Supplementary Information

Appendix A. Patch size *vs.* classification accuracy

All analyses in the main text were performed on figure and ground patches of size 16×16 pixels. The size of the patch in relation to the size of the object at the boundary from which it has been extracted determines the amount of local features contained in the patch. The BSDS300 database which has all images of same size, 481×321 is used in this section. In this analysis, we study the variation of figure/ground classification accuracy based on eq 6 with different sizes of figure and ground patches and hence the amount of local features on them.

We extract the figure and ground patches of 5 different sizes from 14×14 pixels up to 18×18 pixels in increments of one pixel. To eliminate confounding effects of size, once we extract a patch of a size different than 16×16 pixels, it is rescaled to 16×16 pixels, using linear interpolation of pixel values. Therefore, even though the patches are of the same size after re-scaling (16×16), they contain different densities of local features surrounding the boundary. Power spectra are computed for each dataset corresponding to 5 different sizes, as was done before in Sections 5.1 and 5.2. Our results indicate that the FG classification accuracy is little changed with ± 2 pixel change in size of the figure and ground patches (Table A.4).

Appendix B. Image size *vs.* classification accuracy

In an approach that is complementary to that in Appendix A, we now measure the effect of image size on SA when patch size is kept constant,

patch size	Classification accuracy
18×18	60.63%
17×17	60.61%
16×16	62.57%
15×15	60.47%
14×14	61.96%

Table A.4: Classification accuracy *vs.* size of figure and ground patches.

Image size	Number of patch pairs	Classification accuracy
$0 - 0.5$	786	63.36%
$0.5 - 2$	360	68.14%
$2 - 4$	264	68.3%
> 4	348	67.05%

Table B.5: Classification accuracy *vs.* image size (in Mega Pixels) for the LabelMe dataset. Figure and ground patches are 16×16 pixels. The second column shows the numbers of figure/ground patch pairs available for the respective image sizes.

at 16×16 pixels. We use patches from the LabelMe dataset as it has a wide range of image sizes which we quantify in terms of the number of pixels present in the image.

From Table B.5, we see that image size only has a moderate impact on figure/ground classification accuracy. There is a tendency that image sizes in the range of 0.5 - 4 mega pixels give somewhat better classification results (68%) for patch size of 16×16 pixels than other image sizes. Our interpretation of these results is that when the images are very small (< 0.5 mega pixels), more global information is included in the SA analysis window, hence poorer classification accuracy results. On the other hand, when the

image size is very large (> 4 mega pixels), classification accuracy is slightly reduced because the relatively small patch size compared to the foreground area makes it difficult to capture enough information about SA, since the variation in surface curvature of the underlying figure will be small.

Appendix C. Plots related to the LabelMe database

Section 5 in the main text showed results for the BSDS300 database, Figures 3 and 4. The corresponding plots for the LabelMe database are shown in Figures C.8 and C.9.

Appendix D. Patch Extraction Procedure

As described in Section 3, for our analysis we require all image patches to have a common reference frame in which x varies along the OB and y varies orthogonal to it. Patches are therefore rotated into this common reference frame.

Let us denote the location on the OB between figure and ground (yellow dot in Fig. 2A) from which figure and ground patches are to be extracted, as $\mathbf{u}_{center}$. We select a point $\mathbf{u}_{tangent}$ on the tangent to the OB, with $\mathbf{u}_{tangent} \neq \mathbf{u}_{center}$. The distinction between figure and ground is made by visual inspection and a point $\mathbf{u}_{figure}$ is located anywhere on the figure side.

The angle θ_{rot} by which patch $\psi(x, y)$ is rotated is computed as,

$$\theta_{rot} = \angle \mathbf{v}_{FG} + \pi/2 \tag{D.1}$$

where the symbol $\angle \mathbf{v}$ stands for the angle between vector $\mathbf{v}$ and a fixed coordinate axis, say a horizontal border of the image. The vector $\mathbf{v}_{FG}$ is

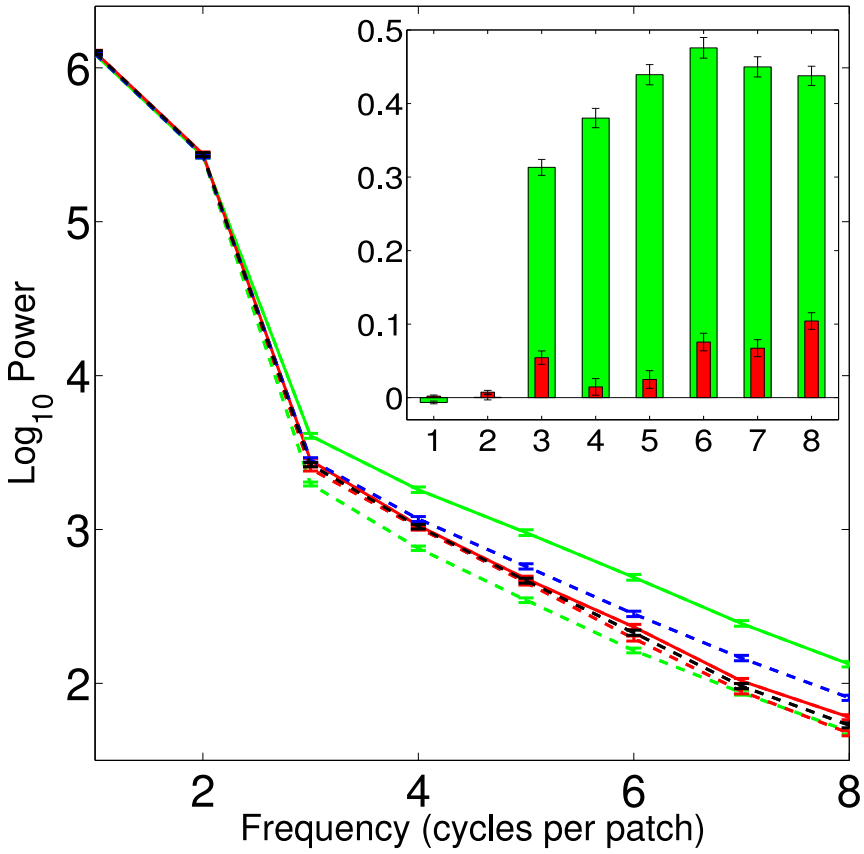

Figure C.8: Average power spectra of all patches of LabelMe data as function of spatial frequency. The unoriented spectra are represented by dashed blue (figure) and black (ground) lines. The oriented spectra in the plot are: $\overline{E}_{f\perp}$ (solid green line), $\overline{E}_{f\parallel}$ (dashed green line), $\overline{E}_{g\perp}$ (solid red line) and $\overline{E}_{g\parallel}$ (dashed red line). Inset: The difference in power ($\log_{10}(\bar{E}_{s\perp} - \bar{E}_{s\parallel})$) in each frequency bin. Axes same as in main figure. Green and red bars represent figure ($s = f$) and ground ($s = g$) differences respectively. Error bars are standard errors in figure and inset. For complementary BSDS results, see Figure 3

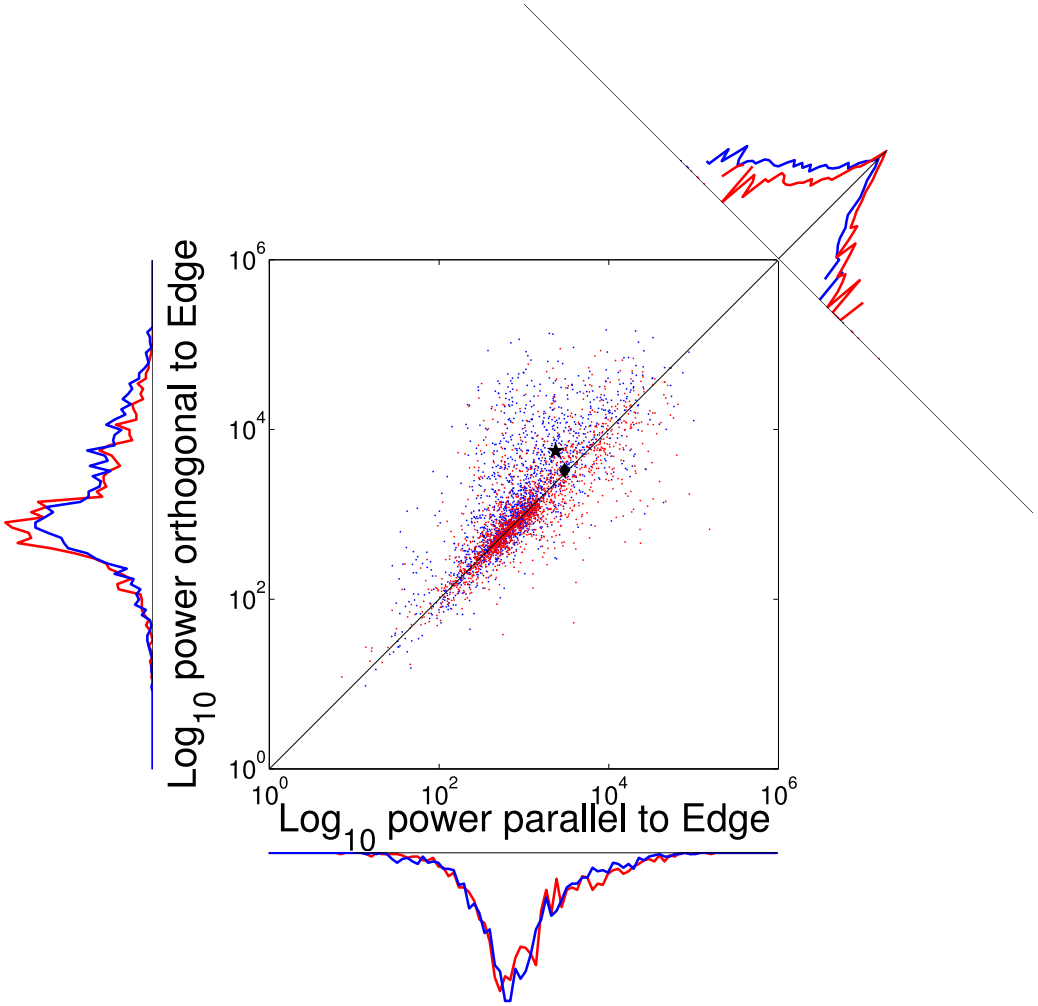

Figure C.9: Two-dimensional distribution of spectral power ($\log_{10} - \log_{10}$ axes) in bins 3–8 orthogonal *vs.* parallel to the OB for all LabelMe patches. Red, background side; ($[T_{g\perp}]_3^8$ *vs.* $[T_{g\parallel}]_3^8$), blue, figure side ($[T_{f\perp}]_3^8$ *vs.* $[T_{f\parallel}]_3^8$). The black diamond, very close to the identity line, shows the mean of the background. The black asterisk, above the identity, shows the mean of the figure. The distance between the figure-side mean and the identity line is larger compared to BSDS (Figure 4). The marginal distributions share their abscissas with the axes of the scatter plot and they have linear ordinates. The marginal distribution at the top right collapses data along the diagonal and has a logarithmic ordinate since the values of the central bins vastly surpass those of other bins. This marginal clearly shows the presence of spectral anisotropy, for BSDS, see Figure 4.

defined as,

$$\mathbf{v}_{FG} = \mathbf{u}_{tangent} + \lambda^*(\mathbf{u}_{center} - \mathbf{u}_{tangent}) \tag{D.2}$$

with λ^* is chosen to ensure that the projection of $\mathbf{u}_{figure}$ is located on the line segment connecting $\mathbf{u}_{center}$ and $\mathbf{u}_{tangent}$. Defining $\langle \mathbf{a}, \mathbf{b} \rangle$ as the scalar product of vectors $\mathbf{a}$ and $\mathbf{b}$, it is obtained from the quantity

$$\lambda = \frac{\langle (\mathbf{u}_{figure} - \mathbf{u}_{tangent}), (\mathbf{u}_{center} - \mathbf{u}_{tangent}) \rangle}{\langle (\mathbf{u}_{center} - \mathbf{u}_{tangent}), (\mathbf{u}_{center} - \mathbf{u}_{tangent}) \rangle} \tag{D.3}$$

as

$$\lambda^* = \max(0, min(\lambda, 1)) \tag{D.4}$$

Appendix E. SA of sharply focused patch pairs

In order to account for the photographer controlling the depth of field and excessively focusing on the foreground objects leaving the background smoothed and blurred out, we re-perform the entire analysis described in Sections 3 and 5 by discarding all blurry patches. Patch pairs were removed from the original datasets if either the foreground or background was blurred out due to limited depth of field. This yielded 1025 non-blurry patch pairs for BSDS and 1716 for LabelMe. The mean spectra corresponding to Figure 3 of the main text are plotted in Figures, E.10 and E.12 for LabelMe and BSDS respectively. Similarly, the plots corresponding to Figure 4 of the main text are shown in Figures, E.11 and E.13 respectively for LabelMe and BSDS. The statistical results for both databases are summarized below:

- Comparison of unoriented spectral power in bin 1 (figure *vs.* ground): Wilcoxon signed-rank tests (BSDS300: $p = 0.20$; LabelMe: $p = 0.66$)

- Comparison of $\bar{T}_f(1,8,1,8)$ *vs.* $\bar{T}_g(1,8,1,8)$: Wilcoxon signed-rank tests (BSDS300: $p = 0.30$; LabelMe: $p = 0.66$)
- Comparison of $\bar{T}_f(3,8,3,8)$ *vs.* $\bar{T}_g(3,8,3,8)$: Wilcoxon signed-rank tests (BSDS300: $p = 4.11 \times 10^{-14}$; LabelMe: $p = 4.63 \times 10^{-4}$)
- Comparison of $[T_{f\perp}]_3^8$ *vs.* $[T_{f\parallel}]_3^8$: Wilcoxon signed-rank tests (BSDS300: $p = 3.24 \times 10^{-23}$; LabelMe: $p = 1.49 \times 10^{-82}$)
- Comparison of $[T_{g\perp}]_3^8$ *vs.* $[T_{g\parallel}]_3^8$: Wilcoxon signed-rank tests (BSDS300: $p = 0.26$; LabelMe: $p = 0.69$)
- Figure-ground discrimination accuracy: (BSDS300: 63.41%; LabelMe: 64.16%)

Appendix E.1. Linear regression results

Please see Table E.6 for results related to linear regression of spectral powers when blurry patch pairs were not included in the analysis.

Appendix F. Two dimensional spectra

An assumption underlying our work is that power varies in specific ways relative to the OB. To show that our observations are not due to some noise effect caused by random power fluctuations in arbitrary directions, we compute the 2D Fourier (power) spectra relative to the OB, separately for foreground and background patches. A 2D Hamming window is applied to each patch before DFT computation. The $\log_{10}$-transformed power spectra are averaged over all samples in each database (LabelMe: 1761; BSDS: 1475) to obtain the mean 2D spectra which are plotted in Figure F.14. The colormaps are

		slope(radians)	**CI(low)**	**CI (high)**	R^2
BSDS300	**Figure** (orthogonal *vs.*. parallel)	1.038	1.030	1.046	0.48
	Ground (orthogonal *vs.*. parallel)	1.00	0.994	1.007	0.66
LabelMe	**Figure** (orthogonal *vs.*. parallel)	1.071	1.064	1.079	0.53
	Ground (orthogonal *vs.*. parallel)	1.000	0.994	1.006	0.57

Table E.6: Regression of $\log_{10}$-transformed high-frequency spectral power in orthogonal and parallel orientations with slope as the only parameter for non-blurry patches only (analogous to Table 2 where all patches were included). Results for both datasets show the slope is higher in the foreground compared to background even after removing the blurry patch pairs. Again more oriented spectral power in orthogonal orientation in figure compared to the parallel orientation is observed. This indicates anisotropy of the figure cannot be caused by the photographer. Note that the confidence intervals of figure and ground are non-overlapping. Please see Appendix E for other results when blurry patches were discarded from analysis.

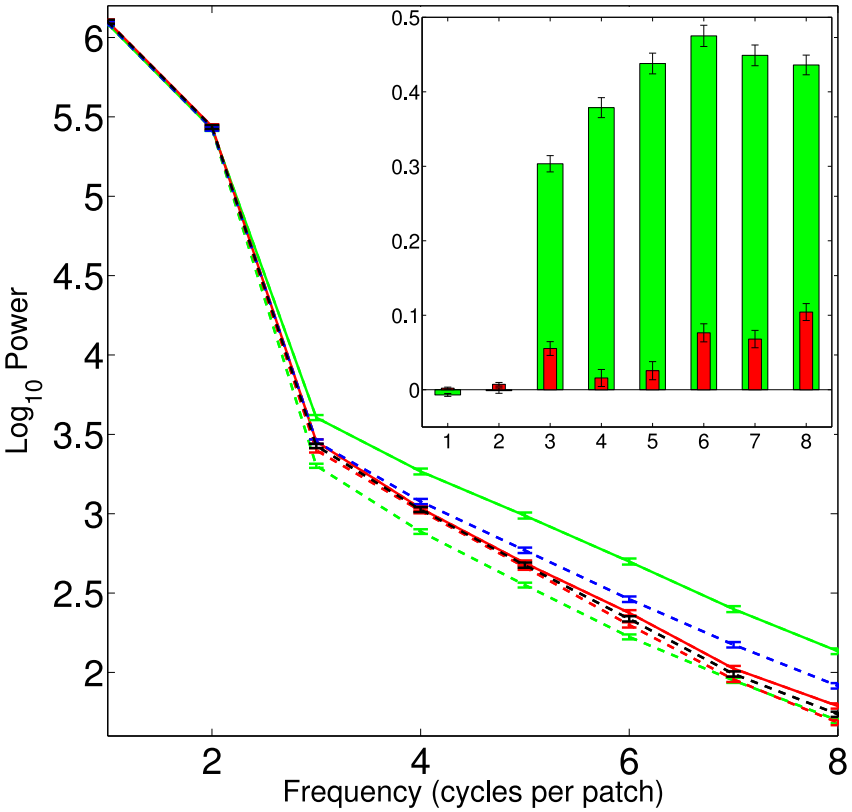

Figure E.10: Average power spectra of the 1716 non-blurry patch pairs of LabelMe dataset as function of spatial frequency. The unoriented spectra are represented by dashed blue (figure) and black (ground) lines. The oriented spectra in the plot are: $\overline{E}_{f\perp}$ (solid green line), $\overline{E}_{f\parallel}$ (dashed green line), $\overline{E}_{g\perp}$ (solid red line) and $\overline{E}_{g\parallel}$ (dashed red line). Inset: The difference in power ($\log_{10}(\bar{E}_{s\perp} - \bar{E}_{s\parallel})$) in each frequency bin. Axes same as in main figure. Green and red bars represent figure ($s = f$) and ground ($s = g$) differences respectively. Error bars are standard errors in figure and inset. Results from the BSDS database for non-blurry patches are similar, see Figure E.12.

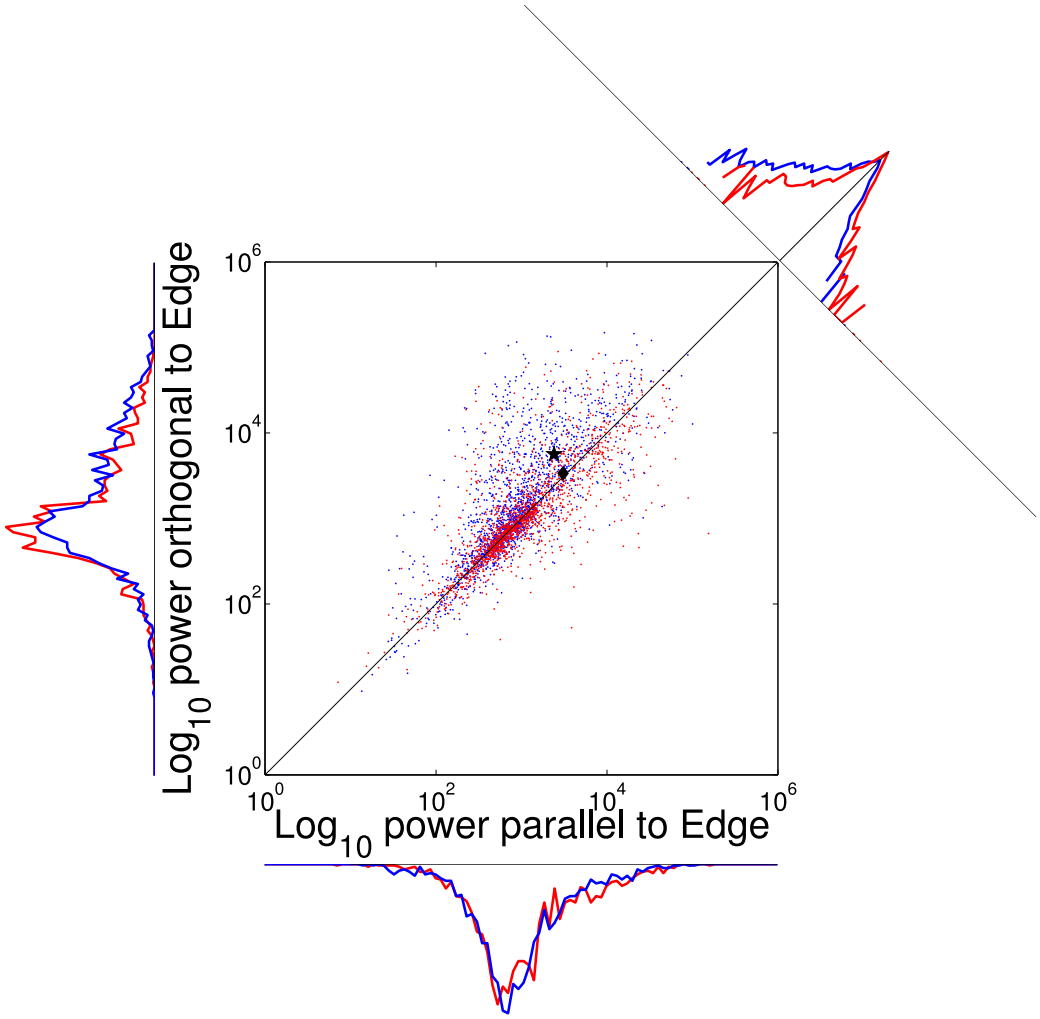

Figure E.11: Two-dimensional distribution of spectral power ($\log_{10} - \log_{10}$ axes) in bins 3–8 orthogonal *vs.* parallel to the OB for 1716 non-blurry LabelMe patches. Red, background side; ($[T_{g\perp}]_3^8$ *vs.* $[T_{g\parallel}]_3^8$), blue, figure side ($[T_{f\perp}]_3^8$ *vs.* $[T_{f\parallel}]_3^8$). The black diamond, very close to the identity line, shows the mean of the background. The black asterisk, above the identity, shows the mean of the figure. The distance between the figure-side mean and the identity line is larger compared to BSDS (Figure E.13). The marginal distributions share their abscissas with the axes of the scatter plot and they have linear ordinates. The marginal distribution at the top right collapses data along the diagonal and has a logarithmic ordinate since the values of the central bins vastly surpass those of other bins. This marginal clearly shows the presence of spectral anisotropy, for BSDS, see Figure E.13.

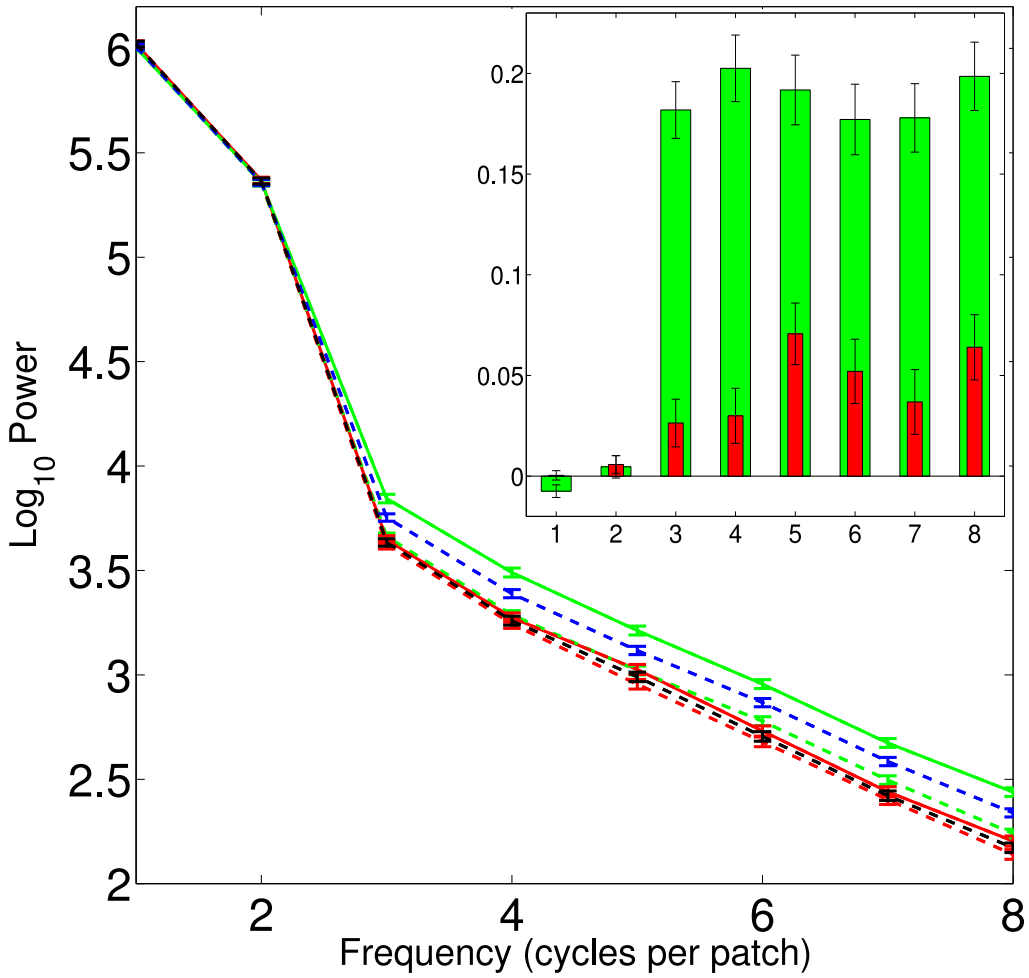

Figure E.12: Average power spectra of the 1025 non-blurry patch pairs of BSDS300 dataset as function of spatial frequency. The unoriented spectra are represented by dashed blue (figure) and black (ground) lines. The oriented spectra in the plot are: $\overline{E}_{f\perp}$ (solid green line), $\overline{E}_{f\parallel}$ (dashed green line), $\overline{E}_{g\perp}$ (solid red line) and $\overline{E}_{g\parallel}$ (dashed red line). Inset: The difference in power ($\log_{10}(\bar{E}_{s\perp} - \bar{E}_{s\parallel})$) in each frequency bin. Axes same as in main figure. Green and red bars represent figure ($s = f$) and ground ($s = g$) differences respectively. Error bars are standard errors in figure and inset. Results from the LabelMe database for non-blurry patches are similar, see Figure E.10.

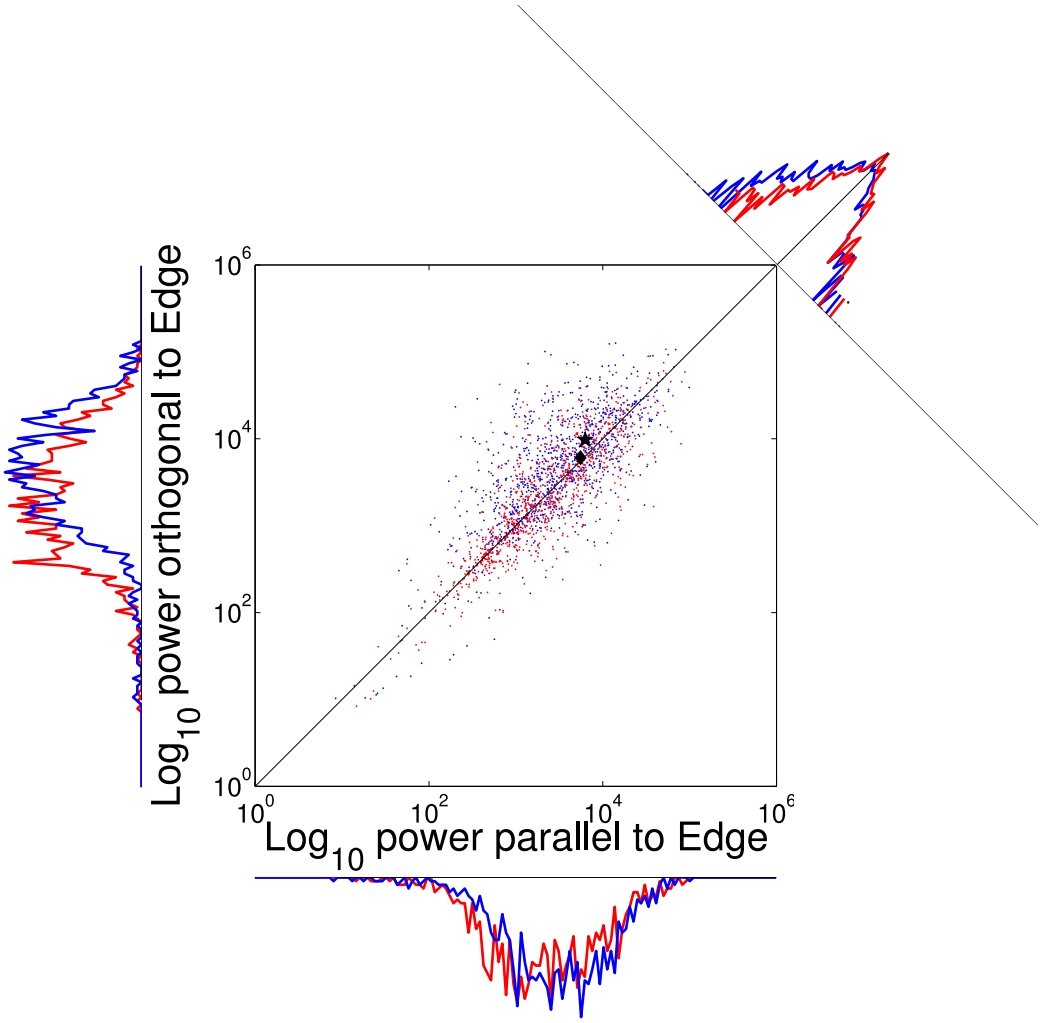

Figure E.13: Two-dimensional distribution of spectral power ($\log_{10} - \log_{10}$ axes) in bins 3–8 orthogonal *vs.* parallel to the OB for the 1025 non-blurry BSDS patches. Red, background side; ($[T_{g\perp}]_3^8$ *vs.* $[T_{g\parallel}]_3^8$), blue, figure side ($[T_{f\perp}]_3^8$ *vs.* $[T_{f\parallel}]_3^8$). The black diamond, very close to the identity line, shows the mean of the background. The black asterisk, above the identity, shows the mean of the figure. The distance between the figure-side mean and the identity line is even larger for LabelMe (Figure E.11). The marginal distributions share their abscissas with the axes of the scatter plot and they have linear ordinates. The marginal distribution at the top right collapses data along the diagonal and has a logarithmic ordinate since the values of the central bins vastly surpass those of other bins. This marginal clearly shows the presence of spectral anisotropy, and again the effect is stronger in the LabelMe data, see Figure E.11.

scaled to enhanced visual clarity. The maximal power difference is observed between orthogonal and parallel orientations relative to the OB, and that is the case on the figure side only.

Appendix G. Maximum likelihood classification

Let the figure and ground parts of a patch pair be denoted by s_1 and s_2, where either s_1 or s_2 can be figure or ground. Let ρ_{s_1} and ρ_{s_2} be the SAs of s_1 and s_2 respectively. A patch pair can be in one of two configurations: (1) Correct (C=1), when figure is on top and ground is on bottom, and (2) Incorrect (C=0) when positions of figure and ground are reversed. Let s_1 and s_2 be the top and bottom parts, respectively, of a patch pair. Let γ be defined as the ratio of SAs of top and bottom parts, *i.e.* $\gamma = \frac{\rho_{s_1}}{\rho_{s_2}}$. The conditional distributions of γ for the two configurations (C=1 and C=0), also the *likelihoods* of the two configurations respectively are, $p(\gamma \mid C = 1)$ and $p(\gamma \mid C = 0)$.

The posterior probabilities of the two configurations are,

$$P(C = 1 \mid \gamma) = \frac{p(\gamma \mid C = 1)P(C = 1)}{p(\gamma \mid C = 1)P(C = 1) + p(\gamma \mid C = 0)P(C = 0)} \tag{G.1}$$

$$P(C = 0 \mid \gamma) = \frac{p(\gamma \mid C = 0)P(C = 0)}{p(\gamma \mid C = 1)P(C = 1) + p(\gamma \mid C = 0)P(C = 0)} \tag{G.2}$$

The correct configuration (C=1) is chosen when,

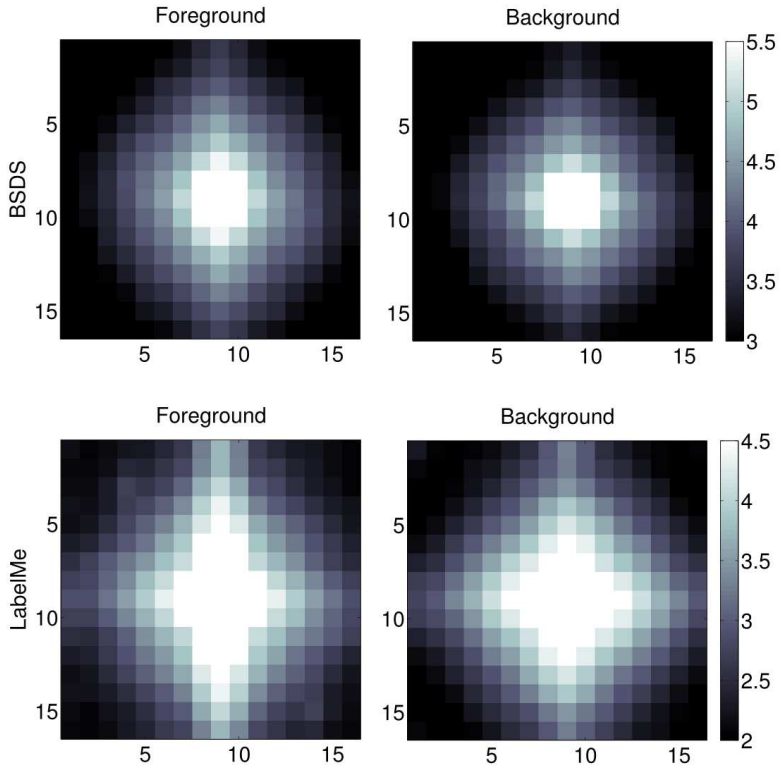

Figure F.14: Two dimensional power spectra ($\log_{10}$-transformed) of LabelMe (bottom two) and BSDS (top two) databases in figure (left) and ground (right). Same colormaps (scaled) are used for both figure and ground within each image set, but colormaps are different for different image sets. Patches are oriented as in Figure 2B. The spectral power is maximal along the vertical (corresponding to the orientation orthogonal to the occlusion boundary) in the figure. A minor anisotropy is noted even in the background which may be due to shadows cast by the foreground object, occasional minor errors in the labeling of occlusion borders, curvature effects of the boundary *etc.*

$$\frac{p(\gamma \mid C=1)P(C=1)}{p(\gamma \mid C=1)P(C=1)+p(\gamma \mid C=0)P(C=0)} > \frac{p(\gamma \mid C=0)P(C=0)}{p(\gamma \mid C=1)P(C=1)+p(\gamma \mid C=0)P(C=0)} \tag{G.3}$$

or,

$$\frac{p(\gamma \mid C=1)}{p(\gamma \mid C=0)} > \frac{P(C=0)}{P(C=1)} \tag{G.4}$$

But, we do not have any *a priori* reason to assume one of the configurations is more likely (hence prior probabilities, $P(C=0)$ and $P(C=1)$ will be the same). In that case, the choice is only based on the likelihoods, hence the test becomes a *maximum likelihood* test. Based on the *maximum likelihood* test, the correct configuration, C=1 is chosen when:

$$p(\gamma \mid C=1) > p(\gamma \mid C=0) \tag{G.5}$$

For the correct and incorrect configurations, the likelihood distributions are: $p(\gamma \mid C=1) = \frac{\rho_{s_1}}{\rho_{s_2}}$ and $p(\gamma \mid C=0) = \frac{\rho_{s_2}}{\rho_{s_1}}$, which are the distributions of ratios of SAs for the two configurations. So, Eq. G.5 can be written as,

$$\frac{\rho_{s_1}}{\rho_{s_2}} > \frac{\rho_{s_2}}{\rho_{s_1}} \tag{G.6}$$

which is equivalent to,

$$\rho_{s_1} > \rho_{s_2} \tag{G.7}$$

Hence, based on the *maximum likelihood* test, top part of the patch pair (s_1), is correctly chosen as figure and bottom as ground when $\rho_{s_1} > \rho_{s_2}$, which is the rule we use to perform figure-ground classification in Eq. 6 of Section 5.3.