LETTERS

Identifying natural images from human brain activity

Kendrick N. Kay[1], Thomas Naselaris[2], Ryan J. Prenger[3] & Jack L. Gallant[1,2]

A challenging goal in neuroscience is to be able to read out, or decode, mental content from brain activity. Recent functional magnetic resonance imaging (fMRI) studies have decoded orientation[1,2], position[3] and object category[4,5] from activity in visual cortex. However, these studies typically used relatively simple stimuli (for example, gratings) or images drawn from fixed categories (for example, faces, houses), and decoding was based on previous measurements of brain activity evoked by those same stimuli or categories. To overcome these limitations, here we develop a decoding method based on quantitative receptive-field models that characterize the relationship between visual stimuli and fMRI activity in early visual areas. These models describe the tuning of individual voxels for space, orientation and spatial frequency, and are estimated directly from responses evoked by natural images. We show that these receptive-field models make it possible to identify, from a large set of completely novel natural images, which specific image was seen by an observer. Identification is not a mere consequence of the retinotopic organization of visual areas; simpler receptive-field models that describe only spatial tuning yield much poorer identification performance. Our results suggest that it may soon be possible to reconstruct a picture of a person's visual experience from measurements of brain activity alone.

Imagine a general brain-reading device that could reconstruct a picture of a person's visual experience at any moment in time[6]. This general visual decoder would have great scientific and practical use. For example, we could use the decoder to investigate differences in perception across people, to study covert mental processes such as attention, and perhaps even to access the visual content of purely mental phenomena such as dreams and imagery. The decoder would also serve as a useful benchmark of our understanding of how the brain represents sensory information.

How do we build a general visual decoder? We consider as a first step the problem of image identification[3,7,8]. This problem is analogous to the classic 'pick a card, any card' magic trick. We begin with a large, arbitrary set of images. The observer picks an image from the set and views it while brain activity is measured. Is it possible to use the measured brain activity to identify which specific image was seen?

To ensure that a solution to the image identification problem will be applicable to general visual decoding, we introduce two challenging requirements[6]. First, it must be possible to identify novel images. Conventional classification-based decoding methods can be used to identify images if brain activity evoked by those images has been measured previously, but they cannot be used to identify novel images (see Supplementary Discussion). Second, it must be possible
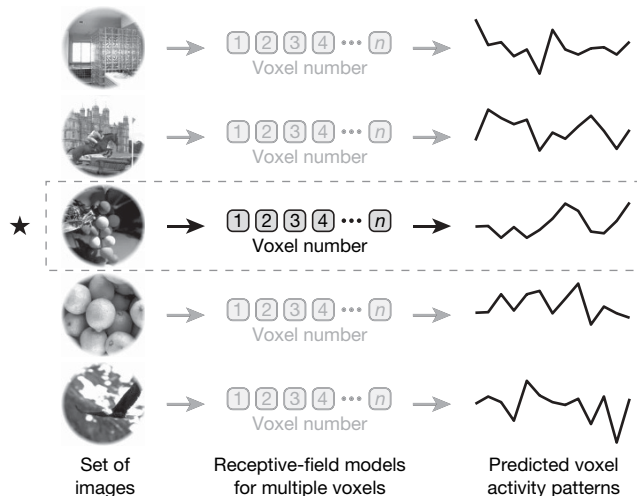
Stage 1: model estimation
Estimate a receptive-field model for each voxel
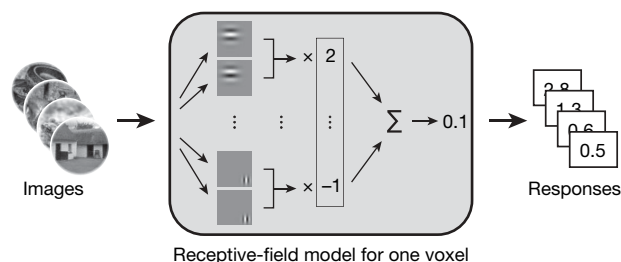
Stage 2: image identification
(1) Measure brain activity for an image
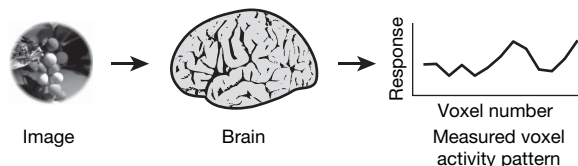
(2) Predict brain activity for a set of images using receptive-field models

(3) Select the image (★) whose predicted brain activity is most similar to the measured brain activity

Figure 1 | Schematic of experiment. The experiment consisted of two stages. In the first stage, model estimation, fMRI data were recorded while each subject viewed a large collection of natural images. These data were used to estimate a quantitative receptive-field model[10] for each voxel. The model was based on a Gabor wavelet pyramid[11–13] and described tuning along the dimensions of space[3,14–19], orientation[1,2,20] and spatial frequency[21,22]. In the second stage, image identification, fMRI data were recorded while each subject viewed a collection of novel natural images. For each measurement of brain activity, we attempted to identify which specific image had been seen. This was accomplished by using the estimated receptive-field models to predict brain activity for a set of potential images and then selecting the image whose predicted activity most closely matches the measured activity.

[1]Department of Psychology, University of California, Berkeley, California 94720, USA. [2]Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA. [3]Department of Physics, University of California, Berkeley, California 94720, USA.
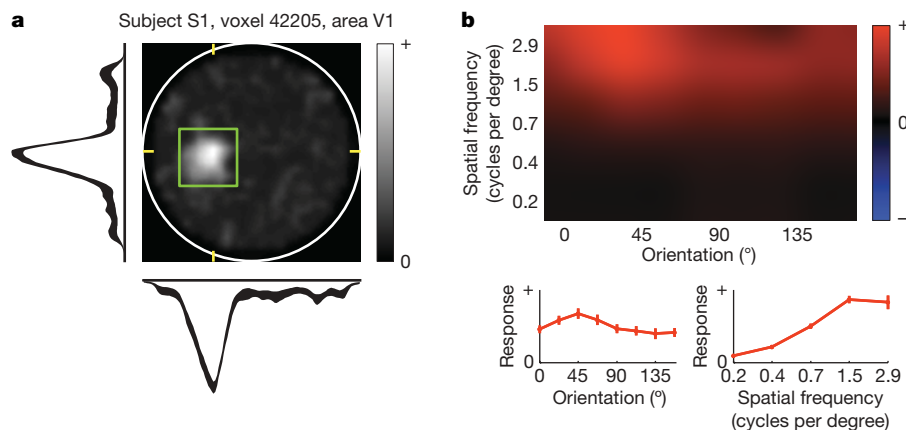
Figure 2 | Receptive-field model for a representative voxel. a, Spatial envelope. The intensity of each pixel indicates the sensitivity of the receptive field to that location. The white circle delineates the bounds of the stimulus ($20° \times 20°$) and the green square delineates the estimated receptive-field location. Horizontal and vertical slices through the spatial envelope are shown below and to the left. These intersect the peak of the spatial envelope, as indicated by yellow tick marks. The thickness of each slice profile indicates ± 1 s.e.m. This receptive field is located in the left hemifield, just below the horizontal meridian. **b,** Orientation and spatial frequency tuning curves. The top matrix depicts the joint orientation and spatial frequency tuning of the receptive field, and the bottom two plots give the marginal orientation and spatial frequency tuning curves. Error bars indicate ± 1 s.e.m. This receptive field has broadband orientation tuning and high-pass spatial frequency tuning. For additional receptive-field examples and population summaries of receptive-field properties, see Supplementary Figs 9–11.

to identify natural images. Natural images have complex statistical structure[9] and are much more difficult to parameterize than simple artificial stimuli such as gratings or pre-segmented objects. Because neural processing of visual stimuli is nonlinear, a decoder that can identify simple stimuli may fail when confronted with complex natural images.

Our experiment consisted of two stages (Fig. 1). In the first stage, model estimation, fMRI data were recorded from visual areas V1, V2 and V3 while each subject viewed 1,750 natural images. We used these data to estimate a quantitative receptive-field model[10] for each voxel (Fig. 2). The model was based on a Gabor wavelet pyramid[11–13] and described tuning along the dimensions of space[3,14–19], orientation[1,2,20] and spatial frequency[21,22]. (See Supplementary Discussion for a comparison of our receptive-field analysis with those of previous studies.)

In the second stage, image identification, fMRI data were recorded while each subject viewed 120 novel natural images. This yielded 120 distinct voxel activity patterns for each subject. For each voxel activity pattern we attempted to identify which image had been seen. To do this, the receptive-field models estimated in the first stage of the experiment were used to predict the voxel activity pattern that would be evoked by each of the 120 images. The image whose predicted voxel activity pattern was most correlated (Pearson's r) with the measured voxel activity pattern was selected.

Identification performance for one subject is illustrated in Fig. 3. For this subject, 92% (110/120) of the images were identified correctly (subject S1), whereas chance performance is just 0.8% (1/120). For a second subject, 72% (86/120) of the images were identified correctly (subject S2). These high performance levels demonstrate the validity of our decoding approach, and indicate that our receptive-field models accurately characterize the selectivity of individual voxels to natural images.

A general visual decoder would be especially useful if it could operate on brain activity evoked by a single perceptual event. However, because fMRI data are noisy, the results reported above were obtained using voxel activity patterns averaged across 13 repeated trials. We therefore attempted identification using voxel activity patterns from single trials. Single-trial performance was 51% (834/1620) and 32% (516/1620) for subjects S1 and S2, respectively (Fig. 4a); once again, chance performance is just 0.8% (13.5/1620). These results suggest that it may be feasible to decode the content of perceptual experiences in real time[7,23].

We have so far demonstrated identification of a single image drawn from a set of 120 images, but a general visual decoder should be able to handle much larger sets of images. To investigate this issue, we measured identification performance for various set sizes up to 1,000 images (Fig. 4b). As set size increased tenfold from 100 to 1,000, performance only declined slightly, from 92% to 82% (subject S1,
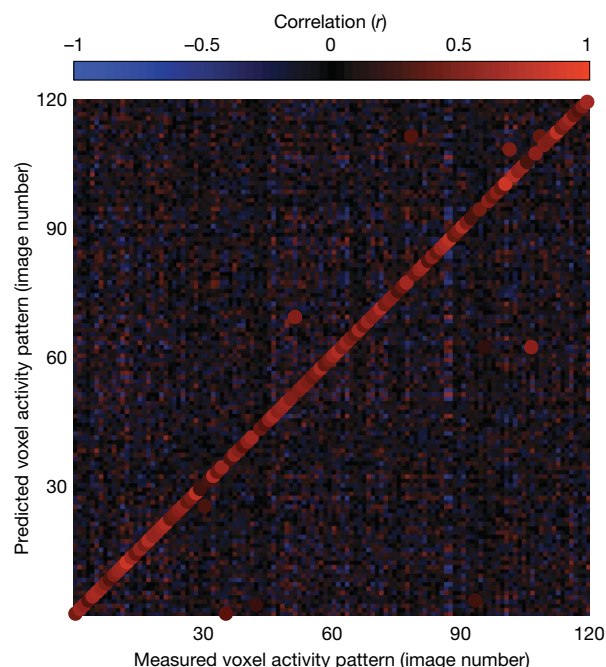

Figure 3 | Identification performance. In the image identification stage of the experiment, fMRI data were recorded while each subject viewed 120 novel natural images that had not been used to estimate the receptive-field models. For each of the 120 measured voxel activity patterns, we attempted to identify which image had been seen. This figure illustrates identification performance for one subject (S1). The colour at the mth column and nth row represents the correlation between the measured voxel activity pattern for the mth image and the predicted voxel activity pattern for the nth image. The highest correlation in each column is designated by an enlarged dot of the appropriate colour, and indicates the image selected by the identification algorithm. For this subject 92% (110/120) of the images were identified correctly.
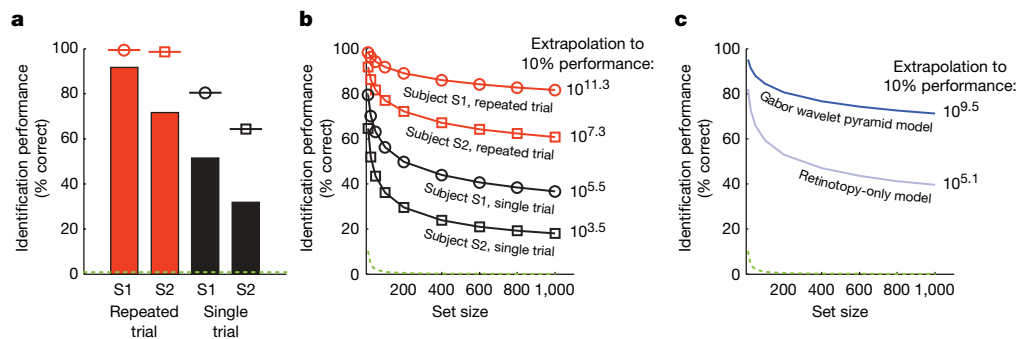
Figure 4 | Factors that impact identification performance. a, Summary of identification performance. The bars indicate empirical performance for a set size of 120 images, the marker above each bar indicates the estimated noise ceiling (that is, the theoretical maximum performance given the level of noise in the data), and the dashed green line indicates chance performance. The noise ceiling estimates suggest that the difference in performance across subjects is due to intrinsic differences in the level of noise. **b,** Scaling of identification performance with set size. The x axis indicates set size, the y axis indicates identification performance, and the number to the right of each line gives the estimated set size at which performance declines to 10% correct. In all cases performance scaled very well with set size. **c,** Retinotopy-only model versus Gabor wavelet pyramid model. Identification was attempted using an alternative retinotopy-only model that captures only the location and size of each voxel's receptive field. This model performed substantially worse than the Gabor wavelet pyramid model, indicating that spatial tuning alone is insufficient to achieve optimal identification performance. (Results reflect repeated-trial performance averaged across subjects; see Supplementary Fig. 5 for detailed results.)

repeated trial). Extrapolation of these measurements (see Supplementary Methods) suggests that performance for this subject would remain above 10% even up to a set size of $10^{11.3}$ images. This is more than 100 times larger than the number of images currently indexed by Google ($10^{8.9}$ images; source: http://www.google.com/whatsnew/, 4 June 2007).

Early visual areas are organized retinotopically, and voxels are known to reflect this organization[14,16,18]. Could our results be a mere consequence of retinotopy? To answer this question, we attempted identification using an alternative model that captures the location and size of each voxel's receptive field but discards orientation and spatial frequency information (Fig. 4c). Performance for this retinotopy-only model declined to 10% correct at a set size of just $10^{5.1}$ images, whereas performance for the Gabor wavelet pyramid model did not decline to 10% correct until $10^{9.5}$ images were included in the set (repeated-trial performance extrapolated and averaged across subjects). This result indicates that spatial tuning alone does not yield optimal identification performance; identification improves substantially when orientation and spatial frequency tuning are included in the model.

To further investigate the impact of orientation and spatial frequency tuning, we measured identification performance after imposing constraints on the orientation and spatial frequency tuning of the Gabor wavelet pyramid model (Supplementary Fig. 8). The results indicate that both orientation and spatial frequency tuning contribute to identification performance, but that the latter makes the larger contribution. This is consistent with recent studies demonstrating that voxels have only slight orientation bias[1,2]. We also find that voxel-to-voxel variation in orientation and spatial frequency tuning contributes to identification performance. This reinforces the growing realization in the fMRI community that information may be present in fine-grained patterns of voxel activity[6].

To be practical our identification algorithm must perform well even when brain activity is measured long after estimation of the receptive-field models. To assess performance over time[2,4,6,23] we attempted identification for a set of 120 novel natural images that were seen approximately two months after the initial experiment. In this case 82% (99/120) of the images were identified correctly (chance performance 0.8%; subject S1, repeated trial). We also evaluated identification performance for a set of 12 novel natural images that were seen more than a year after the initial experiment. In this case 100% (12/12) of the images were identified correctly (chance performance 8%; subject S1, repeated trial). These results demonstrate that the stimulus-related information that can be decoded from voxel activity remains largely stable over time.

Why does identification sometimes fail? Inspection revealed that identification errors tended to occur when the selected image was visually similar to the correct image. This suggests that noise in measured voxel activity patterns causes the identification algorithm to confuse images that have similar features.

Functional MRI signals have modest spatial resolution and reflect haemodynamic activity that is only indirectly coupled to neural activity[24,25]. Despite these limitations, we have shown that fMRI signals can be used to achieve remarkable levels of identification performance. This indicates that fMRI signals contain a considerable amount of stimulus-related information[4] and that this information can be successfully decoded in practice.

Identification of novel natural images brings us close to achieving a general visual decoder. The final step will require devising a way to reconstruct the image seen by the observer, instead of selecting the image from a known set. Stanley and co-workers[26] reconstructed natural movies by modelling the luminance of individual image pixels as a linear function of single-unit activity in cat lateral geniculate nucleus. This approach assumes a linear relation between luminance and the activity of the recorded units, but this condition does not hold in fMRI[27,28].

An alternative approach to reconstruction is to incorporate receptive-field models into a statistical inference framework. In such a framework, receptive-field models are used to infer the most likely image given a measured activity pattern. This model-based approach has a long history in both theoretical and experimental neuroscience[29,30]. Recently, Thirion and co-workers[3] used it to reconstruct spatial maps of contrast from fMRI activity in human visual cortex. The success of the approach depends critically on how well the receptive-field models predict brain activity. The present study demonstrates that our receptive-field models have sufficient predictive power to enable identification of novel natural images, even for the case of extremely large sets of images. We are therefore optimistic that the model-based approach will make possible the reconstruction of natural images from human brain activity.

METHODS SUMMARY

The stimuli consisted of sequences of $20° \times 20°$ greyscale natural photographs (Supplementary Fig. 1a). Photographs were presented for 1 s with a delay of 3 s between successive photographs (Supplementary Fig. 1b). Subjects (S1: author T.N.; S2: author K.N.K.) viewed the photographs while fixating a central white square. MRI data were collected at the Brain Imaging Center at University of California, Berkeley using a 4 T INOVA MR scanner (Varian, Inc.) and a quadrature transmit/receive surface coil (Midwest RF, LLC). Functional BOLD data were recorded from occipital cortex at a spatial resolution of $2\,mm \times 2\,mm \times 2.5\,mm$ and a temporal resolution of 1 Hz. Brain volumes were

reconstructed and then co-registered to correct differences in head positioning within and across scan sessions. The time-series data were pre-processed such that voxel-specific response time courses were deconvolved from the data. Voxels were assigned to visual areas based on retinotopic mapping data[17] collected in separate scan sessions.

In the model estimation stage of the experiment, a receptive-field model was estimated for each voxel. The model was based on a Gabor wavelet pyramid[11–13] (Supplementary Figs 2 and 3), and was able to characterize responses of voxels in early visual areas V1, V2 and V3 (Supplementary Table 1). Alternative receptive-field models were also used, including the retinotopy-only model and several constrained versions of the Gabor wavelet pyramid model. Details of these models and model estimation procedures are given in Supplementary Methods.

In the image identification stage of the experiment, the estimated receptive-field models were used to identify images viewed by the subjects, based on measured voxel activity. The identification algorithm is described in the main text. For details of voxel selection, performance for different set sizes, and noise ceiling estimation, see Supplementary Fig. 4 and Supplementary Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Haynes, J. D. & Rees, G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neurosci.* **8**, 686–691 (2005).
2. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nature Neurosci.* **8**, 679–685 (2005).
3. Thirion, B. *et al.* Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* **33**, 1104–1116 (2006).
4. Cox, D. D. & Savoy, R. L. Functional magnetic resonance imaging (fMRI) ''brain reading'': detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **19**, 261–270 (2003).
5. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
6. Haynes, J. D. & Rees, G. Decoding mental states from brain activity in humans. *Nature Rev. Neurosci.* **7**, 523–534 (2006).
7. Hung, C. P., Kreiman, G., Poggio, T. & DiCarlo, J. J. Fast readout of object identity from macaque inferior temporal cortex. *Science* **310**, 863–866 (2005).
8. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. & Livingstone, M. S. A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674 (2006).
9. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
10. Wu, M. C., David, S. V. & Gallant, J. L. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* **29**, 477–505 (2006).
11. Daugman, J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* **2**, 1160–1169 (1985).
12. Jones, J. P. & Palmer, L. A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58**, 1233–1258 (1987).
13. Lee, T. S. Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal.* **18**, 959–971 (1996).
14. DeYoe, E. A. *et al.* Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl Acad. Sci. USA* **93**, 2382–2386 (1996).
15. Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *Neuroimage* **39**, 647–660 (2008).
16. Engel, S. A. *et al.* fMRI of human visual cortex. *Nature* **369**, 525 (1994).
17. Hansen, K. A., David, S. V. & Gallant, J. L. Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. *Neuroimage* **23**, 233–241 (2004).
18. Sereno, M. I. *et al.* Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893 (1995).
19. Smith, A. T., Singh, K. D., Williams, A. L. & Greenlee, M. W. Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb. Cortex* **11**, 1182–1190 (2001).
20. Sasaki, Y. *et al.* The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron* **51**, 661–670 (2006).
21. Olman, C. A., Ugurbil, K., Schrater, P. & Kersten, D. BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Res.* **44**, 669–683 (2004).
22. Singh, K. D., Smith, A. T. & Greenlee, M. W. Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* **12**, 550–564 (2000).
23. Haynes, J. D. & Rees, G. Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.* **15**, 1301–1307 (2005).
24. Heeger, D. J. & Ress, D. What does fMRI tell us about neuronal activity? *Nature Rev. Neurosci.* **3**, 142–151 (2002).
25. Logothetis, N. K. & Wandell, B. A. Interpreting the BOLD signal. *Annu. Rev. Physiol.* **66**, 735–769 (2004).
26. Stanley, G. B., Li, F. F. & Dan, Y. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J. Neurosci.* **19**, 8036–8042 (1999).
27. Haynes, J. D., Lotto, R. B. & Rees, G. Responses of human visual cortex to uniform surfaces. *Proc. Natl Acad. Sci. USA* **101**, 4286–4291 (2004).
28. Rainer, G., Augath, M., Trinath, T. & Logothetis, N. K. Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr. Biol.* **11**, 846–854 (2001).
29. Salinas, E. & Abbott, L. F. Vector reconstruction from firing rates. *J. Comput. Neurosci.* **1**, 89–107 (1994).
30. Zhang, K., Ginzburg, I., McNaughton, B. L. & Sejnowski, T. J. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79**, 1017–1044 (1998).

Author Contributions K.N.K. designed and conducted the experiment and was first author on the paper. K.N.K. and T.N. analysed the data. R.J.P. provided mathematical ideas and assistance. J.L.G. provided guidance on all aspects of the project. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.L.G. (gallant@berkeley.edu).

METHODS

Stimuli. The stimuli consisted of sequences of natural photographs. Photographs were obtained from a commercial digital library (Corel Stock Photo Libraries from Corel Corporation), the Berkeley Segmentation Dataset (http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/) and the authors' personal collections. The content of the photographs included animals, buildings, food, humans, indoor scenes, manmade objects, outdoor scenes, and textures. Photographs were converted to greyscale, downsampled so that the smaller of the two image dimensions was 500 pixels, linearly transformed so that the 1/10th and 99 9/10th percentiles of the original pixel values were mapped to the minimum (0) and maximum (255) pixel values, cropped to the central 500 pixels × 500 pixels, masked with a circle, and placed on a grey background (Supplementary Fig. 1a). The luminance of the background was set to the mean luminance across photographs, and the outer edge of each photograph (10% of the radius of the circular mask) was linearly blended into the background.

The size of the photographs was $20° × 20°$ (500 pixels × 500 pixels). A central white square served as the fixation point, and its size was $0.2° × 0.2°$ (4 pixels × 4 pixels). Photographs were presented in successive 4-s trials; in each trial, a photograph was presented for 1 s and the grey background was presented for 3 s. Each 1-s presentation consisted of a photograph being flashed ON–OFF–ON–OFF–ON where ON corresponds to presentation of the photograph for 200 ms and OFF corresponds to presentation of the grey background for 200 ms (Supplementary Fig. 1b). The flashing technique increased the signal-to-noise ratio of voxel responses relative to that achieved by presenting each photograph continuously for 1 s (data not shown).

Visual stimuli were delivered using the VisuaStim goggles system (Resonance Technology). The display resolution was 800 × 600 at 60 Hz. A PowerBook G4 computer (Apple Computer) controlled stimulus presentation using software written in MATLAB 5.2.1 (The Mathworks) and Psychophysics Toolbox 2.53 (http://psychtoolbox.org).

MRI parameters. The experimental protocol was approved by the University of Caifornia, Berkeley Committee for the Protection of Human Subjects. MRI data were collected at the Brain Imaging Center at University of California, Berkeley using a 4 T INOVA MR scanner (Varian, Inc.) and a quadrature transmit/receive surface coil (Midwest RF, LLC). Data were acquired using coronal slices that covered occipital cortex: 18 slices, slice thickness 2.25 mm, slice gap 0.25 mm, field-of-view 128 mm × 128 mm. (In one scan session, a slice gap of 0.5 mm was used.) For functional data, a T2*-weighted, single-shot, slice-interleaved, gradient-echo EPI pulse sequence was used: matrix size 64 × 64, TR 1 s, TE 28 ms, flip angle 20°. The nominal spatial resolution of the functional data was 2 mm × 2 mm × 2.5 mm. For anatomical data, a T1-weighted gradient-echo multislice sequence was used: matrix size 256 × 256, TR 0.2 s, TE 5 ms, flip angle 40°.

Data collection. Data for the model estimation and image identification stages of the experiment were collected in the same scan sessions. Two subjects were used: S1 (author T.N., age 33) and S2 (author K.N.K., age 25). Subjects were healthy and had normal or corrected-to-normal vision.

Five scan sessions of data were collected from each subject. Each scan session consisted of five model estimation runs and two image identification runs. Model estimation runs (11 min each) were used for the model estimation stage of the experiment. Each model estimation run consisted of 70 distinct images presented two times each. Image identification runs (12 min each) were used for the image identification stage of the experiment. Each image identification run consisted of 12 distinct images presented 13 times each. Images were randomly selected for each run and were mutually exclusive across runs. The total number of distinct images used in the model estimation and image identification runs was 1,750 and 120, respectively. (For additional details on experimental design, see Supplementary Methods.)

Three additional scan sessions of data were collected from subject S1. Two of these were held approximately two months after the main experiment, and consisted of five image identification runs each. The third was held approximately 14 months after the main experiment, and consisted of one image identification run. The images used in these additional scan sessions were randomly selected and were distinct from the images used in the main experiment.

Data pre-processing. Functional brain volumes were reconstructed and then co-registered to correct differences in head positioning within and across scan sessions. Next, voxel-specific response time courses were estimated and deconvolved from the time-series data. This produced, for each voxel, an estimate of the amplitude of the response (a single value) to each image used in the model estimation and image identification runs. Finally, voxels were assigned to visual areas based on retinotopic mapping data[17] collected in separate scan sessions. (Details of these procedures are given in Supplementary Methods.)

Model estimation. A receptive-field model was estimated for each voxel based on its responses to the images used in the model estimation runs. The model was based on a Gabor wavelet pyramid[11–13]. In the model, each image is represented by a set of Gabor wavelets differing in size, position, orientation, spatial frequency and phase (Supplementary Fig. 2). The predicted response is a linear function of the contrast energy contained in quadrature wavelet pairs (Supplementary Fig. 3). Because contrast energy is a nonlinear quantity, this is a linearized model[10]. The model was able to characterize responses of voxels in visual areas V1, V2 and V3 (Supplementary Table 1), but it did a poor job of characterizing responses in higher visual areas such as V4.

Alternative receptive-field models were also used, including the retinotopy-only model and several constrained versions of the Gabor wavelet pyramid model. Details of these models and model estimation procedures are given in Supplementary Methods.

Image identification. Voxel activity patterns were constructed from voxel responses evoked by the images used in the image identification runs. For each voxel activity pattern, the estimated receptive-field models were used to identify which specific image had been seen. The identification algorithm is described in the main text. See Supplementary Fig. 4 and Supplementary Methods for details of voxel selection, performance for different set sizes, and noise ceiling estimation. See Supplementary Discussion for a comparison of identification with the decoding problems of classification and reconstruction.