Visual Sensitivity, Blur and the Sources of Variability in the Amplitude Spectra of Natural Scenes

DAVID J. FIELD,*‡ NUALA BRADY†

A number of researchers have suggested that in order to understand the response properties of cells in the visual pathway, we must consider the statistical structure of the natural environment. In this paper, we focus on one aspect of that structure, namely, the correlational structure which is described by the amplitude or power spectra of natural scenes. We propose that the principle insight one gains from considering the image spectra is in understanding the relative sensitivity of cells tuned to different spatial frequencies. This study employs a model in which the peak sensitivity is constant as a function of frequency with linear bandwith increasing (i.e., approximately constant in octaves). In such a model, the "response magnitude" (i.e., vector length) of cells increases as a function of their optimal (or central) spatial frequency out to about 20 cyc/deg. The result is a code in which the response to natural scenes, whose amplitude spectra typically fall as $1/f$, is roughly constant out to 20 cyc/deg. An important consideration in evaluating this model of sensitivity is the fact that natural scenes show considerable variability in their amplitude spectra, with individual scenes showing falloffs which are often steeper or shallower than $1/f$. Using a new measure of image structure (the "rectified contrast spectrum" or "RCS") on a set of calibrated natural images, it is shown that a large part of the variability in the spectra is due to differences in the sparseness of local structure at different scales. That is, an image which is "in focus" will have structure (e.g., edges) which has roughly the same magnitude across scale. That is, the loss of high frequency energy in some images is due to the reduction of the number of regions that contain structure rather than the amplitude of that structure. An "in focus" image will have structure (e.g., edges) across scale that have roughly equal magnitude but may vary in the area covered by structure. The slope of the RCS was found to provide a reasonable prediction of physical blur across a variety of scenes in spite of the variability in their amplitude spectra. It was also found to produce a good prediction of perceived blur as judged by human subjects. © 1997 Elsevier Science Ltd

Natural scenes   Wavelet   Image processing   Redundancy   Visual system   Blur

INTRODUCTION

The psychophysical and neurophysiological methods of this last century have provided us with many important insights into the behavior and function of single cells in the visual pathway. However, over the last few years, a number of researchers have taken the position that a complete understanding of visual information processing also requires a better understanding of the nature of the information available in our natural environment (e.g, Barlow, 1961; Srinivasan, Laughlin, & Dubs, 1982; Field, 1987, 1993, 1994; Atick, 1992; Atick & Redlich, 1992; Ruderman, 1994). This approach considers the statistical structure of the environment (e.g., natural scenes) in relation to the known properties of sensory coding. The principle assumption of this approach is that the visual system has evolved and/or developed to produce an efficient representation of its natural environment. Considering the great number of studies that have explored the information processing capabilities of the mammalian visual system, it is surprising that only a few have taken a close look at the sorts of information and redundancy that is available around us. This may be due partly to the bias that many researchers have in assuming that our environment is relatively random and does not produce statistics that are consistent across scenes. The work of the last several years has shown that this is far from the case.

In characterizing the structure of natural scenes, a majority of researchers have concentrated on the redundancy described by the pairwise correlation between

*Department of Psychology, Cornell University, Ithaca, NY 14853, U.S.A.
†Department of Psychology, Manchester University, Manchester, U.K.
‡To whom all correspondence should be addressed [*Fax:* +1 607 255 8433; *Email* djf3@cornell.edu].

luminance values at different points in an image. One such measure of spatial correlation is the autocorrelation function (acf), which describes pairwise correlations as a function of the distance between pixels. If the image statistics are "stationary" (i.e., the correlation as a function of distance is the same at all image locations for the population of images) then the acf provides a complete description of all pairwise correlations, as does the image power spectrum, which is given by the Fourier transform of the acf (Field, 1987). A number of studies have demonstrated that the amplitude spectrum falls with frequency (f) by a factor of about $1/f$ as would be expected from a scale-invariant environment (hence, the power spectrum falls by about f^{-2}). However, the falloff across scenes shows significant variability ranging from approximately $f^{-0.6}$ to $f^{-1.6}$ with averages from different studies in the range $f^{-0.9}$ to $f^{-1.2}$. (Burton & Moorehead, 1987; Field, 1987, Field, 1993; Tolhurst, Tadmor, & Chao, 1992; Ruderman & Bialek, 1994; van der Schaaf & van Hateren, 1996). However, it should be noted that Carlson (1978) noticed the falloff with amplitude as a function of frequency with his images and even Kretzmer (1952) noted that television signals might be efficiently compressed because of the predictable falloff in the correlations in images as a function of distance (Fig. 1).

There have been a number of attempts to account for properties of visual neurons in terms of this form of redundancy (Srinivisan *et al.*, 1982; Atick & Redlich, 1992; van Hateren, 1992; van der Schaaf & van Hateren, 1996; Switkes *et al.*, 1978; Hancock, Baddeley, & Smith, 1992). However, we argue here that the amplitude spectrum provides us primarily with insights into the overall sensitivity of visual neurons (Field, 1987). To account for why cells in the early visual pathway have

their particular bandwidths and spatial profiles, it was proposed that one needs to consider higher-order statistics as represented by the phase spectra (Field, 1987, 1993, 1994; Olshausen & Field, 1996). The argument presented in this work is that the particular parameters found in the primary visual cortex are near to optimal, if optimal is defined in terms of "sparseness". That is, these properties provide a means of representing any particular natural scene with a minimal number of active neurons (i.e., a minimal description length). However, the subset that is active will change from image to image resulting in a low probability that any particular cell is active. This sparse representation is theorized to be the most independent representation that is possible, given a semi-linear representation like that found with simple cells. Olshausen & Field (1996) have recently reported that a neural network designed to increase statistical independence by finding the most sparse representation produces "receptive fields" like those found in primary visual cortex. These receptive fields form even when the spectrum is "whitened" to remove all correlations. The amplitude spectra of the images is largely irrelevant to whether such a sparse representation is possible. In Fourier terms, it is the phase spectra that is relevant to the question of sparseness (Field, 1987, 1994) and therefore it follows that the amplitude spectra of natural scenes provides little insight into understanding why visual neurons have their particular bandwidths and spatial profiles.

If this is the case, then does one gain any insight into visual coding by understanding the statistical regularity revealed by the amplitude spectra of natural scenes? In the following sections, we propose that the amplitude spectra provide insights into the absolute sensitivity of
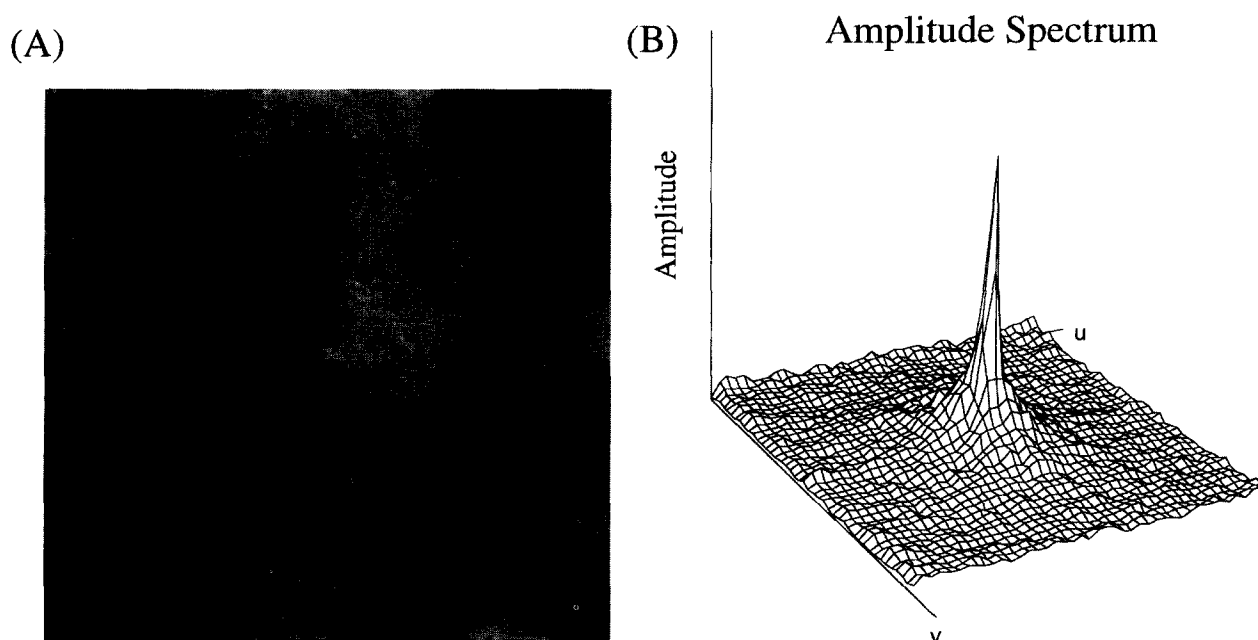
(A)

(B) | **Amplitude Spectrum**

FIGURE 1. An image and its amplitude spectrum. The spectrum peaks at the low frequencies and falls with increasing frequency at all orientations. This falloff has been found to be consistent across a wide range of natural scenes, typically falling with frequency (f) by a factor of approx. f^{-1} with falloffs ranging from approximately $f^{-0.6}$ to $f^{-1.6}$.

visual neurons as a function of spatial frequency. In the following sections we:

1. Describe a model of the sensitivity of visual neurons that allows us to apply that sensitivity to the amplitude spectra of scenes.
2. Consider the sources of variability in the slopes of natural scenes: demonstrating that both the relative contrast and the relative sparseness at different frequencies (i.e., different scales) contribute to the slope.
3. Describe a theory of blur in images which proposes that blur is related to the relative contrast as a function of frequency—not the relative sparseness.
4. Demonstrate that a model which separates relative sparseness and relative contrast can identify blur in natural scenes despite considerable variability in the falloff of their amplitude spectra.

The first section describes a model of sensitivity along the lines of that proposed by Field (1987), Kingdom & Mouldon (1992) and Brady & Field (1995) and Brady et al. (1997). We use this general model to describe the relative contrast of structure as a function of frequency. This will provide a "language" for discussing the sources of variability in the slopes of natural scenes and also provide the basis of the model to predict blur in images. The final section will compare this model to judgments of perceived blur by human subjects.

In the last study of this paper, we look at the ability of subjects to "identify" blur (as opposed to "discriminate" blur). Subjects appear to be capable of identifying whether a novel complex image appears blurred even without a reference image and even though there is insufficient information in the amplitude spectrum to perform this task. Tadmor & Tolhurst (1994), for example, found that when subject's were asked to adjust the slopes of natural scenes to their "best quality", they were accurate at selecting the slope of the original image. Three recent studies have explored human observer's sensitivity to changes in the amplitude spectra of images in relation to visual coding (Knill, Field, & Kersten, 1990; Tadmor & Tolhurst, 1994; Tolhurst, Tadmor, & Arthurs, 1996). Such studies provide insights into how the visual system discriminates complex images and may be related to blur discrimination (Hammerly & Dvorak, 1981; Watt & Morgan, 1983; Walsh & Charman, 1988; Hess, Pointer, & Watt, 1989; Peli et al., 1981). However, it should be noted that when discriminating images with shallow slopes, the term "blur discrimination" may be inappropriate, since neither of the two images in the discrimination task may appear blurred. Furthermore, as Tadmor & Tolhurst (1994) point out, there is considerable variability in slope discrimination from image to image, suggesting the phase structure of the image plays a significant role.

However, before we consider blur and the sources of variability in the amplitude spectra of images, we describe an approach to visual sensitivity that can be directly applied to the natural scenes.

AMPLITUDE SPECTRA AND VISUAL SENSITIVITY

As noted above, researchers have found that the amplitude spectrum shows variability across scenes. However, the average falloff of any particular collection of images appears to average approximately $1/f^{1.1}$. An image with a $1/f$ amplitude spectrum has equal contrast energy in each octave (each octave has the same variance). As previously noted, an image with a f^{-1} spectrum will produce equal average responses across an array of mechanisms that have equal peak spatial frequency sensitivity but increasing linear bandwidth (Field, 1987). A model of this type is shown in Fig. 2. This model of the visual system has recently been shown to be capable of accounting for the contrast matching behavior in human observers (Brady & Field, 1995). Furthermore, recent neurophysiology by Croner & Kaplan (1995), has shown that ganglion cell receptive fields in macaque have a peak sensitivity which is inversely proportional to the square of the area covered by the receptive field, as expected from this model.

(a) Spatial Response

Amplitude = k / width²

Frequency Response

Log Sensitivity

Log Spatial Frequency

Sensitivity

Spatial Frequency

(b) A vector length description of visual sensitivity

Log Sensitivity |V|
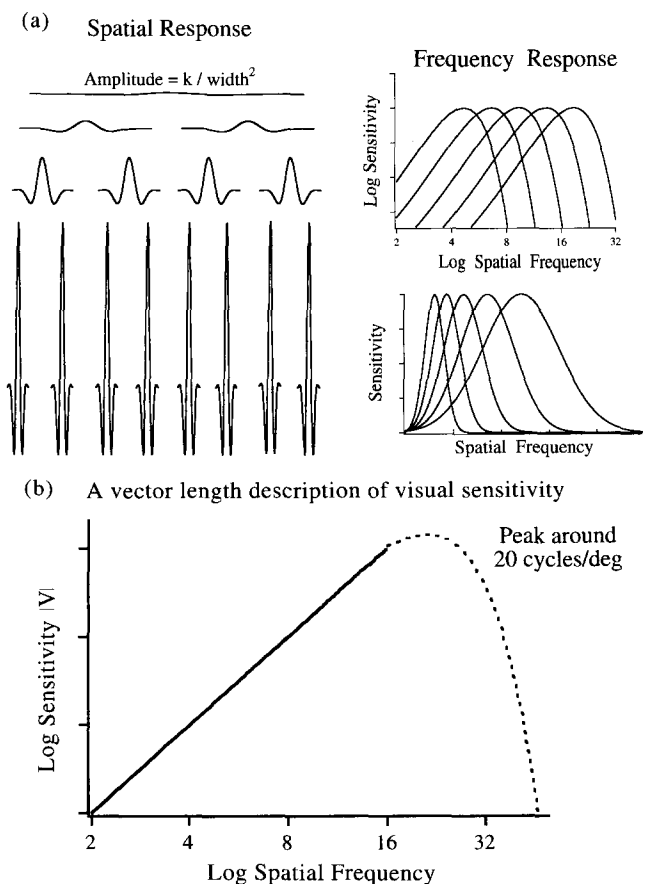
Peak around 20 cycles/deg

Log Spatial Frequency

FIGURE 2. (a) Proposed model of sensitivity. In frequency, the peak sensitivity is constant as a function of frequency. In space, the peak sensitivity increases with the inverse square of the width (i.e., increases with the square of the frequency). (b) If sensitivity of the model in (a) is defined in terms of vector length (V) then the sensitivity increases in proportion to frequency out to the peak of the highest frequency channel (e.g., approximately 20 cyc/deg, according to Brady & Field, 1995). The advantage of this description is that it can be applied directly to broadband stimuli like natural scenes [e.g., equation (2)] to determine the average response magnitude as a function of frequency.

How should we define sensitivity in a model like that shown in Fig. 2? As one can see, the peak sensitivity of the mechanisms is constant across scale (i.e., across the different spatial frequency bands) when measured in the frequency domain. However, in the spatial domain, peak sensitivity increases with the square of frequency. Thus "peak sensitivity" provides a measure of relative sensitivity that depends on the units of measurement. We believe the best way to describe sensitivity is to treat each cell as a vector and consider relative sensitivity in terms of the relative vector length. The length of the vector is determined by its L2 norm—(i.e., the Euclidean sum of the basis vectors). Because the length of a vector is preserved under an orthonormal transformation, this measure of sensitivity is equivalent in the space and frequency domains. In either domain, vector length is defined as:

$$V = \sqrt{(\sum g(x_i)^2)}. \qquad (1)$$

In the frequency domain, the vector length is, therefore, equal to the square root of the volume under the power spectrum. For the model shown in Fig. 2(a), this volume (and hence the vector length) increases with the square of the frequency out to the highest frequency channel and therefore $V(f)$ increases with frequency. Based on the results of the contrast matching experiments, this highest channel is at approximately 20 cyc/deg (Brady & Field, 1995) for human observers. Thus, the model of sensitivity illustrated in Fig. 2(a) has a sensitivity profile, as described by vector length, as shown in Fig. 2(b).

One of the advantages of this approach to describing sensitivity as a function of frequency is that it allows us to compare the average response of cells at different frequencies by simply multiplying the vector length function by the amplitude spectrum of the image

$$R(f) = V(f) * Amp(f). \qquad (2)$$

It also follows that the magnitude of the response to white noise is proportional to this vector length. This model therefore predicts that with white noise input, the visual system produces the largest response at the highest frequencies (around 20 cyc/deg). However, with a $1/f$ image, this model will produce a response which will be flat out to 20 cyc/deg, as shown in Fig. 3. Such a description also seems to match our perceptual observations. The white noise image, as shown below, appears to be dominated by high frequency structure. The image with the $1/f$ spectrum, on the other hand, appears to have structure at a variety of scales.

Does this model of sensitivity conflict with the model of threshold sensitivity revealed by the contrast sensitivity function which peaks at 2–4 cyc/deg? As Field & Brady (1997) point out, the two are not incompatible. Consider what this model predicts about detecting a sinusoid in broadband noise. The response to the sinusoid is relatively constant across frequency, as described by the peaks of the spectral sensitivity curves. But since the bandwidth is increasing with frequency, the response to broadband noise increases with frequency. Therefore, the signal to noise ratio will be decreasing with frequency. The signal to noise ratio for sinusoids would be highest at the peak of the lowest frequency channel—i.e., around 2–4 cyc/deg. Therefore, if we assume that noise that limits threshold sensitivity is relatively flat (as suggested by the equivalent noise measure of Pelli, 1990), then the threshold curve should look something like the contrast sensitivity function (Field & Brady, 1997).

It should also be noted that in this particular model, the bandwidths increase in proportion to frequency (and are constant in octaves). The physiological evidence actually suggests that the bandwidths do not quite increase in proportion to frequency and therefore have smaller bandwidths in octaves at higher frequencies (Tolhurst & Thompson, 1981). For the vector length model to hold, this would require that peak sensitivity actually increase at the highest frequency. We are currently exploring this possibility.

What is the advantage of coding an environment with vectors that have equal average response magnitude? The most obvious advantage is that it allows images to be represented by cells with the same dynamic range. That is, if the cells coding the different frequencies have the same maximum response and same threshold, then this method allows the typical contrasts in the image to be matched to the range of responses that can be produced by the cells.

It should be noted that this "response equalization" (e.g., Field, 1987) differs from the whitening hypothesis proposed by Atick (1992) and Atick & Redlich (1992). They note that the basic shape of the contrast sensitivity function and the spatial frequency tuning of ganglion cells are well described by a model which assumes that the ganglion cells are decorrelating the $1/f$ signal and must also deal with the presence of broadband noise (e.g., photon noise). However, the point we are making is not about whether cells in different spatial locations are uncorrelated. Rather, the argument here relates to the relative sensitivity of cells tuned to different frequencies. We suggest that the cells tuned to different frequencies will have the same response magnitude, on average, in the presence of a $1/f$ image allowing the information in the images to be matched to the response range of the cells. Increasing the overall response magnitude of the high frequency cells in this way will have no effect on the correlations between cells.

In the following section, we use this model of sensitivity to develop an understanding of the effective contrast of natural scenes. In particular, this model suggests that for a $1/f$ image, the average response will be approximately constant as a function of frequency. Similarly, the response to edges (which have $1/f$ spectra) will be approximately constant as a function of frequency.

SOURCES OF VARIABILITY

As mentioned, the amplitude spectra of natural scenes shows significant variability ranging from approximately $f^{-0.6}$ to $f^{-1.8}$ with averages from different studies in the

Response to noise

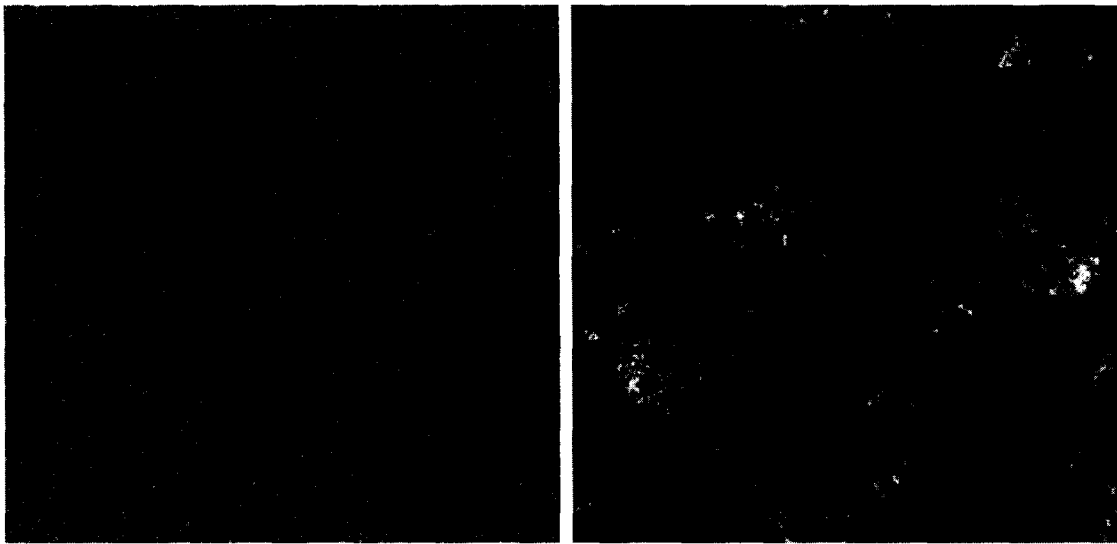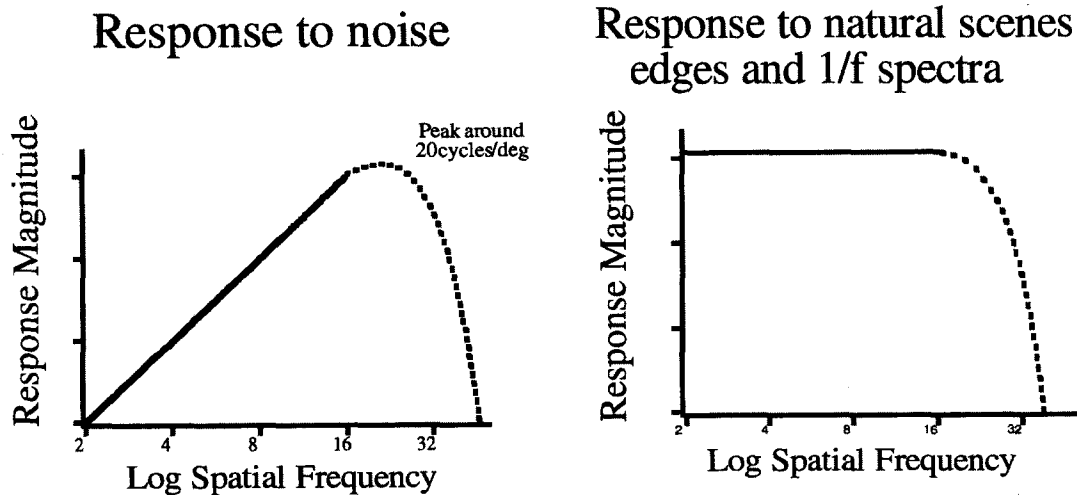## Response to natural scenes edges and 1/f spectra

FIGURE 3. The figure shows the response of the model in Fig. 2 to two types of scenes. The figure on the left shows the average response magnitude as a function of frequency to images with flat spectra such as noise. The figure on the right shows the average response magnitude to images with 1/f spectra, such as the image shown below on the right. The reader will probably agree that the noise image on the left appears dominated by high frequency structure while the image on the right has visible structure at a variety of scales.

range $f^{-0.9}$ to $f^{-1.2}$. Should visual sensitivity be adapted to take into account the fact that the falloff is often steeper than f^{-1}? Or suppose that a particular image has a falloff of $f^{-1.5}$. Would that mean that the visual system's sensitivity is poorly matched to this image? The answer depends on the underlying cause of the steeper spectrum. In this section, we will look at the structure that results in a 1/f spectrum and propose two sources that underlie the variability in the spectra. We will then return to the issue of visual sensitivity to see how it applies to the two sources of variability (Fig. 4).

Previously, it was proposed (Field, 1993, 1994) that a first approximation of scenes by considering the image to be a sum of bandpass functions (e.g., wavelet basis functions), where in correspondence with the wavelet, we sum according to:

$$\text{Image}(x, y) = \sum_{m=0}^{\alpha} \sum_{n=1}^{\beta \sigma 2^m} g[\sigma^m x - x_{nm}, \sigma^m y - y_{nm}] \quad (3)$$

where "m" corresponds to the scale, "α" is the number of scales in the image, "β" controls the density of the vectors at each scale, "σ" controls the spectral distance between scales (and also determines the relative size at each

FIGURE 4. Two images created according to equation (3). Such images are scale invariant with regard to the contrast at different scales (the amplitude spectra fall as f^{-1}) and are also scale invariant with regard to local structure. This approach was proposed (Field, 1993, 1994) as a first approximation to understanding the self-similar nature of natural scenes. However, see Ruderman (1996, 1997) for a related approach that uses occluding surfaces rather than additive bandpass functions like those shown here.

scale). $g(x,y)$ represent a localized function (e.g., a Gabor function) and x_{nm} and y_{nm} signify that each element is randomly positioned within the image.

With this procedure, the image is represented as a sum of functions where the number of functions increases with the square of the scale (i.e., the spatial frequency), the size becomes smaller with scale and the average amplitude remains constant. Such images have the property of being scale invariant such that one can zoom anywhere in the image and maintain the same statistical structure. Examples of two such images created in this manner are shown in Fig. 4. As previously noted (Field, 1994), this sum of functions will produce an image which has a $1/f$ amplitude spectrum which has constant energy in each octave. It will also be equally sparse at all frequencies since the number of elements is inversely related to the area they cover. For the following discussion, it is important to recognize that the energy in any given frequency band is dependent on both the amplitude of the elements and the number of elements. Reducing either will reduce the energy in the band. It should be noted that Ruderman (1996, 1997) takes a similar approach proposing that the appropriate sum of occluding surface elements will also produce $1/f$ behavior and suggests that this provides a more appropriate account of the source of self-similarity in natural scenes.

Of course, a real image will have many forms of redundancy not found in such synthetic images. Nonetheless, this simple model can help us to understand two ways in which an image might have variability in the spectrum. Although images do not consist of sums of independently placed elements, we can think of images in terms of the number of edges at different scales and the amplitudes of the edges at different scales. Each of these factors contributes to the slope of the amplitude spectrum (Fig. 5).

Source 1: blur—amplitude reduction at high frequencies

For images with $1/f$ amplitude spectra created in the manner shown above, the elements at each scale have the same contrast. Changing the contrast as a function of scale will change the slope of the amplitude spectrum. Thus, the first reason that the spectrum might be steeper than f^{-1} in many scenes is that the amplitude of the elements decrease with increasing frequency. This would be the expected result for an image that is out of focus. An increase in the slope of the amplitude spectrum would be expected to occur to some degree for any set of images from a three-dimensional world using a finite aperture, since the aperture limits the depth of field. With a range of depths, at least some part of the image would be out of focus. Furthermore, if there was any motion in the image over the period the shutter was open (e.g., wind in the trees) this will also result in blur. Consider the response of a model like that described in the previous section (Fig. 2). For an image with a $1/f$ amplitude spectrum, the average response across the different frequency bands will be flat. With a blurred image (i.e., the amplitude spectrum steeper than $1/f$), the amplitude of the responses

will be lower at higher frequencies. If we consider the amplitudes of edges, for example, as seen through these filters, we will find that the amplitudes of these edges fall with frequency (Fig. 6).

Source 2: variable density in structure at different frequencies

In the second case, rather than altering the amplitude of the high frequency elements, one can also alter the slope of the spectrum by changing the relative density of the elements at different frequencies. If the number of elements increases by a factor of less than f^2 then the spectrum will also have a slope steeper than f^{-1} (Field, 1994). Figure 7 shows examples of five images that have spectra significantly steeper than f^{-1} but are not out of focus.

For an image to be scale invariant at all locations, the amplitude spectrum must fall as f^{-1} (i.e., one can zoom into any region and find similar statistical structure). The images shown in Fig. 7, therefore, have spectra that show that they are not scale invariant. For these images, it is proposed that the number of zero crossings does not scale appropriately. Indeed, the images shown in Fig. 7(a, b and c) have nearly the same low frequency structure. However, the the image in Fig. 7(a) has a larger area covered by high frequency structure (i.e., there are relatively more zero-crossings at the high frequencies). The result is that the image in Fig. 6(a) has a shallower slope. In a scale-invariant image, we would expect the number of zero-crossings to increase with frequency such that any sector could be enlarged to discover the same number of zero crossings per unit area.

If the slope is steep because there are fewer regions of the image that have high frequency structure, then with the model described earlier, the high frequency neurons will have less average activity. One must note, however, that there is less activity because fewer respond, not because they respond at a low level. Indeed, those responding at the edges will presumably be responding at the same level as those selective to low frequencies. In the following section, we propose a technique for determining the relative contribution of these two factors in accounting for the variability in the spectra of natural scenes (amplitude vs the density of structure at different frequencies).

THE RECTIFIED CONTRAST SPECTRUM MODEL (RCS)

The principal idea behind this approach is to measure the magnitude of the structure (e.g., edge contrast) that exists at different scales independent of the density of that structure. To achieve this, we measure the contrast of all structures that exceed a predetermined level. If a particular image has only a small region covered by structure, then this measure will evaluate only the magnitude of the structure in that region without the large region of the image not covered by structure reducing the estimate of the contrast.

Figure 8 shows the basic notion behind the threshold-
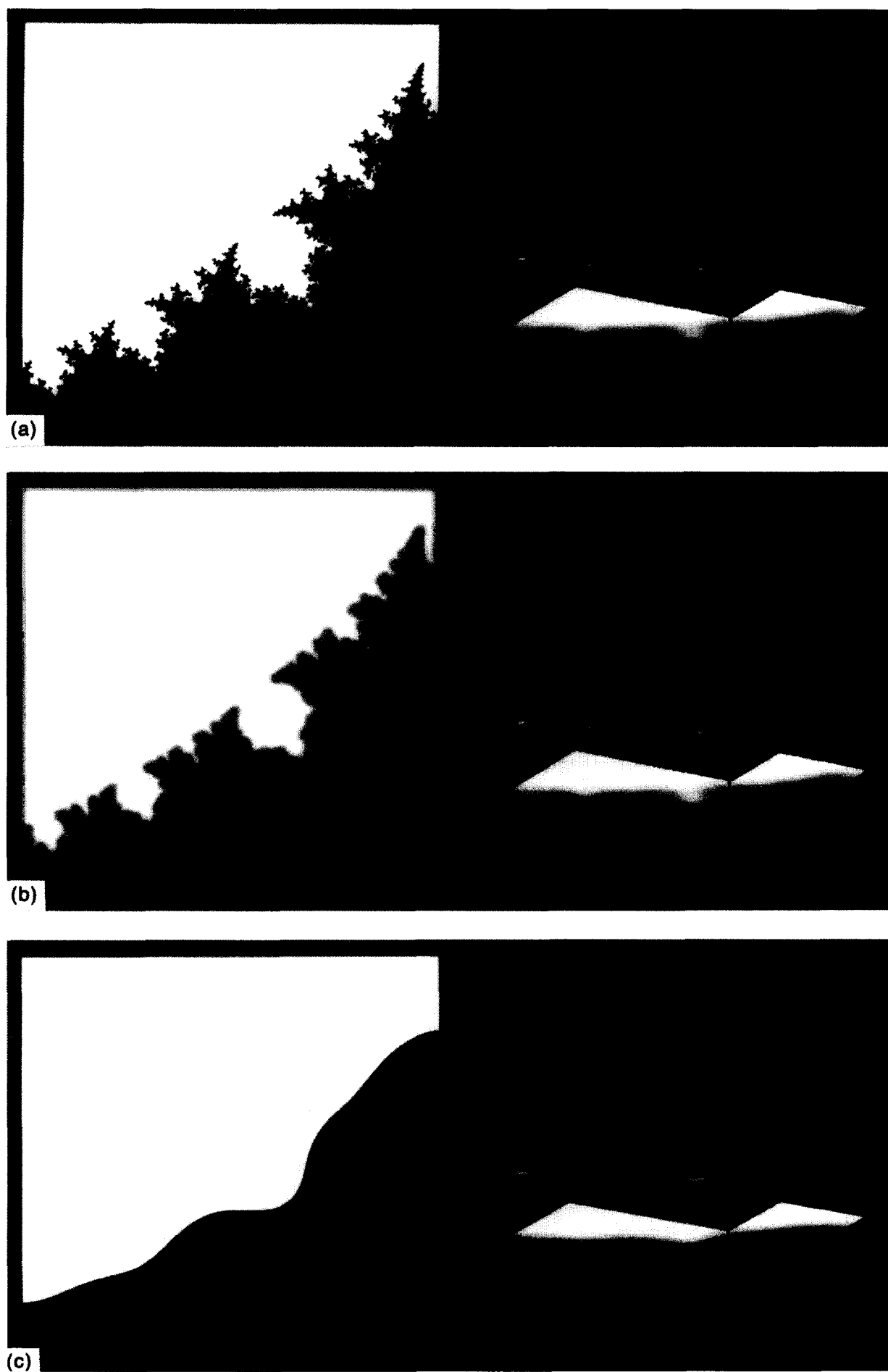
FIGURE 5—*Legend opposite.*

ing. If the area covered by an edge is larger, but has the same amplitude on average, then the variance will be relatively high at the high frequencies. However, if we look at only the region of the image covered by the edge, then we will find the same average variance. By considering only the regions that exceed the threshold, we can find the average amplitude of the areas that contain structure.

Images

The images in this particular study were photographed with a conventional 35 mm camera using Ilford XP1 film along the lines discussed previously (Field, 1994). Twenty natural scenes were used in this study. Each was calibrated for the optics and the luminance non-linearities of the photographic and scanning process. This was done by photographing calibrated sources over a wide intensity range and inverting the non-linearities introduced by a film and scanning process. The compressive non-linearity of the film was advantageous in that many images had an intensity range greater than 1000:1. Although an effort was made to use the smallest aperature of the camera (i.e., the largest depth of field), any image with a finite aperature will probably have some variation in focus across the different depth planes. Also, an effort was made to avoid images that had appreciable movement during the course of the exposure, since this would also introduce spatial blur. This particular set included a number of images from around New York, Canada and Alaska and included blank regions like sky and water but did not contain man-made structures. We limited our set to these 20 natural scenes to allow ourselves to compare the results with the measurements of blur discussed in the final section of this paper. Indeed, the goal here was not to provide a measure of the "typical" falloffs in the spectra of natural scenes; rather, the goal was to find a set of scenes with a variety of amplitude spectra to determine whether the current model could predict the blur in these images despite their variation. Each of these images was 512×512 pixels, and was scanned to a depth of 12 bits using a Nikon LS3510AF scanner and then windowed with a circular aperature with a sinusoidal edge 24 pixels wide.

Method

The outline of the method is as follows.

1. Calculate the slope of the amplitude spectra of the images in the standard way (Measurement 1).
2. Filter the image into 1.5 octave wide bands centered 1.0 octaves apart.

3. Calculate the variance/octave for each of the filtered images and calculate the slope of the RMS contrast (i.e., σ of the filtered image) as a function of frequency (Measurement 2).
4. Threshold each of the filtered images at $\sigma/2$, where σ is defined over the five channels (σ = one standard deviation in the distribution of responses defined across the five octaves of the image).
5. Measure the average magnitude of all responses exceeding the threshold for each of the five frequency bands and calculate the slope of the RMS contrast as a function of frequency (Measurement 3).

A depiction of this method is shown in Fig. 9. The first step is to calculate the falloff in the amplitude spectrum of the calibrated images in the usual manner (e.g., Field, 1987). Using the model shown in Fig. 2, the second step is to filter the image into the 1.5 octave frequency bands. To reduce edge effects, we limit the analysis to the range of 8 cycles/picture to 128 cycles/picture and to the region within a 64 pixel border. As discussed earlier, this filtering results in a set of bandpass, filtered images that have approximately the same RMS contrast for an image with a $1/f$ falloff. We can plot the RMS contrast per frequency band as a function of the octave-band and plot the slope of the spectrum at this stage. As noted earlier, this process adds, on average, a value of approximately 1.0 to the amplitude spectrum since it is a measurement of the contrast/octave (i.e., roughly whitens the spectrum as noted earlier). The fourth step is to threshold these filtered images. For an image that has scale-invariant statistics, the threshold has roughly the same effect on the different scales. We impose the same threshold across each of the filtered images corresponding to half the standard deviation of the average contrast ($\sigma/2$) defined over the set of filtered images (e.g., Fig. 6). The third step is to calculate the average contrast of the structure that exceeded threshold. If the contrast in these images was gaussian distributed in each of the different frequency bands, this process will have little change in the slope after thresholding. On the other hand, if the edges at high frequencies had the same contrast, but were relatively sparse (causing a steeper spectrum), then this method will flatten the spectrum.

Results

Figure 10 shows the results after making the three measurements discussed above. Figure 10(a) shows the distributions of the falloffs in the spectra of the 20 images. In line with previous studies, there is a distribu-

FIGURE 5. This figure demonstrates two reasons an image may have an amplitude spectrum steeper than $1/f$. (a) A fractal edge with a 2D amplitude spectrum that falls by roughly $1/f$. The contrast of the edge at each scale, as seen by the model in Fig. 2, will remain constant, as well as the total energy in each band. In (b), the image is blurred reducing the contrast of the edge at higher frequencies and reducing the total contrast energy at the high frequencies. In (c), the edge is less "jagged". This also reduces the total energy in the higher frequency bands since there are fewer regions covered by the edge. However, the edge maintains a constant contrast through the different frequency bands. Although there is less total power at the high frequencies with the smooth edge, the edge is not blurred.
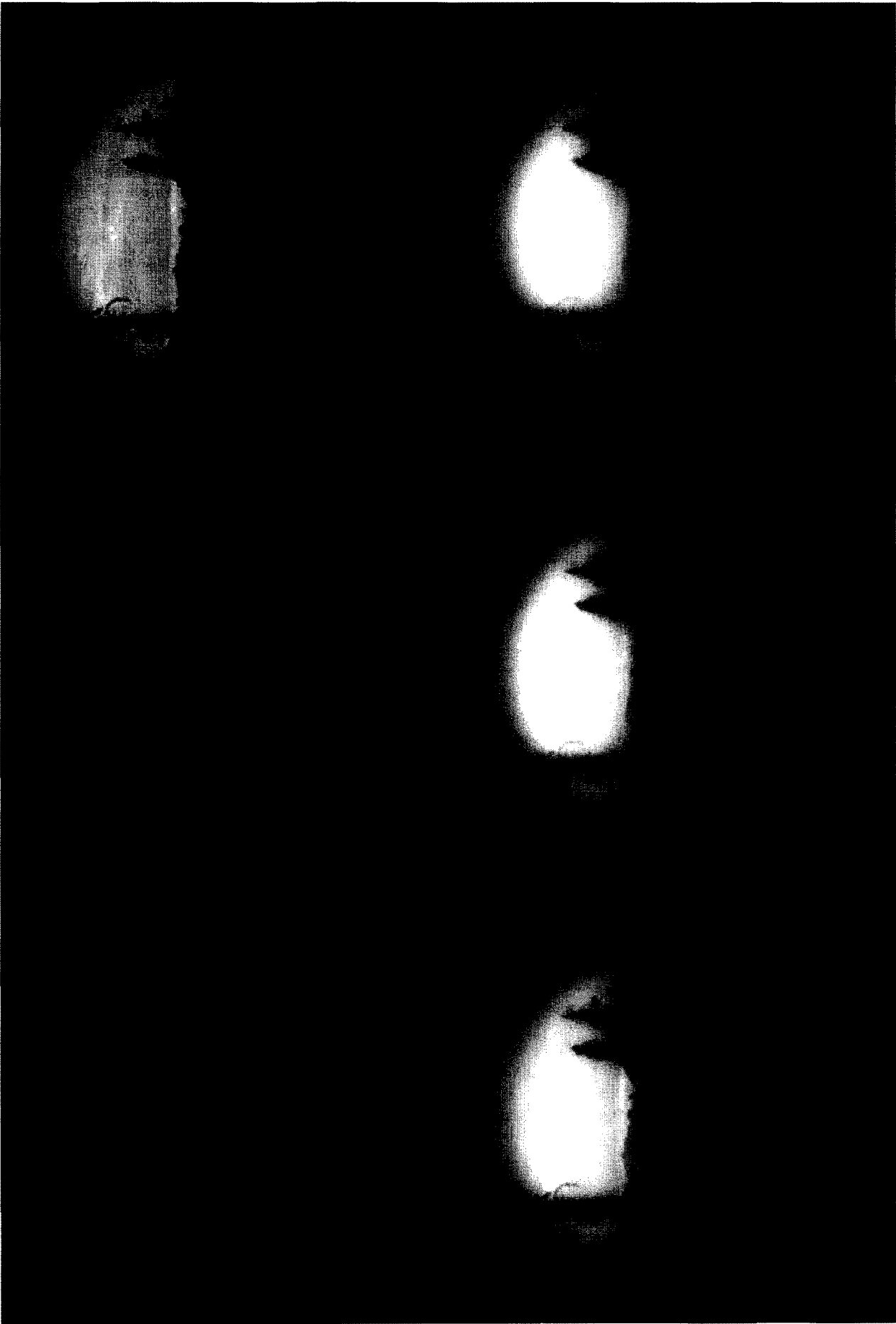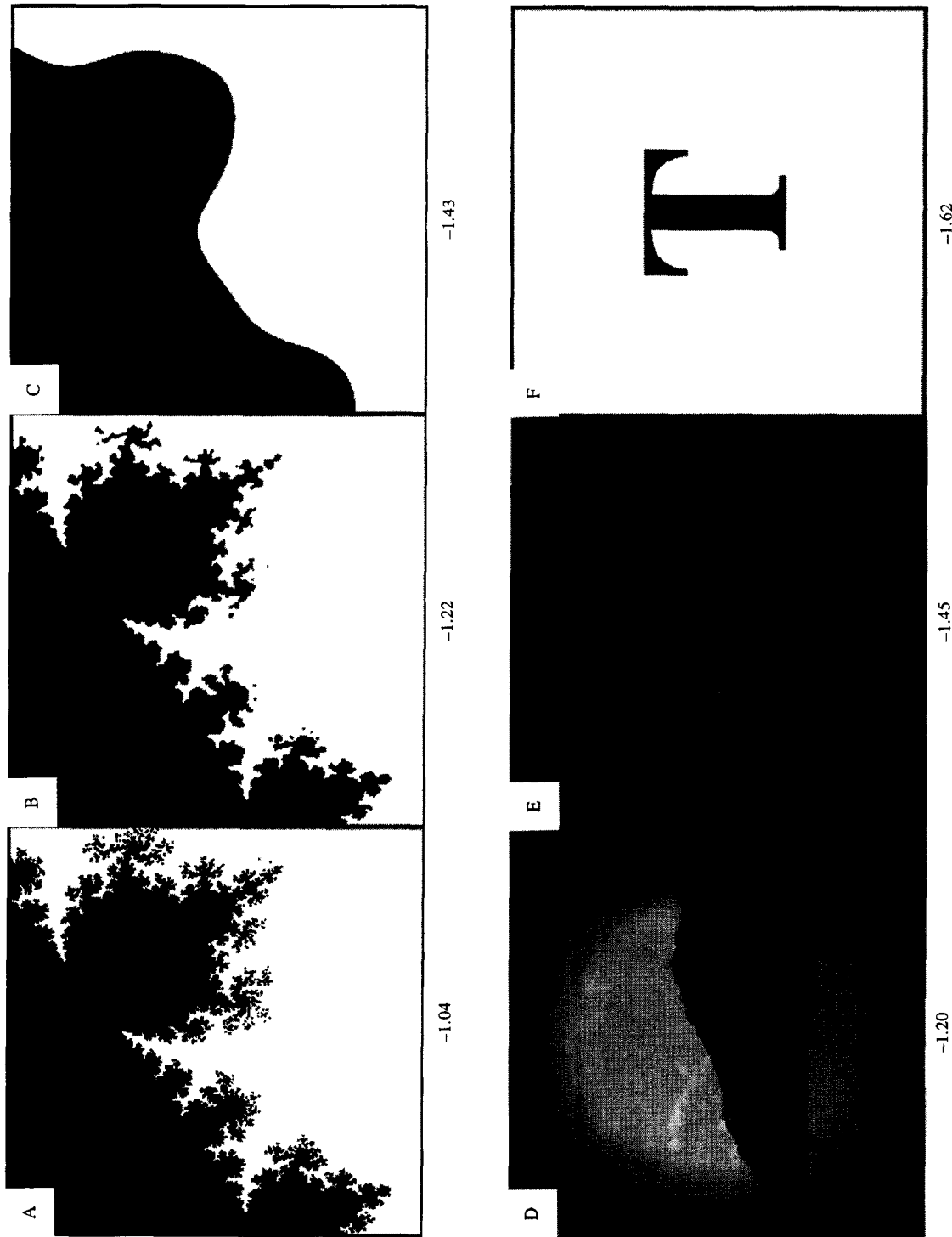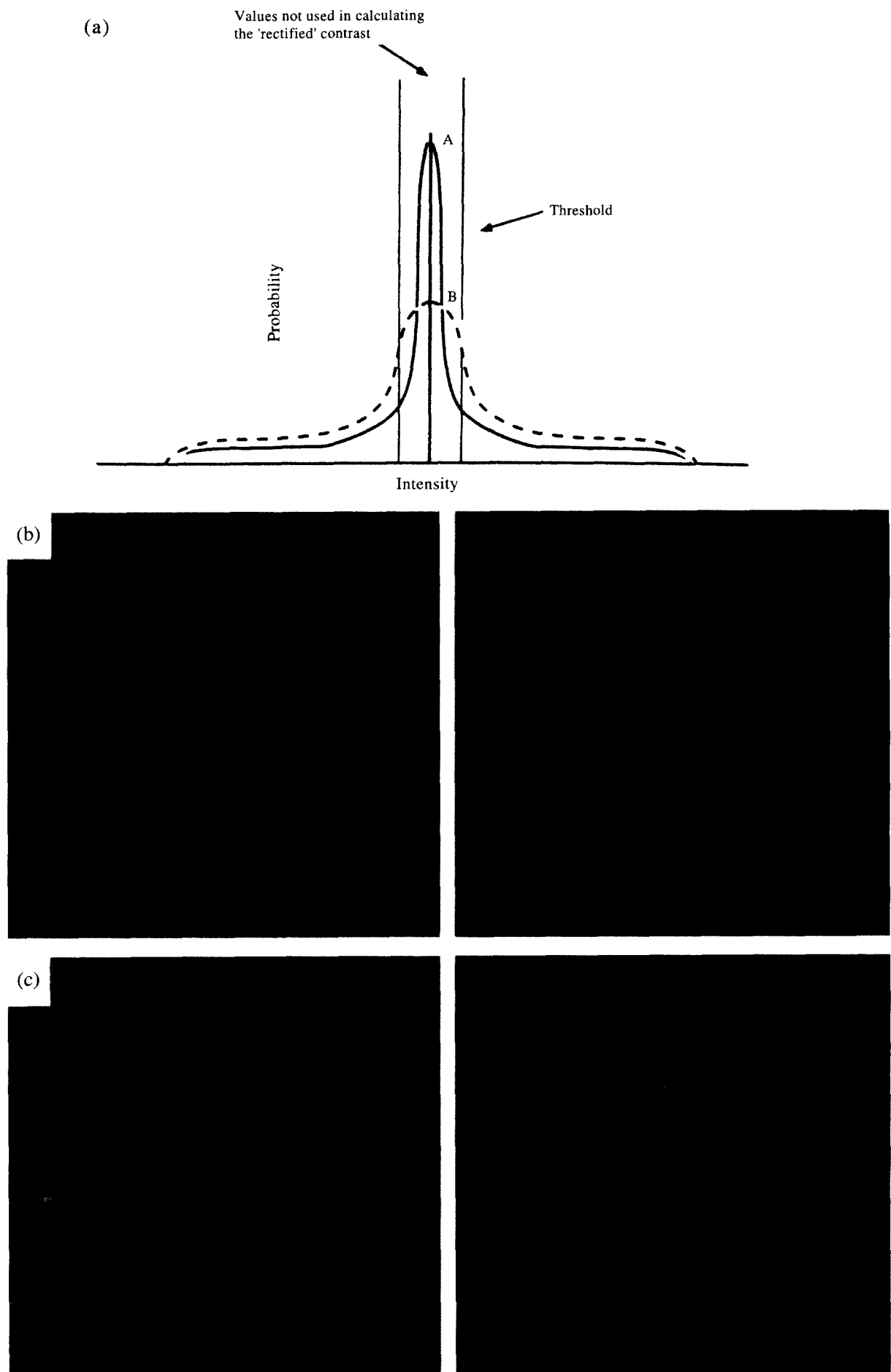
FIGURE 6—*Legend opposite.*

FIGURE 7. This figure provides examples of images that have a variety of slopes in their spectra. (a), (b) And (c) were synthetic images. (d), (e), And (f) were as "in focus" as possible, given the limits of the photographic process. The number beneath each figure shows the slope of the amplitude spectrum.

FIGURE 6. This figure illustrates the result of manipulating the slope of the amplitude spectrum of an image. The slopes of the amplitude spectra of these images, from top-left to bottom-right range from −0.5 to −1.75 in 0.25 steps. For this image, when the slope is altered to be steeper than −1.1, the image appears blurred.

(a)

Values not used in calculating
the 'rectified' contrast

A

Threshold

B

Probability

Intensity

(b)

(c)
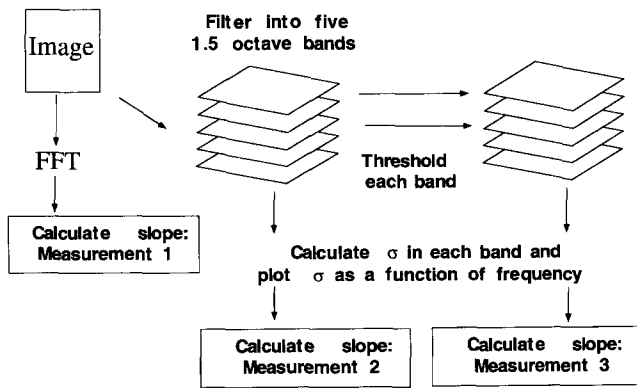
FIGURE 8—*Legend opposite.*

FIGURE 9. This figure portrays the method of making the three slope measurements described in the outline and shown in Fig. 10. See text for details.

tion of slopes with a mean of -1.10 and a standard deviation of 0.14. Figure 10(b) shows the slopes of the contrast spectra which has a mean of -0.04 and a standard deviation of 0.17. As expected, the distribution is roughly the same but the mean has increased by 1.0.

Figure 10(c) shows the results after thresholding. The variability in the spectra has been significantly reduced. The standard deviation of slopes has been significantly reduced from 0.17 to 0.07 ($n = 20$; $P < 0.01$). This suggests that a significant portion in the variability of the spectra in these images is due to the variability in the density of the structure, and not to blur. What is left may be due to the amount of blur or motion in the image. However, it is difficult to tell from this analysis alone. As noted earlier, any image of a three-dimensional environment made with a camera that has a finite degree of blur will probably have some reduction in energy at high frequencies. However, one might argue that some of the drop in variability in slopes is due only to the act of thresholding. To address this point, a study was completed which determined whether this measure was capable of predicting blur in these images in which blur was artificially introduced by altering the slope of the spectrum. The study determines whether both human subjects and the the RCS measure can discriminate changes in the spectrum as due to blur, from changes due to the density of structure at different frequencies.
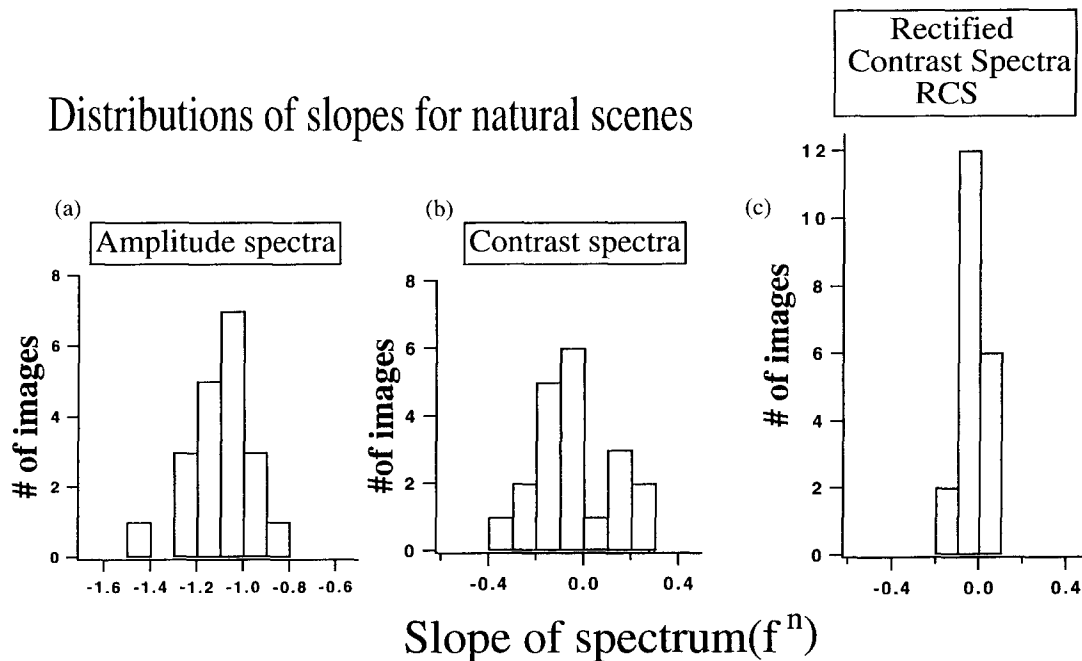

FIGURE 10. Distribution of slopes of (a) the original 20 images; (b) contrast spectra; and (c) the rectified contrast spectra (RCS). The standard deviations of the population of slopes are 0.14, 0.17, and 0.07, respectively. The RCS measure shows less variability than the spectra of the original images. This suggests that a significant proportion of the variability found in the spectra is due to the relative sparseness as a function of frequency. See text for details.

FIGURE 8. This figure portrays how the rectified contrast spectrum was calculated for a particular frequency band. The image shows the result of thresholding a particular frequency band from two different images [Fig. 7(a) and Fig. 7(c)]. The figure at the top provides a characterization of the histograms of the two filtered images shown on the left. The contrast energy as measured by the total variance of the two images on the left would show that Image C had higher variance than Image B, even though the contrast magnitude of the edge was the same. These images are thresholded to find the values in the filtered image that exceed a particular value. The two images on the right show what parts of the image are left after thresholding. The variance per unit area of only the region that exceeds this threshold is then calculated, which provides a measure of the "rectified contrast" at this particular scale. This measure is calculated at different scales (i.e., different spatial frequencies), and the slope of this "spectrum" is defined as the RCS. In the example shown, the edge of image C covers a greater area than the edge in B. Nonetheless, this thresholded measure will determine that the magnitude of the edge was the same for the two images. By this means, the RCS provides a measure of the average amplitude of the edge, independent of the area covered by the edge.

EXPERIMENT: PREDICTING BLUR

In this section we investigate whether the RCS can be used to predict perceived blur in images. As noted earlier, the issue of how the visual system detects blur has been the subject of a number of studies (e.g., Walsh & Charman, 1988; Hess et al., 1989; Tolhurst & Tadmor, 1997). However, these studies concentrate on how the visual system discriminates different degrees of blur. Here, we ask the question as to how the visual system identifies absolute blur. That is, given a complex image such as a natural scene, it is clearly possible at some level to identify the image as out of focus without a reference image. There may be various cues to indicate blur under conditions of actual optical blur (e.g., chromatic aberrations, contrast as a function of accomodation). However, here we concentrate on the spatial information available to an observer in a static image. In this study, we investigate whether the RCS measure is sufficient to predict observers' judgments of blur when the blur is computer generated. As shown above, the slope of the spectra of the natural scenes is quite variable, even when efforts are made to have the image in focus and, therefore, the amplitude spectrum alone should not be a good prediction of blur across scenes.

In this study, we introduce blur by altering the slope of the amplitude spectra of the images and attempt to determine whether the RCS analysis is capable of predicting blur in a scene, despite the variation in the slopes of the amplitude spectra between images. If the RCS is a good measure of the amplitude of structure as a function of frequency, then the RCS should be capable of predicting blur in spite of the variability in the amplitude spectra of these images.

Method

Stimuli. The 20 natural scenes described above were adjusted in their luminance range to fit within the limits of the computer screen. The images were windowed as described above to limit distortions at the edges. In addition to the natural scenes, three images consisting of a disc, the letter T (shown above), and a "plus sign" were analyzed. These three were selected because they had relatively steep spectra. An image of white noise was also added, creating a set of 24 images. Each of these 24 images was filtered to have amplitude spectra with slopes that varied between -2.0 and -0.1 in steps of 0.05. This was achieved by multiplying each of the original spectra by a factor of f^{-n}, where n was equal to the sum of the slope of the original image and the desired slope factor. The total variance (RMS contrast) of each image sequence was set to a constant which maximized the image contrast for the screen. This created 24 sequences of 20 filtered images along the lines of the sequence shown in Fig. 6. Images were presented at a distance of 0.5 m and subtended an angle of 17 deg. The mean luminance of the screen was 6.5 cd/m^2.

Procedure. For the estimates of blur by the model, the RCS was calculated for each of the images in the sequence described above. An image was judged to be blurred when the RCS had a slope steeper than -0.1. To estimate blur by human subjects, 12 subjects were asked to select the point in the image sequence at which the images appeared blurred. For each image sequence, the subjects began at a random point in the sequence and by means of the mouse button, could move up or down through the sequence until they had reached the point at which they were confident that the image "just began to appear blurred". Each subject thus provided a single data point for a given image sequence. As described above, the natural scenes were given a sinusoidal edge of 24 pixels to reduce the edge effects. All calculations and the judgments of blur were performed on this windowed image. Twelve subjects with normal or corrected to normal vision participated. It should be noted that although the subjects had little difficulty with this task, this method is not equivalent to optical blur. Since the total power remains constant for all the images in a set, an increase in the contrast at high frequencies is accompanied by a reduction in the contrast at low frequencies. However, as the results will show, the total power in any given band is not a sufficient cue to judge blur.

Results

Figure 11(A) shows the slope at which the images appeared plotted against the slope of the image before filtering. The error bars represent the standard errors for the 12 subjects. As one can see, the original "in focus" images have considerable variation in their slope. Nonetheless, with the exception of the "noise image", the subjects fairly reliably described the image as blurred when the slope was slightly steeper than the slope of the original image. This is in line with the results of Tadmor & Tolhurst (1994). For example, the original slope of the "letter T" stimulus was -1.62, but was not considered blurred until, on average, the slope was -1.68. It is clear that subjects are not simply using the slope of the amplitude spectrum to determine blur. With the noise image, however, there is no defined point of where the image can be described as "in focus". Although the variability among subjects was higher than for natural scenes, subjects, on average, judged the image to be blurred when it exceeded $1/f$. Introducing correlations in the noise does not produce the perception of blur until the correlations exceed the correlations found in typical scenes of the natural environment.

Figure 11(B) shows the prediction of blur from the model plotted against the results of the 12 subjects. Although the model is not as accurate as the subjects in predicting blur, the model does make reasonable predictions for the different images. Again, the amplitude spectrum would provide a poor prediction of blur since it would provide the same predicted blur slope for all the images. For each image, the model estimated which images in the sequence of filtered images should be judged as blurred, based on the slope of the RCS. It should be remembered that the model does not have the original stimulus as a reference. It is judging blur entirely by the spatial structure within each image—not by
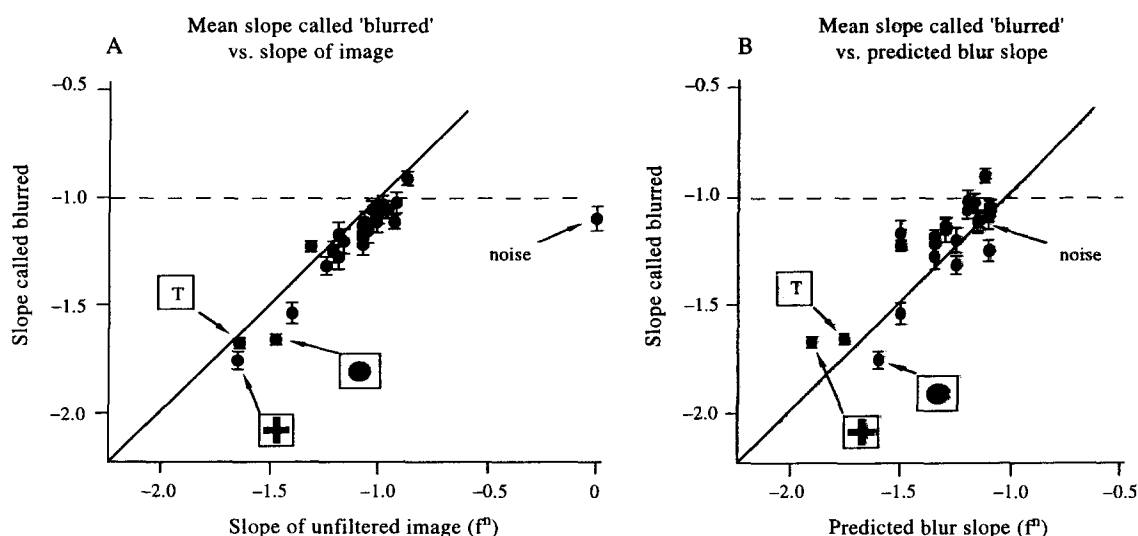
FIGURE 11. (A) The slope at which subjects judged the 24 images as beginning to appear blurred is plotted against the slope of the original image. Results are for 12 subjects and error bars represent standard errors. The three computer-generated images with steep spectra are noted. These results suggest that with the exception of the noise image, the images appeared blurred when the slopes were slightly steeper than the original image. Clearly, the slope of the amplitude spectrum does not predict blur since the slope judged to be blurred varies from image to image. (B) Results comparing perceived blur for the 12 subjects with the calculations of blur by the RCS. Although these images had significant variability in their slopes, the measure provided a reasonable prediction of the psychophysical results. Notice that the model predicts that a white noise pattern will appear blurred when its spectrum becomes steeper than $1/f$. This is in agreement with the subject's responses.

comparing it to the original. Although the slope of the amplitude spectrum alone is a poor predictor of blur, the RCS slope appears to be able to capture at least a portion of the information relevant to blur (i.e., the contrast of structure across scale). These results therefore show that this RCS measure is capable of using information other than that shown by the amplitude spectra. The model appears to make a reasonable prediction of blur in these images.

What is also quite interesting is the result with the noise pattern. This is an image without edges or other phase structure. Notice that for the white noise image, the slope of the original image was 0.0 (i.e., flat) but the model predicts that the image is not blurred until it becomes steeper than -1.0, in agreement with the subjects' estimates.

DISCUSSION

The results above show that the amplitude spectrum of natural scenes is not sufficient to predict when an image is in focus or when human observers will judge an image to be blurred. The slope of the amplitude spectrum is a result of both the amplitude of the structure at different frequencies (e.g., contrast of edges) as well as the density of the structure (number of edges). Blur, however, appears to depend entirely on the amplitude of the structure. This may help to explain why images under scotopic conditions or low contrast conditions do not appear blurred. It is not the absolute amplitude of the structure at high frequencies which determines blur. Rather, it is the relative amplitude of the detectable structure at different frequencies. Only when the contrast of the detectable

structure (e.g., edges) at high frequencies falls below that of the low frequencies is blur perceived.

It is not clear whether these result can be applied to those studies that have looked at a human observer's ability to discriminate blur. Tolhust et al. (1996) and Tadmor & Tolhurst (1994) have shown that discrimination thresholds of different natural scenes varied considerably from scene to scene. It is possible that the RCS measure described here would provide a better account of this variability. However, the task of judging which images in sequence appear blurred may well involve a different set of constraints than those required to discriminate instances of the sequence.

It should also be noted that the RCS method described above is fundamentally limited in that it is a global measure. A more accurate measure of blur will certainly involve local measures and will probably be best calculated on an edge by edge basis (i.e., by tracking the amplitude of the individual edge across scale: e.g., Bergholm, 1987; Elder & Zucker, 1997). Since our visual system can certainly accommodate to a local region of the image, the global measure will certainly be limited in predicting blur. Indeed, the fact that the subjects were more accurate at predicting blur than the model demonstrates that more information than that described in the model must be involved. We believe that by using the basic model to calculate blur locally and comparing the local edge magnitude across scale will produce good results. Indeed, one of the advantages of such an approach is that it allows the individual edges that are blurred to be discriminated from edges that are low contrast. In comparison, a measure which calculates only the energy at the highest frequencies, will confound contrast with blur.

It has been suggested that the chromatic amplitude spectra of natural scenes appears to be steeper than the achromatic spectra like that described here (Burton & Moorehead, 1987). This is certainly an area that deserves further study. However, the results here provide two possible explanations. First, of course, the chromatic aberrations of the lens may limit the extent to which different wavelengths might be simulataneously in focus. However, a second interesting possibility is that the chromatic amplitude spectra may be more sparse at the higher frequencies. For example, consider an image of a tree. The leaves will all have similar chromatic properties. However, the luminance will depend on the angle each leaf has, relative to the lightsource. Therefore, we might expect the chromatic edges to be of much lower amplitude within the tree than the achromatic edges. The high amplitude edges would be mainly found along the perimeter of the tree. Thus, the chromatic spectra may be steeper because there will be fewer high contrast chromatic edges in images that have regions of uniform chromaticity.

We are certainly not claiming that this approach provides a complete account of how observers calculate blur. For example, a wide range of factors are important in understanding both accommodation and emmetropization (e.g., chromatic aberrations, the phase of fringes etc.). Furthermore, the method of producing blur by altering the slope of the spectrum is not a good model of optical blur. And as noted, this is not a model of how observers "discriminate" blur. Rather, the goal of these studies was to determine the source of the variability found in the slopes of the amplitude spectra of natural scenes, and determine whether this information could be used to identify blur in these scenes.

rectified contrast spectrum (RCS) provides a means of separating the amplitude of the structure from the density of the structure. By calculating the amplitude of only those regions which exceed a predetermined threshold, our results show that one can provide a reasonable prediction of when an image is blurred. This approach suggests that a significant portion of the variability found in the amplitude spectra of natural scenes is due to the variability in the density of structure at different scales— not simply the amplitude of that structure. Ruderman's (1996, 1997) approach of modeling the $1/f$ structure of the environment in terms of a sum of occluding surfaces will likely provide further insights into the interactions of element density as a function of scale.

Although the results predicting blur with this model appear promising, it is acknowledged that the global measure is fundamentally limited. In the real world, the limited depth of focus of the eye will produce some regions that are in focus, with others out of focus. Clearly, the method derived here must be extended to track individual edges across scale in order to provide a more accurate account of how observers judge blur. However, there are some cases where a global model might be useful. In emmetropization (e.g., Troilo & Wallman, 1991), eye growth appears to be modulated by the optical image quality (i.e., focus). The information that the visual system uses to judge overall image focus is still a matter of debate. Although further research is required, we believe it an interesting question whether this system uses anything analogous to the methods in this paper to estimate the general state of defocus found within the retinal image.

SUMMARY

We conclude that the main insight one gains into the visual system by understanding the amplitude spectrum of natural scenes is in understanding the overall sensitivity of visual neurons. In line with Brady & Field (1995), it is proposed that the "vector magnitude" of visual neurons increases in proportion to frequency. With images that have spectra which decrease in proportion to frequency (i.e., $1/f$), this results in a flat response across the spectrum (out to approximately 20 cyc/deg). It is proposed that the principle advantage of this flat response is that it allows the information at different scales to be represented by neurons with similar dynamic range (i.e., the same level of noise).

However, the amplitude spectrum provides only a partial description of the relative activity of cells at difference scales. Some images have steep slopes because the structure at high frequencies has relatively low contrast. In such a case the image is blurred and the cells responding at high frequencies will have relatively low responses. For some images with steep spectra, the high frequency regions have relatively few regions covered by high frequency structure, even though that structure may have the same contrast as lower frequency structure. The

REFERENCES

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network, 3*, 213–251.

Atick, J. J. & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation, 4*, 449–572.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W. A. (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

Bergholm, F. (1987). Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 9*, 726–741.

Brady, N. & Field, D. J. (1995). What's constant in contrast constancy?: the effects of scaling on the perceived contrast of bandpass patterns. *Vision Research, 35*, 739–756.

Brady, N., Bex, P. J. & Fredericksen, R. E. (1997). Independent coding across spatial scales in moving fractal images. *Vision Research, 37*, 1873–1883.

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

Carlson, C. R. (1978). Thresholds for perceived image sharpness. *Photographic Science and Engineering, 22*, 69–71.

Croner, L. J. & Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Research, 35*, 7–24.

Elder, J. H., & Zucker, S. W. (1997). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press.

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, 4*, 2379–2394.

Field D. J. (1993). Scale-invariance and self-similar "wavelet" trans-

forms: an analysis of natural scenes and mammalian visual systems. In Farge, M., Hunt, J. & Vassilicos, J. C. (Eds), *Wavelets, fractals and Fourier transforms*. Oxford: Oxford University Press.

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

Field, D. J. and Brady, N. (1997). "Where is the peak of visual sensitivity" In Preparation.

Hammerly, J. R. & Dvorak, C. A. (1981). Detection and discrimination of blur in edges and lines. *Journal of the Optical Society of America, 71*, 448–452.

Hancock, P. J., Baddeley, R. J. & Smith, L. S. (1992). The principal components of natural images. *Network, 3*, 61–7011.

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

Hess, R. F., Pointer, J. S. & Watt, R. J. (1989). How are filters used in fovea and parafovea. *Journal of the Optical Society of America A, 6*, 329–339.

Kingdom, F. & Mouldon, B. (1992). A multi-channel approach to brightness coding. *Vision Research, 32*, 1565–1582.

Knill, D. C., Field, D. J. & Kersten, D. (1990). Human discrimination of fractal images. *Journal of the Optical Society of America A, 7*, 1113–1123.

Kretzmer, E. R. (1952). The statistics of television signals. *Bell System Technical Journal, 31*, 751–763.

Olshausen, B. A. & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609.

Peli, E., Goldstein, R. B., Young, G. M., Trempe, C. L. & Buzney, S. M. (1991). Image enhancement for the visually impaired. *Investigative Ophthalmology and Visual Science, 32*, 2337–2350.

Pelli, D. G. (1990). The quantum efficiency of vision. In Blakemore, C. (Ed.), *Visual coding and efficiency* (pp. 3–24). Cambridge: Cambridge University Press.

Ruderman, D. L. (1994). The statistics of natural images. *Network, 54*, 517–548.

Ruderman, D. L. (1996). Origins of scaling in natural images. *SPIE—Human Vision and Electronic Imaging, 2657*, 120–131.

Ruderman, D. L. (1997). Scaling in the Woods. *Vision Research* (in press).

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

van der Schaaf, A. & van Hateren, J. H. (1996). Modeling the power spectra of natural images: statistics and information. *Vision Research, 36*, 2759–2770.

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, 216*, 427–459.

Switkes, E., Mayer, M. J. & Sloan, J. A. (1978). "Spatial frequency analysis of the visual environment: anistropy 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. & Arthurs, G. (1996). The detection of changes in the amplitude spectra of natural images is explained by a band-limited local-contrast model. *SPIE—Human Vision and Electronic Imaging, 2657*, 154–164.

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

Tolhurst, D. J. & Tadmor, Y. (1997). "Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra of natural scenes". *Vision Research* (in press).

Tolhurst, D. J. & Thompson, I. D. (1981). On the variety of spatial frequency selectivity shown by neurons in area 17 of the cat. *Proceedings of the Royal Society of London B, 213*, 183–199.

Troilo, D. & Wallman, J. (1991). The regulation of eye growth and refractive state: an experimental study of emmetropization. *Vision Research, 31*, 1237–1250.

Walsh, G. & Charman, W. N. (1988). Visual sensitivity to temporal change in focus and its relevance to the accommodation response. *Vision Research, 28*, 1207–1221.

Watt, R. J. & Morgan, M. J. (1983). The recognition and representation of edge blur: evidence for spatial primitives in human vision. *Vision Research, 23*, 1465–1477.