Image statistics underlying natural texture selectivity of neurons in macaque V4

Gouki Okazawa[a,1], Satohiro Tajima[b,c], and Hidehiko Komatsu[a,d]

[a]Division of Sensory and Cognitive Information, National Institute for Physiological Sciences, Okazaki, Aichi 444-8585, Japan; [b]RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan; [c]Japan Society for the Promotion of Science, Chiyoda, Tokyo 102-0082, Japan; and [d]Department of Physiological Sciences, The Graduate University for Advanced Studies (Sokendai), Okazaki, Aichi 444-8585, Japan

Edited by Tony Movshon, New York University, New York, NY, and approved November 24, 2014 (received for review August 7, 2014)

Our daily visual experiences are inevitably linked to recognizing the rich variety of textures. However, how the brain encodes and differentiates a plethora of natural textures remains poorly understood. Here, we show that many neurons in macaque V4 selectively encode sparse combinations of higher-order image statistics to represent natural textures. We systematically explored neural selectivity in a high-dimensional texture space by combining texture synthesis and efficient-sampling techniques. This yielded parameterized models for individual texture-selective neurons. The models provided parsimonious but powerful predictors for each neuron's preferred textures using a sparse combination of image statistics. As a whole population, the neuronal tuning was distributed in a way suitable for categorizing textures and quantitatively predicts human ability to discriminate textures. Together, we suggest that the collective representation of visual image statistics in V4 plays a key role in organizing the natural texture perception.

texture perception | material perception | visual area V4 | single-cell recording | image analysis

In the visual world, objects are characterized in part by their shapes, but also by their textures (1). The wide variety of textures we experience enables us to segment objects from backgrounds, perceive object properties, and recognize materials. It is well established that the representation of complex shapes and contours is gradually built up along the ventral visual pathway (2–5). On the other hand, how textural information is processed in the cortex is largely unknown, although some recent studies have examined the representation of natural textures and surface properties in the macaque ventral visual areas (6–11). Because, unlike contours, textures cannot be described based on combinations of edge fragments, we need to consider different underlying cortical processing.

In contrast to the limited knowledge available from physiology, computational descriptions of textures have been extensively developed in the fields of psychophysics (12–16) and computer vision (17–19). In one such description, Portilla and Simoncelli (20) proposed that textures could be represented using an ensemble of summary statistics, including features derived from the luminance histogram and the amplitudes of the outputs of Gabor-like filters, as well as higher-order statistics such as the correlations across the filter outputs (see Fig. 3 for details; hereafter, we call this collection of statistics "PS statistics," using the authors' initials). Portilla and Simoncelli successfully generated new textures that were visually indistinguishable from the originals solely by making their PS statistics identical. Their algorithm is particularly inspiring because PS statistics use filters and computations that share biological properties. It was recently shown, for example, that a version of their synthesis algorithm can generate perceptually indistinguishable visual images (visual metamers) (21) and that naturalistic textures incorporating these summary statistics strongly activate neurons in V2, compared with noise images lacking these features (7).

Given the previous successes in the parametric description of textures, it is tempting to expect that promoting this framework will reveal a full picture of texture representation in the subsequent stages in the ventral pathway, such as area V4, where many neurons show selectivity for textures (6, 22). However, such an attempt has been hampered by a fundamental difficulty of data sampling due to the high dimensionality of natural texture space. Indeed, previous studies on neural texture representation were typically based on strictly limited samples (6–8)—e.g., 15 different textures were used in Freeman et al. (7)—despite a potentially large number of parameters describing the neural selectivity for various natural textures. In this study, we overcome this problem of dimensionality by combining the parametric texture synthesis with an adaptive sampling procedure (23–25) that efficiently sampled a portion of the stimuli that were expected to evoke stronger responses in the recorded neuron, as well as linear regression with a dimension-reduced version of the PS statistics. As a result, we successfully fitted neural responses to hundreds of textures using PS statistics. The fitted result revealed an unprecedentedly detailed picture of how the summary statistics shape the texture selectivity of V4 neurons. Notably, the tuning of each single neuron was well described using a small number of PS statistics. At the same time, the population-level neural representation was shown to be functionally relevant to humans' ability to discriminate and categorize textures. The present results provide direct evidence that the visual system extracts a variety of sparsely combined higher-order image statistics, suggesting that these representations bridge

Significance

Our visual world is richly decorated with a great variety of textures, but the brain mechanisms underlying texture perception remain poorly understood. Here we studied the selectivity of neurons in visual area V4 of macaque monkey with synthetic textures having known combinations of higher-order image statistics. We found that V4 neurons typically respond best to particular sparse combinations of these statistics. We also found that population responses of texture-selective V4 neurons can explain human texture discrimination and categorization. Because the statistics of each image can be computed from responses of upstream neurons in visual area V1, our results provide a clear account of how the visual system processes local image features to create the global perception of texture in natural images.

Author contributions: G.O., S.T., and H.K. designed research; G.O. performed research; G.O. and S.T. analyzed data; and G.O. and H.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 942.

[1]To whom correspondence should be addressed. Email: okazawagouki@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1415146112/-/DCSupplemental.

the gap between local image features and global perception of visual textures.

Results

Adaptive Sampling and Texture-Selective V4 Neurons. The stimuli for each neuron were adaptively sampled from 10,355 synthetic textures defined in a sampling space (Fig. 1*A*). They were all grayscale (mean: 15 cd/m^2; SD: 6 cd/m^2), square (6.4° wide), and presented for 200 ms at the center of the receptive fields of the recorded V4 neurons. These neurons were recorded from the prelunate gyrus and their receptive fields typically extended from 1° to 10° (average receptive field size: 6.1 ± 3.1°). All 10,355 textures were synthesized using a texture synthesis algorithm (20) based on synthesis parameters (PS statistics) extracted and interpolated from 4,400 natural textural images. These images were derived from photographs of eight material categories (bark, sand, fabric, fur, leather, stone, water, and wood; 550 images for each). We generated the sampling space (Fig. 1*A*) from the PS statistics of the images (740 dimensions) by extracting the seven-dimensional subspace using Fisher's linear discriminant analysis (Fig. S1). This procedure finds the subspace that maximally separates images of different material categories. We explored neuronal responses to textures along this space, although the stimuli themselves were synthesized using all 740 parameters. To efficiently collect preferred textures for individual neurons, we adopted an adaptive sampling procedure (25) (Fig. 1*B*). For each neuron, we first randomly selected 50 textures from the entire set and recorded the responses to those textures. We then produced another 50 textures by selecting more stimuli from among the textures nearby the strongly active ones in the sampling space (such as groups A and B in Fig. 1*B*) and selecting fewer stimuli from among those nearby weakly active textures (such as groups C, D, and E in Fig. 1*B*). These procedures were repeated for 5–10 generations (average 7.2), recording responses to 250–500 textures per neuron.

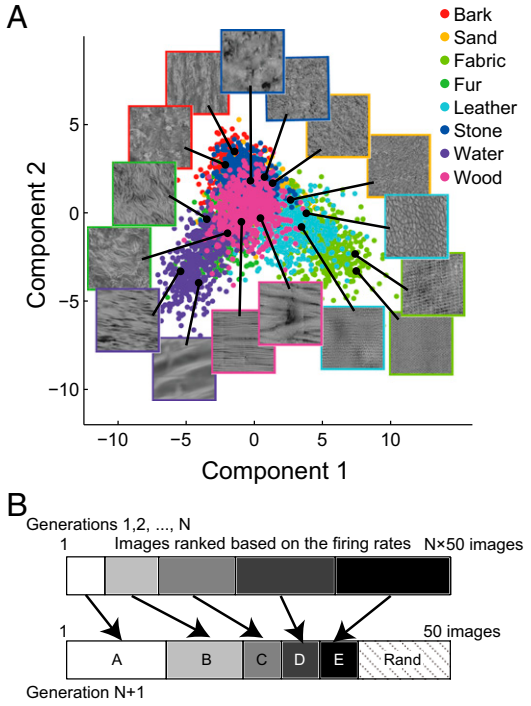

Fig. 1. The parametric sampling space and the adaptive sampling procedure. (*A*) The sampling space and locations of example textures. The panel shows the first two dimensions of the seven-dimensional sampling space. The textures were synthesized based on synthesis parameters extracted from photographs of one of the eight material categories (bark, sand, fabric, fur, leather, stone, water, and wood). Each dot represents one textural image derived from a photograph of material textures, and the color indicates its material category. The *Insets* show example textural stimuli located at the black dots. The frame colors indicate the material categories. There are also many other textures generated by interpolating the parameters of the textures neighboring in this space. These textures cannot be categorized into the materials and are not shown in the figure. Including the interpolated ones, a total number of 10,355 textures was derived. (*B*) Adaptive sampling. Textures in preceding generations were classified into five groups (indicated by different gray levels), depending on the rank of the firing rates they elicited. Then, 38 of 50 textures in the subsequent generation were selected from among the neighbors of the textures for each of the five groups, with priority given to the higher ranked groups. The remaining stimuli (12) were selected randomly ("Rand").

In two macaque monkeys, we recorded from 109 V4 neurons that selectively responded to the presented textures ($P < 0.0001$, Kruskal–Wallis test; 64 neurons from monkey SI, 45 neurons from monkey EV). Of these, 90 neurons that responded to a sufficiently large number of textures (sparseness index < 0.75; for the definition of sparseness index, see *SI Methods*, *Electrophysiological Recording*) were further analyzed (53 neurons from monkey SI; 37 neurons from monkey EV). Typically, each neuron vigorously responded to textures with similar appearances (Fig. 2*A* for an example cell) but not to textures with different appearances (Fig. 2*B*), and different neurons preferred textures with different appearances (Fig. S2*A*). In a subset of neurons ($n = 13$), we examined whether two independent adaptive samplings starting from different first generations converged to similar texture selectivity. The preferred textures obtained from two independent samplings appeared to be visually similar (Fig. S2*B*). The minimum distance between the five most effective textures obtained from two samplings was on average 2.44 in the sampling space, which is close to the average distance between adjacent textures in the space (1.65) and is significantly smaller than that obtained when the five textures were randomly sampled (4.87; $P = 0.0018$; $n = 13$; two-tailed t test). We also confirmed that the texture selectivity is largely invariant, irrespective of the positions or sizes of the images (Fig. S2*C*). Textures with appearances similar to their effective parents were preferentially selected by the adaptive sampling, and those descendants usually produced comparable levels of neuronal activity (Fig. 2*C*). When we averaged the firing rates evoked by all descendant textures sorted according to the ranks of the parents' firing rates (Fig. 1*B*), we found that the offspring selected from highly responsive parents generally evoked stronger responses than those selected from the weakly responsive parents (group A vs. C–E: $P < 0.005$, Mann–Whitney test; Fig. 2*D*) or randomly chosen textures (group A vs. Rand: $P < 0.0001$, Mann–Whitney test). We also found that, as the generation proceeded, the average distance between the sampled stimuli and the optimal stimuli for each neuron (here, we regarded stimuli evoking more than 90% of the maximum firing rate as optimal) became smaller (Fig. 2*E*, orange line), and that this effect was more evident than in the average distances computed from simulated data assuming that the neurons were randomly tuned to the textures (blue line, vs. "Data": generation 2; $P = 0.002$, generation 3–10; $P < 0.001$, Mann–Whitney test; *SI Methods*). This indicates that the stimuli near the optimal ones were more densely sampled as generation proceeded. The convergence of the sampled stimuli was also evident when we computed the distance between the sampled stimuli and the optimal stimuli extracted from two independent samplings starting from different first generations (Fig. S2*D*). Together, the results suggest that the procedure for adaptive sampling worked successfully, and the neurons' preferences could be characterized by the texture parameters used to generate the sampling space.

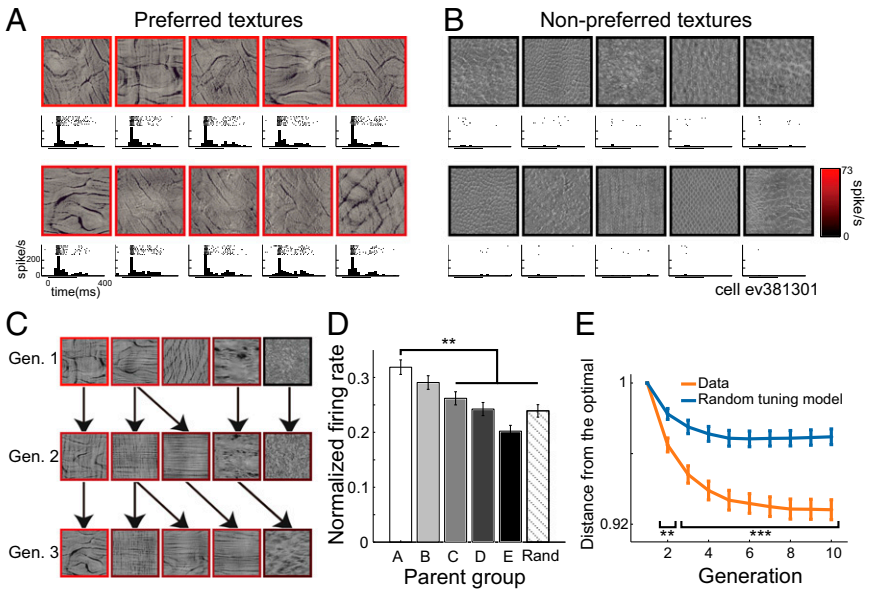

Fig. 2. Neural texture preferences and effects of adaptive sampling. (*A* and *B*) Textures evoking the 10 strongest (*A*) and the 10 weakest (*B*) responses by an example V4 neuron. The frame colors represent average firing rates. Panels under each image depict raster plots and poststimulus time histograms (PSTHs). The horizontal bars under the PSTHs indicate the stimulus presentation period. (*C*) Example cascade in the adaptive sampling for the same neuron. Textures in the same row were from the same generation. Textures at the arrowheads were selected from among the neighbors of the texture at the origin of arrow as descendants. As in *A*, the frame colors represent firing rates. Gen., generation. (*D*) Mean firing rates of all presented textures sorted according to the parent groups ($n = 90$ neurons). The firing rates were normalized to the maximum response of each neuron. Different grayscale shades represent different groups of the ancestors and correspond to those in Fig. 1*B*. The error bars indicate SEM across the neurons. $^{**}P < 0.01$. Rand, randomly selected. (*E*) Average distance in the sampling space of all presented textures up to a given generation from the optimal stimuli for each neuron ($n = 90$ neurons). The orange line (Data) was computed using the data from all neurons, whereas the blue line (Random tuning model) was computed from the simulated data assuming that the neurons were randomly tuned to the textures. Even if a neuron was randomly tuned, the distance could decrease because the adaptive sampling procedure densely samples stimuli around effective textures. Significant departure from this "random tuning model" indicates that the adaptive sampling procedure effectively captured the meaningful texture tunings of the neurons. The error bars indicate SEM across the neurons. $^{**}P < 0.01$, $^{***}P < 0.001$.

Fitting Using Higher-Order Image Statistics. To understand texture selectivity in terms of tuning to textural parameters, we tried to linearly regress neuronal responses using the textural parameters (PS statistics; Fig. 3) to determine how well this simplest fitting explains texture selectivity. Because the large number of parameters in the PS statistics (740; Table S1) makes the fitting unreliable, we reduced the number by removing redundant features from the individual groups of parameters in the PS statistics. In short, we reduced the resolutions of scale and orientation from 4×4 to 2×2 and also applied principal-component analysis to parameters in the groups of statistics called "Linear cross position" and "Energy cross position." For "Marginal" statistics, only the skewness of the luminance histogram was used because the mean and SD had been equalized across the textures. Through these operations, we reduced the number of parameters to 29. These 29 parameters, which we will call "minimal PS statistics" (minPS; Table S2), were used to fit each neuron's responses to 250–500 textures by adopting the regularized (L1-penalized) linear least-squares regression (26). We evaluated the fitting performance using 10-fold cross-validation (*Methods*, *Fitting to Neuronal Responses*) and found that the firing rates of most neurons were successfully fit using minPS [r between the observed and predicted responses = 0.46 ± 0.18 for the cross-validated set; $P < 0.05$ for 83 of 90 neurons (92%), permutation test; Fig. 4*A*]. Similar levels of fitting performance were obtained from both monkeys ($r = 0.43 \pm 0.18$ for monkey SI; $r = 0.50 \pm 0.18$ for monkey EV). The cells proved to be significant ($n = 83$) have also been shown to be significant using the test of correlation coefficient ($P < 0.05$, Pearson's test). On average, the fittings explained $29 \pm 17\%$ of the explainable variance of the firing rates, computed by subtracting the trial-by-trial variance from the whole variance. Furthermore, in a subset of neurons ($n = 13$), we examined the prediction performances to responses obtained in a separate adaptive sampling session started with a different first generation and found the performance levels to be comparable ($r = 0.49 \pm 0.20$), again supporting the validity of the fittings.

We carefully inspected the fitting weights for each neuron (Fig. 4*B*) because they should tell us the statistical parameters that are critical for activation of these neurons. We found that a relatively small number of parameters had higher weights (Fig. 4*C*). The number of parameters with weight absolute values that exceeded the half-maximum weight was 3.4, on average (Fig. 4*D*), suggesting the neurons are sparsely tuned to a small number of parameters. Fig. 4*E* shows the numbers of neurons tuned to each of the different groups of statistical parameters in the minPS with greater than half-maximum weight (the groups are listed in Fig. 3). Of these, "Spectral" is a lower-order statistic that corresponds to the tuning to the amplitude of particular spatial frequency/orientation subbands and is already represented in V1. That a large number of neurons are tuned to Spectral statistics is consistent with earlier studies showing the spatial frequency/orientation selectivity of V4 neurons (27, 28). More importantly, a large fraction of the neurons were also selective for other higher-order statistics (72 of 83 neurons had greater than half-maximum weight on at least one parameter in groups other than Spectral). To further validate the significance of the higher-order statistics, we tried to fit the responses using only Spectral statistics and found that this operation significantly degraded the fitting performance ($r = 0.35$; vs. the minPS: $P < 0.0001$, Wilcoxon test). We also examined whether models of statistical image representation other than the minPS could better explain the neural responses, and found that no model outperformed the minPS (Fig. S3).

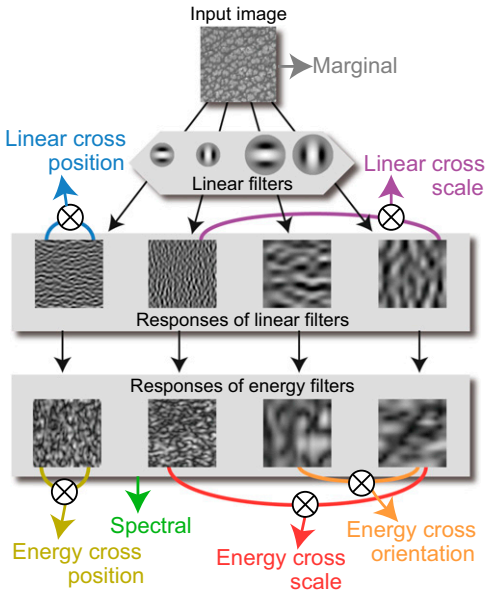

Fig. 3. PS statistics used in the texture synthesis. A similar description can be found in McDermott and Simoncelli (51). The statistics are grouped into seven categories represented by different colors. The terms used for each group are based on those used by Freeman et al. (7). The uppermost image represents the input texture. Marginal statistics were directly computed from that texture and include statistics from the luminance histogram, including mean, variance, skewness, and kurtosis; in the minPS, only the skewness was incorporated. The image was convolved using Gabor-like filters with different scales and orientations to generate "Responses of linear filters." From these responses, the correlations between spatially neighboring filters (Linear cross position) and the correlations between filters with neighboring scales (Linear cross scale) were computed. The Responses of linear filters were then converted to "Responses of energy filters" by taking the amplitudes of the responses. From those, average amplitudes of filter outputs (Spectral), correlations between filters with neighboring orientations (Energy cross orientation), correlations between spatially neighboring filters (Energy cross position), and correlations between filters with neighboring scales (Energy cross scale) were extracted.

Examples of Tuning to the Image Statistics. We observed tight correspondence between the responses of individual neurons and the statistical features on which they had large weights. Here, we present six example neurons that showed good fitting performance and were heavily weighted for each group of minPS (Fig. 5). We rarely encountered neurons heavily weighted for "Linear cross scale" statistics and no example is shown. For each neuron, we depicted three example effective/ineffective textures and their corresponding image statistics selected from the best and worst five textures for that neuron. The neuron shown in Fig. 5*A* had the highest weight for Spectral statistics; the weight was on the amplitude of low-frequency vertical components, and the preferred textures contained coarse vertical stripes. The neuron depicted in Fig. 5*B* had the highest negative weight for skewness of the luminance histogram in the Marginal statistics. This neuron clearly preferred images containing black lines, which induce stronger negative skewness, whereas it did not respond to images containing bright white lines or spots. The neurons in Fig. 5 *C* and *D* had maximum weights for Linear cross position and Energy cross position, respectively. They were heavily weighted for particular patterns of spatial autocorrelation in the filter outputs, which were frequently observed in their preferred textures. The weight of the neuron shown in Fig. 5*E* indicates that it preferred the correlation between fine-scale vertical and horizontal subbands in "Energy cross orientation." This feature is characteristically observed in images with fine white spots or grids. Finally, the neuron shown in Fig. 5*F* was weighted heaviest for the correlation between fine- and coarse-scale vertical subbands in "Energy cross scale." This correlation is observed in images containing vertical lines. It is noteworthy that some neurons preferentially responded to visually dissimilar textures (e.g., neurons in Fig. 5 *A*, *D*, and *F*). This is because, although these dissimilar textures share neurons' preferred statistical features in common, other features to which neurons were not sensitive have affected the appearances to make them different. Thus, a neuron's preferences for textures could be better interpretable in terms of its tuning to the image statistics rather than to the image appearances.

The Model Predicts Responses to Manipulated Images. To verify the neurons' tuning to the PS statistics, we examined the responses of the neurons (n = 90) to four types of control images: (*i*) images in which the phase of the Fourier transform was randomized (Scramble), (*ii*) images made by rotating the original image 90° (Rotation), (*iii*) images synthesized using the same PS parameters as the original textures but from different random noises (Same), and (*iv*) grayscale photographs used to extract the synthesis parameters of the original image (Photo) (Fig. 6*A*). The first two manipulations partially deform the PS statistics, whereas the second two manipulations retain the PS statistics of the original images even though the pixel-level features are considerably different. Five textures equally sampled based on the evoked firing rates in the adaptive sampling experiment were used to generate each of the four control images (Fig. 6*A*). The responses to Same and Photo were similar to those to Original, as would be expected from the tuning to the PS statistics (three

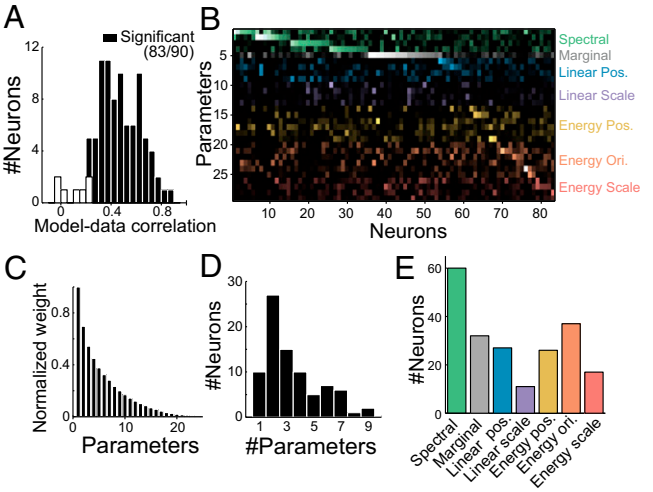

Fig. 4. Fitting V4 neuronal responses with minPS. (*A*) Fitting performance using minPS. The histogram shows the distribution of correlation coefficients between the observed and predicted firing rates across all neurons with cross-validation. The significance was examined using a permutation test with data in which correspondences between images and firing rates were shuffled. (*B*) Weights of individual neurons (horizontal axis, n = 83) for the minPS parameters (vertical axis, n = 29) obtained through the linear fit. The brightness of the dots represents the absolute values of the weights. Dot colors represent the groups of statistical parameters indicated in Fig. 3. Neurons were sorted according to their most preferred group of PS statistics. The abbreviated names of groups are shown to the *Right*. Ori., orientation; Pos., position. (*C*) Absolute values of the weights of individual parameters sorted according to their rank order for each neuron and averaged across all neurons significantly fit by the minPS. The values are normalized to the maximum weight for each neuron (n = 83 neurons). (*D*) The distribution of numbers of parameters that have weights greater than the half-maximum weight for all significant neurons (n = 83 neurons). (*E*) Numbers of neurons in which each group of statistical parameters has weights greater than the half-maximum for at least one of the parameters in a given group.

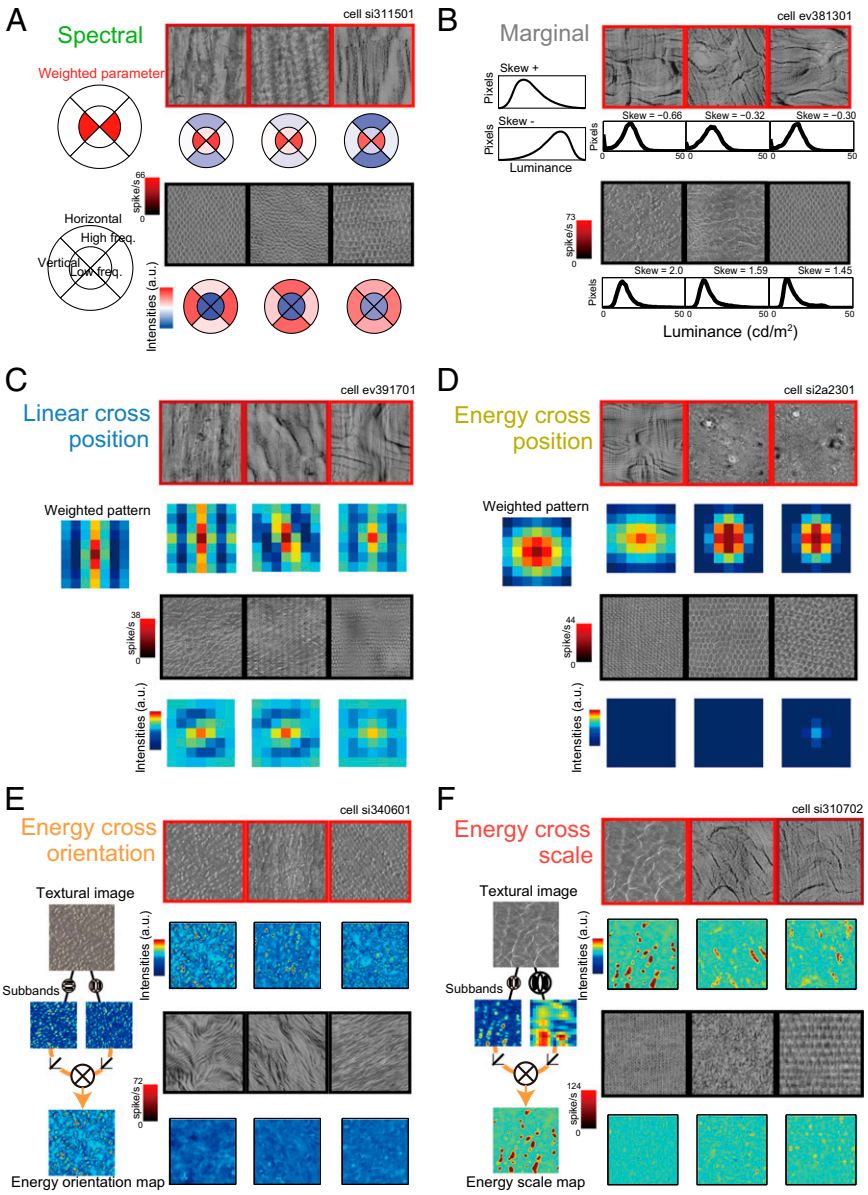

Fig. 5. Visualization of the PS parameters with the largest weight for example neurons. Each panel corresponds to the results from one neuron. Three examples out of the five most preferred and least preferred textures are shown. The frame colors represent firing rates evoked by the textures. Below each texture image, the statistics with the largest fitting weight for that neuron are depicted. (*A*) Neuron with the largest weight on Spectral statistics. The upper left panel shows its weighted parameter. The explanations for the symbols are shown in the lower left panel. (*B*) Neuron with the largest negative weight on the skewness of luminance histogram in the Marginal statistics group. The left panel schematically illustrates the luminance histogram of positive (*Top*) and negative (*Bottom*) skewness. (*C*) Neuron with the largest weight in the Linear cross position statistics group. Spatial autocorrelations in the filter output were computed using the neighboring 7 × 7 pixels as shown in the panels. The leftmost image represents the neuron's preferred pattern. (*D*) Neuron with the largest weight in the Energy cross position statistics group. The conventions are the same as in *C*. (*E*) Neuron with the largest weight in the Energy cross orientation statistics group. The weight was on the correlation between outputs of fine-scale vertical and horizontal filters. The leftmost schema illustrates the procedure to generate these statistics. (*F*) Neuron with the largest weight in the Energy cross scale statistics group. The weight was on the correlation between the outputs of vertical fine- and coarse-scale filters.

example cells in Fig. 6*B*). By contrast, the responses to Scramble and Rotation were cell dependent; some neurons showed similar selectivity for Rotation but not Scramble (Fig. 6*B*, *Top*), some others showed the opposite (Fig. 6*B*, *Middle*), and still others showed similar selectivity to neither Scramble nor Rotation (Fig. 6*B*, *Bottom*). At the population level, the correlations between the Original and Same/Photo conditions were significantly higher than between the Original and Scramble/Rotation conditions (Same vs. Scramble/Rotation: $P < 0.003$; Photo vs. Scramble/Rotation: $P < 0.020$, Mann–Whitney test, Fig. 6*C*). The firing rates averaged across all five textures in the Same and Photo conditions were also higher than in the Scramble and Rotation conditions (Fig. 6*D* and Fig. S4).

The variation in the outcomes of the Scramble and Rotation manipulations can be explained by the fact that these image manipulations deform different elements of the PS statistics. For example, the Spectral statistics are affected by Rotation but not by Scramble (Fig. 7*A*). Accordingly, a neuron selective for Spectral statistics (Fig. 5*A*) showed less selectivity in the Rotation condition but retained its selectivity in the Scramble condition (Fig. 7*A*; the

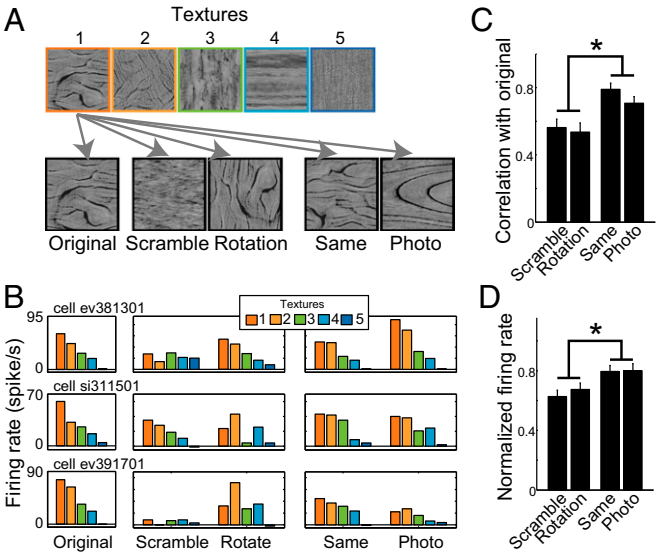

Fig. 6. Responses to the control stimuli. (*A*) Five control images were prepared for each of five textures selected based on the rank (indicated by frame color) of the elicited firing rates in the main experiment. Original: same images used in the main experiment. Scramble: images whose Fourier transform phases were randomized. Rotation: images that were rotated by 90°. Same: synthetic texture generated using the same synthesis parameters as the original. Photo: the original photograph used to extract the synthesis parameters. (*B*) Responses of three example neurons to all stimuli presented under the control conditions. Each row indicates the responses of one neuron. Bar colors indicate different textures corresponding to those in *A*. The rank order of the responses to the original image was basically retained for the "Same" and "Photo" images, but only partially retained or not retained for the "Scramble" and "Rotation" images, depending on the neuron. (*C*) Comparison between the neuronal responses to different control conditions in terms of the similarities of the responses to the originals (n = 90 neurons). The vertical axis indicates the correlation coefficients between the responses to the original five textures and those to five textures in each control condition. The bars indicate the median correlations across all neurons. (*D*) Firing rates averaged across all five textures in each condition (n = 90 neurons). The firing rates were normalized to those averaged across the five textures in the Original condition. In *C* and *D*, the error bars indicate SEM across the neurons. *$P < 0.05$, Mann–Whitney test.

correlation coefficient with the Original condition is shown above the image). Likewise, the preferred pattern of the spatial autocorrelation for the neuron weighted for Linear cross position (Fig. 5*C*) was largely modulated in the Rotation condition, and the neuron changed its response in this condition (Fig. 7*C*). Conversely, the skewness of the luminance histogram in Marginal statistics was unaffected by Rotation, but was affected by the Scramble condition, and changes in the neuron's response weighted on the skewness (Fig. 5*B*) coincided with it (Fig. 7*B*). The neurons weighted on Energy cross position, Energy cross orientation, and Energy cross scale also showed those trends (Fig. 7 *D*–*F*). To validate these tendencies at the population level, we classified neurons according to whether their fitting weights were larger for parameters tolerant or intolerant to Scramble or Rotation (Fig. 7*G*; *SI Methods*), and computed the correlations with Original for each manipulation separately (Fig. 7*H*). In this analysis, only neurons with high fitting performances ($r > 0.4$; $n = 53$) were used. Neurons heavily weighted for tolerant parameters had clearly higher correlations than neurons with large weights on intolerant parameters in both types of control conditions (Scramble: $P = 0.047$; Rotation: $P = 0.0026$; Mann–Whitney test). Note that neurons were classified based on the fitting weights obtained in the adaptive sampling experiment, whereas the correlations with the Original stimuli were examined using the responses obtained in the control experiment. In addition, we found that the responses to the control images could be directly predicted using the fitting weights to the minPS of each neuron (Fig. S5). Thus, the results further support the validity of the fitting using the minPS.

Functional Significance of the Observed Tuning. The final question addressed is how the tuning to minPS in V4 is related to perception of texture and material of objects. Freeman et al. (7) psychophysically compared human sensitivity to textures and corresponding noise images having the same Fourier spectrum and linearly fitted these sensitivities using PS statistics to estimate the contributions of each group of the PS statistics to the psychophysical sensitivity. They found that the higher-order statistics, especially "Energy" statistics, are critical for explaining human sensitivity (Fig. 8*A*, *Right*). We examined whether the sensitivity of V4 neurons to discriminate textures from noise images is also capable of reproducing this psychophysical observation. To do that, we calculated the sensitivity of each V4 neuron to a texture by contrasting the firing rates elicited by the texture and its corresponding noise image predicted from the observed fitting weights (Fig. 4*B*), and averaged these sensitivities across all 83 significant neurons (*SI Methods*). We then assessed the contribution of each statistical parameter group to the neuronal sensitivity by comparing the power of the linear fit (R^2) before and after removing a particular parameter group (averaging-over-orderings technique) (29) (Fig. 8*A*, *Left*). The results obtained from our V4 data exhibited striking correspondence with the psychophysical data reported by Freeman et al. (7) ($r = 0.90$; Fig. 8*A*, *Right*), and this correspondence was significantly better than in cases where the fitting weights were shuffled within individual neurons ($P = 0.016$, permutation test). These results suggest that V4 tuning potentially contributes to the perceptual discriminability of textures.

Different materials have characteristic textures, and how the texture tuning of V4 neurons is related to material categorization is another important issue. To address this, we defined category separability (J) as an index to quantify how well neurons separate the eight material categories used to define the textural stimuli (see Fig. 1*A* for these categories). The index corresponds to the between-category variance divided by the within-category variance within a given space. From the obtained fitting weights of the neurons (Fig. 4*B*), we computed the predicted responses to all of our textures tagged with the categories (n = 4,400) and computed the neurons' category separability (J) as a function of the number of neurons. The separabilities were significantly larger than was the case when we shuffled the fitting weights of individual cells (number of neurons = 3–27: $P < 0.01$, Mann–Whitney test; Fig. 8*B*), which indicates the observed weights are also desirable for classifying the categories of materials. To explore why the observed weights produced better category segregation, we quantified the degree to which image statistics separate different material categories using a measure similar to "J," but in this case the parameter values in each group of PS statistics were used instead of neural responses (Ji in Fig. 8*C*, black line). We found that parameters in some groups were more separable across different categories of textures than were others. Importantly, the absolute values of the fitting weights summed across all 83 significant neurons closely matched the differences in separability across groups ($r = 0.79$; Fig. 8*C*, color bars). The correlation is not an outcome of the adaptive procedure in the sampling space because the stimuli were synthesized using all PS statistics rather than the parameters defining the sampling space. This suggests that neurons are able to separate the categories well because they are likely tuned to the statistics that effectively differentiate textures from different material categories.

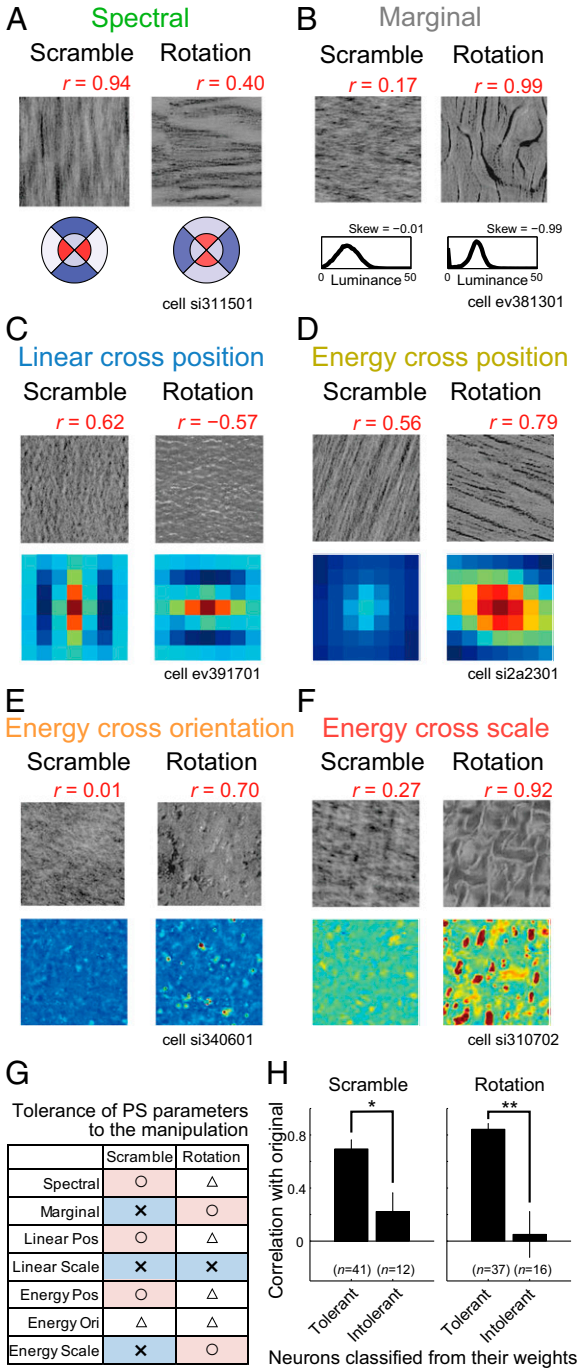

Fig. 7. Effects of image scramble and rotation on the neuronal responses. (*A–F*) Control stimuli in the Scramble and Rotation conditions and their statistics for neurons depicted in Fig. 5 *A–F*, respectively. The images correspond to the stimuli that evoked the highest firing rates in the Original condition among the five selected textures. Red numbers above the images indicate the correlation between the neural responses in the Original condition and those in the Scramble/Rotation conditions. The conventions used to describe the statistics are the same as in Fig. 5 *A–F*. (*G*) The table summarizes the tolerance of each parameter to the image manipulations. To measure the tolerance, we computed the minPS parameters for our 10,355 textures and their corresponding control stimuli, and calculated the correlation coefficients between them for each parameter in the minPS. Parameters with correlation greater than 0.4 were regarded as tolerant of the manipulations. ○: all parameters in the group are tolerant; △: some parameters in the group are tolerant; ×: no parameters in the group are tolerant. (*H*) Correlation coefficients between the responses to the original and control conditions averaged across neurons classified as tolerant or intolerant of the control manipulations based on their fitting weights. Neurons with large weights on statistics that were not greatly affected by the control manipulation were classified as "Tolerant"; otherwise, they were classified as "Intolerant" (*SI Methods*). $*P < 0.05$, $**P < 0.01$, Mann–Whitney test.

Discussion

In the present study, we attempted to reveal the neuronal underpinnings of natural texture perception by examining the texture selectivity of area V4—midlevel area in the ventral pathway—in terms of the higher-order image statistics such as those used in the texture synthesis (PS statistics; Fig. 3). The fundamental difficulty of data sampling in the great diversity of natural textures was overcome by extending an adaptive sampling procedure (25) to explore the manifold of natural textures in a high-dimensional sampling space. We successfully modeled the obtained responses of individual cells as a linear tuning to sparse combinations of image statistics. The models provided an unprecedentedly clear picture of texture representation at the intermediate level of visual processing. At the population level, the observed tuning was found to be suitable to explain perceptual texture discriminability and categorization of materials by human observers. These results allow us to explain the natural texture perception in terms of the collective representation of visual image statistics.

Here, we performed only a linear fit using the synthesis parameters; consequently, there remains the possibility that V4 neurons could be better explained by introducing nonlinear terms into the parameters, although determining those parameters would be more difficult because of the increase in dimensionality. In addition, we do not think that the exact forms of the parameters used in the PS statistics are important, as we may be able to develop variants of the PS statistics that better account for the neuronal data. An essential finding here is that the texture selectivity of V4 neurons can be better understood by introducing higher-order statistics, such as those used in the texture synthesis, than by simple spatial-frequency/orientation tuning. We also suggest that, because the neural responses could be fit using a small number of parameters, individual V4 neurons respond to limited parts of the higher-order statistics rather than to the whole.

Recently, Freeman et al. (7) demonstrated that V2 neurons are more strongly driven by textural stimuli containing the higher-order features of PS statistics than by noise images, whereas responses in V1 do not distinguish between the two. The present study significantly extended these observations by explaining the neurons' "selectivity" for textures in terms of their tuning to those higher-order parameters. Both the activities of V4 neurons in the present study and those of the V2 neurons studied by Freeman et al. (7) correlated with the psychophysically measured texture sensitivity, which suggests tuning properties for natural textures are shared between V2 and V4. The results of several studies emphasize the role of texture segmentation in areas downstream of V2 (30, 31), which suggests V4 is also involved in the texture-defined shape perception. Another earlier study demonstrated that responses of V4 neurons were less affected by the removal of global structures from natural scene images by using the texture synthesis algorithm than neurons in the inferior temporal cortex (32). Analogously, V4 might be more sensitive to changes in global structures than V2, a property that could not be captured by textural stimuli. Such differences between V2 and V4 are an important issue remaining for future research.

Previous physiological studies have shown that neurons in V4 respond to various visual attributes (33) including shape (27, 34), color (35, 36), disparity (37, 38), and texture (6, 22). In particular, many studies have demonstrated that V4 neurons show

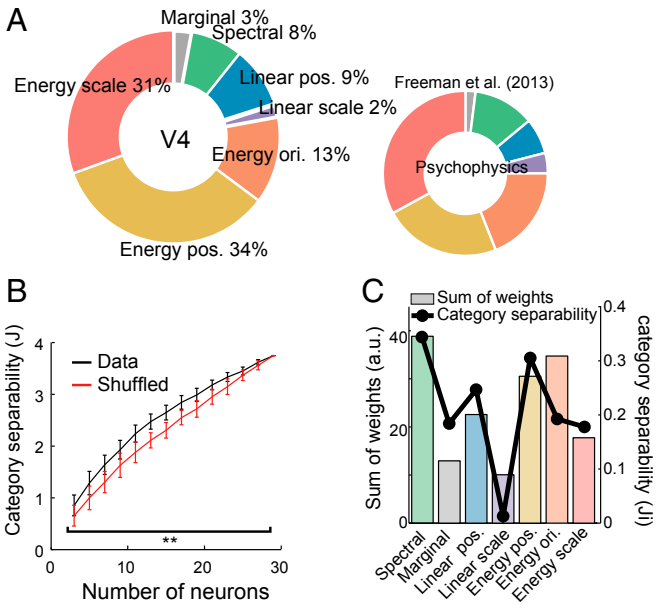

Fig. 8. Functional significances of the observed tuning. (*A*) Percent contribution of each group of PS statistics to the ability of V4 neurons to discriminate textures from the noise images (*Left*) was compared with human sensitivity to textures (*Right*; reprinted with permission from ref. 7). The contributions were computed from the fitting weights for all significant V4 cells (n = 83) (*SI Methods*). It should be noted that this computation was based on the sensitivity to textures vs. noise and should differ from the proportions of neurons shown in Fig. 4*E*. For both V4 neural sensitivity (*Left*) and human sensitivity (*Right*), the same groups of statistical parameters are painted with the same colors. The order of the groups follows that in Freeman et al. (7). (*B*) The degree of category separability (*J*) computed from the neural data (black) and shuffled data (red) are shown as a function of the number of neurons. *J* reaches a plateau when the number of neurons is 29 because the degrees of freedom of the predicted firing rates are constrained by the number of the statistical parameters used for the fitting (i.e., the minPS). The error bars indicate the SD for 100 repetitions of random neuronal samplings. $^{**}P < 0.01$. (*C*) Sum of fitted weights across all significant neurons (n = 83) for each group of PS statistics (color bars). A black line shows the category separability computed from the image statistics of each group in the minPS (*Methods*).

tuning to curvature in object contours (39, 40). Considering that curvature can also be regarded as a higher-order combinatory feature of Gabor filter output, it may be that there is shared tuning between these contour fragments and textures. However, texture synthesis does not capture global image structures and cannot reproduce inhomogeneous image features such as object contours (20), although it may explain visual discriminability of those features in peripheral visual field (21, 41, 42). It is therefore also plausible that tuning to geometric shapes or curvatures is hidden in the unexplained variances in the neural responses recorded in this study, or that they can be attributed to a different population of V4 neurons tuned to those features. It may be possible to construct a unified model to explain responses to both textures and shapes by introducing even higher-level features or hierarchical networks (43).

We have shown that the representations of statistical features in V4 neurons can explain several perceptual features of visual texture processing including the sensitivity to textures (Fig. 8*A*) and the classification of materials (Fig. 8*B*). The utility of higher-order correlation statistics for texture classification has been shown in theoretical terms (44). In addition, our previous functional MRI study showed that cortical responses gradually change from the representation of low-level image properties in V1 to those of material properties in higher visual cortices (11, 45), which is consistent with the finding that V4 is suitable to separate the material categories (Fig. 8 *B* and *C*). Recent psychophysical studies have suggested that statistics, including skewness of the luminance histogram and congruence of local orientations, play important roles in the perception of surface properties such as gloss (46, 47). Because these features are partially incorporated into PS statistics, V4 neurons can also be considered a reasonable intermediate step for the perception of these properties. Using their outputs, higher-order visual areas, such as the inferior temporal cortex, would be able to build up the representations of material categories or surface properties (8, 9, 11, 48), ultimately leading to our perceptual experience of the material-rich visual world.

Methods

Stimulus Presentation and Electrophysiology. Neurons were recorded from area V4 in two macaque monkeys. All procedures for animal care and experimentation were in accordance with the National Institutes of Health *Guide for the Care and Use of Laboratory Animals* (49) and were approved by the Institutional Animal Care and Use Committee of the National Institute of Natural Sciences. Stimuli were presented on a cathode ray tube monitor (frame rate: 100 Hz; Totoku Electric) situated at a distance of 57 cm from the monkeys. See *SI Methods* for further details.

Visual Stimulus Generation. Stimuli in the main experiment were generated using the texture synthesis procedure described in Portilla and Simoncelli (20) with a program provided by the authors (www.cns.nyu.edu/~lcv/texture/). The program includes codes to extract a set of synthesis parameters (PS statistics) from an image and codes to synthesize an image based on the PS statistics. In the synthesis, starting from a white-noise image, the algorithm iteratively modifies the image to match its PS statistics with the desired one. Before the synthesis of our stimuli set, we collected 4,400 texture images from photographs of eight material categories (bark, sand, fabric, fur, leather, stone, water, and wood; 550 images for each category) from commercial databases (SOZAIJITEN; Datacraft) and the Internet (Fig. S1*A*). We then computed the PS statistics of each image using four scales and four filter orientations. For the parameters describing correlations between spatially neighboring filter outputs, positional shifts within a seven-pixel square were taken into account. Under these conditions, the algorithm yielded in total 740 parameters of PS statistics for each image (Table S1). To generate a sampling space, we normalized individual parameters across the 4,400 natural texture images, denoised them using principal-component analysis, which reduced the dimensions to 300, and finally projected them into a seven-dimensional space using Fisher's linear discriminant analysis (LDA). LDA finds the linear subspace that maximally separates different categories. Mathematically, it finds the vectors $\mathbf{w}$ that maximize the following objective function, $J(\mathbf{w})$:

$$J(\mathbf{w}) = \frac{\mathbf{w}\boldsymbol{S_B}\mathbf{w}^T}{\mathbf{w}\boldsymbol{S_W}\mathbf{w}^T}, \quad [1]$$

where $\boldsymbol{S_B}$ is the between-category covariance matrix and $\boldsymbol{S_W}$ is the within-category covariance matrix. $\boldsymbol{S_B}$ is defined as follows:

$$\boldsymbol{S_B} = \sum_c N_c(\boldsymbol{m_c} - \boldsymbol{m})(\boldsymbol{m_c} - \boldsymbol{m})^T, \quad [2]$$

where N_c is a number of samples in category c, $\boldsymbol{m_c}$ is the vector averaged across samples in category c, and $\boldsymbol{m}$ is the vector averaged across samples in all categories. $\boldsymbol{S_W}$ is defined as follows:

$$\boldsymbol{S_W} = \sum_c \sum_{n \in c} (\boldsymbol{x_n} - \boldsymbol{m_c})(\boldsymbol{x_n} - \boldsymbol{m_c})^T, \quad [3]$$

where $\boldsymbol{x_n}$ is a vector corresponding to an individual sample in category c. We extracted the first seven dimensions that maximized the objective function J. This can be done by computing the generalized eigenvalues of $\boldsymbol{S_B}$ and $\boldsymbol{S_W}$. For each texture, we connected the 20 nearest textures in this seven-dimensional space as its neighbors, and its descendants were chosen from among these neighbors. To ensure the uniformity of sampling in this parameter space, we removed 230 points that correspond to a part of the 4,400 natural texture images by which the sampling space was constructed, and added 6,185 points by interpolating the parameters of 4,400 textural images such that the distance between each pair of textures fell within 3SD of the distance distribution before the interpolation. The interpolation was done

by linearly averaging the PS parameters of two neighboring textures. These procedures yielded in total 10,355 sets of PS parameters and corresponding textural images among which the adaptive sampling was performed. Note that for image synthesis we used all 740 parameters rather than the seven parameters extracted above. Individual images were synthesized from a white-noise image in an iterative manner to match its PS statistics with the desired one. The size of images was 128 × 128 pixel (corresponding to 6.4°). The number of synthesis iterations was 50, and we confirmed that the PS statistics of images converged to the desired values. The mean and SD of the luminance histogram were equalized to 15 and 6 cd/m^2, respectively, to avoid the effects of these low-level factors. The images were presented on a gray background (10 cd/m^2).

Adaptive Sampling Procedure. We introduced the adaptive sampling procedure to efficiently search parameter space for finding effective textural stimuli (25). In this procedure, we first randomly selected 50 textures (the first generation) from the 10,355-image set and recorded the single-cell responses they elicited. Then in subsequent generations we selected stimuli from among neighbors of ancestor stimuli selected from earlier generations based on the ranks of the elicited firing rates: 13 from the top 10% of stimuli, 10 from the next 10–24%, 5 from 24–44%, 5 from 44–70%, and 5 from 70–100% (Fig. 1*B*). Each subsequent generation also included 12 new, randomly selected stimuli. We repeated these procedures at least 5 times and at most 10 times to record neuronal responses to 250–500 textures.

Dimension Reduction of the Synthesis Parameters. To avoid overfitting to the neuronal responses, we reduced the number of parameters of the PS statistics. This was possible because the parameters were highly redundant. We took two basic approaches to reducing dimension number: averaging across neighboring subbands and extracting principal components. Given that the outputs of the neighboring filters were highly correlated, we averaged the statistics extracted from different subbands, which were originally 4 scales × 4 orientations resolution, into 2 scales × 2 orientations. The averaging was performed only within each group of statistical parameters (Fig. 3). In the second approach, we performed principal-component analysis of Linear cross position and Energy cross position, which are related to spatial pattern information, to extract patterns prevalent across images and subbands. Reduction was accomplished in each group as follows. For the Spectral group, we averaged the amplitudes of neighboring subbands to leave only 2 scales × 2 orientations. For the Marginal group, we left only the skewness of the luminance histogram because the mean and SD were equalized across textures in advance. We did not incorporate the kurtosis of the luminance histogram because it was highly correlated with the skewness in our texture stimuli. For the Linear cross position group, all subbands were expressed as combinations of four major spatial patterns using principal-component analysis. Because different subbands extracted from the same image generally had similar spatial patterns, we averaged the values of all subbands. For the Linear cross scale group, we left 2 scales × 2 orientations. Only combinations between the same orientations in different scales were incorporated. For the Energy cross position group, all subbands were expressed as combinations of three major spatial patterns extracted using principal-component analysis. Patterns obtained in different scales were averaged and the number of orientations was reduced to two. For the Energy cross orientation group, we extracted three combinations of orientations: vertical vs. horizontal, vertical vs. oblique, and horizontal vs. oblique. The number of scales was two. For the Energy cross scale group, the reduction was performed in the same manner as the Linear cross scale. As a result of the reductions, the number of parameters ultimately became 29, and we called the obtained parameter set minimal PS statistics (minPS; Table S2).

Fitting to Neuronal Responses. We used L1-penalized linear least-squares regression (known as lasso) (26) to fit the firing rates of neurons elicited by 250–500 textures. The regression minimizes the following loss function:

$$L = \frac{1}{N}\sum_{i=1}^{N}(FR_i - \boldsymbol{PS}_i \cdot \boldsymbol{W})^2 + \lambda|\boldsymbol{W}|, \quad [4]$$

where N indicates the number of stimuli, FR_i and $\boldsymbol{PS}_i$ indicate the observed firing rate and the PS statistics of image i, $\boldsymbol{W}$ indicates the fitting weights, and λ indicates the regularization coefficient. A 10-fold cross-validation was performed to estimate the fitting performance. In that analysis, we partitioned the presented textures into 10 subsamples and estimated fitting weights ($\boldsymbol{W}$) using the responses to 9 of the 10 subsamples. We then calculated the correlation coefficients between the observed responses of the remaining one subsample and the predicted responses. We repeated this procedure for all 10 combinations and averaged the correlation coefficients. The statistical significance was tested using a permutation test in which we shuffled the combinations between the textures and the firing rates, and repeated the same analyses to calculate a correlation coefficient. We repeated the procedure 2,000 times to obtain the distribution of the correlation coefficients under the null hypothesis for each cell. In lasso, we set a hyperparameter λ in Eq. **4** to perform the fitting. For that, we performed a fivefold cross-validation within the training set (i.e., the 9 of 10 subsamples) to obtain the λ that achieved the best fitting performance within the set. For the final estimation of fitting weights (displayed in Fig. 4*B*), we used the whole dataset without the cross-validation. We also determined the fraction of the variance that was explained in the fitting to the whole variance of the neuronal responses, taking into account the noise in the responses (50). For this, we first computed the amount of noise by calculating the trial-by-trial variances of the neuronal responses to a single stimulus. We then subtracted this noise variance from the whole variance of the neuronal responses to all stimuli. The resultant subtracted variance corresponds to the true variance that should be perfectly explained if there is the true model of the neuron. We therefore divided the coefficient of determination obtained with the fitting by this subtracted variance and regarded the resulting value as the percent explained variance of the fittings.

Analyses of Category Separability. As described above, 4,400 texture stimuli were tagged with eight different material categories. To estimate whether given parameters could separate those categories, we formulated category separability (J) using Eq. **1**. In short, J denotes the between-class variance divided by the within-class variance, which can be obtained from a given parameter set. For the neuronal data, we first computed the predicted responses to the 4,400 images using the weights obtained by fitting to the minPS. We then randomly selected n neurons and calculated the J in this n-dimensional space. We repeated this procedure 100 times to estimate the average J as a function of the number of neurons (n). As a control, we shuffled the weights of each neuron and conducted the same analysis. To compute the category separability using image statistics (Ji), we used the values of the minPS of the same 4,400 images to calculate the J for each group of parameters in the same manner (Fig. 8*C*).

ACKNOWLEDGMENTS. We thank M. Takagi and T. Ota for technical assistance, N. Goda for valuable comments on the manuscript, and persons at the Center for Neural Science, New York University, for valuable discussion. This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas "Shitsukan" (22135007) from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to H.K.), and by a Grant-in-Aid for Japan Society for the Promotion of Science Fellows from the Japan Society for the Promotion of Science (to S.T.).

1. Adelson EH (2001) On seeing stuff: The perception of materials by humans and machines. *Proceedings of the SPIE. Volume 4299: Human Vision and Electronic Imaging VI*, eds Rogowitz BE, Pappas TN (SPIE, Bellingham, WA), pp 1–12.
2. Riesenhuber M, Poggio T (2000) Models of object recognition. *Nat Neurosci* 3(Suppl): 1199–1204.
3. Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71(3):856–867.
4. Kourtzi Z, Connor CE (2011) Neural representations for object perception: Structure, category, and adaptive coding. *Annu Rev Neurosci* 34:45–67.
5. Orban GA (2008) Higher order visual processing in macaque extrastriate cortex. *Physiol Rev* 88(1):59–89.
6. Arcizet F, Jouffrais C, Girard P (2008) Natural textures classification in area V4 of the macaque monkey. *Exp Brain Res* 189(1):109–120.
7. Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. *Nat Neurosci* 16(7):974–981.
8. Köteles K, De Mazière PA, Van Hulle M, Orban GA, Vogels R (2008) Coding of images of materials by macaque inferior temporal cortical neurons. *Eur J Neurosci* 27(2):466–482.
9. Nishio A, Goda N, Komatsu H (2012) Neural selectivity and representation of gloss in the monkey inferior temporal cortex. *J Neurosci* 32(31):10780–10793.
10. Okazawa G, Goda N, Komatsu H (2012) Selective responses to specular surfaces in the macaque visual cortex revealed by fMRI. *Neuroimage* 63(3):1321–1333.
11. Goda N, Tachibana A, Okazawa G, Komatsu H (2014) Representation of the material properties of objects in the visual cortex of nonhuman primates. *J Neurosci* 34(7):2660–2673.
12. Julesz B (1981) Textons, the elements of texture perception, and their interactions. *Nature* 290(5802):91–97.
13. Bergen JR, Adelson EH (1988) Early vision and texture perception. *Nature* 333(6171): 363–364.
14. Graham NV (2011) Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years. *Vision Res* 51(13): 1397–1430.

15. Rao AR, Lohse GL (1996) Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Res* 36(11):1649–1669.
16. Victor JD, Conte MM (2012) Local image statistics: Maximum-entropy constructions and perceptual salience. *J Opt Soc Am A Opt Image Sci Vis* 29(7):1313–1345.
17. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987.
18. Varma M, Zisserman A (2005) A statistical approach to texture classification from single images. *Int J Comput Vis* 62(1-2):61–81.
19. Heeger DJ, Bergen JR (1995) Pyramid-based texture analysis/synthesis. *Proceedings of SIGGRAPH* (ACM, New York), pp 229–238.
20. Portilla J, Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. *Int J Comput Vis* 40(1):49–71.
21. Freeman J, Simoncelli EP (2011) Metamers of the ventral stream. *Nat Neurosci* 14(9): 1195–1201.
22. Hanazawa A, Komatsu H (2001) Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J Neurosci* 21(12): 4490–4497.
23. Carlson ET, Rasquinha RJ, Zhang K, Connor CE (2011) A sparse object coding scheme in area V4. *Curr Biol* 21(4):288–293.
24. Hung CC, Carlson ET, Connor CE (2012) Medial axis shape coding in macaque inferotemporal cortex. *Neuron* 74(6):1099–1113.
25. Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE (2008) A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* 11(11):1352–1360.
26. Tibshirani R (2011) Regression shrinkage and selection via the lasso: A retrospective. *J R Stat Soc Series B Stat Methodol* 73(3):273–282.
27. Desimone R, Schein SJ (1987) Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *J Neurophysiol* 57(3):835–868.
28. David SV, Hayden BY, Gallant JL (2006) Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96(6):3492–3505.
29. Grömping U (2007) Estimators of relative importance in linear regression based on variance decomposition. *Am Stat* 61(2):139–147.
30. El-Shamayleh Y, Movshon JA (2011) Neuronal responses to texture-defined form in macaque visual area V2. *J Neurosci* 31(23):8543–8555.
31. Merigan WH (2000) Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. *Vis Neurosci* 17(6):949–958.
32. Rust NC, Dicarlo JJ (2010) Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30(39):12978–12995.
33. Roe AW, et al. (2012) Toward a unified theory of visual area V4. *Neuron* 74(1):12–29.
34. Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259(5091):100–103.
35. Zeki SM (1973) Colour coding in rhesus monkey prestriate cortex. *Brain Res* 53(2): 422–427.
36. Kotake Y, Morimoto H, Okazaki Y, Fujita I, Tamura H (2009) Organization of color-selective neurons in macaque visual area V4. *J Neurophysiol* 102(1):15–27.
37. Hinkle DA, Connor CE (2001) Disparity tuning in macaque area V4. *Neuroreport* 12(2): 365–369.
38. Watanabe M, Tanaka H, Uka T, Fujita I (2002) Disparity-selective neurons in area V4 of macaque monkeys. *J Neurophysiol* 87(4):1960–1973.
39. Pasupathy A, Connor CE (2002) Population coding of shape in area V4. *Nat Neurosci* 5(12):1332–1338.
40. Nandy AS, Sharpee TO, Reynolds JH, Mitchell JF (2013) The fine structure of shape tuning in area V4. *Neuron* 78(6):1102–1115.
41. Rosenholtz R, Huang J, Raj A, Balas BJ, Ilie L (2012) A summary statistic representation in peripheral vision explains visual search. *J Vis* 12(4):14.
42. Balas B, Nakano L, Rosenholtz R (2009) A summary-statistic representation in peripheral vision explains visual crowding. *J Vis* 9(12):1–18.
43. Yamins DLK, et al. (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 111(23):8619–8624.
44. Karklin Y, Lewicki MS (2009) Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 457(7225):83–86.
45. Hiramatsu C, Goda N, Komatsu H (2011) Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage* 57(2):482–494.
46. Kim J, Marlow P, Anderson BL (2011) The perception of gloss depends on highlight congruence with surface shading. *J Vis* 11(9):1–19.
47. Motoyoshi I, Nishida S, Sharan L, Adelson EH (2007) Image statistics and the perception of surface qualities. *Nature* 447(7141):206–209.
48. Orban GA, Zhu Q, Vanduffel W (2014) The transition in the ventral stream from feature to real-world entity representations. *Front Psychol* 5:695.
49. Committee on Care and Use of Laboratory Animals (1996) *Guide for the Care and Use of Laboratory Animals* (Natl Inst Health, Bethesda), DHHS Publ No (NIH) 85-23.
50. Pasupathy A, Connor CE (2001) Shape representation in area V4: Position-specific tuning for boundary conformation. *J Neurophysiol* 86(5):2505–2519.
51. McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron* 71(5):926–940.