



Texture sparseness, but not local phase structure, impairs second-order segmentation



Elizabeth Zavitz^{a,b,*}, Curtis L. Baker^a

^a McGill Vision Research, Department of Ophthalmology, McGill University, Montreal, Quebec, Canada

^b Department of Physiology, Monash University, Clayton, Victoria, Australia

ARTICLE INFO

Article history:

Received 28 March 2013

Received in revised form 8 July 2013

Available online 11 August 2013

Keywords:

Texture

Segmentation

Higher-order statistics

Natural images

Second order

ABSTRACT

Texture boundary segmentation is typically thought to reflect a comparison of differences in Fourier energy (i.e. low-order texture statistics) on either side of a boundary. However in a previous study (Arsenault, Yoonessi, & Baker, 2011) we showed that the *distribution* of energy within a natural texture (i.e. its higher-order statistical structure) also influences segmentation of contrast boundaries. Here we examine the influence of specific higher-order texture statistics on segmentation of contrast- and orientation-defined boundaries. Using naturalistic synthetic textures to manipulate the sparseness, global phase structure, and local phase alignments of carrier textures, we measure segmentation thresholds based on forced-choice judgments of boundary orientation. We find a similar pattern of results for both contrast and orientation boundaries: (1) randomizing all structure by globally phase scrambling the texture reduces segmentation thresholds substantially, (2) decreasing sparseness also reduces thresholds, and (3) removing local phase alignments has little or no effect on segmentation thresholds. We show that a two-stage filter model with an intermediate compressive nonlinearity and expansive output nonlinearity can account for these data using synthetic textures. Furthermore, the model parameter fits obtained using synthetic textures also predict the segmentation thresholds presented in Arsenault, Yoonessi, and Baker (2011) for natural and phase-scrambled natural texture carriers.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The segmentation of boundaries is an important problem that the visual system must solve before any more complex object processing can occur. Boundaries between objects result in discontinuities in a variety of image properties, among which changes in texture are a particularly interesting example because the means by which they are segmented is not yet well understood. Texture can be represented in terms of spatial statistics, but it is unclear what subset of these statistics is actually employed by segmentation mechanisms. Much previous research has aimed to determine the precise statistical differences that *enable* segmentation when they differ on either side of a boundary (Julesz, 1962; Julesz, Gilbert, & Victor, 1978; Beck, 1983), such as contrast or orientation. Although textures contain, and their neuronal representation may encode, many other statistics that are constant on either side of a boundary, the potential *influence* of these statistics has been largely unexamined. For example in Fig. 1 the texture statistics of the bark (A) and leaves (B) do not vary across the boundary, so they cannot *enable* segmentation, but the modulation defined over the

leaves is easier to segment – thus the nature of the texture *influences* segmentation. Although this demonstration makes the influence of the texture structure apparent, it is unclear which specific aspects of the structure exert this influence.

An early study demonstrating the influence of texture properties was that of Caelli (1980) who examined the influence of a box-shaped feature common throughout the stimulus on segmentation of a boundary defined by a difference in the orientation of line segments within the boxes. He found that segmentation was more difficult when the boxes were present than when the line segments were presented alone. Arsenault, Yoonessi, and Baker (2011) used contrast modulations applied to natural textures to show that higher-order texture statistics can impair contrast boundary segmentation, even though those statistics are not relevant to the segmentation task. We noticed that textures with a greater difference in threshold between the intact and phase-scrambled conditions appeared to be more sparse. We applied a number of image statistical measures that have been used in the literature to quantify the density of textures or natural scenes, and found that edge density (Bex, 2010) correlated strongly with the difference between thresholds. From this, we suggested local edge structure and sparseness as two candidates for texture properties that might influence segmentation, resulting in such a performance difference.

* Corresponding author. Present address: Department of Physiology, Monash University, Building 13F, Clayton, Victoria 3800, Australia.

E-mail address: elizabeth.arsenault@mail.mcgill.ca (E. Zavitz).

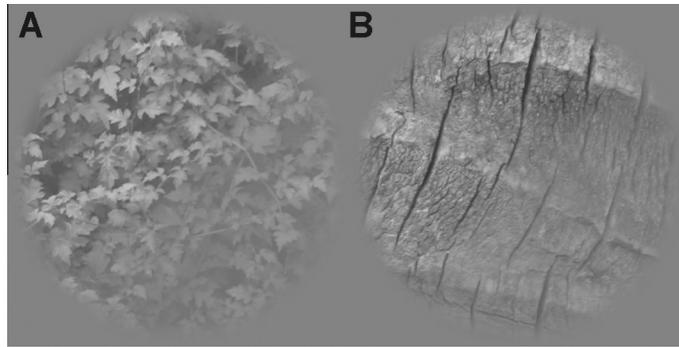


Fig. 1. In both of these textures, a contrast difference enables the percept of a right oblique boundary. The properties of the materials (leaves and bark) forming the carrier textures are different in structure, which results in a difference in the strength of the boundary percept (the modulation of the leaves (A) is easier to see than the modulation of the bark (B)), even though the difference in contrast across the boundary is identical in both stimuli. In this example, the characteristics of the textures can be said to influence segmentation.

It is difficult to assess the roles of specific statistics using natural textures because most individual properties cannot be varied independently or manipulated parametrically. While we have reason to suspect that sparseness or local edge structure might be important influencing statistics, our previous results are only correlational. In the following experiments, we address these challenges by creating synthetic textures consistent with observations of the statistical properties of natural textures using broadband “edgelet” micropatterns. These textures allow us to not only manipulate global structure through phase scrambling as in our previous study (Arsenault, Yoonessi, & Baker, 2011), but also to control the presence of local structure (by phase-scrambling individual micropatterns) and sparseness (by changing the number of micropatterns).

In this paper, we explore not only which higher-order texture properties may be at the root of the observed threshold reduction following phase scrambling, but also why these properties might have the impact they do on segmentation. It could be the case that the overall contrast-defined boundary was masked by local contrast modulations caused by variegated regions of high-contrast features that form the structure of sparse textures (Allard & Faubert, 2007). Alternatively, it could be that the presence of broadband contours in the texture distracts observers from the less-salient contrast boundary, or some combination of these two effects.

A widely accepted model for contrast boundary segmentation, the filter–rectify–filter (FRF) model, is a helpful starting point when thinking about how texture properties can affect segmentation. The most general form of the model consists of a stage of relatively high spatial frequency linear filtering, followed by a pointwise rectification (typically implemented with a square law), and a second stage of linear filtering, on the scale of the boundary to be segmented. Depending on the shape of the rectification (expansive, compressive), images with the same global contrast but different local structure could produce different responses. For example, a texture with its contrast energy concentrated in locally high-contrast regions will produce a greater response for expansive nonlinearities than a texture with an even distribution of contrast energy over space. Given this possibility, sparseness and local broadband edges are particularly logical statistical properties of interest, because both result in localized concentrations of image energy.

Here we first aim to verify that these synthetic textures contain the relevant properties of natural textures by demonstrating again the effect of phase-scrambling on contrast boundary segmentation thresholds as in Arsenault, Yoonessi, and Baker (2011). By varying texture density and phase structure, we are also able to differentiate the influence of local phase alignments, global phase relationships, and sparseness in segmentation of both contrast and orientation boundaries. We chose to study contrast boundaries

because they are the simplest kind of texture boundary, and orientation boundaries because we have observed that natural textures are frequently narrowband for orientation and this type of boundary has been widely studied (e.g. Landy & Oruç, 2002; Meso & Hess, 2011). We implement a filter–rectify–filter model and fit the shape of the rectification to account for the pattern of both our contrast and orientation boundary segmentation results. Having fit the results using synthetic textures, we assess how well this model can also predict the thresholds obtained using contrast modulations of natural textures in Arsenault, Yoonessi, and Baker (2011).

2. General methods

2.1. Stimuli

Each stimulus consisted of a single texture pattern that was contrast-modulated with a half-disc envelope, or two texture patterns ‘quilted’ together to form a disc with distinct halves (a procedure illustrated in Fig. 2A). The textures we used were designed to mimic the image statistics of natural textures, while allowing for control of specific texture properties, by randomly scattering a large number of *edgelet* micropatterns.

2.1.1. Micropatterns

To emulate the local edge structure of natural textures, we used *edgelet* micropatterns each of which contained a spatially localized edge composed of phase-aligned Fourier components. The edge of a micropattern of size s was created by adding together the Fourier components of a half-cycle of a square wave ($f, 3f, 5f, \dots, nf$ where $n = s/4$), with decreasing amplitudes (scaled by $1/f$), of a given orientation (θ) and aligned in sine-phase ($\phi = 0$). One cycle of the lowest spatial frequency pattern was combined with like-oriented in-phase harmonics of gratings (\mathbf{G}) to form a square wave “edge” (\mathbf{D}):

$$\mathbf{D}_{x,y}(\theta, s) = \sum_{k=0}^{\frac{s}{4}} \frac{1}{f} \mathbf{G}_{x,y}(\theta, \phi, f, s), \quad f = 2k \quad (1)$$

These edges were tapered by a Gaussian window, whose sigma was $1/8$ of the size of the micropattern ($\sigma = s/8$), for the final *edgelet* (\mathbf{D}') (Fig. 2B – top):

$$\mathbf{D}' = \mathbf{D}_{x,y} e^{-\left(\frac{(x-\frac{s}{2})^2}{2\sigma^2} + \frac{(y-\frac{s}{2})^2}{2\sigma^2}\right)} \quad (2)$$

To generate novel textures rapidly, we created a library of 48 such ‘intact’ micropatterns at four sizes (16, 32, 64, 128 pixels, or 0.22, 0.44, 0.87, and 1.74 degrees of visual angle), each at twelve orientations evenly spaced in 30° increments.

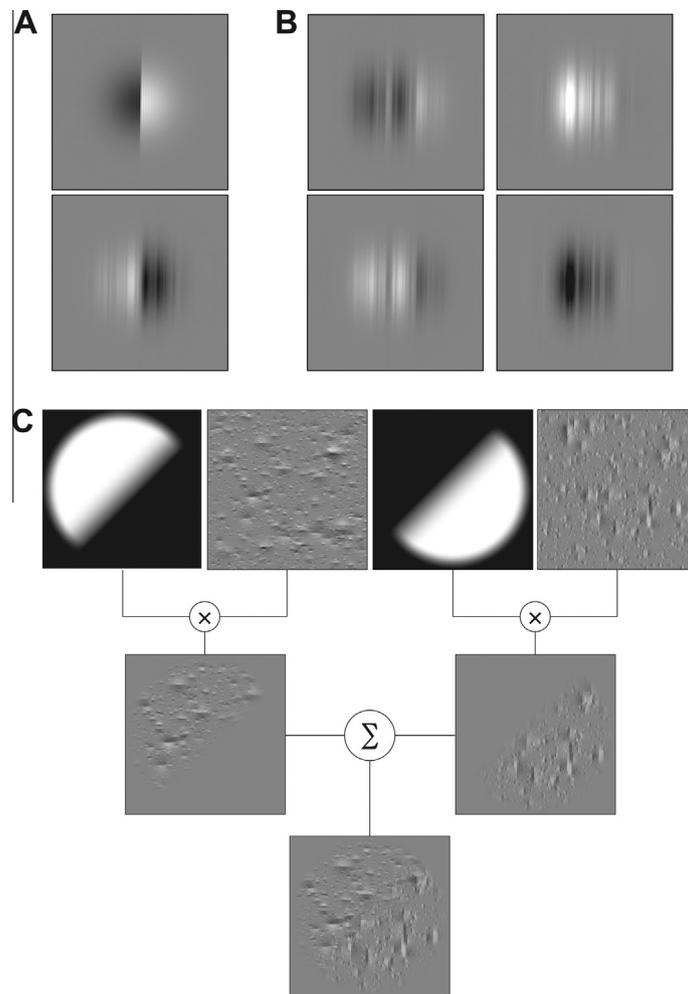


Fig. 2. Procedure for constructing naturalistic textures and orientation-modulated boundaries. (A) Examples of micropattern types used to create synthetic textures. Top: intact Gaussian-enveloped half cycle of a square wave (“edgelet”); bottom: phase-scrambled square wave within the same Gaussian envelope. (B) Top: variations of phase-scrambled micropatterns, Bottom: same instances, polarity-reversed. (C) Procedure for quilting stimuli. Windowed half-disc envelopes (product of \mathbf{W}_{xy} and \mathbf{E}_{xy} , see text) are multiplied with their corresponding carrier textures (\mathbf{C}_{xy}), yielding contrast-modulated boundaries. These modulated halves are then combined, in this case to produce an orientation-modulated boundary. The modulation depth of the stimulus shown is 100%.

In addition to square-wave edgelets, we created locally phase-scrambled edgelets in a similar manner but with the components’ phases (ϕ) randomized rather than aligned (Fig. 2B – bottom). A library of ‘locally scrambled’ micropatterns was created with 50 versions of the phase-scrambling (e.g., Fig. 2C) at each micropattern size and orientation.

2.1.2. Textures

Edgelets drawn from the library were randomly positioned on a 544×544 pixel canvas and summed where they overlapped. While the square-wave edgelets were luminance-balanced with equally sized light and dark regions, the random phases of the scrambled edgelets could result in net mean luminance differences from the grey background. To ensure that the texture stimuli were approximately luminance-balanced, each phase-scrambled edgelet was polarity-reversed (or not) with a 50% probability before being added to the canvas (Fig. 2C). To obtain an approximately $1/f$ amplitude spectrum, four sizes of micropatterns (16, 32, 64, and 128 pixels) were added in proportions necessary to achieve equal coverage for each spatial frequency (Kingdom, Hayes, & Field, 2001). Thus for each 128-pixel micropattern, 4, 16, and 64 of the progressively smaller micropatterns were added – so each texture contained an integer multiple of 85 micropatterns. The possible positions of the micropatterns were constrained to lie entirely within a 544×544

canvas, which was subsequently cropped to the central 480×480 region after all the micropatterns had been drawn.

Three density conditions were created by varying the number of micropatterns within each texture stimulus: the low density condition had 595 micropatterns, medium density 1530, and high density 2975. These conditions were chosen to result in textures that were qualitatively different in appearance, while satisfying the above constraints to produce an approximately $1/f$ amplitude spectrum.

For each density, three *structure* conditions were created: intact, locally scrambled, and globally scrambled. *Intact* (INT) textures were composed from square-wave edgelets as described above – these textures were rich in both global structure (arrangement of micropatterns) and local structure (broadband edges). By one-dimensionally phase-scrambling the individual edgelets we produced a *locally scrambled* (LS) texture that had an equivalent amount of global structure but lacked local phase alignments. We created *globally scrambled* (GS) textures by applying a Fourier transform to both the intact texture and a white-noise image of the same size. The phase values in the original texture were replaced with those of the white noise and then inverse-transformed, thus leaving the power spectrum unchanged while completely randomizing the phases (Dakin et al., 2002).

Due to the random arrangement of micropatterns, some of the INT and LS textures exhibited substantial inhomogeneity, and thus

were unsuitable to use as carrier patterns. To circumvent this problem we rejected textures having differences in luminance or RMS contrast greater than 3 dB between quadrants of the texture (Arsenault, Yoonessi, & Baker, 2011). In the low density condition, only about 12% of the generated textures passed this test; in the medium density condition, about 47% of textures passed; and in the high density condition, about 76% passed. Each texture was scaled to have a mean value of 0, and its extreme luminance values were clipped at ± 3 standard deviations and scaled to fit in the range of intensities between ± 1.0 .

2.1.3. Boundary creation

2.1.3.1. Contrast boundaries. To create contrast-defined boundaries, textures were contrast-modulated by an envelope pattern, consisting of an obliquely oriented half-disc, graduated with a cosine taper. The final stimulus, $S_{x,y}$, is the product of the texture carrier, $C_{x,y}$, the tapered window $W_{x,y}$, and the envelope, $E_{x,y}$, scaled by the modulation depth, m :

$$S_{x,y} = L_o \{1 + cC_{x,y}W_{x,y}(1 + mE_{x,y}/2)\} \quad (3)$$

where $|C_{x,y}| \leq 1.0$, $|E_{x,y}| \leq 1.0$, L_o is the mean luminance, and c is a contrast scaling factor which is adjusted to produce the desired RMS contrast.

2.1.3.2. Orientation boundaries. Orientation-defined boundaries were created between different textures using a method of ‘quilting’ described by Watson and Eckert (1994) and Landy and Oruç (2002), and illustrated in Fig. 2C. To modulate two texture carriers (C_A and C_B) with respect to one another we used a half-disc envelope function ($E_{x,y}$), scaled to create the carrier-specific modulators (E_A and E_B):

$$E_A = W_{x,y} \sqrt{(1 + mE_{x,y})/2} \quad (4)$$

$$E_B = W_{x,y} \sqrt{(1 - mE_{x,y})/2} \quad (5)$$

The modulation depth (m) parameter specifies the difference in contrast between the envelope halves. The luminance-balanced carrier textures (C_A and C_B) are scaled to yield the desired contrast with scaling factor c , and their means are adjusted so that the final stimulus will be luminance balanced after the envelope has been applied:

$$C'_A = cC_A - \frac{\int \int cC_A E_A - 0.5}{\int \int E_A} \quad (6)$$

$$C'_B = cC_B - \frac{\int \int cC_B E_B - 0.5}{\int \int E_B} \quad (7)$$

The latter adjustment is necessary because of the stochastic nature of the textures. Even though each original (entire) texture is zero-balanced, the portion of it enclosed by the envelope might have a slight non-zero mean, which could result in an artifactual luminance boundary. The adjustment in Eqs. (6) and (7) is designed to prevent such luminance artifacts.

The final stimulus, $S_{x,y}$, is the sum of the two carriers, each spatially weighted by their respective envelopes:

$$S_{x,y} = L_o \{1 + C'_A E_A + C'_B E_B\} \quad (8)$$

The weighting is specified by the modulation depth, m in Eqs. (3)–(5). At a modulation depth of zero, the resulting stimulus is a homogeneous blend of the two textures. At a modulation depth of 100%, one half of the disc is entirely C_A , and the other half entirely C_B , with a smooth taper between them at the boundary.

2.2. Apparatus

The stimuli were presented on a CRT monitor (Sony Trinitron Multiscan G400, 81 cd/m², 75 Hz, 1024 × 768 pixels), gamma-linearized with a digital video processor (Bits++, Cambridge Research Systems) for greater bit-depth at low contrasts. Stimulus patterns appeared in a central 480 × 480 pixel patch on a mean grey background. Observers viewed the stimuli from a distance of 114 cm, resulting in a stimulus visual angle of approximately 6.5°. The experiment was run on a Macintosh (Desktop Pro, MacOSX) using Matlab and PsychToolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

2.3. Task

Observers were presented with a central fixation point and initiated each 100-ms stimulus presentation with a button press. The stimulus contained a boundary that was oriented 45°, either left or right oblique, and observers indicated the perceived orientation with a button press. Feedback was not provided. The screen was maintained at the mean grey background between stimulus presentations.

We determined an appropriate range of testing values from pilot experiments for each observer, and used a method of constant stimuli over five logarithmically spaced level values of modulation depth to measure each threshold. All stimuli were presented at a suprathreshold carrier RMS contrast of 14.5%. A minimum of three blocks of 100 trials, with 20 trials per level, were run for each condition to yield a total of at least 60 trials per level.

These experiments conformed to McGill University’s ethical guidelines for human experimentation, and all subjects provided informed consent.

2.4. Data analysis

Percent-correct data from a total of 600 trials were fit with a logistic function, and a threshold was interpolated at the 75% correct point. Curve-fitting was performed using the statistics package Prism (GraphPad Software, Inc.), and standard errors were estimated with its bootstrapping algorithm.

We measured the effect size (Klein, 2005) using Cohen’s d with the standardizer s computed as:

$$s = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{2} \quad (9)$$

where σ_1 and σ_2 are the sample standard deviations of the compared conditions.

3. Experiment 1: Contrast boundary segmentation

In this experiment we set out to (1) test whether our finding that higher-order statistics impair contrast boundary segmentation in natural textures (Arsenault, Yoonessi, & Baker, 2011) could be replicated using synthetic edgelet textures and, if so, (2) investigate the influence of sparseness and local edge structure on segmentation thresholds. Given our previous findings, we expect better performance in the globally scrambled (GS) condition than in the intact (INT) condition, at least for some values of density. If local edge structure influences segmentation, we expect a difference between the LS and INT conditions; if sparseness influences segmentation, we expect a decrease in threshold as density is increased in the INT and LS conditions.

3.1. Methods

Modulation depth thresholds for segmentation of contrast-defined boundaries were measured over a number of synthetic textures, created as described in the general methods. We tested intact, locally scrambled, and globally scrambled textures at each of three density levels. These are depicted in Fig. 3 with the structure changes (INT, LS, GS) varying across columns and the density increasing down each column. Thresholds were measured for four experienced psychophysical observers with normal or corrected-to-normal vision, three of whom (JH, AR, YJK) were naive to our hypotheses.

3.2. Results

Contrast boundary segmentation results are shown in Fig. 4, for individual observers in the upper four graphs, and as a group average in the lower graph.

Effect of structure: The globally scrambled condition (GS, white triangles) appears easier than the other conditions, with lower thresholds independent of density. Thresholds in the intact (INT, dark circles) and locally scrambled (LS, grey diamonds) conditions

appear both higher overall, and dependent upon density, with lower thresholds at higher densities.

A two-way ANOVA was run to test for significant differences between the structure conditions (INT, LS and GS) at the three values of density. This confirmed a significant main effect of structure $F(2,18) = 49.4$, $p < .05$. Post-hoc Bonferroni tests (Supplementary material – Table 1) indicated no evidence of a difference between the INT and LS results at any density, but that both are significantly different from the GS condition at all but the highest density. That the effect of density depends on the structure condition was confirmed by a significant interaction ($F(4,18) = 7.69$, $p < .05$). Analysis of the effect-size, measured using Cohen's d , shows that the difference between the intact and globally scrambled conditions (the effect of phase scrambling) decreases as density is increased (Supplementary material – Table 1). The same trend is observed between the locally (LS) and globally (GS) scrambled conditions.

Effect of density: It appears as though density affects psychophysical performance in the intact (INT) and locally scrambled (LS) conditions only, with thresholds decreasing as density increases. Thresholds in the globally scrambled condition (GS) do not appear to change as a function of density, which should be the case, because density is a higher-order statistic that should

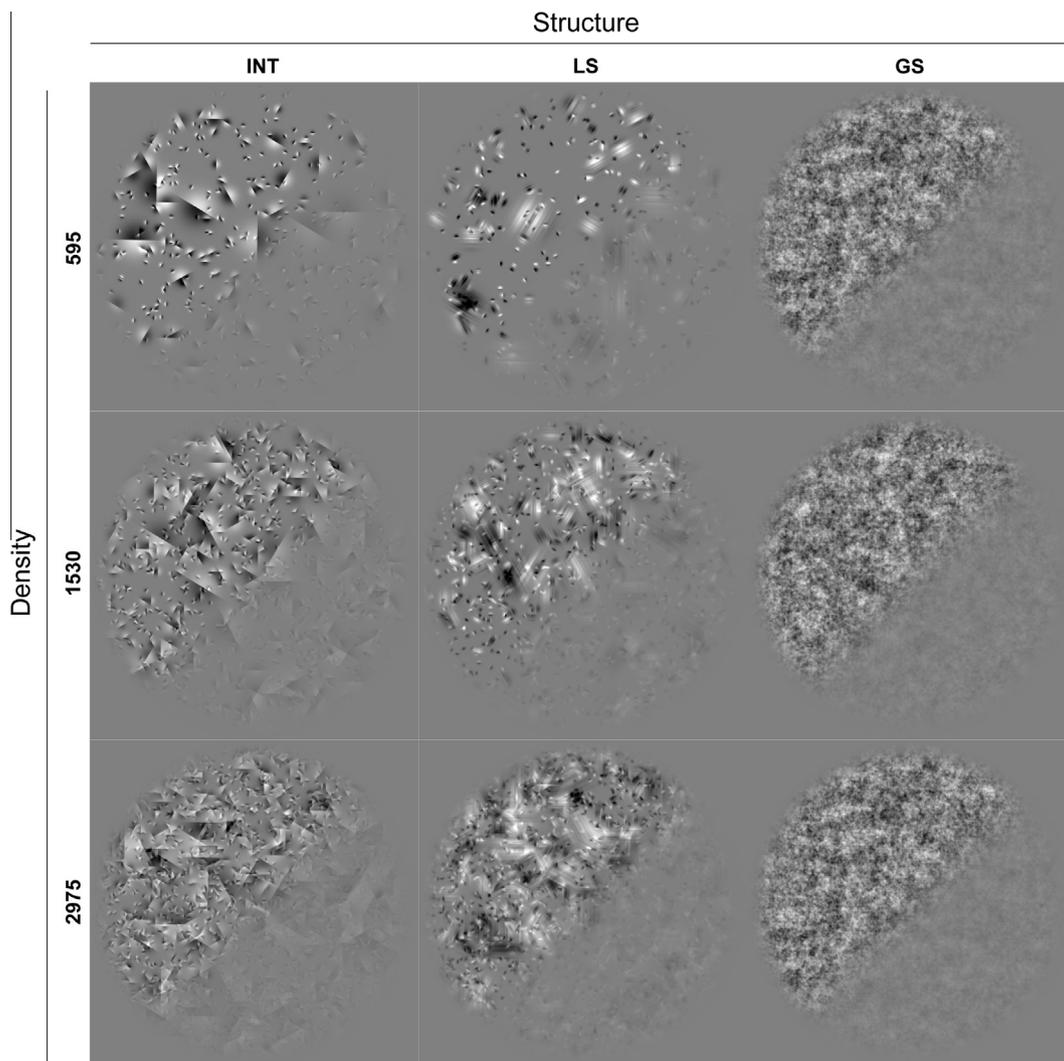


Fig. 3. Examples of contrast modulated stimuli used for Experiment 1, shown at a modulation depth of 50%. The structure conditions are arranged horizontally, while the density increases from top to bottom. The three structure conditions are intact (INT), locally scrambled (LS), and globally scrambled (GS). Notice that the globally scrambled condition does not appear different at varying densities, because density information is destroyed by phase scrambling.

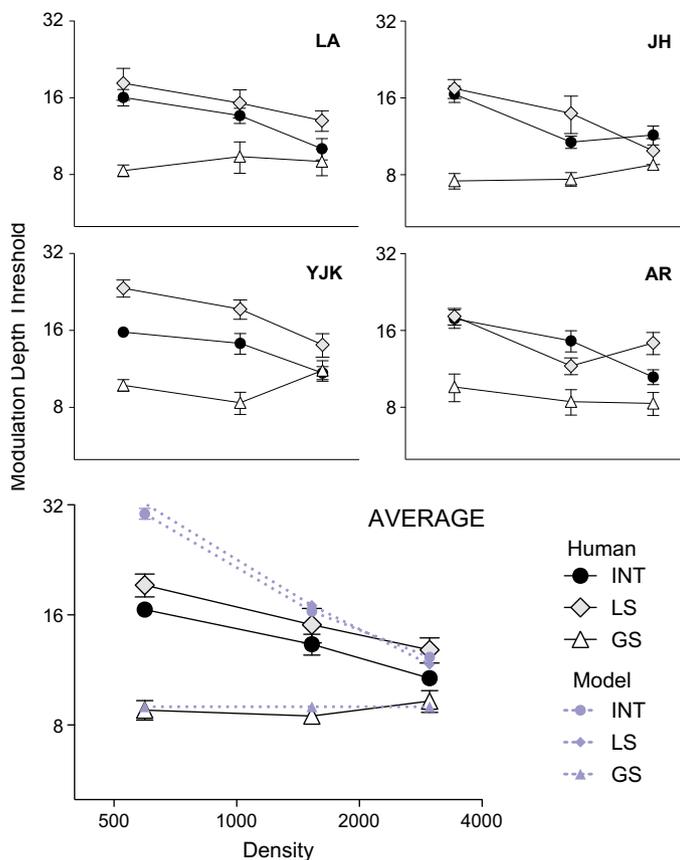


Fig. 4. Experiment 1 (contrast boundary segmentation) results for four observers (small graphs), and the average of these observers (large graph). The structure conditions are: intact (INT), filled circles; locally scrambled (LS), grey diamonds; and globally scrambled (GS), open triangles. Density increases along the horizontal axis. Note improved performance for phase scrambled carrier textures (GS), and lack of effect of density in this condition. In contrast, intact and locally scrambled conditions both result in higher thresholds at low densities than at high. Error bars indicate standard errors.

be largely nullified following phase scrambling in broadband images such as these.

The two-way ANOVA showed a significant main effect of density $F(2,9) = 12.25, p < .05$, and a significant interaction between density and structure $F(4,9) = 7.69, p < .05$. A nonlinear regression (linear in log–log space) was performed on the each of the structure conditions as a function of density (INT, LS, and GS). F -tests confirmed that the slope of the globally scrambled condition (GS) with respect to density is not significantly different from a slope of zero ($F(1,10) = 0.37, p > .05$), while the slopes of the intact ($F(1,10) = 51.52, p < .05$) and locally scrambled ($F(1,10) = 12.99, p < .05$) conditions are significantly different from zero.

These results suggest that our synthetic textures capture at least some of the image statistics that cause the phase scrambling effect that we observed earlier in natural textures (Arsenault, Yoonessi, & Baker, 2011). Because performance in the intact and locally scrambled conditions is about the same, it appears that local phase structure may not substantially affect contrast boundary segmentation. Furthermore, because segmentation thresholds decrease as density is increased, it appears that sparseness is a key aspect of global structure that impairs contrast boundary segmentation.

4. Experiment 2: Orientation boundary segmentation

In the previous experiment we found that global phase structure, specifically sparseness, influenced contrast boundary

segmentation. This experiment aimed to extend those findings to texture boundaries defined by orientation. The structure of the models used to segment orientation-defined boundaries is very similar or the same as those used to segment contrast boundaries (Landy & Oruç, 2002) – two stages of linear filtering separated by a pointwise nonlinearity – so it is a reasonable possibility that the two conditions might produce a similar pattern of results.

4.1. Methods

In this experiment, we measured modulation depth thresholds for observers segmenting boundaries defined by differences in texture orientation. To maximize orientation contrast we used pairs of textures, each texture narrowband for orientation with micropatterns oriented at 0° and 90° to form a “herringbone” along the $\pm 45^\circ$ boundary in the quilted stimulus. As in the contrast boundary segmentation task, thresholds were measured for intact and locally scrambled textures at each of three density levels (Fig. 5). The globally scrambled stimulus was only depicted once because density information is destroyed following phase scrambling. In this experiment, modulation depth thresholds were measured for four experienced psychophysical observers with normal or corrected-to-normal vision, three of whom (JB, JH, AR) were naive to our hypotheses.

4.2. Results

Orientation boundary segmentation results are plotted in Fig. 6, with individual thresholds in the four top graphs, and group-average results at bottom. The global scramble (GS) condition was tested only once, because it does not vary with density (as observed in Experiment 1).

Effect of local structure and density: The data suggest that removing only local phase alignments (LS, grey diamonds) has no systematic effect on segmentation, as evidenced by the similarity between the intact (INT, dark circles) and LS conditions. Thresholds in both the intact (INT) and LS conditions appear to decrease as density increases. A two-way 2×3 ANOVA confirmed a main effect of density ($F(2,9) = 51.78, p < .05$), found no effect of structure ($F(1,9) = 0.03, p > .05$), and no evidence of an interaction ($F(2,9) = 0.01, p > .05$).

Effect of global structure: As in the contrast boundary segmentation experiment, the data indicate that removing global phase structure (GS, white triangles) decreased thresholds substantially relative to the intact and locally scrambled texture thresholds at low densities, and progressively less at higher densities. To confirm these effects, we compared the INT and LS conditions to the GS condition at each density using three t -tests with Sidak–Bonferroni corrections for multiple comparisons. The GS condition was significantly different from the INT and LS conditions at each density: 595 ($t(10) = 14.56, p < .05, d = 8.41$), 1530 ($t(10) = 5.74, p < .05, d = 3.54$), and 2975 ($t(10) = 2.92, p < .05, d = 1.81$). The size of the effect (d) decreases as density increases.

Comparison of orientation and contrast boundary segmentation: Orientation and contrast boundary segmentation can be compared in three ways: the overall difficulty of the boundary type, the effect of density, and the effect of phase scrambling. (1) Thresholds are much higher in every condition for orientation modulations than for contrast modulations, indicating that orientation boundary segmentation is overall more difficult than contrast boundary segmentation. A one-tailed, paired samples, t -test confirmed that this difference is statistically significant $t(6) = 5.13, p < .05$. This finding is consistent with Motoyoshi and Nishida (2004), who also found a similar quantitative difference when comparing orientation- and contrast-defined boundary segmentation directly. (2) The threshold decline with density for the intact and locally

scrambled conditions appears slightly steeper in the case of orientation than contrast boundary segmentation. A nonlinear regression (linear in log–log space), revealed that although the slopes were steeper for orientation (-0.49 for both INT and LS) than for contrast (-0.26 for both INT and LS), these individual measures of slope did not fit the data significantly better using the same slope for all four functions ($F(3,40) = 1.341, p > .05$). (3) The influence of phase scrambling can be examined by looking at the change in log threshold between the intact (INT) and globally scrambled conditions (GS) at each density. A comparison of this difference at each density between orientation and contrast boundary segmentation using a paired, two-tailed t -test shows that the influence of phase scrambling was significantly greater for orientation modulated stimuli than for contrast modulated stimuli $t(2) = 4.45, p < .05$.

In sum, orientation boundary segmentation, like contrast boundary segmentation, is affected by global phase structure, but not local phase structure, and thresholds decline with density at a similar rate. However, it is more difficult than contrast boundary segmentation, and exhibits a greater difference in performance between intact and global phase scrambling.

5. Model

Second-order boundary segmentation has often been understood in terms of a “filter–rectify–filter” model using early high spatial-frequency filters to capture the texture, followed by a nonlinearity and a late, low spatial-frequency filter that recovers the boundary (e.g. Chubb & Sperling, 1988; Landy & Graham, 2004; Malik & Perona, 1990). Here we implement a model with a filter–rectify–filter architecture in order to see whether the observed effects of structure and density can be accounted for using such a model, and if so, for what configuration and parameter values.

5.1. Filter–rectify–filter model

We implemented a basic filter–rectify–filter model (Fig. 7A), as described below. First the stimulus ($S_{x,y}$) was convolved ($*$) with a bank of linear filters (G_1) that varied in orientation (θ), spatial frequency (ω), and phase (ϕ):

$$F_1(\theta, \omega, \phi, x, y) = G_1(\theta, \omega, \phi) * S(x, y) \quad (10)$$

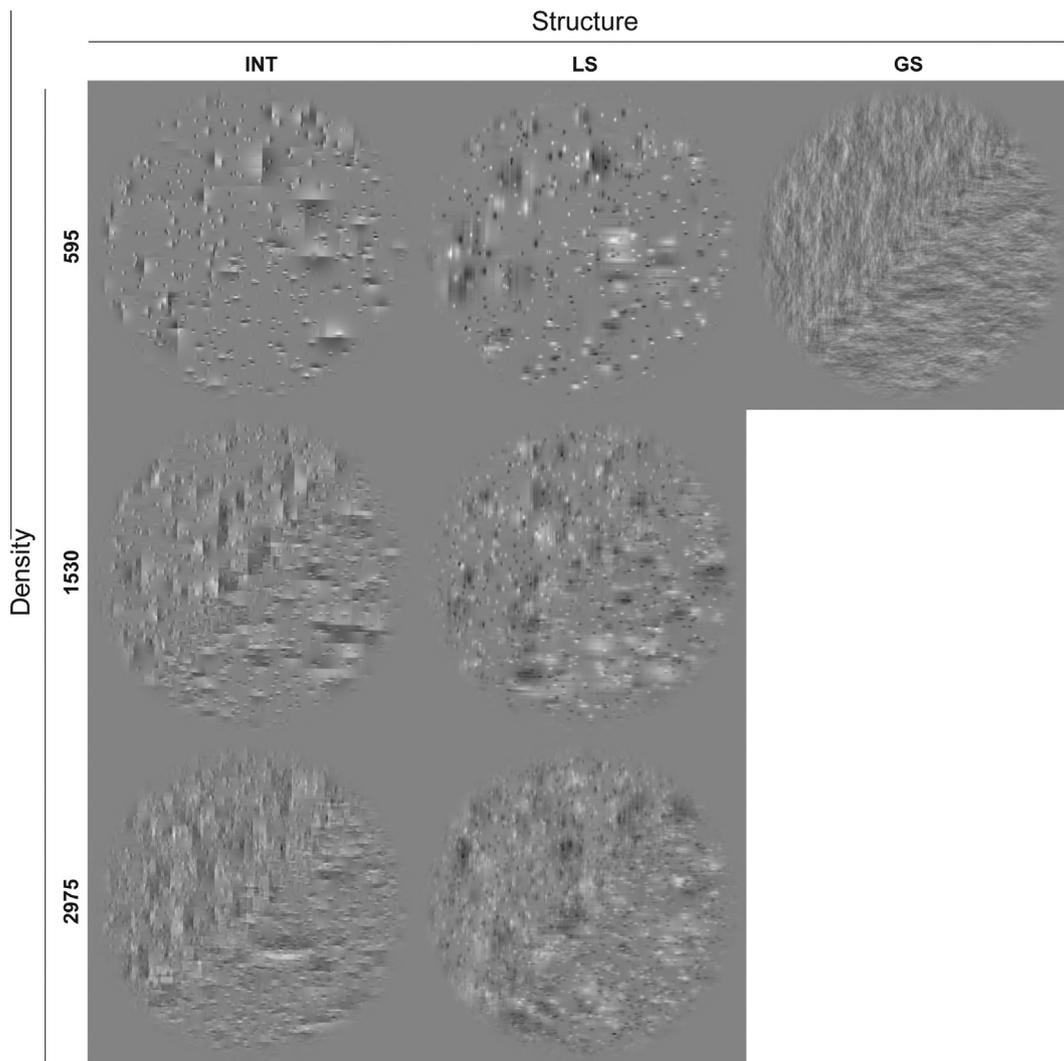


Fig. 5. Examples of orientation modulated stimuli used for Experiment 2, shown at a modulation depth of 100%. The structure conditions are arranged horizontally, while the density is varied vertically. The globally scrambled condition (GS) was only tested once, because density information is destroyed by phase scrambling and (as expected) no systematic effect of density was observed in the previous experiment.

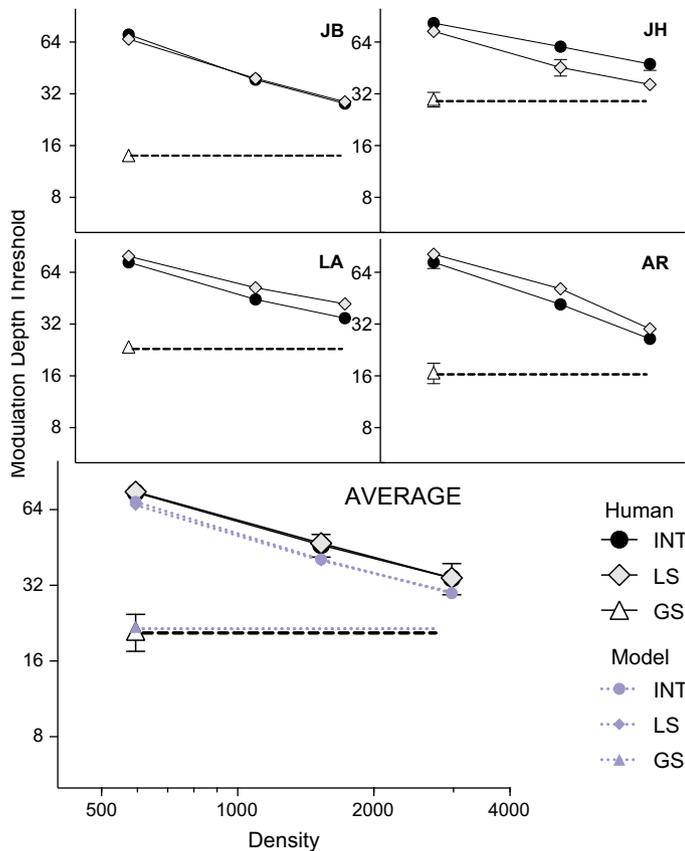


Fig. 6. Experiment 2 (orientation boundary segmentation) results for four observers (small graphs), and the average of these observers (large graph). The structure conditions are: intact (INT), filled circles; locally scrambled (LS), open circles; and globally scrambled (GS), open triangles. These results show improved performance for phase scrambled carrier textures (GS) as previously found for contrast segmentation. There is no systematic difference between the intact and locally scrambled conditions, as both result in higher thresholds at low densities than at high. Error bars indicate standard errors.

The filters were log-gabors (Kovesi, 2000), at two phases (even and odd), six orientations (evenly spaced with their bandwidths chosen for approximately uniform coverage), and four spatial frequencies (160, 80, 40, and 20 cpi, each with a bandwidth of approximately 1.5 octaves). The output of each of these filters (F_1) was weighted (w_f), full-wave rectified and raised (pointwise) according to a power-law of order k , then pooled over phase:

$$\mathbf{R}(\theta, \omega, x, y) = \sum_{\phi} |\mathbf{F}_1(\theta, \omega, \phi, x, y) \cdot w_f|^k \quad (11)$$

The values of w_f were chosen to equalize responses across spatial scales for a stimulus with a $1/f$ spectral falloff (Field, 1987). To this end, the responses to the higher spatial frequency channels were increased relative to the responses to lower spatial frequencies, using a weighting function $w(f) = 2^f$, where f is an index of spatial frequency with $f = 1$ designating the lowest frequency. Note that this is functionally equivalent to multiplying by the frequency (f). Dot products were computed between these responses ($\mathbf{R}_{x,y}$) and two second-stage filters (\mathbf{G}_2) in the form of low spatial frequency sine-phase gabor functions that match the two possible orientations of the boundary in the stimulus ($+45^\circ$ and -45°), as well as its central position:

$$\mathbf{F}_2(+45, \theta, \omega) = \mathbf{G}_2(+45) \cdot \mathbf{R}(\theta, \omega, x, y) \quad (12)$$

$$\mathbf{F}_2(-45, \theta, \omega) = \mathbf{G}_2(-45) \cdot \mathbf{R}(\theta, \omega, x, y) \quad (13)$$

The outputs (o) were computed by pooling the magnitudes of the late-stage filter responses across the orientations and spatial frequencies of the early-stage filters. The pooled response magnitudes were raised to a power (the reciprocal of k), and then combined with additive decision noise (n) for the final output value. (Eqs. (14) and (15)). The noise values n_1 and n_2 were drawn from a normal distribution with a mean of 0 and whose standard deviation, or amplitude, a is a free parameter of the model.

$$o_{+45} = (\sum_{\theta, \omega} |\mathbf{F}_{2,+45}(\theta, \omega)|)^{1/k} + n_1 \quad (14)$$

$$o_{-45} = (\sum_{\theta, \omega} |\mathbf{F}_{2,-45}(\theta, \omega)|)^{1/k} + n_2 \quad (15)$$

These outputs were compared, and the late-stage filter (left- or right-oblique) with the strongest response determined the decision (d) of the model:

$$d(o) = \begin{cases} +45 & \text{if } o_{+45} \geq o_{-45} \\ -45 & \text{if } o_{-45} > o_{+45} \end{cases} \quad (16)$$

5.2. Simulation

We tested the model to determine its segmentation thresholds in much the same manner as we tested our human participants. The model made left- or right-oblique decisions in 60 trials for each of the stimulus conditions illustrated in Figs. 3 and 5 on 12 logarithmically spaced modulation depth levels that spanned chance to perfect performance. We measured the percent-correct for each level and stimulus condition, and then fit a logistic function using Matlab to determine the model's threshold. Because the stimuli are randomly generated on each trial, model results varied from one simulation to another. For this reason, we simulated the experiment four times and averaged the thresholds. Standard errors were determined based on variability between the four runs.

5.3. Optimization

The model has two free parameters: k , the order of the power-law nonlinearity, and a , the amplitude of the decision noise distribution from which n_1 and n_2 are sampled. We optimized the model to best estimate the thresholds in both the contrast and orientation tasks by performing an exhaustive search of the parameter space. We simulated the experiment as described above for five power-law exponents, each at 400 noise levels. For each pair of noise amplitude and power-law exponent values, we computed a sum of squares error from the difference between the log human threshold (h), and log model performance (m) (Fig. 7B).

$$\text{SSE} = \sum_i (\log h_i - \log m_i)^2 \quad (17)$$

5.4. Results

The minimum SSE was obtained with $k = 0.5$, and $a = 10^{7.25}$. The variance accounted for (VAF), computed using Pearson's r , with these parameters was 96.3% (Fig. 8A). The threshold estimates, using the optimized values of the two free parameters, were averaged over four runs of the model and plotted in Figs. 4 (contrast) and 6 (orientation). Fig. 4 shows the contrast boundary segmentation results for the model (dashed lines) and humans (solid lines). The model, like human observers, shows a decrease in segmentation thresholds for globally scrambled textures relative to the similar intact and locally scrambled texture conditions, and matches the human thresholds well for most conditions. Thresholds for intact and locally scrambled textures are again very similar and the

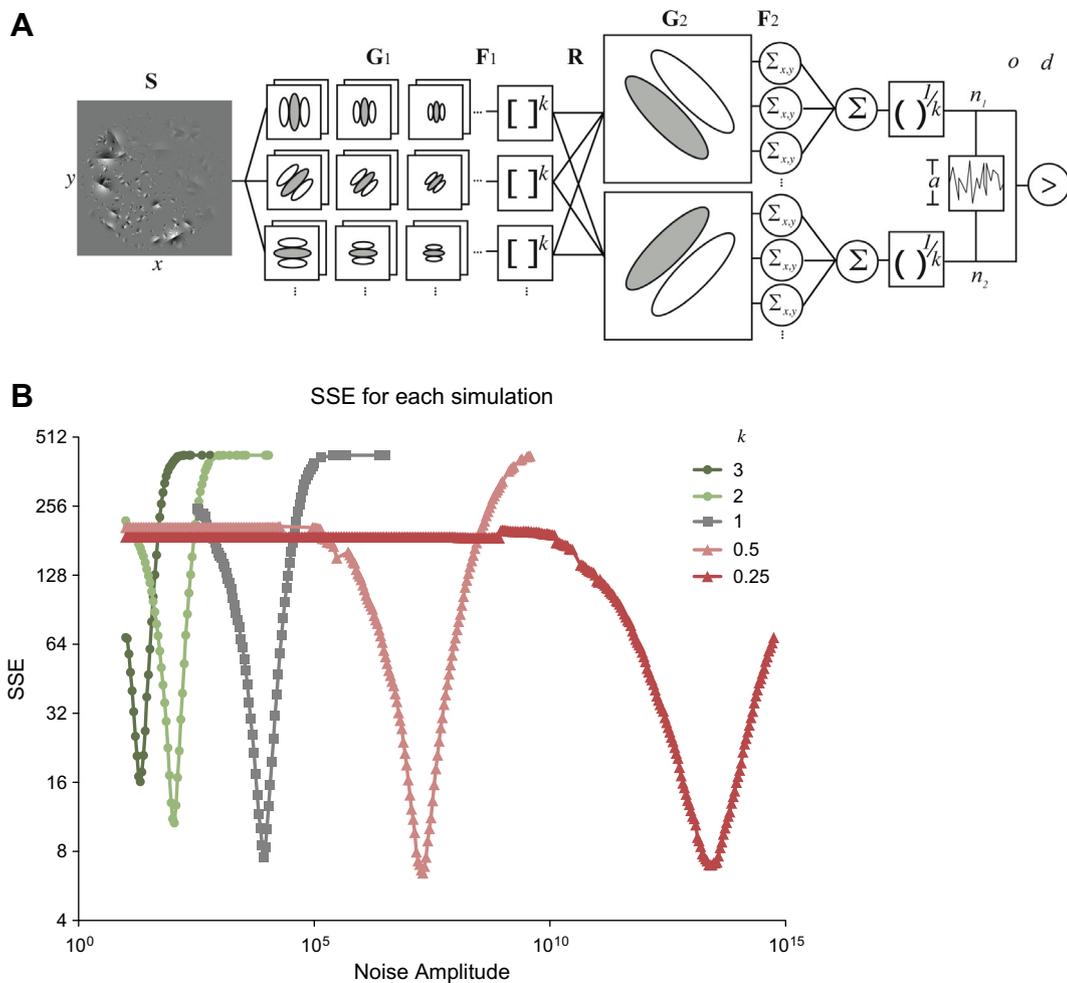


Fig. 7. Model architecture and parameter optimization. (A) Architecture of our FRF-style model. Symbols at top correspond to equations in text. In this model, a stimulus image (S) is filtered by a bank of quadrature pair log gabor functions (G_1). These early-stage filter outputs (F_1) are rectified and raised to some power, k . The rectified outputs (R) are then filtered by two large gabor functions (G_2) matched to the scale and possible orientations of the boundary. For each second-stage filter, responses (F_2) are pooled over space and across first-stage channels. The scalar output for each second-stage filter is raised to the reciprocal of k , and decision noise is added. The filter (left- or right-oblique) with the greatest response indicates the model's 'perceived' boundary orientation. (B) Sum-of-squares error for each simulation, as a function of noise amplitude parameter – each curve corresponds to a different value of k , the power-law exponent. For each value of k there is a local minimum with respect to noise amplitude – the global minimum of these occurring when $k = 0.5$. For the sake of clarity, only points where SSE was less than 500 are illustrated.

model consistently predicts declining threshold with density, but the model considerably overestimates the thresholds in the lowest density condition, and thus the rate at which thresholds decrease with density in the INT and LS conditions. The orientation boundary segmentation results are shown in Fig. 6. The model matches the quantitative human data in the GS condition perfectly. As in the human data, the model shows little difference between the INT and LS conditions at any density, and the rate at which thresholds in these conditions decrease with density is very similar to the human results. The model appears to slightly underestimate thresholds for the INT and LS conditions at all densities. A detailed examination of the character and robustness of the model fit is provided in the Supplementary material (Figs. S1 and S2).

Both the model and human observers had higher thresholds for orientation than for contrast boundary segmentation, which is readily explainable in terms of the model. The information for orientation segmentation is only in two orientation channels of the early filters (vertical and horizontal), while contrast information is available in all of the first-stage spatial frequency and orientation channels. Overall, the model performs very well given that only two free parameters were optimized. With a compressive nonlinearity and an appropriate amount of decision noise, the

model performance depends upon global higher order statistics in the same way that human performance does, yet also (like humans) is insensitive to local phase structure.

After fitting the model to thresholds obtained using synthetic textures, we used it to predict the average thresholds for each texture published for contrast boundary segmentation in natural textures (Arsenault, Yoonessi, & Baker, 2011) (Fig. 8B). These 40 thresholds (20 intact, 20 scrambled) were reasonably well-predicted by the model, with VAF = 0.86%.

6. Discussion

Our previous work (Arsenault, Yoonessi, & Baker, 2011) indicated that higher-order image statistics impair segmentation, and suggested that sparseness and/or local edge structure might be important texture statistics for this task. Here we have identified sparseness as a critical texture property, ruled out a role for local phase alignments, and extended our findings on contrast boundary segmentation to orientation boundaries. In both cases the presence of higher-order structure was found to impair segmentation; among textures with structure (i.e., excluding globally scrambled

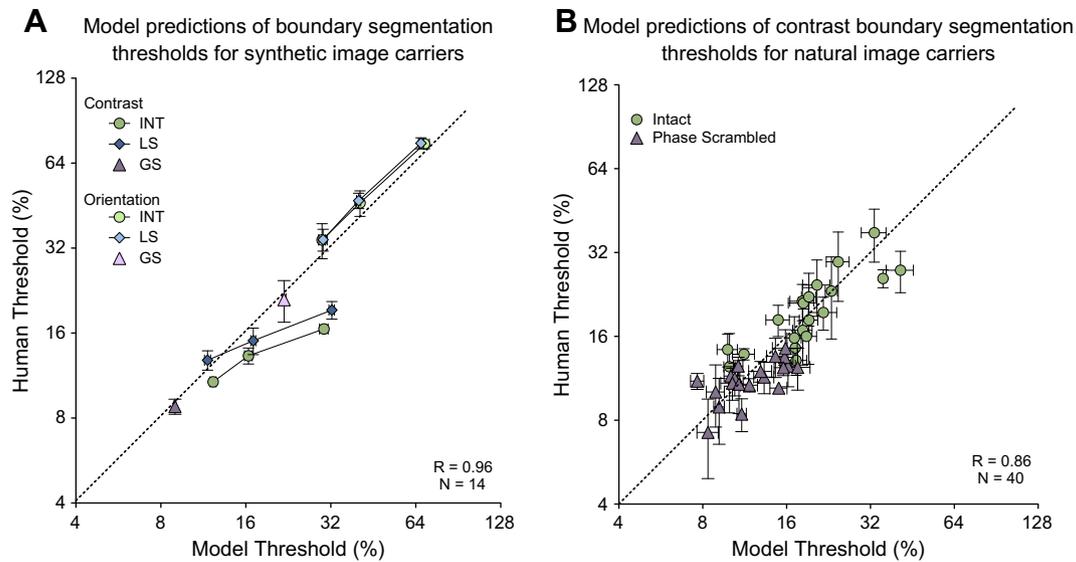


Fig. 8. Model performance following parameter optimization. (A) Comparison of human and model segmentation thresholds for synthetic textures for all 14 conditions. The dashed line indicates 1:1 correspondence between thresholds. The variance accounted for, calculated with Pearson's R , is 96%. The model correctly predicts the thresholds of the globally scrambled textures (triangles) and the high density textures, but somewhat overestimates the difficulty of the low-density textures in the contrast boundary segmentation task (dark diamonds and circles). It correctly predicts no difference between the intact (circles) and locally scrambled (diamonds) conditions. (B) Comparison between human and model thresholds for the 20 natural textures in both natural and phase scrambled conditions (for a total of 40 thresholds) tested in Arsenault, Yoonessi, and Baker (2011). The model does a good job predicting segmentation thresholds (VAF = 86%) for most of the intact and phase scrambled natural textures.

textures), segmentation was progressively impaired by increasing sparseness. For both orientation- and contrast-defined boundaries, the presence of local phase alignments did not affect segmentation. These results were accounted for using an FRF-style model with a compressive intermediate nonlinearity, an expansive output nonlinearity, and decision noise.

It can be instructive to examine why the model behaves as it does for sparse textures. When sparseness is increased, with overall RMS contrast held constant, energy is clumped into local, higher contrast regions separated by lower contrast regions. These changes in local contrast are themselves contrast modulations that could act as masking noise for the mechanism segmenting the main boundary (Allard & Faubert, 2007). We examined the potential impact of these local contrast modulations on the model's behaviour by measuring the average second-stage filter response to ten randomly generated, unmodulated textures in each condition (Fig. S3 – see Supplementary material). Even though the second-stage filter was narrow-band and the local contrast modulations relatively high spatial frequency (Hutchinson & Ledgeway, 2004), its response was affected by the texture's sparseness. Second-stage filter responses were lowest (less second-order interference) in the GS condition, and highest (more second-order interference) in the INT & LS conditions at the lowest density (595). The average response also progressively declined with increasing density. Because the model processes orientation and contrast boundaries in the same pathway, local contrast modulations are detectable by the second-stage filter in the same way, and thus could similarly interfere with orientation boundary segmentation. Thus it appears as though the effect of sparseness in the model is due to the second-order noise inherent in sparse textures. This result suggests that a similar masking might be happening in human observers.

A compressive intermediate nonlinearity provided the best fit for our results. In most earlier texture segmentation models, a square law has been the conventional intermediate nonlinearity, probably because it conceptually corresponds to the Fourier energy (Malik & Perona, 1990). Graham and Sutter (1998) used a local contrast summation paradigm to estimate this nonlinearity, finding that it was expansive with k between 2 and 4. The role of the

compressive nonlinearity in fitting our results was mainly to minimize the effect of local structure, and to a lesser extent, to determine the magnitude of the effect of sparseness. Such a compressive nonlinearity is consistent with some recent analyses of visual cortex neurons. Mineault et al. (2012) estimated the nonlinearity transforming MT outputs before they are combined by MST neurons is usually compressive, with values of k approximately 0.2–0.4. Nishimoto and Gallant's (2011) model that best predicted responses of MT neurons used a compressive nonlinearity ($k = 0.5$) between V1 outputs and MT. This compressive summation could be the result of normalization mechanisms, for example contrast gain control or surround suppression (Carandini & Heeger, 2011).

Recent work by Westrick, Henry, and Landy (2013) demonstrated that a noise masking method of estimating second-stage filter bandwidth produces qualitatively different results from a discrimination at threshold task. They showed that both kinds of result could be explained by a model with an intermediate-stage normalizing nonlinearity, using a "winner-take-all" operation between early-stage filter responses. Due to the early noise in their model, this nonlinearity behaves on average like a sigmoid-shaped function with a steep slope, which is like a thresholding function. Such a function would have a similar effect as our simple compressive nonlinearity, by producing an approximate binarization of early filter responses.

A notable discrepancy of our model is its prediction that sparseness impairs performance on contrast boundaries more than is seen for human observers (Fig. 4). This difference between observed and simulated results could occur if human observers employ a second-stage filter with a narrower bandwidth than implemented in our model – such a filter would pass less of the "contrast noise" generated by texture sparseness. However in that case it is unclear why a similar discrepancy is not seen for orientation boundaries (Fig. 6).

The best model for our results employed an (expansive) output nonlinearity following the second stage of filtering and pooling, as well as an intermediate (compressive) nonlinearity between the first- and second-stage filters. The sigmoidal function that is commonly used to represent the contrast response function of a neuron is expansive at low contrasts and compressive at high contrasts. It

is possible that the first-stage responses are strong enough to be affected by gain control mechanisms, and thus fall on the compressive part of such an S-shaped contrast response function, while the second-stage responses fall within the lower, expansive, portion of such a function.

An important kind of sensitivity to higher-order image statistics might be some sort of ‘conjunction-detection’, i.e. an enhanced (supralinear) response to combinations of signals from earlier stages. Methods of conjunction detection have been proposed on the feature level, using products of early filter responses and clustering methods (Freeman & Simoncelli, 2011; Martin, Fowlkes, & Malik, 2004). Peirce (2007) suggests that an intermediate compressive nonlinearity can serve to create conjunction selectivity – however, see also May and Zhaoping (2011), Peirce (2011).

The model fits our data reasonably well with two free parameters, but there are a number of biologically relevant changes that might improve the model further if a more complex data set were available for fitting additional model parameters. Interactions across space in the form of surround suppression (Tanaka & Ohzawa, 2009), or between channels in the form of cross-orientation inhibition (Motoyoshi & Nishida, 2004) are suggested by the compressive nonlinearity, but their nature is not revealed in detail. Spatial interactions are critical to incorporate, because spatial arrangement of information inherently reflects higher-order image statistics. Likewise, cross-channel interactions might be expected to play an important role for responses to broadband stimuli (Bex, Mareschal, & Dakin, 2007; David, Vinje, & Gallant, 2004).

There are some image properties known to be relevant to texture perception that our edgelet-based textures did not allow us to manipulate. For instance, we did not consider local contrast polarity, though it can enable segmentation (Malik & Perona, 1990; Motoyoshi & Kingdom, 2007), or the higher-order spatial properties that are known to be relevant to texture appearance (Portilla & Simoncelli, 2000). However, edgelet micropattern textures provided a useful tool for studying the impact of specific texture properties on segmentation in a parametric way. They allowed us to produce textures with a naturalistic shape of the amplitude spectrum ($1/f$ rolloff) and to vary sparseness while separating out the effects of global from local phase alignments. This enabled us to characterize the distinct effects (or lack thereof) of local phase alignments, global phase relationships, and density on second-order segmentation mechanisms.

From this work, we can conclude that higher-order statistical information is important to texture segmentation mechanisms, and that texture sparseness in particular plays a large role our ability to see boundaries defined by that texture.

Acknowledgments

The authors would like to thank James Elder, Fred Kingdom and Aaron Johnson for helpful discussions and suggestions. We are also grateful to Ahmad Yoonessi, who provided useful contributions to programming. And we wish to especially thank our psychophysical observers who so generously donated their time. This research was funded by an NSERC Grant OPG0001978 to C.B.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.visres.2013.07.018>.

References

Allard, R., & Faubert, J. (2007). Double dissociation between first-and second-order processing. *Vision Research*, *47*, 1129–1141.

- Arsenault, E., Yoonessi, A., & Baker, C. Jr., (2011). Higher order texture statistics impair contrast boundary segmentation. *Journal of Vision*, *11*.
- Beck, J. (1983). Textural segmentation, second-order statistics, and textural elements. *Biological Cybernetics*, *48*, 125–130.
- Bex, P. J. (2010). Sensitivity to spatial distortion in natural scenes. *Journal of Vision*, *10*.
- Bex, P. J., Mareschal, I., & Dakin, S. C. (2007). Contrast gain control in natural scenes. *Journal of Vision*, *7* (article no. 12).
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Caelli, T. M. (1980). Facilitative and inhibitory factors in visual texture discrimination. *Biological Cybernetics*, *39*, 21–26.
- Carandini, M., & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62.
- Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli – A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America*, *A*, *5*, 1986–2007.
- Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology*, *12*, 476.
- David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *Journal of Neuroscience*, *24*, 6991–7006.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, *A*, *4*, 2379–2394.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*, 1195–1201.
- Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision Research*, *38*, 231–257.
- Hutchinson, C. V., & Ledgeway, T. (2004). Spatial frequency selective masking of first-order and second-order motion in the absence of off-frequency ‘looking’. *Vision Research*, *44*, 1499–1510.
- Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory*, *8*, 84–92.
- Julesz, B., Gilbert, E. N., & Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, *31*, 137–140.
- Kingdom, F. A. A., Hayes, A., & Field, D. J. (2001). Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision Research*, *41*, 585–598.
- Klein, D. F. (2005). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Am Psychiatric Assoc.
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3. *Perception*, *36*.
- Kovesi, P. D. (2000). MATLAB and octave functions for computer vision and image processing.
- Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1106–1118). MIT.
- Landy, M. S., & Oruc, I. (2002). Properties of second-order spatial frequency channels. *Vision Research*, *42*, 2311–2329.
- Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America*, *A*, *7*, 923–932.
- Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*, 530–549.
- May, K. A., & Zhaoping, L. (2011). Exploring the roles of saturating and supersaturating contrast-response functions in conjunction detection and contrast coding. *Journal of Vision*, *11* (article no. 11).
- Meso, A. I., & Hess, R. F. (2011). Orientation gradient detection exhibits variable coupling between first-and second-stage filtering mechanisms. *Journal of the Optical Society of America*, *A*, *28*, 1721–1731.
- Mineault, P. J., Khawaja, F. A., Butts, D. A., & Pack, C. C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proceedings of the National Academy of Sciences*, *109*, E972–E980.
- Motoyoshi, I., & Kingdom, F. A. A. (2007). Differential roles of contrast polarity reveal two streams of second-order visual processing. *Vision Research*, *47*, 2047–2054.
- Motoyoshi, I., & Nishida, S. (2004). Cross-orientation summation in texture segregation. *Vision Research*, *44*, 2567–2576.
- Nishimoto, S., & Gallant, J. L. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *Journal of Neuroscience*, *31*, 14551–14564.
- Peirce, J. W. (2007). The potential importance of saturating and supersaturating contrast response functions in visual cortex. *Journal of Vision*, *7* (article no. 13).
- Peirce, J. (2011). Nonlinear summation really can be used to perform AND operations: Reply to May and Zhaoping. *Journal of Vision*, *11*.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–70.
- Tanaka, H., & Ohzawa, I. (2009). Surround suppression of V1 neurons mediates orientation-based representation of high-order visual features. *Journal of Neurophysiology*, *101*, 1444–1462.
- Watson, A. B., & Eckert, M. P. (1994). Motion-contrast sensitivity: Visibility of motion gradients of various spatial frequencies. *Journal of the Optical Society of America*, *A*, *11*, 496–505.
- Westrick, Z. M., Henry, C. A., & Landy, M. S. (2013). Inconsistent channel bandwidth estimates suggest winner-take-all nonlinearity in second-order vision. *Vision Research*, *81*, 58–68.