Bestelldat: 23-10-2003-16:43
Bestell-Nr: 2003102304093

NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

normal

SUBITO

Tel: 0641-9926108

Mail: Miriam.Spering@psychol.uni-giessen.de

Spering
Miriam
Otto-Behaghel-Str. 10F
35390 Giessen

USER-GROUP-1

Kontaktperson

Miriam.Spering@psychol.uni-giessen.de

Kunden-/Zugangsnummer
STD03X14309

Verfasser: Kersten, D.
(Aufsatz)
Titel: Predictability and redundancy of natural images
(Aufsatz)
Seiten: 2395-2400
Titel (Monographie/Zeitschrift)
Journal of the Optical Society of America / A
Washington, DC
Soc.
1084-7529/n

Standort:
ZB 59499 / BBP ZB 59499

Band/ Heft	**Jahrgang**
4/	1987

Lieferform:	**Lieferart:**
PDF	E-Mail

Bemerkungen:

Einzelrechnung
Rechnungsanschrift
.
.
.
. .

LIEFERSCHEIN

Abrechnung über Subito-Zentralregulierung

Lieferbibliothek: BBPhysik / SUB Göttingen
Bearbeitungsdatum: 24.10.03

☐ Eildienst (24 STD/hrs) ☐ Altbestand

☐ Kopie / hardcopy ☒ E-mail ☐ Ariel ☐ Fax

☐ Ausleihe eines Buches (Mikroform) / lending of a book (microform)

Anzahl Kopien: 6

Hausadresse:
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Bereichsbibliothek Physik Tammanstr. 1 37077 Göttingen
GERMANY

Telefon: +49 (0)551 / 39-2360
Fax: +49 (0)5 51/ 39-13146
E-Mail: bbphys@sub.uni-goettingen.de
Ust-IdNr: DE152336201

Predictability and redundancy of natural images

Daniel Kersten

Departments of Psychology and Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island 02912

Received April 17, 1987; accepted August 14, 1987

One aspect of human image understanding is the ability to estimate missing parts of a natural image. This ability depends on the redundancy of the representation used to describe the class of images. In 1951, Shannon [Bell. Syst. Tech. J. **30**, 50 (1951)] showed how to estimate bounds on the entropy and redundancy of an information source from predictability data. The entropy, in turn, gives a measure of the limits to error-free information compaction. An experiment was devised in which human observers interactively restored missing gray levels from 128 × 128 pixel pictures with 16 gray levels. For eight images, the redundancy ranged from 46%, for a complicated picture of foliage, to 74%, for a picture of a face. For almost-complete pictures, but not for noisy pictures, this performance can be matched by a nearest-neighbor predictor.

One of the distinguishing characteristics of intelligent systems is the ability to make accurate and reliable predictions from partial data. Our own ability to interpret the images that our eyes receive involves making inferences about the environmental causes of image intensities, often from incomplete data. This ability to make predictions or inferences depends on the existence of statistical dependencies or redundancies in natural images. Despite the fact that the prediction of information from natural images plays an important role in image understanding, there have been relatively few quantitative studies of the ability of humans to do this. In this paper a simple example of this ability, that of restoring missing pixel gray levels in natural images, is explored. These results are, in turn, related to quantitative estimates of the redundancy of natural images. Although this is a simple prediction task, the technique should be easy to extend to the investigation of more-complicated aspects of our ability to predict picture information.

Some years ago, Attneave[1] and Barlow[2,3] pointed out that a principal task of biological vision may be to encode the visual image into a less redundant form. In this context, rather than searching for features in an image, the visual system codes a given image with regard to its relation to the statistical properties of the set of natural images. Because the space of possible pictures is so great, it makes good sense to utilize naturally occurring redundancy to recode image information into a less redundant form. Efficient coding can result in the transmission of the same amount of information with fewer neurons or with smaller dynamic range. In addition to numerous communication-engineering applications to image compaction,[4] there have been recent quantitative explanations of nonlinear transduction, lateral inhibition, and opponent-color processing as redundancy-reduction mechanisms.[5–8] Further upstream, the image-understanding tasks that the cortex faces may be simplified by a redundancy reduction in the image specification. For example, eigenvector transformation of pictures of faces can make possible a large reduction of dimensionality, which may be useful for economical representation and retrieval.[9] Recent work on autoassociative networks is providing tools for searching for compact image or shape codes.[10,11]

In order to demonstrate the relationship between predictability and redundancy, Fig. 1(a) shows a 128 × 128 pixel image with 16 possible gray levels, where about 150 pixels have been deleted. Here, deletion means that the original gray levels were replaced at random by one of the 16 gray levels chosen from a uniform distribution. The reader should have little difficulty spotting the deleted pixels, and, as will be shown below, it is fairly easy to make good guesses about what the gray levels should be. In Fig. 1(c) 150 pixels have also been deleted. Not only is it impossible to determine which pixels were deleted; it is also impossible to determine what gray levels should be used to replace them. Figure 1(a) belongs to the class of natural images, which is highly redundant. Figure 1(c) is an example of white visual noise with uniformly distributed gray levels, a class of pictures that has no redundancy.

The redundancy for an information source was originally defined quantitatively by Shannon.[12] Consider a class of digitized natural images that might be presented on a graphics display. Suppose that they are specified by k pixels with m bits of gray-level resolution per pixel. The nth-order conditional entropy for this class of pictures, F_n, is the expected value of the negative log (base 2) of the probability of gray level i conditional on the values of n neighbors (over some defined neighborhood structure):

$$F_n = -\sum_{i,j} p(i, b_j) \log_2 p(i|b_j), \tag{1}$$

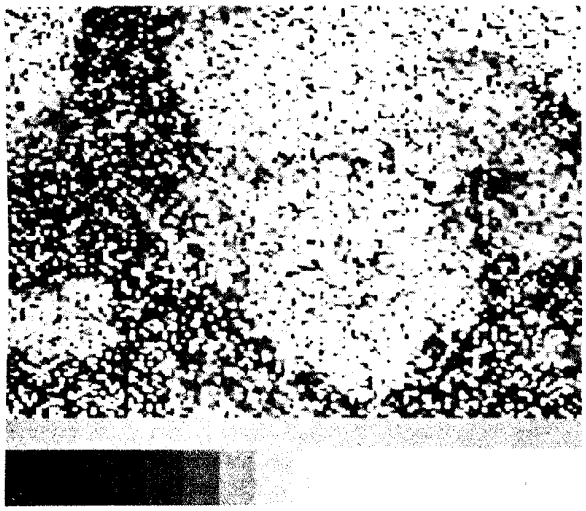
where b_j is the jth block of the n neighborhood pixels ($j = 1$ to 2^{mn}, $i = 1$ to 2^m). As n approaches k, F_n approaches the minimum average number of bits per pixel required to code this class, for arbitrarily small error. If the probability of pixel gray levels is constant and independent of all others, the entropy is a maximum value of m bits per pixel. This provides a useful baseline to quantify predictability and redundancy. Redundancy is

$$1 - \frac{F_n}{m}. \tag{2}$$

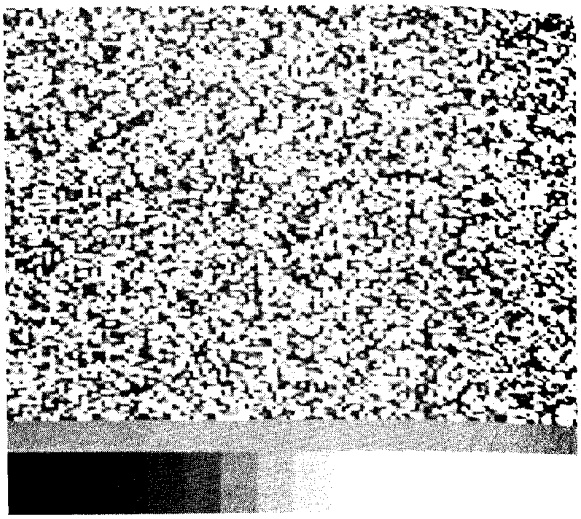
In actual practice, it is impractical to calculate high-order conditional probabilities and thus redundancy. However, in the 1950's Shannon[12] showed that if a device exists that can predict unknown alphabet members from known ones in

0740-3232/87/122395-06$02.00

(a)



(b)



(c)

Fig. 1. Picture hbb quantized to 128 × 128 × 4 bits. (a), (b), and (c) have increasing fractions of deleted pixels. About 1% and 100% of the pixels have been deleted from (a) and (c), respectively.

a text, it is possible to compute bounds on the entropy and the redundancy of a language. The redundancy estimates get better as the predictor approaches ideal. Although we are probably not ideal predictors, human observers implicitly possess an enormous store of knowledge about natural images. In this paper the human ability to predict missing pixel gray levels is measured and used to estimate image redundancy for a particular image quantization.

In contrast to numerous applications of Shannon's guessing game to language studies,[13] there have been only a few studies of the redundancy of pictures by using human prediction. Two such studies[14,15] investigated gray-level predictability for a small set of natural images, and a third[16] measured predictability in simple contour line drawings. In 1965, Parks[14] reported on the predictability of half-tone gray-level pictures covered by a 36 tile × 44 tile grid. Starting with a completely covered picture, the subject chose a tile and guessed the gray level until the correct answer was obtained (binary guessing was used if the subject was unsure). This tile was removed, and the subject went on to the next tile. The gray level was estimated subjectively (for both the subject and the scorer) by comparison with a quantized gray-level card. Entropy was estimated as the ratio of the number of guesses to the number of tiles. For a picture of a girl (2.5 bits/tile) and a picture of sailor (3 bits/tile), the redundancy estimates were 60 and 74%, respectively. In another study, Tzannes *et al.*[15] used the same measure of entropy for a 50 × 50 pixel lunar-surface photograph quantized to 8 levels. Two subjects were familiarized with samples of images from a class of lunar-surface photographs. For the two subjects, the redundancies were 39 and 56%. It is shown below that the measure of entropy used in these studies typically underestimates the lower bound on redundancy.

In this study we extend previous work in several ways. Computer graphics makes it easy to improve the guessing game, over previous studies, by using interactive substitution. Here, the observer can see the results of a particular choice before making a commitment. This makes it reasonable to use higher spatial resolution and more gray levels. The technique also promises to be a useful tool for future studies of the predictability of image features that are not pixel based. The performance of several simple nearest-neighbor models are compared with human prediction per-

formance. One of these nearest-neighbor models does well when the image is relatively noise free but breaks down for images that have a large fraction of deleted pixels.

METHOD

The experiment was set up as follows. Eight pictures were digitized to 128 × 128 pixels.[17] The pictures were a close-up of foliage (leaves), a stream in a woods (woods/stream), a cityscape consisting of skyscrapers (city), a woman's face (face), Half-dome at Yosemite Park (half-dome), a picture of four elderly people in a shack (four elders), a man's face (hbb), and a Gaussian pseudofractal image (fractal) with a power spectrum corresponding to a fractal dimension of 2.5 and a rms contrast of 31%.[18] Pictures with large areas of open sky or other regions of uniform gray level were not used. Each pixel subtended a 10 min × 7 min rectangle at the eye. The gray-level histogram was stretched to full range (0 to 255 gray levels), and then the gray-level scale was quantized to 16 levels (4 bits). The dimmest and brightest pixels were 0.3 and 34 nits, respectively. The effective gamma of the display was 1.3. The alphabet or basis set consisted of these quantized pixels. The reason for using only 16 levels was that this was judged to be the right compromise between having enough gray levels for image intelligibility and not so many as to complicate the results by making the viewer unable to discriminate contrast. When 5 bits/pixel were used, it was difficult in some instances to discriminate one gray level from a nearby value in the picture. At the other extreme, binary pictures are often perceptually difficult to interpret. Under certain conditions, 3-bit/pixel quantization is adequate for recognition, so 4-bit/pixel quantization was about right.[19] However, both spatial and gray-level quantization alter the statistical structure of a class of pictures in a way that may make it difficult to generalize redundancy to other quantizations (see the Discussion section).

Before the observer was allowed to see the picture, a predetermined fraction of the 16,384 pixels was deleted (Fig. 1). For observer DJK, deletion was defined as above. For observer DCK, deletion meant setting the gray level to zero. The observer's task was to set the level of a deleted pixel to what it had been before deletion. To do this, the observer requested the computer (by pushing the right-hand button of a mouse) to show which pixel was to be reset. This pixel would blink. When ready, the observer would try out various gray levels from the palette shown beneath the picture. The computer painted the pixel with the gray level behind the cursor cross hairs on the palette. When satisfied with the choice, the observer would click the left-hand button. If the choice was right, the observer was informed of that fact and went on to the next pixel. If the choice was wrong, the observer kept guessing until he got it right. A marker was placed on the palette indicating wrong choices, so that the observer would not pick those again. Because there were only 4 bits/pixel, the maximum number of guesses for a pixel was 16. Several deletion levels were used, but most data were collected with only 1% deleted. There were 100 trials for each percent deletion. Both observers had normal acuity. The author was one of the observers.

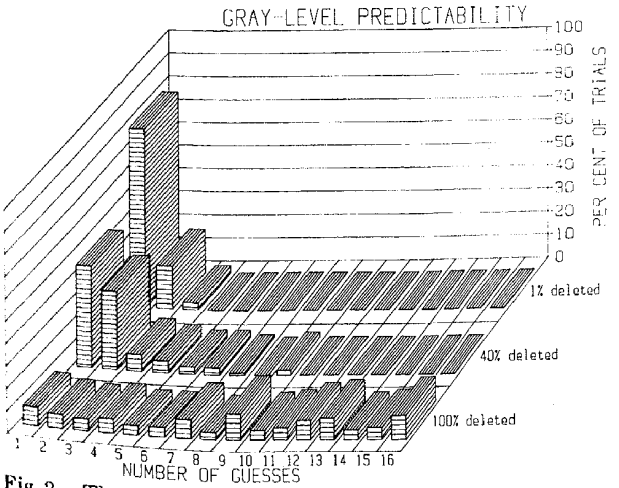


Fig. 2. The percentage of trials for which the observer got the right answer as a function of the per cent deleted and the number of tries (1 to 16) for observer DJK for the picture of the man's face (hbb).

RESULTS

Figure 2 shows the percentage of trials in which the observer guessed the right answer as a function of the percent deleted and the number of tries for observer DJK for the picture of the man's face (hbb) in Fig. 1. When 100% of the pixels were deleted, the observer had no clue about what to guess, and to guess correctly took anywhere from 1 to 16 tries for a given pixel on a given trial.[20] At the other extreme, when only 1% of the pixels were deleted, the observer guessed the correct gray level on the first guess 78% of the time. More than three guesses were never required.

Shannon[12] showed that upper and lower bounds on the entropy, F, can be calculated from predictability data such as those in Fig. 2. If q_i^N is the proportion of trials in which the observer guessed right on the ith try, the bounds are given by

$$\text{upper bound} = -\sum_{i=1}^{2^m} q_i^N \log_2 q_i^N, \tag{3}$$

$$\text{lower bound} = -\sum_{i=1}^{2^m} i(q_i^N - q_{i+1}^N)\log_2 i. \tag{4}$$

For example, q_i^N is the value of the height of the bars in Fig. 2 divided by 100. N is the number of known or undeleted pixels. The sum is taken over the 2^m gray levels for a given N. The upper entropy bound provides an upper limit on the average number of bits per pixel required to encode the class of pictures considered with an arbitrarily small error. The lower bound is provisional in that it is guaranteed to be valid only if the predicting device is ideal. Ideal prediction means choosing, for the first guess, the mode of the probability of the pixel gray-level distribution conditional on the known pixel values. If the first guess is wrong, the second guess should be the next most probable, and so forth. Because of this, the lower entropy bound is of limited meaning. It should be made clear that entropy is an ensemble statistic. In this experiment, the pixel gray-level predictability data are averaged over space for a single picture. Ergodicity is assumed, and ensemble averages are replaced by spatial averages. In two other pixel-based studies, entropy has been estimated by the ratio of the number of guesses to the number of trials.[14,15] However, this estimate, in all cases

Table 1. Upper Entropy Bounds Estimated from Gray-Level Predictability

Observer or Predictor	Upper Entropy Bound (bits/pixel) for the Following Image									
	Leaves	Woods/Stream	City	Face	Half-Dome	Four Elders	hbb	Fractal	Average	SE
DJK	2.06	1.34	1.70	0.96	1.30	1.05	0.89	1.27	1.32	0.13
DCK	2.23	1.66	1.00	1.20	1.47	1.61	1.20	1.38	1.47	0.13
Third-order serial entropy	2.59	2.38	2.44	1.64	2.24	2.19	1.55	1.79	2.15	0.14
Mode predictor	2.70	1.97	2.52	1.36	2.20	2.04	1.27	1.83	1.99	0.17
Average predictor	2.14	1.73	2.00	1.55	1.90	1.69	1.16	1.53	1.71	0.10
Median predictor	2.14	1.45	1.15	1.05	1.68	1.47	1.03	1.38	1.42	0.12

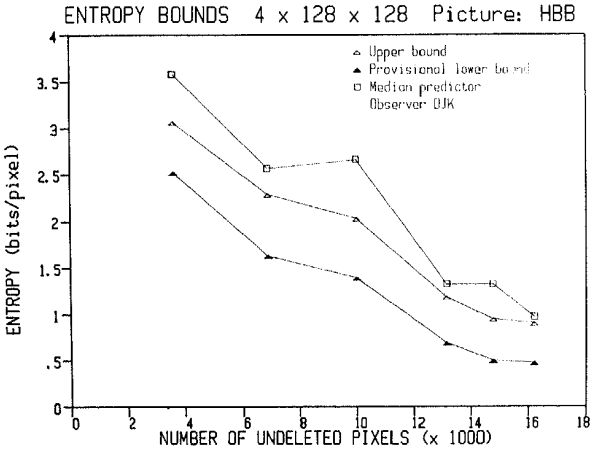


Fig. 3. Upper and lower entropy bounds, in bits per pixel, are shown for the picture of the man's face as a function of the number of undeleted pixels, that is, the number of known pixels for observer DJK (triangles). Upper entropy bounds are also shown (squares) for a median predictor that chooses the second brightest gray-level of the four nearest neighbors as its first guess (see the text for more details).

studied here, produces entropy estimates higher than that of Eq. (3). For example, the data for the eight pictures from observer DCK (Table 1) give a mean ratio of guesses required to trials of 1.7 bits/pixel, an average of 0.23 bit too high.

Figure 3 shows upper and provisional lower entropy bounds, in bits per pixel, for the picture of the man's face (hbb) as a function of the number of undeleted pixels. As more and more pixel information becomes available to the observer, the predictability increases, and the entropy bounds decrease. As one would expect, with all the pixels deleted (left-hand side of the graph), we have upper-bound estimates of about 4 bits/pixel (not shown) and, with 10,000 undeleted pixels, 2 bits/pixel. By the time most of the original pixel values are available (right-hand side of the graph), the estimated entropy is no more than 0.9 bit/pixel. The second observer, DCK, produced similar results. The most interesting entropy bounds are the small values on the far right-hand side of the graph. They can be interpreted as estimates of upper and lower entropy bounds for the class of $128 \times 128 \times 4$ bit natural images with no deleted pixels. Table 1 shows the estimated upper entropy bounds in bits per pixel for the eight pictures with only 1% of the pixels deleted for the two observers. Also shown in Table 1 are the results for three simple nearest-neighbor predictors and the third-order serial entropy (discussed below). The average upper entropy bounds across the pictures are 1.32 and 1.47 bits/pixel for observers DJK and DCK, respectively. The values range from slightly more than 2 bits/pixel to slightly less than 1 bit/pixel. The average provisional lower bounds were 0.78 (standard error of 0.1) and 0.95 [standard error (SE) = 0.09] bits/pixel for DJK and DCK, respectively.

It is convenient to measure entropy relative to the number of bits per pixel required if there were no statistical dependencies among pixels and if each gray level were equally likely. Recall that redundancy is defined as 1 minus the ratio of the actual entropy to the maximum entropy (which is 4 bits/pixel in our case). The average lower and (provisional) upper bounds on redundancy for observer DJK were 67 and 81%, respectively. For observer DCK, the average lower and upper redundancy bounds were 63 and 76%, respectively. If DJK is an ideal gray-level predictor, then, allowing for sampling error, 81% is an actual upper bound on redundancy for the class of $4 \times 128 \times 128$ bit images. However, it is not clear how close people are to being ideal predictors. We have calculated the actual redundancy for the class of Gaussian pseudofractals from which the fractal picture was drawn to be 69%.[21] The estimated upper and lower redundancy bounds based on predictor DJK are 81 and 68%, respectively. In this case, the lower bound is close to the actual redundancy, and the provisional upper bound is conservative.

DISCUSSION

The processes of spatial averaging and quantization used to generate the $128 \times 128 \times 4$ bit pictures will, in general, alter the conditional probabilities. This makes it difficult to generalize the redundancy estimates obtained to other quantization levels.[22] However, the processes of spatial averaging and quantization are inherent to the degradation of an image by any optical device. Even under the best of conditions, because of the particle nature of light, an ideal sensor effectively quantizes the range of light levels that it receives according to Poisson statistics. Further, because of the wave nature of light, diffraction limits the spatial resolution, causing spatial averaging, and thus reduces the effective number of pixels. In addition, biological eyes spatially quantize images because of discrete receptor spacing, and intensity quantization occurs because of stochastic neural codes. Consider a picture that subtends a 1 deg × 1 deg square at the eye and is on for 33 msec, with a luminance range of 1–10 nits (at 555 nm) and viewed with a pupil size of

3 mm. It is effectively quantized to about 125 × 125 × 4 bits, about the same quantization used in these experiments.[23] The problem of spatial resolution used in a predictability experiment would be of greater importance if we wanted to estimate the entropy of a more narrowly defined source, for example, a collection of human faces viewed from 1 m with a specified imaging system. Here, the spatial scale would be important. On the other hand, a property of many natural images is scale invariance.[18]

We can use the redundancy bounds to answer the somewhat whimsical question, How many meaningful pictures could be presented on a 128 × 128 × 4 bit screen? To the extent that the reader is willing to believe that the pictures used were a representative sample of the class of natural images (and any natural image is meaningful), an upper entropy bound of about 1.4 bits/pixel, that is, 65% redundancy, translates into not more than 10^{6905} ($2^{1.4\times128\times128}$) natural pictures of a total possible number of $10^{19,728}$. Because of the coarse quantization, these should all be discriminable, albeit with painstaking scrutiny. This is a substantial reduction over the number of possible pictures but, of course, still a ridiculously big number.

There are at least two ways in which predictability measures may be useful for understanding human vision. One approach is to understand what limits human performance at pixel gray-level predictability. Here, the goal is to seek models that can account for human performance. A second method is to measure predictability for alternative basis elements. The goal, in this case, would be to search for a minimal but perceptually complete scheme for describing images, a scheme that yields chance performance at the predictability test. If found, this scheme would provide a better answer to the question of the number of meaningful pictures.

To pursue the first direction, Table 1 shows results from three simple predictors that come close to matching human performance under some conditions. Table 1 shows mode, mean, and median predictors that use gray-level data from the nearest neighbors. For comparison, third-order entropies have been directly computed from adjacent pixels along a horizontal raster. This calculation makes use of the gray levels of the two previous adjacent pixels.[24] The average third-order entropy is 2.15 bits/pixel.[25] The mode predictor compiles a histogram based on the gray levels of the eight nearest neighbors and then guesses the most-frequent level, the next-most-frequent level, and so forth. Because there are more possible levels than neighbors, this predictor makes random guesses once it reaches the zero entries. Its average upper bound is 1.99 bits/pixel. The average predictor uses the quantized average over the four nearest neighbors as the first guess. If wrong, it chooses the next-dimmer value. If this is wrong, it chooses the next-brighter value above its initial guess and continues alternating until it is correct. The average upper entropy bound for the average predictor is 1.71. The median predictor resembles the average predictor except that its first guess is the third brightest gray level of the four nearest neighbors. The usual convention of taking the mean of the two middle values yielded slightly poorer prediction. The median predictor produces an upper entropy bound of 1.42 bits/pixel, which is quite close to the estimates based on data from the human predictors. The median predictor seems to cope with edges better than the average predictor. At first glance, the close match of the median predictor's performance to that of human predictors suggests that it may be a good model of human gray-level prediction. Although this may be true for relatively noise-free images, Fig. 3 shows that when a large fraction of pixels is deleted, human observers do better than the median predictor, probably because of the ability of humans to make use of long-range information. We have collected some data on the predictability of gray levels as a function of the size of the neighborhood. For human subjects, there is little improvement in performance beyond the eight nearest neighbors. Observer DCK scored average upper-bound entropy estimates of 1.43 [standard error (SE) = 0.1] and 1.47 (SE = 0.19) with 8 and 1224 nearest neighbors visible, respectively. Although the information from the nearest neighbors may be sufficient to match human performance in the relatively noise-free case, it is not necessary for better-than-chance prediction. When a 3 × 3 black annulus surrounded each of the 100 pixels to be estimated, the upper entropy bound for observer DJK (picture hbb) rose from 0.89 to 2.43 bits. In contrast to the first-order entropy of 3.76, this is substantially better.

The fact that a simple median predictor does so well when the image is relatively noise free suggests that more-sophisticated predictors may give us even better estimates of redundancy. Communication-engineering research has produced image-compaction algorithms that can reduce 8-bit/pixel images to less than 1 bit/pixel.[4] Much of this work depends on the high correlations between nearby pixel gray levels exemplified by Gaussian Markov-source models.[26,27] In addition to reducing redundancy, these algorithms achieve part of their success by tolerating distortion to which human observers are insensitive. Thus it is difficult to predict the success of these algorithms when adapted to the gray-level predictability test. One error-free scheme presented in Ref. 28 yielded 1.4, 2.1, and 3.8 bits/pixel for 4-bit original pictures of a girl's face, the cameraman photograph, and a crowd, respectively. Although difficult to compare directly, these results correspond to an average redundancy of 39%, somewhat lower than the 65% obtained in this study.

Rather than trying to account for gray-level ability, we could ask, How predictable are alternative basis elements? It is clear that the simple pixel-based description used here is analogous only up to receptor coding. Studies of human and primate vision have suggested minimal but perceptually complete basis sets to describe images and thus to allow for distortion tolerance.[29-31] However, our ability to discriminate two images does not necessarily imply that we can replace a missing basis element in a natural image. This suggests the following challenge of the predictability test. Given a minimal but perceptually complete representation to describe the class of natural images, how predictable is a missing element? If a missing part is very predictable, giving a high redundancy figure, then there is higher-level coding that needs to be understood. On the other hand, if missing components are hard to predict, that is, if one possibility is as good as another, we have arrived at a code, not necessarily unique, that embodies the observer's knowledge of natural images.

ACKNOWLEDGMENTS

Some of these results were reported at the 1986 meeting of the Association for Research in Vision and Ophthalmology

[Invest. Ophthalmol. Vis. Sci. Suppl. **27**, 226 (1986)] and at the Rank Prize Funds Symposium, Cambridge, England, December 1986. This research benefited from discussions with David Knill, Horace Barlow, and Stuart Geman. E. Adelson pointed out that the median predictor might do better than an average predictor. David Knill calculated the entropy of the pseudofractal. Vicki Michela and Peter Spiro helped with programming and data collection. This research was supported by National Institutes of Health grant EY06071-01, National Science Foundation grant BNS 87-8532, and National Institutes of Health Biomedical Research Support grant PHS2 S07 RR07085.

REFERENCES AND NOTES

1. F. Attneave, "Informational aspects of visual perception," Psych. Rev. **61**, 183–193 (1954).
2. H. B. Barlow, "Sensory mechanisms, the reduction of redundancy, and intelligence," in *Proceedings of the National Physical Laboratory Symposium on the Mechanization of Thought Processes* (H. M. Stationery Office, London, 1959), No. 10, pp. 535–539.
3. H. B. Barlow, "Perception: what quantitative laws govern the acquisition of knowledge from the senses?" in *Functions of the Brain*, Wolfson College Lectures, C. Cohen, ed. (Clarendon, Oxford, 1985), pp. 11–43.
4. A. N. Netravali and J. O. Limb, "Picture coding: a review," Proc. IEEE **68**, 336–406 (1980).
5. G. Buchsbaum and A. Gottschalk, "Trichromacy, opponent colour coding and optimum information transmission in the retina," Proc. R. Soc. London Ser. B **220**, 89–113 (1983).
6. S. Laughlin, "A simple coding procedure enhances a neuron's information capacity," Z. Naturforsch. **36**, 910–912 (1981).
7. M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: a fresh view of lateral inhibition," Proc. R. Soc. London Ser. B **216**, 427–459 (1982).
8. W. A. Richards, "A lightness scale from image intensity distributions," Artificial Intelligence Memo 648 (Massachusetts Institute of Technology, Cambridge, Mass., 1981).
9. L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," J. Opt. Soc. Am. A **4**, 519–524 (1987).
10. E. Saund, "Dimensionality-reduction using connectionist networks," Artificial Intelligence Memo 941 (Massachusetts Institute of Technology, Cambridge, Mass., 1987).
11. G. W. Cottrell, P. Munro, and D. Zipser, "Image compression by back propagation: an example of extensional programming," Institute for Cognitive Science Rep. 8702 (University of California, San Diego, La Jolla, Calif., 1987).
12. C. E. Shannon, "Prediction and entropy of printed English," Bell. Syst. Tech. J. **30**, 50–64 (1951).
13. T. M. Cover and R. C. King, "A convergent gambling estimate of the entropy of English," IEEE Trans. Inf. Theory **IT-24**, 413–421 (1978).
14. J. R. Parks, "Prediction and entropy of half-tone pictures," Behav. Sci. **10**, 436–445 (1965).
15. N. S. Tzannes, R. V. Spencer, and A. J. Kaplan, "On estimating the entropy of random fields," Inf. Control **16**, 1–6 (1970).
16. J. G. Simon, "Redondance et complexité," Psychol. Belg. **12**, 255–263 (1972).
17. The experiment was run with an IBM/XT with a Number Nine Revolution graphics board, a Microsoft Mouse, and a Tektronix GMA 301 display. The programs were written in Lattice C with the HALO subroutine library. Pictures were digitized with an 8-bit/pixel Chorus Data Systems digitizer and an RCA camera.
18. B. B. Mandelbrot, *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).
19. J. V. Gaven, Jr., J. Tavitian, and A. Harabedian, "The informative value of sampled images as a function of the number of gray levels used in encoding the images," Photogr. Sci. Eng. **14**, 16–20 (1970).
20. If the gray levels are uniformly distributed, then the histogram for the required number of correct guesses for each gray level is uniformly distributed as well.
21. The entropy for the pseudofractal Gaussian image was calculated by using a discrete form of Shannon's formula for the entropy of a continuous Guassian source. [See C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 371–428 (1948)].
22. Relative entropy is F_n/m. The median nearest-neighbor predictor gave upper bounds of relative entropy, which increased by an average of 0.56 for each 1-bit/pixel increase in quantization from 2 to 6 bits/pixel, for leaves, four elders, and hbb. The relative entropy bound leveled off between 6 and 7 bits/pixel, probably because of the degradation in the original imaging process.
23. An eye's required capacity can be estimated roughly as follows. The photon flux at 555 nm, corresponding to 1 Td, is about 1.25×10^6 photons $\text{sec}^{-1}\,\text{deg}^{-2}$. If we assume a receptor diameter of about 0.008 deg, then 10 nits for 33 msec with a 3-mm pupil corresponds to about 150 photons. We assume an effective discretization of levels according to Poisson statistics, which separates levels by 1 standard deviation. There are then about 17 levels, or 4 bits, going from 150 to 15 photons. A 1 deg × 1 deg area divided by the receptor area is 125 × 125 pixels. The actual required capacity would be somewhat less than 125 × 125 × 4 bits, because of a nonrectangular modulation transfer function and nonuniform receptor spacing.
24. If i, j, and k represent the gray levels of three consecutive pixels, the third-order conditional entropy, $F(k|i, j)$, can be estimated from the relative frequencies of sequences of ijk and ij. If these frequencies are denoted by $p(i, j, k)$ and $p(i, j)$, respectively, then the third-order conditional entropy can be calculated from the difference between the third-order and second-order joint entropy:

$$-\sum_{i,j,k} p(i,j,k)\log_2 p(i,j,k) + \sum_{i,j} p(i,j)\log_2 p(i,j).$$

For comparison with nearest-neighbor predictors, it would be interesting to know the eighth-order entropies, based on the eight nearest neighbors. However, this would require gathering data on 16^8-dimensional histograms. This is prohibitive because of the number of samples, and hence the number of pictures, that would need to be measured and because of the computational demands.
25. W. F. Schreiber, "The measurement of third order probability distributions of television signals," IRE Trans. Inf. Theory, **IT-2**, 94–105 (1956). Schreiber measured third-order statistics, based on a raster scan, for a 6-bit/pixel image that gave a third-order entropy of 1.5 bits/pixel or redundancy of 75%. This picture was described as the least complicated picture. The average third-order redundancy for the 4-bit/pixel images measured here was 44%. It is difficult to compare these results directly because they depend on both the picture complexity and the quantization. However, estimates of entropy as a function of the degree of quantization typically show a decrease in redundancy as the number of levels increases.[22]
26. W. K. Pratt, *Digital Image Processing* (Wiley, New York, 1978).
27. H. C. Andrews, *Introduction to Mathematical Techniques in Pattern Recognition* (Krieger, Malabar, Fla., 1983).
28. R. C. Gonzalez and P. Wintz, *Digital Image Processing* (Addison-Wesley, Reading, Mass., 1977).
29. B. Sakitt and H. B. Barlow, "A model for the economical encoding of the visual image in cerebral cortex," Biol. Cybern. **43**, 97–108 (1982).
30. W. S. Geisler and D. B. Hamilton, "Sampling-theory analysis of spatial vision," J. Opt. Soc. Am. A **3**, 62–70 (1986).
31. A. B. Watson, "Efficiency of a model human image code," J. Opt. Soc. Am. A **4**, 2401–2417 (1987).