Pergamon

PII: S0042-6989(96)00002-8

Vision Res., Vol. 36, No. 17, pp. 2759–2770, 1996

Printed in Great Britain
0042-6989/96 $15.00 + 0.00

Modelling the Power Spectra of Natural Images: Statistics and Information

A. van der SCHAAF,*† J. H. van HATEREN*

Received 15 May 1995; in revised form 29 August 1995

Power spectra of an extensive set of natural images were analysed. Both the total power in a spectrum (corresponding to image contrast) and its dependence on spatial frequency vary considerably between images, and also within images when considered as functions of orientation. A series of probabilistic models for power spectra enabled calculating the information obtained from prior knowledge of parameters describing spectra. Most information is gained from contrast, $1/f^2$ spatial frequency behaviour, and contrast as a function of orientation. Variations in spatial frequency behaviour are relatively unimportant. For oriented contrast, a bandwidth of 10–30 deg is sufficient to obtain most information. Copyright © 1996 Elsevier Science Ltd.

Visual system Power spectrum Spatial frequency Contrast Orientation

INTRODUCTION

A major challenge for the early stages of the visual system is handling the huge data load it receives, while its information capacity is limited. Because the information capacity needed depends on the characteristics of the input, it is important to note that natural images are not random, but show a large degree of structure. This structure can be described by the statistics of the image source, and can be considered as prior knowledge. Therefore, a certain amount of the image data is predictable and thus redundant. As suggested by Barlow (1959), reducing some of this redundancy at an early stage of visual processing can help the visual system to optimize the transfer of information to later stages in the visual system. An appropriate processing scheme that reduces redundancy, i.e., removes predictable data, must obviously be carefully tuned to the statistics of the image source.

In this article we will concentrate on the redundancy associated with the second order statistics of images. This is most easily studied through the power spectrum of an image. The power spectrum is directly related to the autocorrelation of an image, which describes how closely related two points in an image are as a function of their distance and orientation. If we want to reduce redundancy as a valuable (pre)processing strategy, an adequate model for the power spectrum of natural images is needed. It has been found that power spectra of natural images tend to depend as $1/f^2$ on the spatial frequency f (Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992a; Tolhurst *et al.*, 1992; Ruderman & Bialek, 1994). This fact has been used successfully to explain and predict the processing early in the visual system of both insects and higher vertebrates (insects: Srinivasan *et al.*, 1982; van Hateren, 1992a,b; vertebrate retina: Bialek *et al.*, 1991; Atick & Redlich, 1990; human psychophysics: Atick & Redlich, 1992; van Hateren, 1993). Notwithstanding the success of this approach, which assumes $1/f^2$ as the model for natural power spectra, the adequacy of this model has not been systematically investigated with respect to its capacity to reduce redundancy. The main purpose of this article is to remedy this situation.

As a first step towards testing models for natural power spectra, we obtained an extensive set of carefully calibrated spectra of natural images, and confirm here several results reported in previous studies. Furthermore, we investigate how contrast and frequency behaviour depend on orientation. The experimental data are subsequently used to evaluate a range of simple models (similar to $1/f^2$) for their capacity to reduce second-order redundancy in images and for their capacity to estimate the contrast of an image. The latter can be utilized by the visual system for contrast normalization. Several of these models not only capture the frequency behaviour of the spectra, but also their orientation behaviour. Just as $1/f^2$ served as a successful model in previous work on circular symmetric neural preprocessing (i.e. non-oriented), we expect that the most successful models we propose here can be used as a tool for analysing the neural processing of orientation by visual systems. Finally, we expect that the methods developed here for analysing variations between the power spectra of different images can serve

* Department of Biophysics, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands.

† To whom all correspondence should be addressed [*Fax* +31 (50) 363 4740].

as a guide for investigating how localized forms of frequency spectra vary within images.

METHODS

Image acquisition

We recorded a sample set of 276 images, using a CCD video camera (77RS, PCO Computer Optics, Kelheim, Germany) with a 16 mm Sony TV lens. The video signal was digitized with a PC-based frame grabber (Data Translation DT2855) at a resolution of 768 × 512 pixels and a depth of 8 bit. Images were temporarily stored on the PC, which was mounted on a small trolley and powered by a battery with 220 V-converter. A HP workstation was used for further analysis.

The images were recorded in a number of different outdoor environments (woods, fields, parks, residential areas), at various times of the day, in various seasons, and in various types of weather (sunny, overcast, foggy, rainy). We tried to ensure the diversity of the sample set by taking images of widely different objects and scenes, at varying distances (with a minimum of 30 cm) and varying elevations (keeping the left–right axis of the camera approximately parallel to the horizon). The images were not taken completely at random, however, since we had a bias towards choosing scenes with at least something of interest to see, thus excluding, e.g., images consisting exclusively of clear sky. We assume here that this is a natural bias, although we can not quantify how representative our set is for the images encountered by a particular visual system during its owner's natural behaviour in its natural habitat.

Calibration

We took great care to properly calibrate the measurements. We used the camera with gamma-correction off and automatic gain control off, because either of these compromised the linearity of the recording system. The exposure was controlled manually by adjusting the electronic shutter speed of the camera, or the aperture of the lens (F = 5.6–16). Using calibrated neutral density filters, we found a linear intensity-response relation for the recording system, with a small and stable offset (dark current of the CCD), which was subtracted from the response. Variation in pixel sensitivity was less than the quantization step, and neglected. The decline of image intensity towards the edges, caused by the camera lens, was less than 20%, and was corrected to within 2.5%. The aspect ratio of the recorded pixels (0.946 horizontal to vertical ratio) was corrected during the calculation of the power spectra (see below). Geometrical distortion caused by the camera lens was less than 1%, and neglected. The maximum amplitude of the jitter in the horizontal synchronization of the video signal was less than 1 pixel, and also neglected.

We measured the response of the entire recording system to a light source, consisting of a pinhole of 30 μm diameter illuminated from behind, and viewed at a distance of 5.6 m. The radiance of this light source was of the same order of magnitude as the radiance encountered in outdoor scenes. We took care not to saturate the camera by keeping the response peak at approximately half the total camera range. The angular extent of the light source was less than 1% of the angular extent of a single pixel of the CCD, and the source can thus be considered as a true point source. From the resulting point spread function we calculated the power spectrum, after masking the noise present in the digital image at non-illuminated parts of the CCD far from the illuminated area. We found that this power spectrum slightly depends on the particular position of the image of the point source relative to the CCD pixel grid. Therefore, we recorded 100 point spread functions evenly distributed over the total extent of a single pixel, and averaged the corresponding power spectra. This average, Γ, was used to correct the power spectra of our set of images. The rationale of this calibration is that it will yield, on average, a flat power spectrum when a point light source is recorded, as theoretically expected from a point source.

To avoid saturation of the camera response at bright spots in the images, we took the images (when necessary) at a low exposure level. The average mean luminance over all images was approximately 40 quantization steps (of a total of 256). The drawback of this procedure is the increase of quantization noise. Assuming that the quantization noise is white, however, we calculated that the quantization noise level in the spectra is at least an order of magnitude below the lowest part of the power spectra, even in extreme cases where the mean luminance of the image was set at only 10 quantization steps. Therefore, we can neglect quantization noise. The little truncation that is still present in some of our images due to saturation of the camera response is unlikely to have a significant effect on the form of the power spectra (measured by the $1/f$-exponent, see Results). We conducted three experiments to test this. First, we made multiple recordings of six scenes with different exposure levels, such that the quantization and truncation were different. The resulting $1/f$-exponents change for different exposure levels, on average, by 4% in magnitude. As the change was in either direction, the change of the average $1/f$-exponent was less than 1%. In a second experiment, we deliberately truncated all recorded images to twice their mean, and found that the average $1/f$-exponent of the truncated images differed less than 1% from the average $1/f$-exponent of the original images. In the last experiment we investigated the truncation and quantization of generated $1/f$-noise. We truncated the generated noise between −2 and +1 standard deviations from its mean, and quantized the range between the truncation levels into 256 levels (8 bit). We found that the $1/f$-exponent of the resulting noise differed less than 1% from the original. This was also found when the number of quantization levels was only 32.

Calculation of power spectra

We took a rectangle of 541 × 512 pixels out of each digitized image. Because of the aspect ratio of the pixels,

this rectangle corresponds to a 512 × 512 image with square pixels. Boundary effects were reduced by applying a Kaiser–Bessel window (with window parameter 2, see Harris, 1978). This window (w) is a weight function that is maximal in the centre of the image, and gradually decreases to zero towards the edges. The contours of the window are approximately circular, so that it does not introduce an orientation bias. To avoid leakage in the spectral transformation from the DC-component of the image, the weighted mean intensity was subtracted before applying the window. The weighted mean intensity (μ) is defined as

$$\mu = \frac{\sum_{(x,y)} l(x, y)\, w(x, y)}{\sum_{(x,y)} w(x, y)}, \quad (1)$$

where $l(x, y)$ denotes the intensity of a pixel at spatial coordinates (x, y) in the 541 × 512 cut-out range. After subtraction of the weighted mean, the images were normalized, windowed, and subsequently Fourier transformed. The Fourier transform (F) is then given by

$$F(u, v) = \sum_{(x,y)} \frac{l(x, y) - \mu}{\mu}\, w(x, y) e^{2\pi i(ux/541 + vy/512)}, \quad (2)$$

where u and v are spatial frequency coordinates in the horizontal and vertical direction, respectively, expressed in cycles per image. The unit of spatial frequency is now equal for the horizontal and the vertical direction, because the aspect ratio of the image boundary equals 1. We calculated $F(u, v)$ for integer values of u and v, using a discrete Fourier transform. The two-dimensional spatial frequencies are in this study usually presented in polar coordinates (f, φ), with $u = f \cos \varphi$ and $v = f \sin \varphi$, where f denotes the (absolute) spatial frequency, and φ the orientation.

For this study, only spatial frequencies (f) up to 127 cycles per image are used for further analysis. Higher frequencies suffer from noise, low modulation transfer of the recording system at these frequencies, and aliasing. At 127 cycles per image, the signal (i.e., before correction) is approximately suppressed three-fold by the modulation transfer function, while any aliasing products are at least four times more suppressed than the non-aliased signal.

Finally, we calculated the power spectrum S of each image as the square amplitude of the Fourier transform:

$$S(u, v) = \frac{1}{L} \frac{|F(u, v)|^2}{\Gamma(u, v)}, \quad (3)$$

with Γ the correction factor mentioned above, and L the number of pixels in the image (L = 541 × 512).

RESULTS AND MODELLING

As described in the previous section, we obtained power spectra of 276 natural images. In this section we will investigate the statistics of several properties of these power spectra, such as their total power (related to the contrast of the images), and their shape as a function of spatial frequency and orientation. As the statistical parameters of individual images vary widely, it is not *a priori* clear how important these parameters may be for further processing. Therefore, we investigate a series of probabilistic models for the power spectra, which enable us to estimate the amount of information that the various statistical parameters provide about the spectra.

Statistics

Before discussing the statistics of individual power spectra, we will take a brief look at the average power spectrum of the complete set. Figure 1(A) shows the spectral power, averaged over orientation, as a function of spatial frequency (f). On a log–log scale, the data points lie approximately on a straight line, which means that spectral power and spatial frequency are related by a power law (see also Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992a; Tolhurst *et al.*, 1992; Ruderman & Bialek, 1994). Figure 1(B) shows the spectral

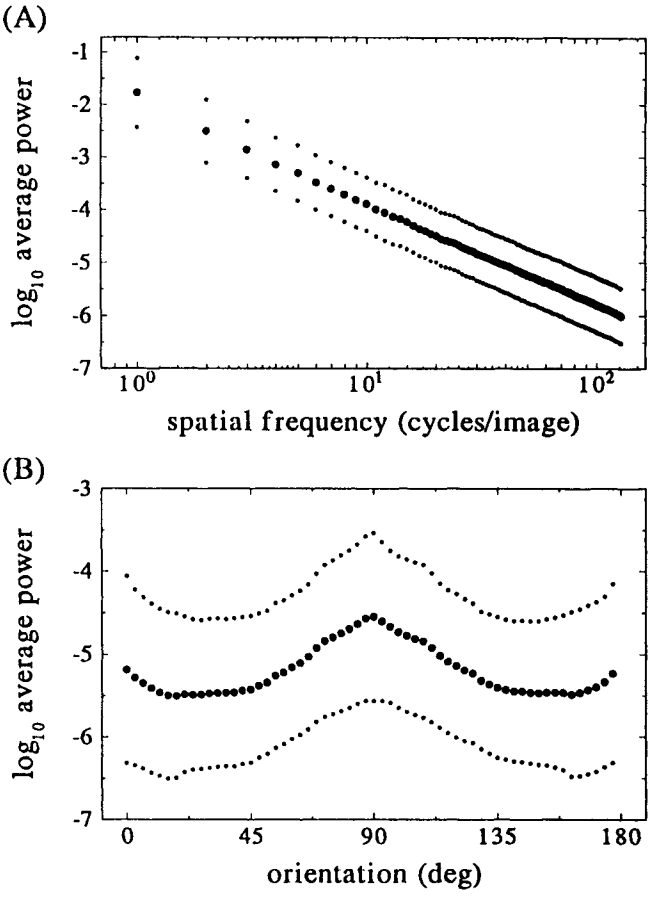

FIGURE 1. (A) The average power as a function of spatial frequency. Large dots show the average over the complete set of the logarithm of the circularly averaged individual power spectra. Small dots give the standard deviation from the average of corresponding plots of individual images. (B) The average power as a function of orientation. Here the power spectra are first averaged over spatial frequency, then the logarithm is taken, and finally the plots are averaged over the complete set (large dots). Small dots as in (A).

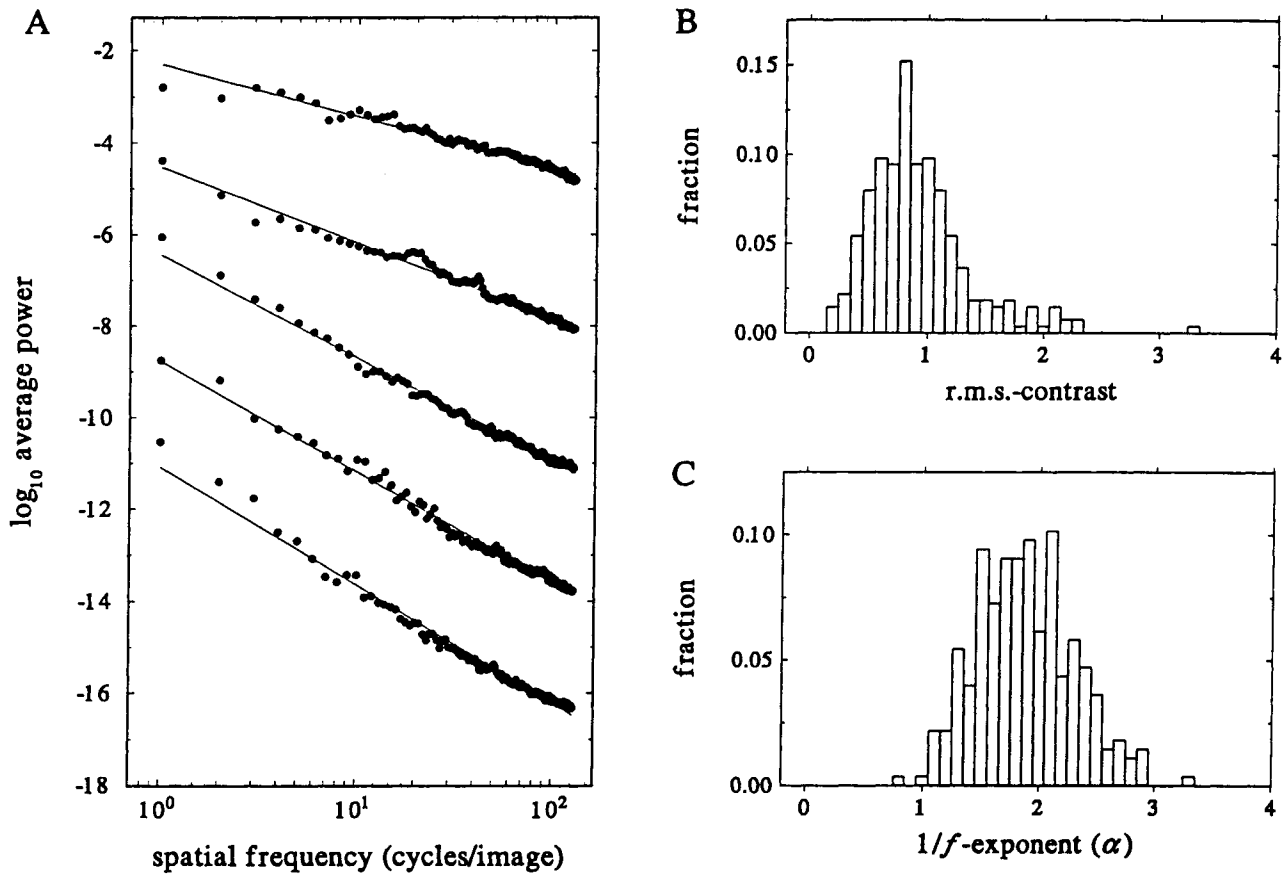

FIGURE 2. (A) Example traces of the power spectra of five individual natural images. Dots show the logarithm of the circularly averaged power spectrum as a function of spatial frequency. The lines show the fits of the $1/f^{\alpha}$ model. The scaling of the vertical axis belongs to the top trace. For clarity, the lower traces are shifted −2, −6, −8, and −10 log-units, respectively. (B) Distribution of r.m.s.-contrasts for the entire set of 276 natural images. (C) Distribution of 1/*f*-exponents (α) for the entire set.

power, averaged over spatial frequency, as a function of orientation (φ). The spectral power is not isotropically distributed, but is larger at horizontal and vertical orientations than at other orientations (see e.g. Switkes *et al.*, 1978). This result is approximately the same for subsets of images showing exclusively man-made objects or exclusively scenes of nature. The effect is not caused by the square geometry of the images, because we apply an almost circular window (see Methods). As a further test, a calculation of the power spectrum of an image with a single edge in an oblique direction, shows that only 1% of its power leaks into horizontal or vertical directions, an effect much smaller than that shown in Fig. 1(B). The effect is also not due to the square pixels, because it is absent in a set of images without dominant orientations (mostly images taken from soil covered with leaves and twigs with the camera pointed vertically at random orientations). The small dots in Fig. 1(A, B) show the standard deviation of the corresponding plots of individual images in the set.

Although the average power spectrum is quite smooth, spectra of individual images are much more irregular and different in shape, as shown by the large standard deviations in Fig. 1(A) and the example traces in Fig. 2(A). The individual power spectra differ not only in shape, but also in the total power of the spectra. This total power is related to the root-mean-square (r.m.s.)-contrast of individual images. The r.m.s.-contrast equals the standard deviation of the intensity values of all pixels in the image divided by the mean intensity. The square of the r.m.s.-contrast is proportional to the total power in the spectral domain under consideration (see Appendix A). Thus the r.m.s.-contrast has a simple and intuitively clear interpretation in both the spatial and the spectral domains, and we will, therefore, use it here to illustrate the variability of contrast. Different definitions of contrast that we will introduce below lead to similar distributions. Figure 2(B) shows the distribution of contrast values for all images. The r.m.s.-contrast varies substantially, with a mean of 0.92 and a standard deviation of 0.44.

Not only the r.m.s.-contrast, and consequently the total power, vary for individual images, but also the shape of the power spectrum. As shown in previous studies and by the almost straight line in Fig. 1(A), the spectral power, averaged over many images, varies approximately as $1/f^{\alpha}$ as a function of spatial frequency, with the 1/*f*-exponent, α, close to 2. If we instead inspect the spectra of individual images, examples of which are shown in Fig. 2(A), we still find that they are roughly proportional to $1/f^{\alpha}$. The root mean square error of the fit of the $1/f^{\alpha}$ model to the circularly averaged power spectra [see Fig.

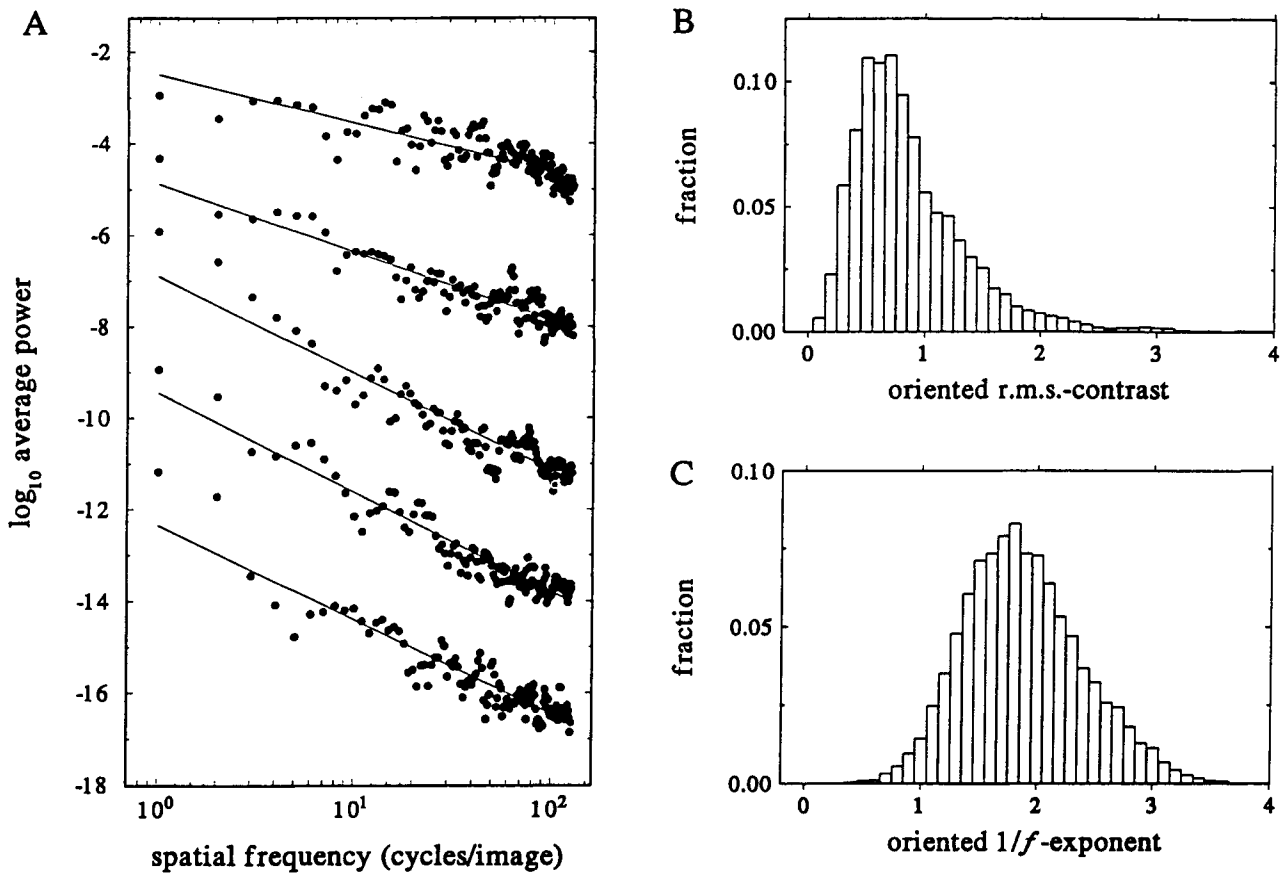

FIGURE 3. (A) Example traces of the power spectra of the same five individual natural images as in Fig. 2, but now at an orientation of 30 deg relative to the horizontal frequency axis. Dots show the logarithm of the power spectrum averaged over an angle of 3 deg as a function of spatial frequency. The lines show the fits of the $1/f^{\alpha}$ model. The traces are shifted vertically as in Fig. 2. (B) Distribution of oriented r.m.s.-contrasts for the entire set of 276 natural images and all orientations. (C) Distribution of oriented $1/f$-exponents (α) for the entire set and all orientations.

2(A)] is on average 0.13 log-units over the complete set. For the individual spectra, α varies considerably from image to image (Tolhurst *et al.*, 1992; van Hateren 1992a; Field, 1993). Figure 2(C) shows how α is distributed for our set of images. We found an average α of 1.88 and a standard deviation of 0.43. There is a correlation of +0.35 between the r.m.s.-contrast and $1/f$-exponent values. This means that steeper spectra tend to have more power. This relationship, however, is very weak and unlikely to be useful as prior knowledge.

The $1/f^{\alpha}$-model of the spectral power distribution is isotropic, i.e., the statistics are the same for every orientation. Figure 1(B) showed, however, that the average spectral power is anisotropically distributed over orientation. This is because there are, on average, more horizontally and vertically oriented structures in natural images than obliquely oriented structures (Switkes *et al.*, 1978). This applies to the average power spectrum of the complete set, but if we examine individual images, we find peaks of spectral power also at other orientations. For oblique orientations, these peaks are smoothed out when the spectra are averaged, but they are very characteristic for individual spectra, and partly account for their large variability [as illustrated by the large standard deviations in Fig. 1(B)].

In order to gain some insight into how spectral power varies as a function of orientation, we define here two additional statistical measures: the oriented r.m.s.-contrast and the oriented $1/f$-exponent. These measures are calculated for the individual spectra in the same way as the related isotropic measures, except that they are not calculated for the complete power spectrum, but only for pie-slices of the spectrum, with orientation φ and width $\Delta\varphi$ (see Appendix A). For most of the present analysis we use 60 pie slices for the 180 deg orientation range ($\Delta\varphi$ = 3 deg), but the consequences of varying $\Delta\varphi$ are investigated in the Discussion. The oriented r.m.s.-contrast is related to the total amount of spectral power at a specific orientation, and the oriented $1/f$-exponent measures the fall-off of spectral power at that orientation. Figure 3(A) shows example traces of the same power spectra of Fig. 2(A) at an orientation of 30 deg to the horizontal frequency axis (with an orientation width of $\Delta\varphi$ = 3 deg). The traces are all roughly proportional to $1/f^{\alpha}$, but differ considerably in their total power (corresponding to the oriented r.m.s.-contrast) and steepness (corresponding to the oriented $1/f$-exponent). The r.m.s. error of the fit of the $1/f^{\alpha}$ model to the power spectra at a specific orientation, with an orientation width of $\Delta\varphi$ = 3 deg, [see Fig. 3(A)] is on average 0.32 log-

units over the complete set of images and all orientations. The distribution of oriented r.m.s.-contrasts and oriented 1/*f*-exponents for all orientations and all images of our set is given in Fig. 3(B,C). The oriented r.m.s.-contrast, with a mean of 0.88 and a standard deviation of 0.52, varies more than the non-oriented contrast. This variation has two components: the variation of non-oriented r.m.s.-contrast between images, and an additional variation between orientations within individual power spectra with a standard deviation of 0.28. Likewise, there is more variation in the oriented 1/*f*-exponent than in the non-oriented 1/*f*-exponent, with a mean of 1.88 and a standard deviation of 0.51, which means an additional variation between orientations within the individual power spectra with a standard deviation of 0.28. The correlation between the oriented r.m.s.-contrast and oriented 1/*f*-exponent values is +0.23; even less than for the non-oriented measures.

Towards a model of the power spectra

Both the r.m.s.-contrast and the 1/*f*-exponent can be viewed as parameters of (simple) models of the typical power spectrum of a natural image. The r.m.s.-contrast gives an indication of the height of the spectrum, and the 1/*f*-exponent of its frequency behaviour. Allowing these parameters to vary as a function of orientation leads to more refined models. How can we decide how good each model is, i.e., how well it describes the typical natural power spectrum? One possibility to evaluate this is to calculate the amount of information that the various models provide about individual spectra. The usefulness of this calculation is based on the following rationale: the more information a model provides about the spectra, the more effective it will be as a means to estimate the contrast and to reduce the second-order redundancy of an image. The capacity to estimate the contrast, i.e., the height of the spectrum, can be important for the following reason: as images vary widely in contrast, it gives the visual system the possibility to optimally use the dynamic range of its processing units through the use of contrast normalization. The capacity of a model to reduce redundancy is in fact the capacity to flatten the power spectrum (if used as an inverse filter for decorrelation), because the model estimates the shape of the power spectrum. Thus, it will be a useful strategy for the visual system to take such a model into account as an effective means to reduce redundancy. In fact, several such models have already been proven to be quite useful for predicting properties of the visual system (e.g., 1/*f*-exponent: Atick 1990, 1992; van Hateren 1992a,b, 1993; Bialek *et al.*, 1991; exponential autocorrelation, which is qualitatively equivalent to a $1/f^2$ power spectrum: Srinivasan *et al.*, 1982; contrast: Laughlin, 1983; orientational variation: Field, 1987, 1993, 1994; Daugman, 1989). Thus, we may hope that our present analysis will provide useful clues for further theoretical and experimental development along these lines.

The models that we propose are very simple, and describe only the rough form of the actual power spectrum. Knowing the model, we have information about the appearance of the power spectrum, but still some uncertainty remains. Analysis of the deviation of the models from the actual spectra results in a probability density function (PDF) for the power spectra, from which the uncertainty (entropy) is calculated. In general, a narrower PDF corresponds to a smaller uncertainty, and it signifies a model that would be very effective in flattening power spectra, i.e., reducing the second-order redundancy in the corresponding image. The reason that this works, is that the first-order redundancy in a power spectrum directly corresponds to the second-order redundancy in the corresponding image. To abolish an arbitrary offset, we compare each uncertainty with the uncertainty of a fixed reference model, which assumes a flat power spectrum with equal contrast for each image. This gives us the information that we have gained from adopting a specific model, which is related to the reduction of redundancy expected from using the model.

In evaluating the models, we assume that the power spectra can be described as the spectra of linearly filtered white noise [i.e., noise with a flat average power spectrum, see e.g. Papoulis (1965)]. Following standard linear system theory [Papoulis (1965)], the power spectrum (S) is separated into a product ($S = M\varepsilon$) of a deterministic modulation function (M) and a probabilistic residue (ε). That this approach is sensible is suggested by the fact that the variability of the spectral components is approximately proportional to the power of that component [see e.g., the errors in Fig. 1(A)]. The modulation function (M) corresponds to the square modulus of the filter's frequency response. It describes the rough appearance of the power spectrum, and it is, therefore, identified with the model. We will investigate here several modulation functions, inferred from what is known about the statistics of the power spectra. The modulation function is calculated separately for each power spectrum. M is generally a function of spatial frequency and orientation, and it may depend upon the statistics of individual power spectra, such as contrast and 1/*f*-exponent. The residue (ε) denotes the power spectrum of an individual white noise image and accounts for the remaining variation in S that is not described by the model. We require that the model M accounts for all known systematic variation of S as a function of (u, v) or image.

We determine the PDF of the residue empirically by making a histogram of the ratio of the measured power spectra and the model function M (see Appendix B for details). The PDF of the power spectra is thus fully specified by the given modulation function M (with the incorporated statistics) and the PDF of the residue. Information theory then allows us to calculate a measure of uncertainty associated with the PDF of the power spectra (Shannon & Weaver, 1949; see also Appendix B). The information that we have gained about the spectra by knowledge of a particular model can now be calculated as the decrease of uncertainty of the particular model, with

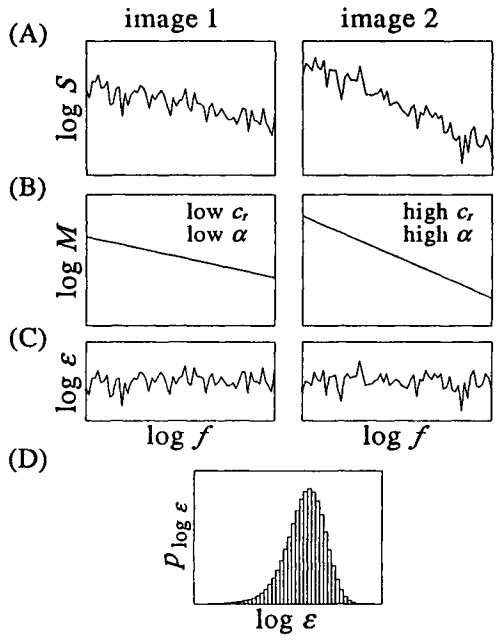

FIGURE 4. Example of uncertainty calculation for the power spectra of two hypothetical images and a model including r.m.s.-contrast (c_r) and the $1/f$-exponent (α). (A) One-dimensional traces of the power spectra (S) of the two images. (B) The modulation function, M, corresponding to (A). (C) The residue (ε) obtained from comparison of (A) and (B). (D) The PDF of the residue (p_ε, here given on a logarithmic scale), approximated by a histogram of all residue values. Together with the modulation function from (B), this determines the PDF of the power spectra, from which the uncertainty is calculated.

respect to the uncertainty of the fixed reference model (a flat power spectrum with equal contrast for each image).

We will illustrate the procedure we follow for modelling the spectra and calculating the informational measures by an example (Fig. 4). In this example, we investigate the power spectra (S) of two different hypothetical images. The spectral power along a single frequency axis is shown on a log–log scale in (A). The model we use here for describing the power spectra has two free parameters, the r.m.s.-contrast (c_r) and the $1/f$-exponent (α). The first image has a low contrast and a slow decrease of spectral power with spatial frequency. The second image, on the contrary, has a high contrast and a steep fall-off of spectral power. The statistics are incorporated into the modulation functions (M) of the two power spectra, which are chosen such that the total spectral power of the modulation function equals the square of the r.m.s.-contrast (c_r^2) and the fall-off of spectral power is proportional to $1/f^\alpha$. These modulation functions, shown in (B), are models of the power spectra shown in (A). The realizations of the residue [ε, see (C)] are obtained, for each spectral coordinate in both images, from the ratio of each power spectrum [from (A)] and the corresponding modulation function [from (B)]. The power spectra are now separated into a structured part, the modulation function in (B), and an unstructured part, the residue in (C). Generally, the distinction between structure and residue depends on the specific model, which includes in this example the r.m.s.-contrast and the $1/f$-exponent. All residue values of all spectral coordinates and both images are collected into a single histogram, which gives an empirical approximation of the PDF of the probabilistic residue (p_ε), as shown on a logarithmic scale in (D). Together with the modulation functions, this determines the PDF, p_s, of each power spectrum (see Appendix B). The p_s is different for each power spectrum, because the modulation function of each is different. From each p_s an associated uncertainty is calculated; the smaller this uncertainty, the better the model for describing the height (related to contrast) and shape (related to second-order redundancy) of the power spectrum. Finally, the entropy is averaged for the two images. Comparing this uncertainty with the uncertainty of other models quantifies the relative merits of the various models in terms of the information obtained from each model.

Different definitions of contrast

In several of the models we will discuss below, the square of the r.m.s.-contrast, c_r^2, is used as an estimate of the total power in the spectrum. This determines the height of the predicted (modelled) power spectrum relative to the actual (measured) power spectrum. Because the actual spectra behave approximately as $1/f^2$, the variation of the measured contrast is, in general, dominated by the variation of the power at only a few low frequency coordinates: although a $1/f^2$-power spectrum has equal power in each octave, the power at high frequency octaves is averaged over more frequency coordinates and thus less noisy than the power in low frequency octaves. Thus, if low frequency components happen to be larger than average, the contrast will be large, and the model prediction of the power spectrum may in fact be larger than the actual spectrum for all but the lowest frequencies. This raises the question of whether it is possible to devise another definition of contrast that is less sensitive to single frequency components, and that takes more frequencies into account. Ideally, we are looking for a simple definition of contrast that estimates the height of the spectrum such that it leads to an uncertainty that is as small as possible. Indeed, we found several alternative measures of contrast that perform better than r.m.s.-contrast as judged by our informational analysis. The most interesting are the whitened contrast and the log-contrast. The whitened contrast, or c_w follows from summing the power spectrum multiplied by f^2 (which makes the spectrum, on average, approximately flat); the log-contrast, or c_l, follows from summing the logarithm of each power spectrum (see Appendix A). In the discussion we will investigate how to interpret these various measures of contrast.

Models and information gain

We shall now give the results of the informational calculations performed on the full set of power spectra for several models incorporating various combinations of statistics. The information gain in Fig. 5 is given, relative

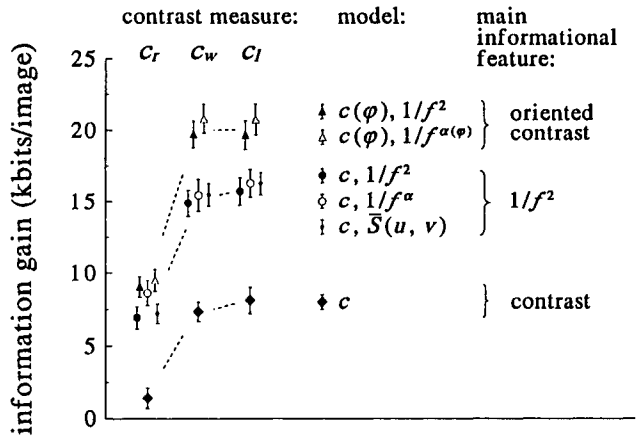

FIGURE 5. Information gain per model. The error bars are estimated from values obtained for several random subsets. Models resulting from different contrast measures are given in the sub-columns in the lefthand column, the middle column shows the various models, and the righthand column the main informational features.

to a minimal model that assumes flat power spectra with equal contrast for each image. The middle column in the figure shows the models we evaluated with the corresponding symbols. The simplest model (c, diamonds) assumes a flat power spectrum with the contrast adjusted to each image separately. As can be seen, the information gain is much larger for c_w and c_l than it is for c_r.

The next group of three models all take the $1/f^2$-behaviour of natural power spectra into account, which results in a much larger information gain. The first of these models (large solid circles) assumes a strict $1/f^2$-behaviour, with the contrast adjusted for each image. The second model (open circles) allows the $1/f$-exponent to differ for each image as well, similarly as in the example of Fig. 4. The third model (small solid circles) takes the average power of all images, $\bar{S}(u, v)$, as a model for the spatial frequency behaviour, again with the contrast adjusted for each image separately. This third model thus takes the anisotropy of the average power spectrum into account [Fig. 1(B)]. Interestingly, the three models all perform about equally well: the error bars show the accuracy of the information gain, as estimated from calculating the information gain for various random partitions of the set of images. Apparently, the $1/f^2$-behaviour that these models have in common is the decisive factor. Neither varying the $1/f$-exponent per image, nor taking the average anisotropy into account, are very effective for gaining information about power spectra of natural images.

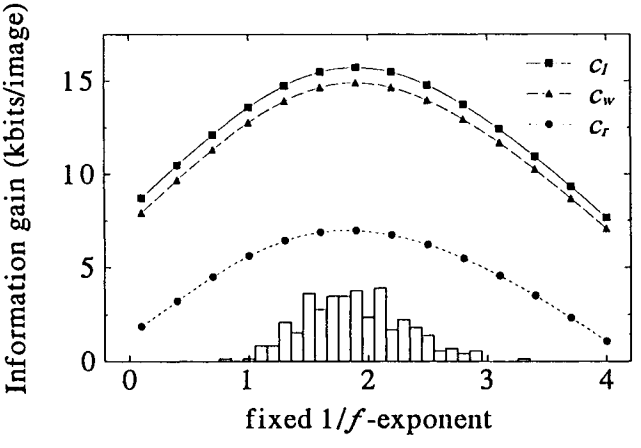

FIGURE 6. Information gain as a function of fixed $1/f$-exponent for models including three different contrast measures (c_r, c_w, and c_l) and a fixed $1/f^\alpha$-frequency behaviour. The distribution of $1/f$-exponents found for our set of natural images [from Fig. 2(C)] is given in the histogram, for comparison.

The final group of two models shown in Fig. 5 also incorporate orientational variations of contrast (solid triangles) or of both contrast and $1/f$-exponent (open triangles). For c_w and c_l, this yields an increase in information gain, but again c_r performs poorly. Figure 5 also shows that making α depend on the orientation hardly matters, compared to making the contrast depend on orientation.

DISCUSSION

In this article we investigated the power spectra of natural images, and found that their main statistical properties, total power (or contrast) and a $1/f^2$-frequency behaviour, are variable, but still typical for natural images. Contrast and $1/f^2$-behaviour were also investigated as a function of orientation, and shown to be similar in variability as the non-oriented measures.

The 1/f-exponent

The $1/f$-exponent, α, is 1.88 ± 0.43 (SD) for our set of 276 images. Previous reports gave $\alpha = 2.05 \pm 0.02$ [SEM; average and two-dimensional extrapolation of one-dimensional data of Burton & Moorhead (1987) on 19 images], $\alpha = 2.13 \pm 0.36$ [SD; van Hateren (1992a), 117 non-calibrated published images], $\alpha = 2.40 \pm 0.26$ [SD; Tolhurst *et al.* (1992), 135 images], $\alpha = 2.2$ [Field (1993), 85 images], and $\alpha = 1.81 \pm 0.01$ [Ruderman & Bialek (1994), 70 images after a logarithmic intensity scaling]. Most of this variability between different studies probably arises from the particular choice of sample set. Subdividing our set into categories depending on content (e.g., type of landscape) or recording conditions (e.g., weather), shows a large variability of the resulting average αs of each subset (the standard deviation of this set of αs is 0.3). Furthermore, the measured $1/f$-exponents are quite sensitive to small calibration errors.

Given the large variation in α [see e.g., Fig. 2(C)], one may wonder how critical the value of α is for obtaining information about natural power spectra. This question is directly addressed in Fig. 6, where the information gain is shown as a function of α, with α fixed for the entire set of images. As the figure shows, the information gain has a maximum near $\alpha = 2$, but this maximum is very shallow. Thus the assumed value of α appears to be relatively unimportant, unless it deviates much from 2, and assumes values unlikely to occur in natural images [see the distribution of α-values, copied from Fig. 2(C)]. In a

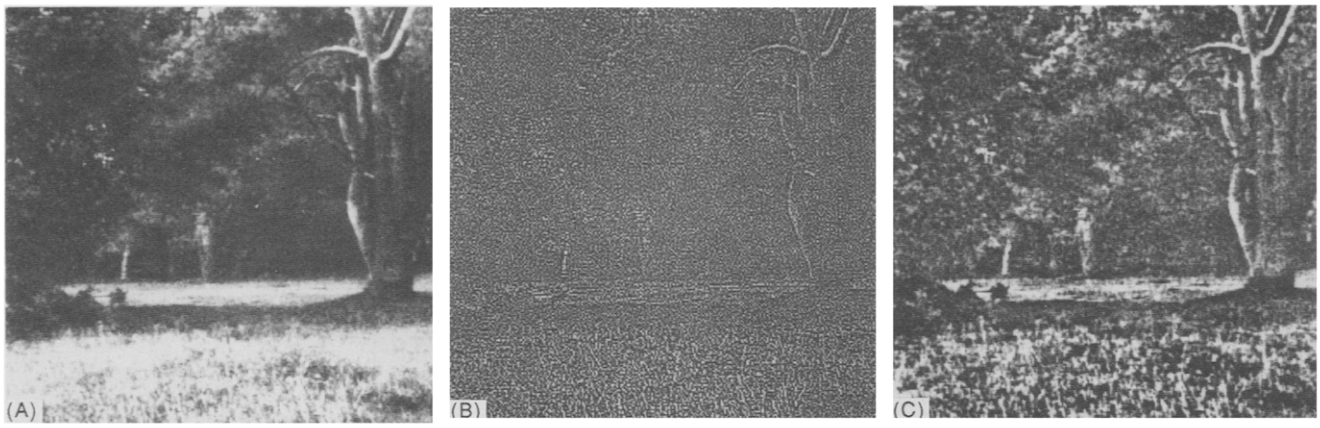

FIGURE 7. Effects of various contrast measures. (A) Image with added noise and all frequency components with power below twice the average noise level clipped. (B) Linear filtered version of (A) with its power spectrum multiplied by f^2. The r.m.s.-contrast of this image is the same as the whitened contrast of image (A). (C) Nonlinear transformation of (A), with a power spectrum equal to the logarithm of that of (A). The r.m.s.-contrast of this image is the same as the log-contrast of image (A).

recent study on the discrimination of second-order statistics by the human visual system, Tadmor and Tolhurst (1994) reach a conclusion which is consistent with our analysis, namely that changes in α are not perceived directly, but only through changes in (local) contrast.

Contrast

The r.m.s.-contrast values we report [Fig. 2(B)] are quite large, and characteristic of outdoor environments. Large intensity variations in the scenes, in particular high intensity peaks, produce a large variance of pixel intensities. In fact, the resulting contrasts are larger than one can easily reproduce in the laboratory with conventional display equipment. Although r.m.s.-contrast is a simple measure of the height of the power spectra, it is neither a good measure for the contrast as subjectively perceived by a human observer [see e.g., Tadmor & Tolhurst (1994), also Peli (1990)] nor is it a particularly good measure for giving information about natural power spectra. As Fig. 5 shows, better measures are whitened contrast, and log-contrast. The former can be considered as the r.m.s.-contrast of a filtered image. The filter here is a differentiating filter (multiplication of the amplitude with f, thus the power with f^2), acting similarly as a naive edge-enhancement algorithm might do.

The log-contrast, on the contrary, is determined by a nonlinear operation, and cannot be produced via a linear filter. Interestingly, it estimates the height of the spectrum identically to what a least squares fit of a straight line to the power spectrum on log–log coordinates would produce. As the variation of the logarithm of the power at a specific frequency coordinate follows in rough approximation a Gaussian distribution, the log-contrast is also close to a Maximum Likelihood estimate of the height of the spectrum.

The log-contrast may seem a strange measure, because it requires taking the logarithm of an image in the spectral domain. As a consequence, the log-contrast does not have a simple interpretation in terms of linear filtering like the whitened contrast. However, there is an interpretation possible in terms of contrast adaptation. Suppose we scale the power spectrum such that the smallest power in the spectrum, or the smallest power larger than the noise present in the image, is slightly larger than 1. Suppose further that we skip all frequencies with power below the noise level, and that we subsequently take the logarithm of the power spectrum. Then the transfer is optimal at frequency coordinates with power somewhat larger than 1. Much larger powers at other frequencies will be attenuated by the logarithmic transformation, the more so, the larger the power is. This can be interpreted as a form of contrast adaptation in the frequency domain: a saturating nonlinearity reducing the impact of high powers. Fourier-transforming this scaled logarithmic spectrum back to the space domain (using the phase components of the original Fourier transform), we obtain an image with enhanced contrast. In this respect it resembles a whitened (i.e., differentiated) image, but it is in fact far less sensitive to noise. The variance of this image is equal to the summation of its power spectrum, which is equivalent to our definition of the log-contrast. Thus, we may interpret the log-contrast as equivalent to the r.m.s.-contrast after a form of contrast adaptation. Figure 7 shows an example of an image with added noise (A), the image constructed from the spectrum multiplied by f^2 (B), and the image constructed from the logarithmic spectrum (C). Image (C) is clearly less sensitive to noise than image (B), and has enhanced contrast compared to (A).

Modelling power spectra

The performance of the various models (Fig. 5) leads to several important conclusions. First of all, the contrast of each image, and the assumption that it behaves as $1/f^2$, are important factors for gaining information about the power spectra. However, leaving the $1/f$-exponent free to

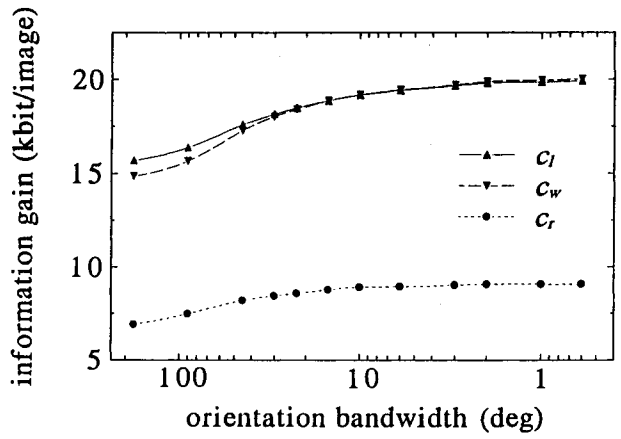

FIGURE 8. The information gain for the model $[c(\varphi), 1/f^2]$ as a function of orientation bandwidth ($\Delta\varphi$), for three different contrast measures (c_r, c_w, and c_l).

vary does not yield much information, nor does including systematic anisotropy in our description.

Of the orientational models we evaluated, the main improvement was for the $c_l(\varphi)$, $1/f^2$-model. Thus it pays to measure contrast as a function of orientation, but, similarly as before, it is only of limited value to allow α to vary, here as a function of orientation. We note that subsets of images containing exclusively man-made objects or exclusively scenes of nature give similar results. As mentioned earlier, the results of Fig. 5 all are at a resolution of $\Delta\varphi$ = 3 deg. Figure 8 shows what happens when we change this orientation bandwidth. The data points on the far left, at $\Delta\varphi$ = 180 deg, are identical to the result in Fig. 5 for a non-oriented contrast. Increasing the orientation resolution increases the information gain, but only up to a certain point: the curve reaches an asymptote, with $\Delta\varphi$ = 3 deg already close to its final value. As can be seen, much of the information to be gained from evaluating contrast in various orientations is already reached for $\Delta\varphi$ = 10–30 deg. Interestingly, this is a value commonly encountered in orientation-sensitive cells in the cortex of higher primates (e.g. De Valois *et al.*, 1982). However, it remains to be seen if this correspondence is retained when the large area of analysis we have used is decreased to the relatively small size of typical cortical receptive fields. Also, it is not clear at present in how far orientation-selective mechanisms play a role in reducing the redundancy associated with oriented structures in the images [see e.g., Zetzsche *et al.* (1993) for a detailed discussion on this issue].

Global or local statistics?

The analysis in this article is of a global nature: the power spectra analysed and modelled are global descriptors of an image, although they describe the correlation between neighbouring points in the image. In natural visual systems, neurons generally have receptive fields of limited size, and thus it might be relevant to know if the statistical properties vary over the image. Nevertheless, we believe our analysis still bears relevance for understanding how much local variation in statistics one can expect within images [see also Zetzsche *et al.* (1993)]. The reason is that even our power spectra concern images of limited extent. As an image remains an image when decreased in angular size (e.g., when seen from a larger distance), its statistical properties will remain similar when it is reduced to a size that would be considered local for a particular visual system. This scaling property of natural images is consistent with the $1/f^2$ distribution of spectral power (Field, 1987), and previous studies (Ruderman & Bialek, 1994; Ruderman, 1994). However, this scaling obviously breaks down when the area considered comes close to the distance between the sampling points (e.g., the photoreceptors) of the visual system. Furthermore, our present analysis does not consider the phase of the Fourier transform (Field, 1993). Therefore, it will be worthwhile to perform a similar analysis as we have performed here for more localized estimates of the frequency contents of an image (e.g., using Gabor functions or wavelets). Similarly, the analysis can be extended to the temporal or spectral (chromatic) domain.

CONCLUSION

Second-order redundancy in images is most conveniently studied through their power spectra. As reducing redundancy appears to be an important (pre)processing strategy utilized by visual systems, it is important to have good models for the power spectra of natural images. To this end we first obtained a large set of power spectra of natural images, and confirmed previous reports on their variability. We also showed a considerable variability of contrast, and showed that this also applies to orientational features of the spectra. Second, we developed a method to evaluate models for natural power spectra; this method shows directly their capacity to estimate contrast and their capacity to reduce second-order redundancy. The results show that a good model for a power spectrum contains: (1) a proper estimate of its height, i.e., contrast; (2) a spatial frequency behaviour specified as $1/f^2$ (no less and no more accurate than that); and (3) a contrast that depends on orientation, with a resolution that need not be more precise than 10–30 deg.

REFERENCES

Atick, J. J. & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation, 2*, 308–320.

Atick, J. J. & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation, 4*, 196–210.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In *Proceedings of the National Physical Laboratory Symposium on the mechanisation of thought processes* (pp. 537–559). London: H. M. Stationery Office.

Bialek, W., Ruderman, D. L. & Zee, A. (1991). Optimal sampling of natural images: A design principle for the visual system? In Lippmann, P. L., Moody, J. E. & Touretzky, D. S. (Eds), *Advances in neural information processing systems 3* (pp. 363–369). San Mateo: Morgan Kaufmann.

Burton, G. J. & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics, 26*, 157–170.

Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering, 36*, 107–114.

De Valois, L. R., Yund, E. W. & Hepler, N. (1982). The orientation and directional selectivity of cells in Macaque visual cortex. *Vision Research, 22*, 531–544.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America, A4*, 2379–2394.

Field, D. J. (1993). Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems. In Farge, M., Hunt, J. C. R. & Vassilicos, J. C. (Eds), *Wavelets, fractals, and Fourier transforms* (pp. 151–193). Oxford: Clarendon Press.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation, 6*, 559–601.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE, 66*, 51–83.

Hateren, J. H. van (1992a). Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A, 171*, 157–170.

Hateren, J. H. van (1992b). Real and optimal neural images in early vision. *Nature, 360*, 68–69.

Hateren, J. H. van (1993). Spatiotemporal contrast sensitivity of early vision. *Vision Research, 33*, 257–267.

Laughlin, S. (1983). Matching coding to scenes to enhance efficiency. In Braddick, O. J. & Sleigh, A. C. (Eds), *Physical and biological processing of images* (pp. 42–52). Berlin: Springer.

Papoulis, A. (1965). *Probability, random variables, and stochastic processes*. New York: McGraw-Hill.

Peli, E. (1990). Contrast of complex images. *Journal of the Optical Society of America, A7*, 2032–2040.

Ruderman, D. L. (1994). The statistics of natural images. *Network: Computation in Neural Systems, 5*, 517–548.

Ruderman, D. L. & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters, 73*, 814–817.

Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.

Srinivasan, M. V., Laughlin, S. B. & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B, 216*, 427–459.

Switkes, B., Mayer, M. J. & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research, 18*, 1393–1399.

Tadmor, Y. & Tolhurst, D. J. (1994). Discrimination of changes in the second-order statistics of natural and synthetic images. *Vision Research, 34*, 541–554.

Tolhurst, D. J., Tadmor, Y. & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic & Physiological Optics, 12*, 229–232.

Zetzsche, C., Barth, E. & Wegmann, B. (1993). The importance of intrinsically two-dimensional image features in biological vision and picture coding. In Watson, A. B. (Ed), *Digital images and human vision* (pp. 109–138). Cambridge: MIT Press.

Acknowledgements—We thank Marten Jansen, Kees Schilstra, Herman Snippe and Doekele Stavenga for comments on the manuscript. This research was supported by the Netherlands Organization for Scientific Research (NWO).

APPENDIX A

Calculating contrast and 1/f-exponent

The r.m.s.-contrast (c_r) of an image is defined as the standard deviation of the intensity values of all pixels in the image divided by the mean intensity (σ/μ). In order to keep in line with the spectral analysis, we use the weighted mean intensity of the images (as defined in Methods) and the weighted standard deviation, which follows from the weighted variance of intensity values in the image $l(x, y)$:

$$\sigma^2 = \frac{\sum_{(x,y)} |(l(x, y) - \mu)\, w(x, y)|^2}{\sum_{(x,y)} |w(x, y)|^2}. \tag{A1}$$

For convenience, the weight function (w) is normalized by setting $\sum_{(x,y)} |w(x, y)|^2 = 1$.

The weighted variance is related to the total power in the weighted Fourier transform (F, as given in the Methods section) through Parseval's theorem:

$$\frac{\sigma^2}{\mu^2} = \sum_{(x,y)} \left| \frac{(l(x, y) - \mu)\, w(x, y)}{\mu} \right|^2 = \frac{1}{L} \sum_{(u,v)} |F(u, v)|^2, \tag{A2}$$

where L is the number of pixels in the image as defined in Methods. The image and its weighted Fourier transform are not yet corrected for the modulation transfer of the recording system. The correct r.m.s.-contrast (c_r) is, therefore, given by:

$$c_r^2 = \frac{1}{L} \sum_{(u,v)} \frac{|F(u, v)|^2}{\Gamma(u, v)} = \sum_{(u,v)} S(u, v), \tag{A3}$$

where the indices of the summation (u, v) run over all orientations and spatial frequencies between 1 and 127 cycles per image. Although the above definition of c_r concerns weighted measures, the contrast for the set of images is, in fact, similar in magnitude and distribution as produced by a direct determination of contrast in the space domain on the non-weighted images.

The 1/*f*-exponent (α) of a single power spectrum is calculated by a least squares fit of a straight line to $\log S(u, v)$ as a function of $\log(1/f)$. The slope of the line equals α, whereas the offset of the line is not further analyzed, but is in fact linearly related to the log-contrast.

To calculate the oriented measures, we divide the frequency plane (u,v) into a number of pie-slices. A pie-slice $(u, v)_\varphi$ contains all spatial frequency coordinates (u, v) with orientation between $\varphi - \frac{1}{2}\Delta\varphi$ and $\varphi + \frac{1}{2}\Delta\varphi$. In order to reduce discretization artefacts, in particular at low spatial frequencies, we subdivide the (u, v)-plane into a much finer grid, (u', v'). We assign to each (u', v')-coordinate the power density of the closest (u, v)-coordinate. The oriented-r.m.s.-contrast, $c_r(\varphi)$, is again defined as the summation of $S(u,v)$ [see equation (A3)], but now for $(u, v)_\varphi$ rather than for the whole (u,v)-range. Similarly, the oriented-1/*f*-exponent $\alpha(\varphi)$ follows from a fit to the average power in a pie-slice vs frequency on a log–log scale.

The log-contrast c_l of an image is defined as

$$c_l^2 = \sum_{(u,v)} \log S(u, v), \tag{A4}$$

with a similar definition for the oriented log-contrast, $c_l(\varphi)$, for the range $(u, v)_\varphi$. Finally, the whitened contrast, c_w, is defined as

$$c_w^2 = \sum_{(u,v)} f^2 S(u, v), \tag{A5}$$

again with a similar definition for the oriented whitened contrast, $c_w(\varphi)$.

APPENDIX B

Modelling of power spectra

To model a set of power spectra, we model each of them as the product

$$S_n(u, v) = M_n(u, v)\varepsilon, \tag{B1}$$

where $S_n(u, v)$ denotes the spectral power of the nth image at spatial frequency coordinates (u, v), $M_n(u, v)$ is a given deterministic modulation function (one for each image), and ε, the probabilistic residue, a random variable. The present analysis will only consider second-order statistics of individual images, not higher order statistics. Therefore, we require that all structure described by the model is

incorporated in M, so we do not know better than that the residue ε is independent of (u, v) and image. Furthermore, we require that the probabilistic residue is identically distributed, with probability density function (PDF) p_ε, for different spatial frequency coordinates or different images. For a particular model, M, the realizations of ε follow from the measured power spectra through equation (B1). The p_ε is estimated by collecting the ε values of all images and all spectral coordinates into a histogram. Because of the large variance of ε, the discrete histogram coordinates are chosen on a logarithmic scale. The histogram spacing can be very fine, because the number of images and spectral coordinates is very large. Having estimated p_ε from the measurements, we can now find the PDF of the power spectra, $p_{sn(u,v)}$ (or p_s for short) from

$$p_{s_n(u,v)}(S_n(u, v)) = \frac{p_\varepsilon(\varepsilon)\,\mathrm{d}\varepsilon}{\mathrm{d}S_n(u, v)} = \frac{p_\varepsilon(\varepsilon)}{M_n(u, v)}, \tag{B2}$$

which follows from equation (B1).

Using Shannon's definition of the uncertainty (or entropy) of a PDF, $H = \int p(x) \log_2 p(x)\,\mathrm{d}x$ (Shannon & Weaver, 1949), we find here for the average uncertainty (H) over all power spectra $S_n(u, v)$:

$$H = -\frac{1}{2N}\sum_{n=1}^{N}\sum_{(u,v)}\int p_{s_n(u,v)}(S_n(u, v)) \times \log_2 p_{s_n(u,v)}(S_n(u, v))\mathrm{d}S_n(u, v), \tag{B3}$$

where N is the number of power spectra, and the factor ½ compensates for the symmetry in the power spectrum. Rather than evaluating equation (B3) directly, it is easier to rewrite it, using equation (B2), as

$$H = -\frac{R}{2}\int p_\varepsilon(\varepsilon) \log_2 p_\varepsilon(\varepsilon)\,\mathrm{d}\varepsilon + \frac{1}{2N}\sum_{n=1}^{N}\sum_{(u,v)} \log_2 M_n(u, v), \tag{B4}$$

where R is the number of spectral coordinates in the range (u, v). Both terms of equation (B4) can be readily computed from the estimated p_ε and the given modulation function $M_n(u, v)$.

It may seem that the uncertainty for different models could as well be calculated from p_ε only, leaving the modulation function (M) out. This is not the case, because the residue measures variation at different scales depending on M, such that their uncertainties are not comparable. As a simple example showing the inadequacy of using p_ε, consider starting with a very inadequate model, M. By multiplying M by an arbitrarily large factor, ε becomes small [through equation (B1)], resulting in an arbitrarily small uncertainty following from p_ε. This contradicts the inadequacy of the model, a problem which does not occur for p_S.