Contrast statistics for foveated visual systems: fixation selection by minimizing contrast entropy

Raghu Raj

Department of Electrical and Computer Engineering and Center for Perceptual Systems, University of Texas at Austin, Austin, Texas 78712

Wilson S. Geisler and Robert A. Frazor

Department of Psychology and Center for Perceptual Systems, University of Texas at Austin, Austin, Texas 78712

Alan C. Bovik

Department of Electrical and Computer Engineering and Center for Perceptual Systems, University of Texas at Austin, Austin, Texas 78712

Received January 7, 2005; accepted February 24, 2005

The human visual system combines a wide field of view with a high-resolution fovea and uses eye, head, and body movements to direct the fovea to potentially relevant locations in the visual scene. This strategy is sensible for a visual system with limited neural resources. However, for this strategy to be effective, the visual system needs sophisticated central mechanisms that efficiently exploit the varying spatial resolution of the retina. To gain insight into some of the design requirements of these central mechanisms, we have analyzed the effects of variable spatial resolution on local contrast in 300 calibrated natural images. Specifically, for each retinal eccentricity (which produces a certain effective level of blur), and for each value of local contrast observed at that eccentricity, we measured the probability distribution of the local contrast in the unblurred image. These conditional probability distributions can be regarded as posterior probability distributions for the "true" unblurred contrast, given an observed contrast at a given eccentricity. We find that these conditional probability distributions are adequately described by a few simple formulas. To explore how these statistics might be exploited by central perceptual mechanisms, we consider the task of selecting successive fixation points, where the goal on each fixation is to maximize total contrast information gained about the image (i.e., minimize total contrast uncertainty). We derive an entropy minimization algorithm and find that it performs optimally at reducing total contrast uncertainty and that it also works well at reducing the mean squared error between the original image and the image reconstructed from the multiple fixations. Our results show that measurements of local contrast alone could efficiently drive the scan paths of the eye when the goal is to gain as much information about the spatial structure of a scene as possible. © 2005 Optical Society of America

OCIS codes: 330.0330, 330.6110, 330.2210, 330.4060, 330.1800, 110.2960.

1. INTRODUCTION

Humans, like many other animals, have a retina with variable spatial resolution. Resolution is highest in a central region, the fovea, and declines smoothly in all directions. High-speed eye movements, and slower head and body movements, are used to direct the fovea at potentially relevant locations in the retinal image of the visual scene. This strategy of combining a variable-resolution retina with eye, head, and body movements is sensible because it minimizes total neural resources while providing both a wide field view and high spatial resolution. However, for this strategy to be effective the visual system needs sophisticated central mechanisms that take into account and exploit the continuously varying spatial resolution of the retina.

There is evidence that visual systems are often matched to the statistical properties of the natural scenes to which they are exposed[1–11] (for reviews see Simoncelli and Olshausen[12] and Geisler and Diehl[13]). Therefore, to gain some insight into the design requirements of the central mechanisms of foveated visual systems, we analyzed the effects of variable spatial resolution on the statistics of local contrast in natural images. (Here, we define the local contrast as the standard deviation of the image intensities within some small region, divided by the mean intensity within that region, i.e., the local rms contrast.)

Contrast is arguably the most fundamental local image property encoded by the retina and transmitted to the brain, and hence its statistics have received considerable attention. A number of studies have been concerned with measuring the distributions of local contrast in natural images and comparing these with the shape of contrast response functions in the eye,[1,6] lateral geniculate nucleus,[14] and primary visual cortex.[15,16] Other studies have characterized the distributions of contrast in different environments[17] and at the center of gaze.[18]

Like most other image properties, contrast is encoded with the greatest precision at the center of the fovea and with decreasing precision as the distance from the center of the fovea (the eccentricity) increases. Specifically, as eccentricity increases, the center sizes of ganglion cell receptive fields increase, blurring the retinal image and

© 2005 Optical Society of America

thereby effectively reducing local contrast and increasing contrast uncertainty. This fact motivated us to directly measure the effect of retinal blur on large numbers of natural images in order to determine the statistical relationship between effective contrast and the true unblurred contrast at different retinal eccentricities. Here, we show that to good approximation the mode ($\hat{c}$) of the posterior probability distribution of the unblurred contrast [i.e., the maximum *a posteriori* (MAP) estimate] is given by the simple formula

$$\hat{c} = kc\varepsilon + c, \tag{1}$$

where ε is the retinal eccentricity and k is a constant that depends on the patch size over which the local contrast (c) is computed. We also show that the average standard deviation (defined later) of the posterior probability distribution is given by

$$\bar{\sigma}^2 = (kc\varepsilon)^2 + {\sigma_0}^2, \tag{2}$$

where σ_0 is a small constant, and thus the contrast uncertainty (the differential entropy of the posterior probability distribution) is given by

$$h = \tfrac{1}{2}\log_2(2\pi e\bar{\sigma}^2). \tag{3}$$

These statistical properties of natural images will be derived and explained in Section 2.

As an example of how these statistical properties of natural images might be exploited by a foveated visual system, we have considered the task of selecting fixation locations, when the organism's goal is to encode images as well as possible with just a few fixations. Specifically, using Eqs. (1)–(3), we derive and evaluate a fixation selection strategy based on the principle of picking fixation locations that minimize the total uncertainty about the contrasts in the image (i.e., minimize the total contrast entropy). We decided to explore an algorithm that minimizes total contrast entropy because minimizing entropy is ideal under some circumstances and has proved useful in other applications.[19–22] We find that our algorithm works very well at reducing total contrast uncertainty and also works well at reducing the mean squared error (MSE) between the original image and the image reconstructed from the multiple fixations.

2. METHODS AND RESULTS

This section describes the measurements of the contrast statistics and the algorithm for fixation selection based on those statistics.

A. Contrast Statistics

The effects of retinal blur on local contrast were measured using a set of calibrated natural images. The image set consisted of 300 rural images (i.e., minimum of man-made objects or animals) obtained from a publicly available image database.[7] The images were selected to be as diverse as possible given the data set. The images were obtained with a Kodak DCS420 digital camera and were calibrated to result in approximately 12 bit values that are linear with respect to the luminance. The 1536 by 1024 images were cropped to the center 1024 by 1024 pixels. Van Hateren and van der Schaaf[7] report that each pixel corresponds to approximately 1 arc min, and thus the cropped images are approximately 17° × 17°.

The contrast sensitivity functions of the human visual system, at different retinal eccentricities, have been measured for transient stimuli.[23–25] Measurements made under transient stimulus conditions are appropriate in the present context because fixation durations are brief (200–300 ms) under most natural viewing conditions. These contrast sensitivity functions are adequately described by the formula[26]

$$C(f,\varepsilon) = C_0 \exp\left(-\alpha f \frac{\varepsilon_2 + \varepsilon}{\varepsilon_2}\right), \tag{4}$$

where α is a constant ($\alpha \cong 0.1$), ε_2 is the retinal eccentricity where spatial resolution falls to half of what it is in the center of the fovea ($\varepsilon_2 \cong 2.3°$), and C_0 is a constant that controls the maximum contrast sensitivity. The contrast sensitivity functions described by Eq. (4) are consistent with the increase in center size of the retinal ganglion cells (midget ganglion cells) with eccentricity,[24,25] and hence Eq. (4) can be used to estimate the reduction in effective contrast as function of eccentricity. Note that the blur produced by the retina (as reflected in ganglion cell center sizes) is a result of both optical and neural factors.

To simulate the blur produced by the retina at different eccentricities, we filtered each of the 300 natural images with radially symmetric transfer functions obtained by setting $C_0 = 1.0$ and $f = (f_x^2 + f_y^2)^{1/2}$ in Eq. (4). Specifically, for each image we padded it appropriately, took the Fourier transform, multiplied the result by Eq. (4), and then took the inverse Fourier transform. Blurred images were obtained for eccentricities (ε) of 0, 1, 2, 4, 8, and 16 deg. The filtered images at an eccentricity of 0 deg were taken to be the unblurred reference images. This was done because the optical transfer function of the camera is unknown (but presumably very good), and hence the raw image cannot be taken to be the effective retinal image in the center of the fovea. Because the unblurred image was taken to be the filtered image with $\varepsilon = 0$, the value of C_0 is irrelevant and hence could be set to 1.0, as we did.

In order to characterize the statistical relationship between effective contrasts at different eccentricities, we measured local contrasts in each image, for all six levels of blur. A large number of local contrasts were sampled randomly from each of the 300 natural images. The locations of the samples were different for each natural image but were the same for each level of blur. The local contrasts were measured in image patches formed by windowing with a circularly symmetric raised-cosine weighting function:

$$w_i = 0.5\left\{\cos\left[\frac{\pi}{p}\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}\right] + 1\right\}, \tag{5}$$

where p is the patch radius, (x_i, y_i) is the location of the ith pixel in the patch, and (x_c, y_c) is the location of the center of the patch. (Note that the half-height diameter of the window equals the patch radius.) The results reported here are for a patch diameter of 32 pixels (0.53 deg), but similar results are obtained with other patch sizes. The

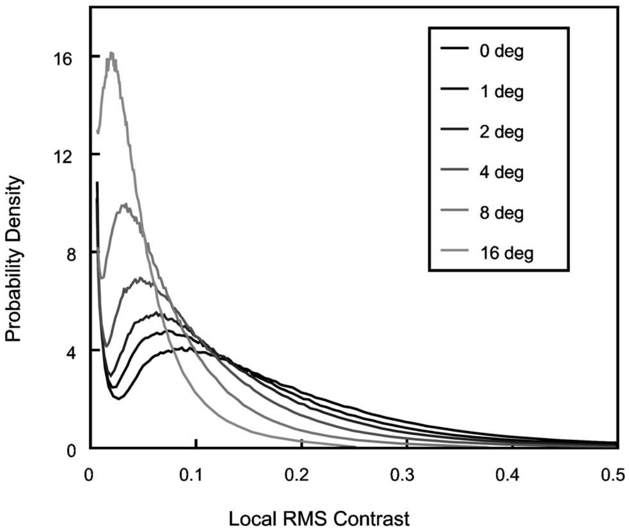

Fig. 1. Probability distributions of local rms contrast for various levels of blur based on the human contrast sensitivity function at different retinal eccentricities. These distributions were obtained by randomly sampling small patches from 300 calibrated natural images.

local contrast was defined by the formula

$$c = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i}\sum_{i=1}^{N} w_i \frac{(L_i - L)^2}{(L + L_0)^2}}, \tag{6}$$

where N is the number of pixels in the patch; L_i is the luminance of the ith pixel; L is the local mean luminance,

$$L = \frac{1}{\sum_{i=1}^{N} w_i}\sum_{i=1}^{N} w_i L_i; \tag{7}$$

and L_0 is a dark light parameter, chosen to be 7 td (1 cd/m^2, assuming a 3 mm pupil), based on human photopic intensity discrimination data.[27] (We note that L_0 had very little effect on the measured contrasts because the mean luminances of the images were generally much higher than 1 cd/m^2.)

Figure 1 shows the estimated probability distributions of local contrast for each level of blur. The distributions have been truncated at a contrast of 0.005 because humans cannot detect contrasts below that value and because the measurements become contaminated by camera or pixel noise. Not surprisingly, as the level of blur (retinal eccentricity) increases, the distributions shift toward lower contrasts. The rise in the function at low contrasts appears to be due to the patches of sky in many of the natural images.

For many visual tasks (including the fixation selection task), one would like to estimate the unblurred contrast from the blurred contrast observed at the given retinal eccentricity. Thus, the statistics of most relevance are the conditional probability distributions for the unblurred contrast given the observed contrast (i.e., the posterior probability distributions). We computed these distributions for a wide range of blurred contrasts, for eccentricities 1, 2, 4, 8, and 16 deg. Several representative distributions are shown in Fig. 2. Each row shows the conditional probability distributions for a different eccentricity, and each plot within a row shows the distribution for a particular value of blurred contrast observed at that eccentricity. There are several clear trends in the data: (1) As eccentricity increases, the peaks of the distributions shift to the right and (2) the widths of the distributions increase; and (3) as the observed blurred contrast increases, the peaks of the distributions shift to the right and (4) the widths of the distributions increase.

To quantify these trends, we fit the empirical distributions with descriptive functions. In general, the distributions are not Gaussian, but they are nicely fit by Gaussian distributions with different standard deviations above and below the mode (skewed Gaussian distributions). The solid curves show the fits to this sample of empirical distributions; the quality of these fits is representative of the whole set. The skewed Gaussian has three parameters: the mode, which we will label $\hat{c}$ because it is the MAP estimate of the unblurred contrast, and two standard deviations, σ_l and σ_h. Figure 3 plots the mode and the average standard deviation, $(\sigma_l + \sigma_h)/2$, for all eccentricities and observed levels of blurred contrast. Measurements outside these ranges were unreliable because the numbers of samples became too small. The solid lines in the figure are best-fitting straight lines through the origin. Although the fits are not perfect, the straight lines summarize the data very well. In other words, to close approximation, both the mode and the standard deviation of all the posterior probability distributions increase in direct proportion to the observed blurred contrast.

What is also clear in Fig. 3 is that the slopes of the best-fitting lines (the proportionality constants) increase with retinal eccentricity. Figure 4 plots the estimated slopes for the modes and the average standard deviations. The straight line in Fig. 4A is the best-fitting line with a intercept of 1.0, and the straight line in Fig. 4B is the best-fitting line through the origin. Again, the fits are not per-

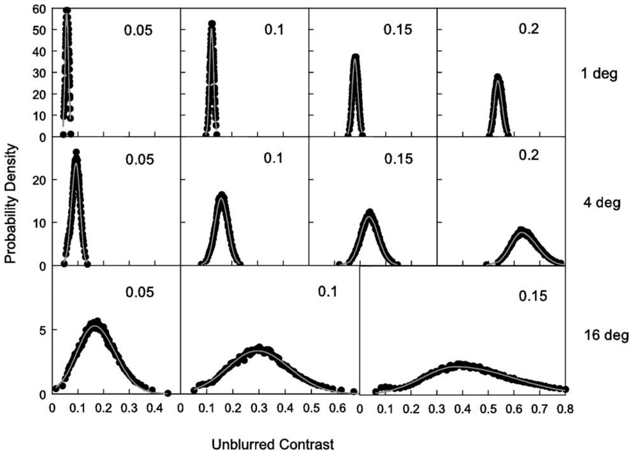

Fig. 2. These plots show examples of the conditional probability distributions of local rms contrast in unblurred images, given the local rms contrast in the blurred versions of the images (columns) and given the retinal eccentricity (rows). The solid symbols are empirical histograms computed from 300 natural images that contained no man-made objects. The smooth curves are the best-fitting skewed Gaussian distribution (a Gaussian with different standard deviations above and below the mode).

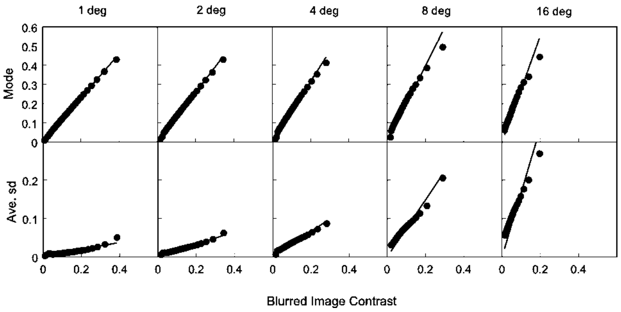

Fig. 3. Modes and average standard deviations of the conditional probability densities are plotted as a function of blurred image contrast and retinal eccentricity. The average standard deviation is the average of the two standard deviation parameters in the skewed Gaussian distribution. See Fig. 2 for examples of the conditional densities and fits of the skewed Gaussian distribution. The curves are best-fitting straight lines through the origin.

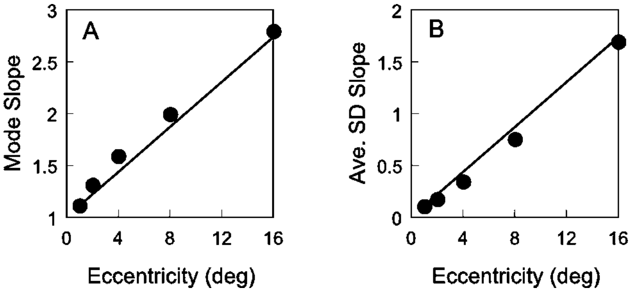

Fig. 4. Slopes of the linear functions in Fig. 3. A, Slope of the contrast versus mode plot as a function of retinal eccentricity. B, Slope of the contrast versus average standard deviation plot as a function of retinal eccentricity. The curves show the predictions of the linear model: $\hat{c}=k\varepsilon c+c$ and $\bar{\sigma}=k\varepsilon c$, where $k=0.105$.

fect, but they do provide a very good summary of the data. Taken together, Figs. 2–4 show that the mode across all conditions is closely approximated by Eq. (1) and the average standard deviation across all conditions is closely approximated by Eq. (2), where $\bar{\sigma}=(\sigma_l+\sigma_h)/2$.

Differential entropy is a fundamental measure of the uncertainty associated with a probability distribution.[28] In Appendix A we show that the differential entropy of a skewed Gaussian distribution is equal to Eq. (3), and hence the differential entropy of the posterior probability distributions (the contrast uncertainty) for the range of eccentricities considered here is closely approximated by substituting Eq. (2) into Eq. (3). The constant σ_0^2 in Eq. (2) reflects the fact there must always be some intrinsic uncertainty about contrast, if for no other reason than photon and sensor or neural noise. Although the constant cannot be estimated from the contrast measurements, it is necessary for it to have a value greater than zero in the fixation selection algorithm; its specific value is not important as long as it is small (see Appendix A).

B. Fixation Selection

We have found a surprisingly simple statistical relationship, for natural images, between the contrast observed at a given retinal eccentricity and the posterior probability distribution of the unblurred true contrast at that location. This relationship, which is described by Eqs. (1)–(3), could be exploited by a visual system to efficiently select fixation locations under certain circumstances. For example, if the goal in some situations is not to search for a particular target or set of targets but simply to gain as much information as possible about the image on each fixation, then a potentially effective strategy would be to pick successive fixations that maximally reduce the total contrast uncertainty about the image. This strategy might be particularly effective if there is a strong correlation between the uncertainty about local contrast and the total uncertainty about the local image structure. To begin exploring this possibility, we have developed an algorithm (a model observer) that selects fixations based on Eq. (1)–(3). Here, we describe the algorithm, then we describe the algorithm's fixation selections on some example images, and finally we compare the algorithm's absolute performance to appropriate ground-truth measurements.

1. Contrast Entropy Minimization Algorithm

We assume that the first fixation is at some arbitrary image location (e.g., at the center of the image). On making this first fixation the observer receives a foveated neural image, where spatial resolution is highest at the fixation point and falls off smoothly in all directions. From this first neural image the observer forms three maps that will be updated after each fixation. The first is an eccentricity map, which stores, for each image pixel, the smallest distance the pixel has been from the center of the fovea. The second is a contrast map, which stores the local rms contrast measured at each pixel, when the pixel was at its smallest distance from the center of fovea. (The contrast at a pixel is defined to be the contrast of the patch centered on that pixel.) The third is an uncertainty map, which stores the contrast uncertainty (entropy) at each pixel [given by Eq. (3)], when the pixel was at its smallest distance from the center of fovea. These three maps cumulate all the relevant information obtained during the sequence of fixations. The sum of all the uncertainties in the uncertainty map is the total contrast uncertainty. The aim of the algorithm is to select the next fixation that will minimize this total contrast uncertainty. To do this, the algorithm considers every possible next fixation location. For each possible fixation location, the algorithm uses the current maps and its knowledge of the posterior probability distributions for contrast [Eqs. (1)–(3)] to estimate the reduction in total contrast uncertainty. It then picks the fixation location with the largest estimated reduction. A formal derivation of the contrast entropy minimization (CEM) algorithm is given in Appendix A.

2. Performance of the Contrast Entropy Minimization Algorithm

The performance of the CEM algorithm was evaluated on 16 natural images selected to be representative of the van Hateren and van der Schaaf[7] data set. Thumbnails of these images are shown in Fig. 5.

We simulated a foveated visual system that approximately matched the human visual system by using radially symmetric transfer functions corresponding to human contrast sensitivity functions [cf. Eq. (4)]:

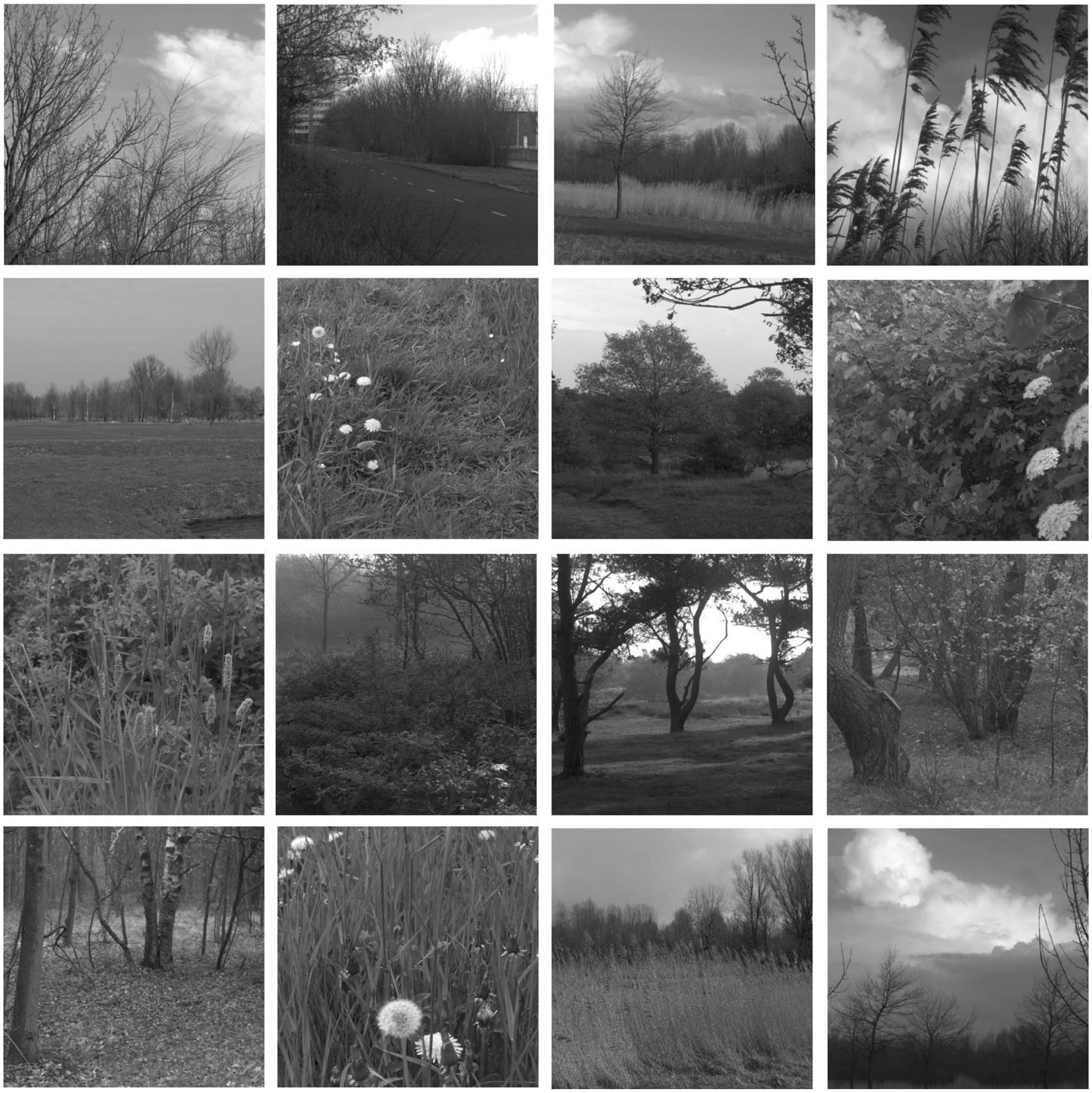

Fig. 5. Images used to test a fixation selection algorithm based on the principle of minimizing contrast entropy.

$$F(f_x,f_y,\varepsilon) = \exp\left(- \alpha \sqrt{{f_x}^2 + {f_y}^2} \frac{\varepsilon_2 + \varepsilon}{\varepsilon_2} \right). \tag{8}$$

For each eccentricity the inverse Fourier transform of this transfer function specifies a linear filter kernel (a Laplacian function) that scales in size with eccentricity.

To speed the calculations, we made use of the fact the resolution of the human visual system declines smoothly as a function of eccentricity. By setting the left side of Eq. (8) to any constant resolution criterion, we see that resolution follows a smooth function of the form

$$r(\varepsilon) \propto \frac{\varepsilon_2}{\varepsilon + \varepsilon_2}.$$

The greatest eccentricity that needs to be considered for our 17° images is 12°, and hence the lowest relevant resolution is approximately 17% of the resolution in the fovea. Therefore, we partitioned the 17%–100% range into eight evenly spaced resolutions and then determined the eccentricity corresponding to each resolution. We then created eight transfer functions by substituting the eight eccentricities into Eq. (8). Before running the algorithm on a natural image, we used the eight transfer functions to obtain eight different resolution versions of the natural image. During the simulation, the foveated (neural) image at any given retinal eccentricity was obtained by linearly interpolating the two images whose resolutions bracketed the resolution at that eccentricity.

On each fixation during the simulation, the local contrasts in the neural image were measured using Eq. (6) for a patch diameter of 32 pixels. To speed the calculations, we sampled the local contrasts on a square lattice with a spacing of 16 pixels (the radius of the raised-cosine window). The overlap of the samples ensured that all image pixels contributed to the local contrast measurements (however, the algorithm performs similarly if there is no overlap between samples). The possible fixation locations and the three maps (contrast, eccentricity, and uncertainty) also corresponded to the same square lattice (i.e., 4096 possible fixation locations).

Figures 6A and 6C show the first nine fixations for two of the natural images. (Recall that the first fixation was always at the center of the image.) There are several trends evident in these fixation patterns. First, the fixations tend to land in or near relatively high-contrast regions. Notice, for example, how there are no fixations into the sky region of the image in Fig. 6A and how the second fixation is near a bright flower in Fig. 6B. This occurs because contrast uncertainty is greater in regions where the effective contrast is higher [see Eqs. (2) and (3)]. Second, the saccade lengths tend to be relatively large and variable in size; the mean and standard deviation of the saccade length for the 16 test images are 8.9° and 2.5°, respectively. The large saccades occur because contrast uncertainty increases with eccentricity [see Eqs. (2) and (3)]. Third, there are few fixations near the edge of the image. This occurs because fixating near the image boundary tends to reduce the total number of image pixels that benefit from being seen at a smaller eccentricity. For example, a fixation on the boundary implies that half the fovea falls outside the image, which tends to reduce the number of image pixels that can benefit from foveal viewing.

Figures 6B and 6D show quantitatively how well the algorithm performs in reducing total contrast uncertainty. The solid circles show the total contrast entropy predicted by the algorithm before the fixation was made, where the total contrast entropy has been normalized by its value after the first fixation in the center of the image. The open circles show the actual total contrast entropy observed after the fixation selected by the algorithm is

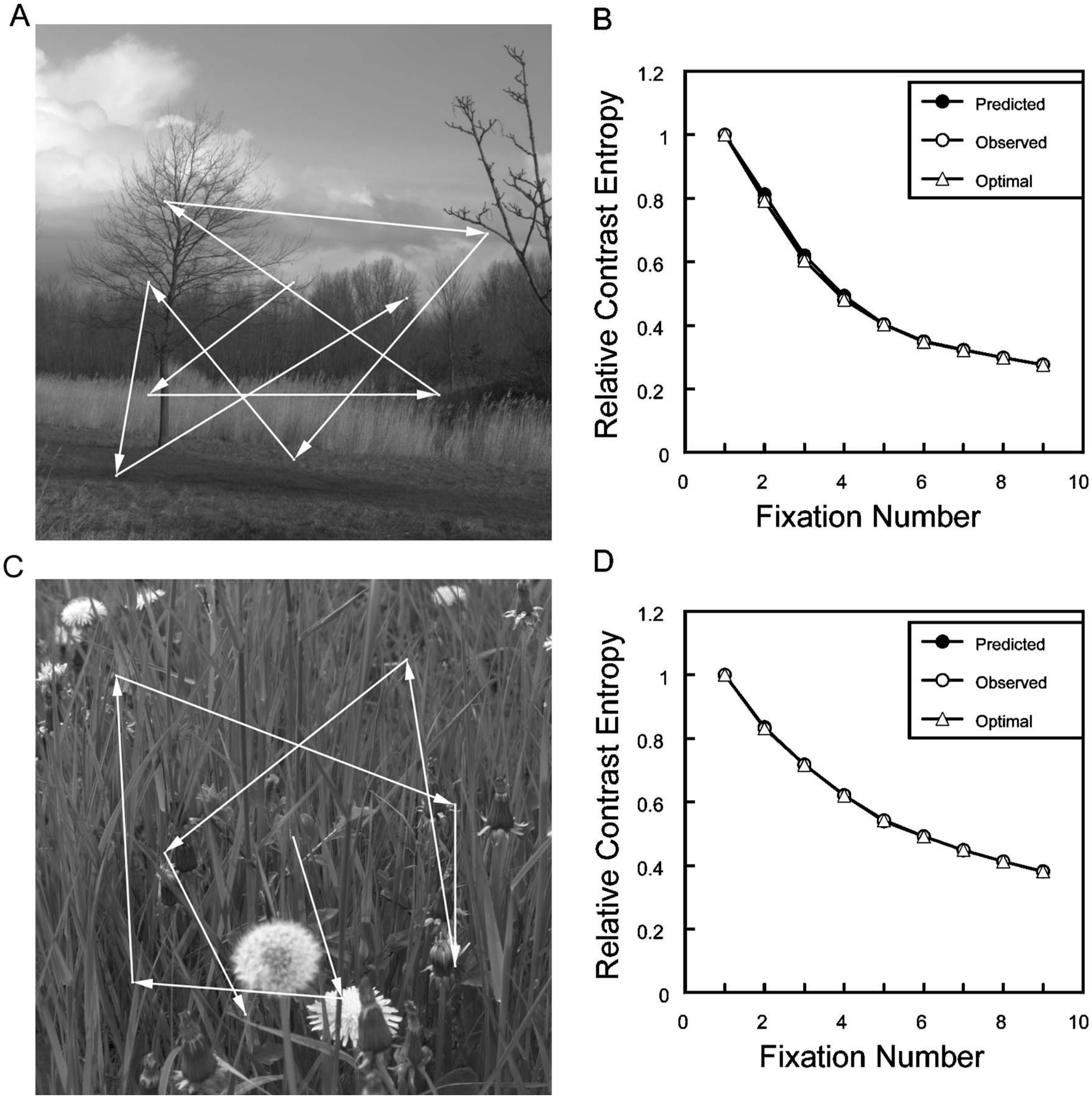

Fig. 6. Fixation points selected by the principle of minimizing total contrast entropy (contrast uncertainty), using the average local contrast statistics of natural images. A, Sequence of nine fixations (eight saccades) for a distant image containing sky, ground, and trees. B, Relative contrast entropy as a function of fixation number for the image in A (open circles), predicted relative contrast entropy before the fixation was made (solid circles), and optimal relative contrast entropy that could be obtained (open triangles). C, Sequence of nine fixations (eight saccades) for a close-up image containing foliage. D, Same type of plot shown in B.

made. In other words, the predicted entropy is the entropy estimated before the next eye movement is made, and the observed entropy is the entropy observed or computed after the next eye movement is made. As can be seen, the predicted and observed entropies are very similar. The open triangles show the lowest possible total contrast entropy that could have been obtained on the fixation. It was determined by literally making every possible fixation and computing the observed entropy. The actual observed entropy obtained by the algorithm is almost indistinguishable from optimal. The average results for all 16 images are shown in Fig. 7A. In general, the reduction in contrast entropy obtained by the CEM algorithm is essentially optimal. This is even more clearly illustrated by the solid circles in Fig. 7B, which plot the ratio of the optimal and observed entropies in Fig. 7A (the first fixation is excluded from the plot because the ratio is necessarily 1.0).

An obvious question is how well the CEM algorithm compares with alternatives. We consider two. The first algorithm tiles the image in a random order without replacement. Specifically, the image is divided into nine square regions (a 3 × 3 grid), and only fixations at the centers of these regions are allowed. During the scan, each square region is fixated only once, with the order of fixations being random. The average performance of this tiling algorithm is given by the open circles in Fig. 7B. It performs substantially worse than entropy minimization. The second alternative is purely random fixation (fixations are selected randomly from the 4096 possible locations). The performance of this algorithm is given by the open triangles in Fig. 7B. The random algorithm performs worse than the tiling algorithm. We conclude that the CEM algorithm does, in fact, optimally reduce the total contrast entropy on successive fixations for natural images and that it substantially outperforms some obvious alternatives.

We have demonstrated that the average contrast statistics of natural images can be used to sequentially select fixations that optimally reduce the total contrast uncer-

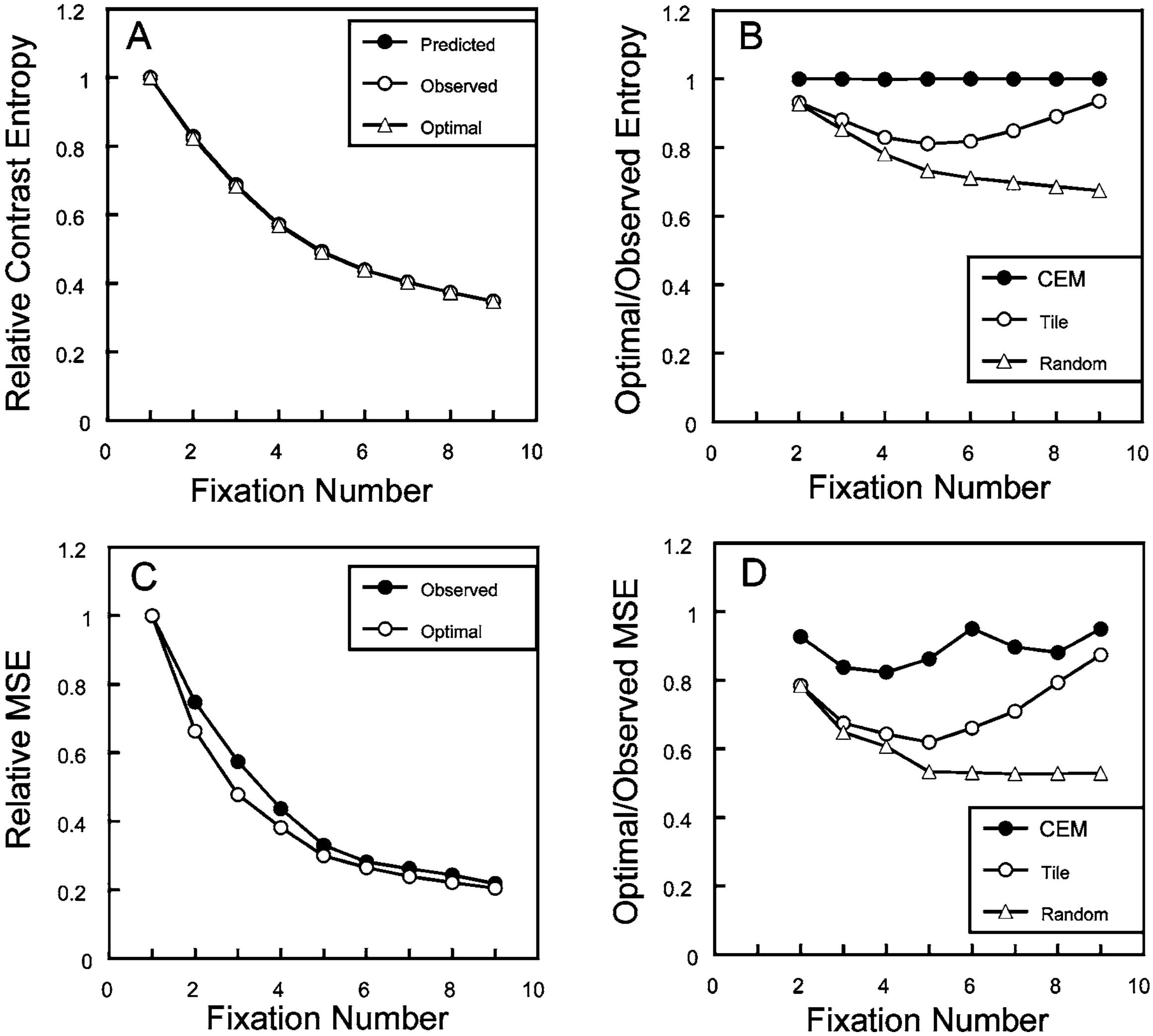

Fig. 7. Average fixation selection performance for the 16 test images in Fig. 5. A, Relative contrast entropy as a function of fixation number (open circles), predicted relative contrast entropy before the fixation was made (solid circles), and optimal relative contrast entropy that could be obtained (open triangles). B, Ratio of the optimal contrast entropy that could be obtained to the contrast entropy that was obtained: CEM algorithm (solid circles), tiling algorithm (open circles), random algorithm (open triangles). C, Relative mean squared error (MSE) between the original (unblurred) image and the image reconstructed from the fixations up to and including the fixation number given on the horizontal axis: CEM algorithm (solid circles), optimal (open circles). D, Ratio of optimal MSE that could be obtained to the MSE that was obtained: CEM algorithm (solid circles), tiling algorithm (open circles), random algorithm (open triangles).

tainty for individual images. Although this is a remarkable fact, contrast is just one local image property. Presumably, humans make fixations not just to reduce uncertainty about contrast but also to reduce uncertainty about many of the other image properties that determine local image structure (e.g., orientation, phase, and spatial frequency). It is not possible to measure the statistics for all local image properties in natural images, and hence it is not practical to develop a rigorous algorithm that selects fixations to reduce total image uncertainty. On the other hand, it is possible that uncertainty in contrast is strongly correlated with uncertainty for other image properties. For example, Schwartz and Simoncelli[29] found that the variances of many local image properties are strongly correlated, even for orthogonal image properties. Therefore, it is possible that minimizing contrast uncertainty would do a good job of minimizing uncertainty about many local image properties.

To evaluate this possibility, we used the mean squared error (MSE), between the original (unblurred) image and the image reconstructed from the sequence of fixations, as a measure of the total image uncertainty. The reconstructed image was obtained using the eccentricity map (the map showing the smallest distance that each pixel has been from the center of the fovea). Specifically, each pixel in the reconstructed image was set to the image gray level that was observed at that pixel for the eccentricity given in the eccentricity map. Thus, in the reconstructed image, every pixel keeps the highest resolution that has occurred so far in the sequence of fixations. (We note that for image reconstruction the eccentricity map was computed for all the 1024×1024 pixels' locations in the image; also, the MSE between the original and reconstructed images was computed over all 1024×1024 pixels.)

For each fixation made by the CEM algorithm, we computed the relative MSE (the MSE after the fixation divided by the MSE after the first fixation). The solid circles in Fig. 7C show the relative MSE as a function of fixation number, averaged across the 16 test images. For ground-truth comparison, we determined, for each fixation made by the CEM algorithm, the fixation that would have minimized the MSE (this was done by making every possible next fixation and computing the resulting MSE). The open circles in Fig. 7C show the optimal values of the MSE that could have been obtained. The solid circles in Fig. 7D show the ratios of the optimal MSE to the observed MSE obtained with the CEM algorithm. The average ratio is 0.9 (i.e., the obtained MSE is about 10% higher than optimal). The open circles and triangles show that the tile and random algorithms perform considerably worse than the CEM algorithm; the average ratio for the tile algorithm is 0.72 and for the random algorithm is 0.59. Thus, it appears that the CEM algorithm does a respectable job of selecting fixations that minimize total image uncertainty.

3. DISCUSSION

To gain insight into the design requirements of visual systems with foveated retinas, we measured the joint distribution of the local contrast in 300 natural images before and after blurring by amounts corresponding to different retinal eccentricities in the human visual system. The joint distribution at each retinal eccentricity is given by the marginal distribution of the blurred contrast (e.g., one of the distributions in Fig. 1) and by the distributions of the unblurred contrast conditional on the blurred contrast (e.g., one of the rows of distributions in Fig. 2). We find that the conditional distributions are described quite well by very simple formulas: The mode of the conditional distribution increases in proportion to the blurred contrast and the eccentricity [Eq. (1)], the average variance of the conditional distribution increases in proportion to the square of the blurred contrast and the square of the eccentricity [Eq. (2)], and the differential entropy of the conditional distribution increases in proportion to the logarithm of the average variance [Eq. (3)].

The image statistics reported here are for one particular analysis patch size (a width of 32 pixels). We find that Eqs. (1)–(3) also summarize the conditional probability distributions for other patch sizes quite well. However, as patch size decreases, the estimated value of the proportionality constant k in Eqs. (1)–(3) increases.

To explore how these natural scene statistics might be exploited by central perceptual mechanisms, we considered the task of selecting successive fixation points to optimize the total contrast information gained about the image (i.e., minimize total contrast entropy). On the basis of the average scene statistics represented by Eqs. (1)–(3), we derived a novel fixation selection algorithm: the CEM algorithm. Remarkably, we found that the average scene statistics for natural images (represented in the CEM algorithm) are sufficient to achieve nearly optimal fixation sequences for individual natural images (see Figs. 6, 7A, and 7B). Presumably, this optimal performance is achieved because each fixation is based on a global pooling of local contrasts from the entire image. In other words, even though there is considerable uncertainty about how much the contrast entropy will be reduced at any particular image location, there is little uncertainty about how much the average contrast entropy from all locations will be reduced. We also examined how well the CEM algorithm performed at reducing the MSE between the original image and the image reconstructed from the sequence of fixations. The MSE serves as a measure of total uncertainty about the original image. We find that the CEM algorithm also does quite well at reducing total uncertainty in individual images: The MSE values average about 10% higher than optimal (see Figs. 7C and 7D).

Although the CEM algorithm is quite simple and is based only on contrast statistics, it performs remarkably well at reducing total image uncertainty, and hence it may be of practical value in certain surveillance and robotic applications involving foveated imaging. For example, if there is time to make only a few fixations with a remote robotic or surveillance camera, then the CEM algorithm could be used to select those few fixations, assuming the goal is to reconstruct the image as accurately as possible. The algorithm is amenable to parallel computing and runs at a respectable speed (a fixation every couple of seconds) on a standard personal computer.

What are the implications of our results for understanding human fixation patterns? The first thing to point out is that human fixation patterns are highly task dependent. In reading, saccade lengths tend to be short and the fixation patterns stereotypical because, for the most part, words must be read sequentially for the communication to be understood.[21] In search tasks where the observer is trying to find a specific target or class of targets, saccade lengths tend to be longer and the fixation patterns more random than in reading because the eye is drawn to any likely target location in the image.[30,31] In general, human fixation patterns are probably different for every kind of perceptual or cognitive task that is performed.[32]

A class of tasks where the CEM algorithm might be a plausible model of human fixation patterns is scene memorization tasks. In such tasks, the goal is to learn as much as possible about a scene in a few fixations, so that the scene can be distinguished from other scenes at a later time. Picking fixations that minimize contrast entropy is a relatively simple and efficient way to gain information about the scene because the fixation selection requires no encoding of spatial structure, no pattern recognition, and little other high-level processing. Minimizing contrast entropy involves only encoding local contrasts and pooling them in a way that is weighted by the eccentricity and the contrast. This is the kind of processing that could be done in a fairly low level and automatic way, without placing great demands on high-level processes that require more attentional resources. What makes minimizing contrast entropy particularly appealing for this class of task is that it also does a good job of reducing total uncertainty about the image. Thus, selecting fixations by minimizing contrast entropy will, to good approximation, maximize the amount of image structure available to the cortex for extraction and storage in memory. The CEM algorithm makes detailed predictions about the statistics of fixation patterns in scene memorization tasks, and hence it should be testable.

A rational survival strategy for an organism might be to continuously work at gaining as much general information as possible about the local environment until situations arise where the organism needs to be engaged in a particular task or until the organism detects a particularly significant object. This background of information could provide the grist for efficient performance in many of the organism's specific tasks. If this principle is correct, then a fixation selection mechanism based on minimizing contrast entropy might work well as an automatic background or default mechanism that is overridden or modulated by more specific task demands or by particularly significant high-level content extracted during the fixations. Obviously, this is speculation, but it points to the real possibility that, in many cases, adequate models of human fixation patterns will require two or more very different fixation selection modules that interleave over time.

A primary aim of this study was to measure the contrast statistics of natural images for foveated visual systems. We have focused on the relevance of these statistics for fixation selection, but it is obvious that they must be of at least some relevance for many tasks that involve information integration or comparison across the visual field. The fact that the posterior probability distribution of the true unblurred local contrast is characterized by very simple formulas should make it possible to incorporate these natural scene statistics into various Bayesian models of perceptual performance.

APPENDIX A

1. Skewed Gaussian Probability Function

We define the skewed Gaussian to be a Gaussian with different standard deviations above (σ_h) and below (σ_l) the mode (u):

$$g(x;u,\sigma_l,\sigma_h)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\left(\dfrac{\sigma_l+\sigma_h}{2}\right)}\exp\left[\dfrac{-(x-u)^2}{2{\sigma_l}^2}\right] & x\leqslant u\\[2ex] \dfrac{1}{\sqrt{2\pi}\left(\dfrac{\sigma_l+\sigma_h}{2}\right)}\exp\left[\dfrac{-(x-u)^2}{2{\sigma_h}^2}\right] & x>u\end{cases}. \tag{A1}$$

2. Differential Entropy of the Skewed Gaussian Distribution

The differential entropy of a probability density function is defined by the integral

$$h(p)=\int_{-\infty}^{\infty}p(x)\ln[p(x)]\mathrm{d}x. \tag{A2}$$

Substituting Eq. (A1) into Eq. (A2) and letting $\phi_l(x)$ and $\phi_h(x)$ be Gaussian density functions with means u and standard deviations of σ_l and σ_h, we have

$$h(g)=\frac{2\sigma_l}{(\sigma_l+\sigma_h)}\int_{-\infty}^{u}\phi_l(x)\left\{\ln\left[\frac{(\sigma_l+\sigma_h)}{2\sigma_l}\right]+\ln(\sqrt{2\pi}\sigma_l)+\frac{(x-u)^2}{2{\sigma_l}^2}\right\}\mathrm{d}x+\frac{2\sigma_h}{(\sigma_l+\sigma_h)}\int_{u}^{\infty}\phi_h(x)\left\{\ln\left[\frac{(\sigma_l+\sigma_h)}{2\sigma_h}\right]+\ln(\sqrt{2\pi}\sigma_h)+\frac{(x-u)^2}{2{\sigma_h}^2}\right\}\mathrm{d}x,$$

$$h(g)=\frac{\sigma_l}{(\sigma_l+\sigma_h)}\left\{\ln\left[\frac{(\sigma_l+\sigma_h)}{2\sigma_l}\right]+\ln(\sqrt{2\pi{\sigma_l}^2})+\frac{1}{2}\right\}+\frac{\sigma_h}{(\sigma_l+\sigma_h)}\left\{\ln\left[\frac{(\sigma_l+\sigma_h)}{2\sigma_h}\right]+\ln(\sqrt{2\pi{\sigma_h}^2})+\frac{1}{2}\right\},$$

$$h(g)=\frac{\sigma_l}{(\sigma_l+\sigma_h)}\left\{\ln\left[\frac{\sqrt{2\pi}(\sigma_l+\sigma_h)}{2}\right]+\frac{1}{2}\right\}+\frac{\sigma_h}{(\sigma_l+\sigma_h)}\left\{\ln\left[\frac{\sqrt{2\pi}(\sigma_l+\sigma_h)}{2}\right]+\frac{1}{2}\right\},$$

$$h(g)=\ln\left[\frac{\sqrt{2\pi}(\sigma_l+\sigma_h)}{2}\right]+\frac{1}{2},$$

$$h(g)=\tfrac{1}{2}\log_2(2\pi e\bar{\sigma}^2).$$

3. Contrast Entropy Minimization Algorithm

Here we formalize the CEM algorithm. To begin with, let (x_i, y_i) represent the location of the ith pixel in the image, and let C_i be the true (unblurred) rms contrast at that location. We note that the term image location refers to a scene location expressed in degrees of visual angle in the horizontal and vertical directions.

Consider a series of fixations, $t=1, 2, \ldots$. Let the location of fixation number t be x_t, y_t, and let the observed local rms contrast at the ith pixel, on that fixation, be c_{it}. The retinal eccentricity, ε_{it}, of the ith pixel location is

$$\varepsilon_{it} = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2}. \tag{A3}$$

Thus, if the observer is currently on fixation number T, then the current eccentricity map is given by

$$\varepsilon_i(T) = \min_{t \leq T} \varepsilon_{it}. \tag{A4}$$

(Note that new values appear in the eccentricity map only if a new fixation happens to bring a pixel closer to the fovea than it has been before.) The current contrast map, $c_i(T)$, is defined to be the contrast that was observed when the eccentricity was at its minimum value, as given by the eccentricity map [Eq. (A4)]. The uncertainty map is given by

$$h_i(T) = \tfrac{1}{2}\log_2(2\pi e\{[k\varepsilon_i(T)c_i(T)]^2 + \sigma_0{}^2\}). \tag{A5}$$

The total uncertainty after fixation number T is made is

$$U(T) = \sum_{i=1}^{n} h_i(T). \tag{A6}$$

To select the next fixation, the observer considers each possible location (x_{T+1}, y_{T+1}) for fixation $T+1$, estimates the total contrast uncertainty that will be obtained if that fixation is made, and then picks the location $(\hat{x}_{T+1}, \hat{y}_{T+1})$ with the minimum estimated total uncertainty:

$$(\hat{x}_{T+1}, \hat{y}_{T+1}) = \arg \max_{x_{T+1}, y_{T+1}} [\hat{U}(T+1, x_{T+1}, y_{T+1})], \tag{A7}$$

where

$$\hat{U}(T+1, x_{T+1}, y_{T+1}) = \sum_{i=1}^{n} \hat{h}_i(T+1, x_{T+1}, y_{T+1}), \tag{A8}$$

$$\hat{h}_i(T+1, x_{T+1}, y_{T+1}) = \tfrac{1}{2}\log_2(2\pi e\{[k\varepsilon_i(T+1, x_{T+1}, y_{T+1}) \times \hat{c}_i(T+1, x_{T+1}, y_{T+1})]^2 + \sigma_0{}^2\}). \tag{A9}$$

To evaluate Eq. (A9), we note that the eccentricity map $\varepsilon_i(T+1, x_{T+1}, y_{T+1})$ for fixation location (x_{T+1}, y_{T+1}) is obtained directly from Eqs. (A3) and (A4). The estimated contrast map, $\hat{c}_i(T+1, x_{T+1}, y_{T+1})$, can be obtained from text equation (1). Specifically, Eq. (1) gives the maximum *a posteriori* (MAP) estimate of the true contrast, $\hat{C}_i(T)$, for each location in the current contrast map:

$$\hat{C}_i(T) = kc_i(T)\varepsilon_i(T) + c_i(T). \tag{A10}$$

If this MAP estimate is relatively stable and unbiased, then approximately the same MAP estimate will be obtained after the next fixation is made,

$$\hat{C}_i(T) \cong kc_i(T+1, x_{T+1}, y_{T+1})\varepsilon_i(T+1, x_{T+1}, y_{T+1}) + c_i(T+1, x_{T+1}, y_{T+1}), \tag{A11}$$

and therefore our prediction of the observed contrast after the next fixation is

$$\hat{c}_i(T+1, x_{T+1}, y_{T+1}) = \frac{\hat{C}_i(T)}{k\varepsilon_i(T+1, x_{T+1}, y_{T+1}) + 1}. \tag{A12}$$

In sum, Eq. (A3), (A4), (A7)–(A9), and (A12) can be used to estimate the fixation that will maximally reduce the total contrast uncertainty. In practice, we find that this estimate of the optimal fixation location is quite accurate.

A minor technical issue that arises in evaluating the CEM algorithm is that differential entropy can be negative. Therefore, we convert the differential entropy into discrete entropy by finely sampling the Gaussian distribution to obtain a discrete probability distribution. Using this discrete distribution guarantees that the uncertainty map is always nonnegative.

ACKNOWLEDGMENTS

Wilson S. Geisler's research was supported by NIH grant R01EY11747.

Correspondence should be addressed to Wilson S. Geisler (geisler@psy.utexas.edu).

REFERENCES

1. S. B. Laughlin, "A simple coding procedure enhances a neuron's information capacity," Z. Naturforsch. C **36**, 910–912 (1981).
2. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).
3. D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," Ophthalmic Physiol. Opt. **12**, 229–232 (1992).
4. J. J. Atick and A. N. Redlich, "What does the retina know about natural scenes?" Neural Comput. **4**, 196–210 (1992).
5. J. H. van Hateren, "Real and optimal neural images in early vision," Nature **360**, 68–70 (1992).
6. D. L. Ruderman, "The statistics of natural images," Network Comput. Neural Syst. **5**, 517–548 (1994).
7. J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," Proc. R. Soc. London, Ser. B **265**, 359–366 (1998).
8. A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," Vision Res. **37**, 3327–3338 (1997).
9. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy by V1?" Vision Res. **37**, 3311–3325 (1997).
10. W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," Vision Res. **41**, 711–724 (2001).
11. D. Purves and R. B. Lotto, *Why We See What We Do: An Empirical Theory of Vision* (Sinauer, 2003).
12. E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1215 (2001).
13. W. S. Geisler and R. Diehl, "Bayesian natural selection and the evolution of perceptual systems," Philos. Trans. R. Soc. London, Ser. B **357**, 419–448 (2002).
14. Y. Tadmor and D. J. Tolhurst, "Calculating the contrasts

that retinal ganglion cells and LGN neurones encounter in natural scenes," Vision Res. **40**, 3145–3157 (2000).

15. N. Brady and D. J. Field, "Local contrast in natural images: normalization and coding efficiency," Perception **29**, 1041–1055 (2000).
16. P. L. Clatworthy, M. Chirimuuta, J. S. Lauritzen, and D. J. Tolhurst, "Coding of the contrasts in natural images by populations of neurons in primary visual cortex (VI)," Vision Res. **43**, 1983–2001 (2003).
17. R. M. Balboa and N. M. Grzywacz, "Power spectra and distribution of contrasts of natural images from different habitats," Vision Res. **43**, 2527–2537 (2003).
18. P. Reinagel and A. M. Zador, "Natural scene statistics at the centre of gaze," Network Comput. Neural Syst. **10**, 1–10 (1999).
19. D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," IEEE Trans. Pattern Anal. Mach. Intell. **18**, 1–14 (1996).
20. T. S. Lee and S. Yu, "An information-theoretic framework for understanding saccadic behaviors," in *Advances in Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K.-R. Muller, eds. (MIT Press, 2000) Vol. 12, pp. 834–840.
21. G. E. Legge, T. A. Hooven, T. S. Klitz, J. G. Mansfield, and B. S. Tjan, "Mr. Chips 2002: new insights from an ideal observer model of reading," Vision Res. **42**, 2219–2234 (2002).
22. L. W. Renninger, J. Coughlan, P. Verghese, and J. Malik, "An information maximization model of eye movements," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, eds. (MIT Press, 2005), pp. 1121–1128.
23. J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," Vision Res. **21**, 409–418 (1981).
24. M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling," J. Opt. Soc. Am. A **8**, 1775–1787 (1991).
25. T. L. Arnow and W. S. Geisler, "Visual detection following retinal damage: predictions of an inhomogeneous retino-cortical model," Proc. SPIE **2674**, 119–130 (1996).
26. W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," Proc. SPIE **3299**, 294–305 (1998).
27. D. C. Hood and M. A. Finkelstein, "Sensitivity to light," in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, 1986), Vol. 1.
28. T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, 1991).
29. O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," Nat. Neurosci. **4**, 819–825 (2001).
30. U. Rajashekar, L. K. Cormack, and A. C. Bovik, "Visual search: structure from noise," in *Proceedings of Eye Tracking Research & Applications, ACM SIGGRAPH* 2002, A. T. Duchowski, ed. pp. 119–123 (www.siggrraph.org).
31. J. Najemnik and W. S. Geisler, "Optimal eye movement strategies in visual search," Nature **434**, 387–391 (2005).
32. A. L. Yarbus, *Eye Movements and Vision* (Plenum, 1967).