.

While our dead leaves discrimination task in natural images is obviously related to texture perception, it is important to note that this is a crowding task rather than a texture discrimination task. It is relatively easy to discriminate an isolated patch of dead leaves presented peripherally on a uniform background from the patch of natural image used to generate it. That would be a texture discrimination task, not a crowding task, because the influence of the surrounding spatial structure has been removed. Our task could therefore be considered crowded texture discrimination—but this does not make crowding about texture discrimination any more than using letters to study crowding makes crowding about reading.

Our analysis revealed that target size, eccentricity, RMS contrast, and edge density are of primary importance for determining the “crowdability”—the likelihood that structure in a given location will be crowded—of an arbitrary natural image viewed at a given eccentricity. The model coefficients for these image predictors demonstrate systematic changes across spatial scales. The threshold detection size of the dead leaves patch was consistent with Bouma’s law, with a mean slope of 0.4. For RMS contrast but not edge density, the peak tuning of absolute predictor weights increases with eccentricity in a manner reminiscent of the increase in receptive field sizes of visual neurons through the visual hierarchy (see Figure 4A–B). We compared the scales of these tuning parameters with receptive field size estimates measured with electrophysiological and imaging methods in other primates (Figure 4C, J. Freeman & Simoncelli, 2011). A Bouma’s law near 0.5 for size dependence as a function of eccentricity is consistent with the receptive field sizes of neurons in area V2 of primate brains. More speculatively, the RMS contrast tuning sizes measured in our behavioral task in human observers in the present study are closer to the receptive field sizes in
cortical area V1. The coefficients for edge density were too uncertain to identify with any given cortical area.

In a task that required observers to discriminate pairs of texturized natural image “metamers” from one another, J. Freeman and Simoncelli (2011) found that the threshold spatial scale corresponded to V2 receptive field sizes. Our task involving the discrimination of a modified patch embedded within a larger unmodified natural image produced threshold size estimates that were in good agreement, but estimates of scaling for RMS contrast were smaller. We propose that this variation in the scale over which different parameters modulate the likelihood of crowding is consistent with recent evidence from fMRI showing that crowding may occur at multiple stages of the visual hierarchy (Anderson, Dakin, Schwarzkopf, Rees, & Greenwood, 2012). Collectively, these data suggest that it is not possible to characterize crowding as specific to one stage of the visual system.

The area under the ROC curve for our model was approximately 0.75 (see Figure 5). It may not be possible to achieve significantly better prediction than this without also considering internal sources of variability (that is, variability not related to the stimulus). Response consistency paradigms across a variety of tasks show that the ratio of internal-to-external noise can vary between 0.75 up to 2.5, but is typically about 1 (Burgess & Colborne, 1988; Gold, Bennett, & Sekuler, 1999; Green, 1964; Murray, Bennett, & Sekuler, 2002), and values around this range have been recently reported for judgments of noisy edges embedded in natural images (Neri, 2011). That is, human discrimination performance is limited approximately equally by stimulus and non-stimulus uncertainty. If performance were completely determined by the stimulus and our model captured all of this information, we would expect an area under the ROC approaching one; conversely, if observers’ responses were completely independent of the stimulus, we would expect our model to be at chance. Therefore, an internal-to-external noise ratio of around 1 is consistent with an area under the ROC of 0.75, if the model captures most of the stimulus information important to the observers’ decisions.

Crowding of orientation is one of the most replicated results in the literature (Levi, 2008; Parkes et al., 2001; Pelli & Tillman, 2008; Pelli et al., 2004; van den Berg et al., 2010; van den Berg, Roerdink, & Cornelissen, 2007), and is stronger than for photometric (Kingdom et al., 2007; To et al., 2011; van den Berg et al., 2007) and structural (Bex, 2010) judgments in natural images. It is therefore interesting that orientation failed to contribute to our model. However, this is simply reconciled by the fact that while crowding impairs fine-scale orientation discrimination, absolute orientation does not determine where crowding will occur.

Centre-surround interactions of orientation differences are captured by orientation variability, and to a lesser extent by edge density.

To restrict the scope of our study we only considered two-way interactions and relatively simple image statistics. Similarly, there is a large space of candidate models not considered by our model selection procedure for reasons of computational tractability. There is complex correlational structure between pixel intensities in natural images (Field, 1987; Lee et al., 2001), and this structure has been shown to be important for human perception (Geisler & Perry, 2011; Kingdom, Hayes, & Field, 2001). Future work could examine this structure explicitly, but our results show that even simple local image statistics incorporated into simple probabilistic models allow reliable predictions about the probability that crowding will occur in arbitrary natural images. In addition, our statistical model could be compared to more mechanistic models such as saliency (Itti, Koch, & Niebur, 1998; Harel, Koch, & Perona, 2006), clutter (Rosenholtz, Li, & Nakano, 2007), and texture segmentation models (Rosenholtz, 2000; Malik, Belongie, Leung, & Shi, 2001; Malik & Perona, 1990). It is worth noting at this point that our model is simply a refined description of the data; we do not claim it is a theory of any visual process (see Roberts & Pashler, 2000 for related discussion). Finally, our approach may be beneficial to consider for image quality assessments and compression: potentially, images could be compressed severely in areas with high crowding probability away from common gaze points with no corresponding loss of perceived quality (Watson, 1993; Gao, Lu, Tao, & Li, 2010).

Conclusions

Traditional studies of crowding with simple stimuli show that features falling halfway between the fovea and a target’s eccentricity will render the target unrecognizable. This arrangement is ubiquitous in natural scenes and suggests that our perception of the natural world should be unrecognizable in the peripheral visual field. Our data show that spatial manipulations of natural scenes can be reliably detected depending on local image features, which challenges this general inference. We used a reverse correlation analysis between performance and local image statistics. This analysis revealed that target size, eccentricity, local RMS contrast and edge density can be used to make reasonable predictions of the likelihood that an observer will experience crowding. The existence of metamers, whether dead leaves or “texturized” natural images, cannot be used to infer that the image processing that generates such metamers is equivalent.
to the mechanisms that represent spatial structure and cause crowding. Instead, we demonstrate that the discrimination of dead leaves in complex images can be used to predict where crowding is likely to occur in natural scenes.

Appendix

Image statistic computation

The 750 pixel square image from each trial was padded with mean intensity out to 1024 × 1024 pixels. The seven feature maps were Gaussian weighted by multiplying the map with a Gaussian kernel $K$ at one of 4 scales ($\sigma$ of 2, 8, 32 or 128 pixels, corresponding to 0.06, 0.25, 1, and 4 degrees of visual angle) in the frequency domain.

Local RMS contrast at each scale was computed as the square root of the difference between the Gaussian weighted estimate of squared intensity and the weighted estimate of raw intensity (the standard deviation of local pixel intensity (see Bex et al., 2009)).

Edge density was computed by first running Matlab’s Sobel edge detector over the image, returning ones at pixels containing edges and zeros elsewhere. Convolving this edge map with the Gaussian kernels returned a local estimate of edges per pixel at each scale.

Local orientation was computed using steered derivative of Gaussian filters (W. Freeman & Adelson, 1991). We started with a set of two dimensional circularly symmetric Gaussians at a number of wavelengths $i$:

$$ G(x,y,\sigma_i) = e^{-\frac{x^2+y^2}{2\sigma_i^2}} $$

where $\sigma$ is the standard deviation of the Gaussian at wavelength $i$. The filter weight was given by the integral of the Gaussian:

$$ w_i = \int G(x,y,\sigma_i) \, dx \, dy $$

We considered four wavelengths of $\sigma_i = \{2, 16, 64, 128\}$ pixels. We took the Hilbert transform of the first $x$ derivative of $G$:

$$ H_i = \text{hilb}\left(\frac{\partial}{\partial x} G_i\right) $$

to obtain a set of quadrature pairs in which the even and odd components are represented in the real and imaginary components of $H_i$ respectively. These filters were then rotated across four orientations $\theta_j = \{0, 90, 180, 270\}$ to produce a set of filters $F_{\theta_j, \sigma_i}$ at each orientation $\theta$ and wavelength $\sigma$. These filters were then convolved with the image. The filter response was taken as the real positive component of this convolution, summed across orientations for each wavelength:

$$ r_{\sigma_i} = \sum_{j=1}^{4} \text{max}(0, \text{real}(F_{\theta_j, \sigma_i} * I)) $$

The amplitude of the filter response was the absolute value of the response summed across orientations:

$$ m_{\sigma_i} = \sum_{j=1}^{4} |F_{\theta_j, \sigma_i} * I| $$

Sine and cosine responses at each scale were also computed:

$$ s_{\sigma_i} = \sum_{j=1}^{4} \sin\theta_j \cdot \text{max}(0, \text{real}(F_{\theta_j, \sigma_i})) $$

$$ c_{\sigma_i} = \sum_{j=1}^{4} \cos\theta_j \cdot \text{max}(0, \text{real}(F_{\theta_j, \sigma_i})) $$

which was then used to compute the local orientation estimate at each wavelength, wrapped across four quadrants using Matlab’s atan2 function.

This orientation estimate at each wavelength was then summed across scales after weighting by the integral of the filter:

$$ s_{\theta} = \sum_{i=1}^{4} \frac{r_{\sigma_i}}{w_i} \cdot \sin \theta_i $$

$$ c_{\theta} = \sum_{i=1}^{4} \frac{r_{\sigma_i}}{w_i} \cdot \cos \theta_i $$

and then used to compute the final orientation estimate across spatial scales using atan2.

The sine and cosine of this orientation feature map were then convolved with the Gaussian weighting kernels $K$ to produce Gaussian weighted estimates of orientation across space. Additional computations performed on orientation, including some summary statistics, were computed using the Circular Statistics Toolbox for Matlab (Berens, 2009).

Orientation variance was calculated as the square root of the variance in sine and cosine components of the local orientation estimates at each scale. This local orientation variance metric is bounded [0–1], where lower values refer to regions of similar local orientation and higher values refer to regions where nearby pixels vary in orientation.

Local amplitude spectrum slope was approximated by computing the slope of the best fitting regression line through the log filter response magnitudes at every
pixel, summed across orientations, as a function of the log spatial frequency of the filter.

Phase congruency (the maximum moment of phase congruency covariance) was calculated using methods described by Kovesi (2003) and associated Matlab routines (phasecong3.m, November 2010 update; run on the unpadded analysis image). The maximum moment of phase congruency provides a measure of local edge strength bounded [0–1]. High values are returned for both edges and textures; this measure is less dependent on local RMS contrast than the Sobel edge detector.

To summarize, seven image statistics were computed for each image presented in the experiment, and these statistics were spatially averaged at four Gaussian scales. The weighted statistic at the centre of each possible dead leaves patch location relative to fixation was then entered into further analysis.

Model selection procedure

We aimed to characterize performance on the 4AFC task as a function of both manipulated parameters, such as patch size, and image parameters that stochastically vary across trials, such as local contrast. To do this we apply a Generalized Linear Model (GLM) in the form of a logistic regression. Logistic regression describes the change in logged odds (a linear transformation of the expected value of proportion correct) as a function of a linear system of predictors and weights:

$$\log \left( \frac{p}{1 - p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

where $p$ is the proportion correct (probability of success), $\beta$ are the regression weights, and $x$ are the values of the predictors.

Psychometric functions are often fit with two additional free parameters: a term that represents the chance rate for the number of response alternatives, and a term that represents the lapse rate (Wichmann & Hill, 2001). To fit psychophysical percent correct data in the context of GLMs, some authors (Yssaad-Fesselier & Knoblauch, 2006) have suggested using a modified link function in which the function is bounded between these additional parameters. However, this procedure violates an important assumption of least-squares fitting, namely that variance is equal across the extent of the function (homoscedasticity). The variance of the binomial distribution decreases as the expected value approaches zero or one. Bounding floor performance to a theoretical chance rate using a modified link function will violate this assumption, since the variance at this asymptote is greater than expected from the binomial distribution. To avoid this problem, here we fit full [0–1] logistic functions to our data, but test quadratic terms to account for the tendency of expected values to fall to chance levels (here 0.25) rather than to zero.

To constrain our model parsimoniously, we performed a hierarchical model selection procedure using the BIC (Kass & Raftery, 1995; Raftery, 1995, 1999; Wagenmakers, 2007; Schwarz, 1978) and associated metrics as our primary selection rule. Model selection was performed for each subject separately, but since the three subjects showed similar final models (see below), we fit one model to all subjects for simplicity.

The BIC is computed as:

$$BIC = -2 \log L + k \log n$$

(1)

where $n$ is the number of observations, $k$ is the number of free parameters, and $L$ is the likelihood of the data given the model. This information criterion therefore penalizes models for additional parameters and takes the number of observations into account ($k \log n$ is a penalty term for these two factors). This is important to do with a large dataset such as ours. The Bayes Factor between two competing models ($H_0$ and $H_1$) is then approximated as the ratio between their prior predictive probabilities (Wagenmakers, 2007):

$$BF_{01} \approx \exp \left( \frac{\text{BIC}_1 - \text{BIC}_0}{2} \right)$$

The approximate posterior probability of $H_0$, assuming equal prior probabilities between $H_0$ and $H_1$, is given by

$$Pr_{\text{BIC}}(H_0 | D) = \frac{1}{1 + \exp \left( -\frac{1}{2} \Delta \text{BIC}_{10} \right)}$$

(Wagenmakers, 2007) where a posterior probability $> .75$ constitutes “positive” evidence for $H_0$, and a posterior probability of $> .99$ constitutes “very strong” evidence for $H_0$ (interpretations as per Raftery, 1995).

Additional metrics of model fit, the Akaike Information Criterion (AIC) and the area under the receiver operating characteristic curve (AROC) are also provided in Supplementary Tables.

Manipulated parameters

The first step we took in model selection was to consider candidate models for describing manipulated parameters on performance. Four factors were manipulated during the experiment: the size of the dead leaves patch, its location relative to fixation, its eccentricity from fixation, and the orientation of the underlying image segment. The candidate models we tested are shown in Table S1.
For two of the three observers the most parsimonious model of the set contained 11 free parameters: patch size, patch size squared, eccentricity, an interaction between patch size and eccentricity, target location, and interaction terms between target location and eccentricity (Table S1). For the remaining observer an 8 parameter model containing patch size, a quadratic patch size term, eccentricity, an interaction between eccentricity and patch size, and visual field location (but no interaction between field location and eccentricity) was preferred. The orientation of the underlying natural image was not significantly predictive of performance for any observer.

Image statistic predictors

What do variations in local image statistics contribute to explaining performance, once manipulated parameters are taken into account? To answer this question, we added each image predictor to the manipulated parameter model for each observer (Table S2). We chose to add the predictor at all four Gaussian scales, plus an interaction term between each scale and eccentricity, for a total of 8 parameters for each image predictor (16 for orientation, since the circular variable must include both sine and cosine terms). For all observers, all image statistics except orientation (either relative to screen or image) improved model likelihood over task parameters alone (Table S2), but some were preferred relative to others (Table S3).

To determine the most parsimonious model based on this, we cumulatively added the image predictors in the order they most improved the manipulated-parameter model (Table S4). Evaluation against the BIC revealed that, once RMS contrast and edge density were included in the model, including parameters for amplitude slope did not significantly improve model fit for TW or N1. For PB, amplitude slope could be usefully included, but the next most important image predictor (phase congruence) could not.

Preferred model

This model selection procedure resulted in a 27 parameter model for TW, a 35 parameter model for PB, and a 24 parameter model for N1. To simplify the interpretation of these fits across observers, we fit all observers’ data with the 27 parameter model (see Table S5 for fit metrics). All fits shown in the paper are derived from this model except for those in Figure 4, where we entered 12 scale parameters rather than 4. The full specification of the preferred model’s design matrix is:

\[
\log \left( \frac{p}{1-p} \right) = \beta_1 + \beta_2(\text{patch}) + \beta_3(\text{ecc}) + \beta_4(\text{patch} \times \text{ecc}) + \beta_5(\text{patch})^2 + \beta_6(\text{vf1}) + \beta_7(\text{vf2}) + \beta_8(\text{vf3}) + \beta_9(\text{vf1} \times \text{ecc}) + \beta_{10}(\text{vf2} \times \text{ecc}) + \beta_{11}(\text{vf3} \times \text{ecc}) + \beta_{12}(\text{RMS1}) + \beta_{13}(\text{RMS2}) + \beta_{14}(\text{RMS3}) + \beta_{15}(\text{RMS4}) + \beta_{16}(\text{RMS1} \times \text{ecc}) + \beta_{17}(\text{RMS2} \times \text{ecc}) + \beta_{18}(\text{RMS3} \times \text{ecc}) + \beta_{19}(\text{RMS4} \times \text{ecc}) + \beta_{20}(\text{Edge1}) + \beta_{21}(\text{Edge2}) + \beta_{22}(\text{Edge3}) + \beta_{23}(\text{Edge4}) + \beta_{24}(\text{Edge1} \times \text{ecc}) + \beta_{25}(\text{Edge2} \times \text{ecc}) + \beta_{26}(\text{Edge3} \times \text{ecc}) + \beta_{27}(\text{Edge4} \times \text{ecc})
\]

where ‘patch’ is \(\log_{10}\) (patch size (°)), ‘ecc’ is eccentricity (°), ‘vf1’ – ‘vf3’ are dummy-coded variables signifying target visual field location relative to fixation, and ‘RMSn’ and ‘Edgen’ refer respectively to \(\log_{10}\) (RMS contrast) and edge density at Gaussian scale \(n\ (0.06, 0.25, 1, \text{and } 4°)\).

Raw data

Raw data (after computation of image statistics) is provided as a supplemental file, with an accompanying guide to column labels and units. Re-use is encouraged with proper attribution.

Acknowledgments

TW and PB designed the experiment, performed the analyses, and wrote the paper. The authors thank Michael Dorr, Tobias Elze, Luis Lesmes, David Muller, Christopher Taylor, and Jennifer Wallis for helpful suggestions and comments. Elements of this work were presented at the Vision Sciences Society Meeting, May 2011 and at the Optical Society of America Fall Vision Meeting, September 2011.

Supported by NIH grant R01EY019281 to PB and NHMRC Research Fellowship 634560 to TW.

Commercial relationships: none.

Corresponding author: Thomas S. A. Wallis.

Email: thomas.wallis@schepens.harvard.edu.

Address: Schepens Eye Research Institute, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA.
References


Motter, B. C. Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a
constant-sized sampling of the V1 surface. Journal of Neuroscience, 29(18), 5749.


