NATURAL CONTRAST STATISTICS AND THE SELECTION OF VISUAL FIXATIONS

Raghu G. Raj[1,2], Wilson S. Geisler[1,3,*], Robert A. Frazor[1,3], Alan C. Bovik[1,2]

Center for Perceptual Systems[1], Department of Electrical and Computer Engineering[2], Department of Psychology[3]
University of Texas at Austin

ABSTRACT

In this paper we address the problem of visual surveillance, which we define as the problem of optimally extracting information from the visual scene with a fixating, foveated imaging system. We are explicitly concerned with eye/camera movement strategies that result in maximizing information extraction from the visual field. Here we demonstrate how a novel characterization of the contrast statistics of natural images can be used for selecting fixation points that minimize the total contrast uncertainty (entropy) of natural images. We demonstrate the performance of the algorithm and compare its performance to ground truth methods. The results show that our algorithm performs favorably in terms of both efficiency and its ability to find salient features in the image.

1. INTRODUCTION

The Human Visual System (HVS) samples the visual field with a spatially non-uniform resolution such that the sampling density is maximal at the point of fixation and decreases radially away from the fixation point. This non-uniform sampling of the visual field is called foveation. By adopting such a strategy (in conjunction with head and body movements), the HVS 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, which take into account and exploit the continuously varying spatial resolution of the retina. Understanding the mechanisms that underlie eye movements is important not only for vision research but also (as we shall see) for the development of visual surveillance algorithms for foveated machine vision systems, wherein the task is to optimally survey the visual scene with the goal of extracting as much information from the image as possible.

To gain insight into some of the design requirements of these mechanisms, we have analyzed the effects of variable spatial resolution on 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 [1]. 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 the ganglion cell receptive fields increase, blurring the retinal image, and thereby effectively reducing local contrast and increasing contrast uncertainty. To analyze the role of contrast in visual surveillance we characterize the contrast statistics of natural images parameterized by eccentricity so that this knowledge can be exploited in devising information maximizing surveillance algorithms. While there has been much prior work on the contrast statistics of natural scenes [2, 3, 4], we demonstrate that our novel characterization of contrast statistics is well suited for addressing the problem of optimum fixation point selection.

Despite a long standing interest in eye movements [5] our understanding of the factors that underlie them in search and surveillance tasks is still limited. The reason is that eye movements are influenced by both low-level and high-level factors. High-level factors include cognitive strategies and complex representations of the features and objects that define the search targets [6]. Low level factors include foveation and simple image features, such as local contrast and spatial frequency content. The importance of these low level factors in simple search tasks has been demonstrated in [7]. In [8] it is shown that high contrast regions tend to attract human fixations. A more quantitative study [9] using point of gaze analysis reveals other simple statistical features that attract human fixations in surveillance tasks.

In this paper we explicitly demonstrate that the contrast statistics of natural images can be used very effectively in visual surveillance tasks. Given the close connection between natural scene statistics and the design of perceptual systems [10], our results suggest potentially important mechanisms that may be used by the HVS in performing visual surveillance.

Futhermore, given the promising role of foveated image processing in accommodating the competing

demands of high data rate and low channel bandwidth that are typical of most image and video processing systems of practical interest [11,12], an efficient means of automatically determining optimal fixation points—such as the one we have proposed in this paper—is important for realizing practical implementations of such foveated image processing systems.

2. CONTRAST STATISTICS OF NATURAL IMAGES

The first goal of this paper is to characterize the local contrast statistics of natural images relevant for the selection of fixation locations in foveated visual systems.

The local contrast function of an image is defined as follows:

$$C_{rms}(j) = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{(I_i - \bar{I})^2}{(\bar{I} + I_{dark})^2}}$$

where, $C_{rms}(j)$ is the RMS (root mean square) contrast at pixel location j. I_{dark} is the "dark light" parameter chosen to be 7 td (1 cd/m^2 assuming a 3 mm pupil), based on human photopic intensity discrimination data. (We note that this parameter has little effect on the measured contrasts because the mean luminances of the images were generally much higher than 1 cd/m^2.)

To compute the contrast we use a circular patch of N pixels about each pixel j in the image. I_i is the intensity level at the i^{th} pixel of the patch and w_i is the windowing function. We chose a raised cosine weighting function:

$$w_i = 0.5\left(\cos\left(\frac{\pi}{p}\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\right) + 1\right)$$

where p is the radius, (x_i, y_i) is the location of the i^{th} pixel of the patch, and (x_j, y_j) is the location of the center pixel of patch. The same weighting function was used to compute the local mean luminance $\bar{I}$. For our simulations the radius of the raised cosine window was taken to be 16 pixels (i.e., N = π x 16x16 pixels) for 1024x1024 images.

We studied the properties of the conditional contrast distributions for natural images $P\left(c \mid c_b(\varepsilon)\right)$, where c and $c_b(\varepsilon)$ are, respectively, the local contrasts of the unblurred original image and the blurred image at eccentricity ε. (The eccentricity is the distance from fixation point expressed in degrees of visual angle.) Note that these conditional distributions can be regarded as the posterior probability distributions of the unblurred contrast given the observed blurred contrast.

The empirical measurements of the conditional contrast distributions were carried out on a database of

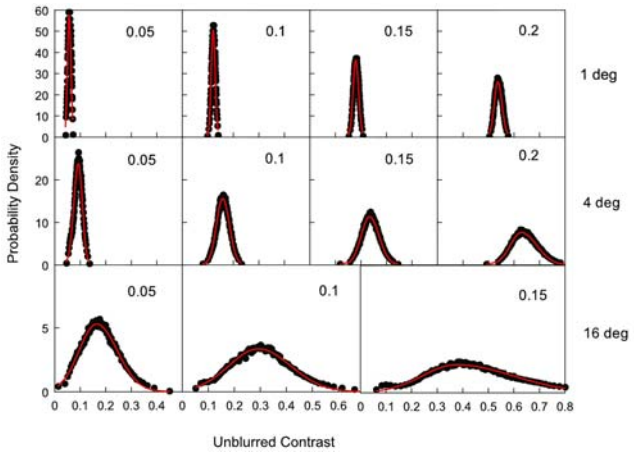

Figure 1. These are examples of the conditional probability distributions of local RMS contrast in unblurred images, given the local RMS contrast in the blurred versions of the images (columns), and given the retinal eccentricity (rows). The smooth curves are the best fitting Skewed-Gaussian distribution.

over 300 calibrated natural images found in [13]. We found that the contrast statistics are characterized by a simple set of formulas. The conditional distributions of contrast given the blurred observed contrast are accurately fit by skewed-gaussian distributions,

$$P\left(c \mid c_b\right) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\left(\dfrac{\sigma_l + \sigma_h}{2}\right)} \exp\left(\dfrac{-(c-u)^2}{2{\sigma_l}^2}\right) & c \le u \\ \dfrac{1}{\sqrt{2\pi}\left(\dfrac{\sigma_l + \sigma_h}{2}\right)} \exp\left(\dfrac{-(c-u)^2}{2{\sigma_h}^2}\right) & c > u \end{cases}$$

where u is the mode and $({\sigma_l}^2, {\sigma_h}^2)$ are the variances of the two halves of the skewed-gaussian distribution with respect to the mode (see Figure 1). Furthermore, the parameters $(u, {\sigma_l}^2, {\sigma_h}^2)$ vary in a simple fashion with the blurred contrast c_b:

$$u(c_b, \varepsilon) = (k\varepsilon + 1)c_b \qquad (1)$$

$$\frac{\sigma_l(c_b, \varepsilon) + \sigma_h(c_b, \varepsilon)}{2} = k\varepsilon c_b \qquad (2)$$

and hence,

$$\bar{\sigma}(c_b, \varepsilon)^2 = (k\varepsilon c_b)^2 + \sigma_0^2 \qquad (3)$$

Where σ_0 is a very small unknown constant, and $k = 0.1082$ is an empirically determined constant (see Figures 2 and 3). It is easily shown that the differential entropy of a skewed-gaussian distribution is given by:

$$h = \frac{1}{2}\log_2\left(2\pi e\left(\frac{\sigma_l + \sigma_h}{2}\right)^2\right) \qquad (4)$$

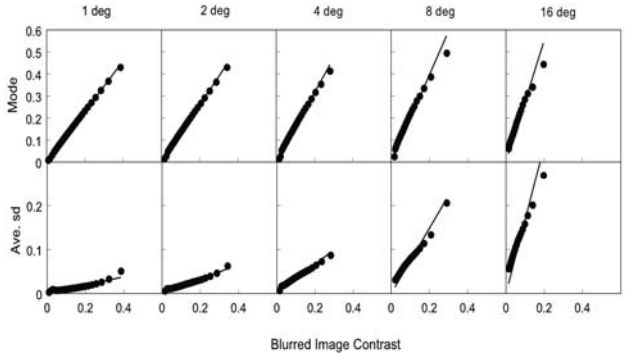

Figure 2. The modes and the average standard deviations of all the conditional probability densities are plotted as a function of blurred image contrast and retinal eccentricity. The curves are best fitting straight lines through the origin

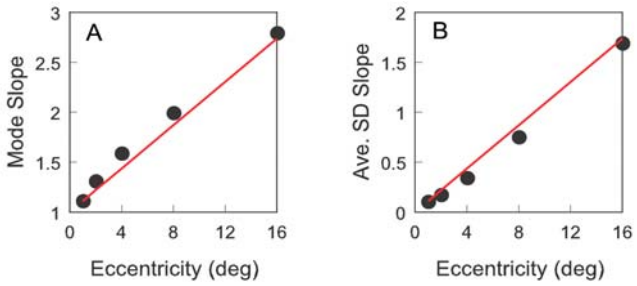

Figure 3. Slopes of linear functions in Figure 2. **A**. Slope of the contrast vs. mode plot as a function of retinal eccentricity. **B**. Slope of the contrast vs. average standard deviation plot as a function of retinal eccentricity. The curves show the predictions of the equations (1) and (2) in the text.

3. FIXATION POINT SELECTION

Having characterized the contrast statistics of natural images as parameterized by eccentricity (i.e., the level of blur), our second goal is to employ this in the design of optimal visual surveillance algorithms for foveated vision systems.

Consider foveated images created by performing linear scale varying (LSV) filtering [14]:

$$I_f(\mathbf{x}) = \frac{1}{[s(\mathbf{x})]^2} \int_{\boldsymbol{R}^2} g[\boldsymbol{a}/s(\mathbf{x})]\, I(\mathbf{x}-\boldsymbol{a})\, d\boldsymbol{a}$$

With the proper choice for kernel shape $g(\bullet)$ and the scaling function $s(\bullet)$, this is fairly accurately model of the foveation in the HVS [11].

To determine the approximately optimal sequence of fixation points we employ a serial optimization (greedy) algorithm. Specifically, our aim is to find the sequence of fixation points $r_0, r_1, r_3, \cdots$ (with $r_j = (x_j, y_j)$) such that the $(T+1)^{th}$ fixation maximally reduces the total contrast entropy:

$$r_{T+1} = \arg\max_r \left(H(r_0, \cdots, r_T) - H(r_0, \cdots, r_T, r) \right)$$

where $H(r_0, \cdots, r_T)$ is the total contrast entropy after the T^{th} fixation, summed over all *n* pixel locations in the image:

$$H(r_0, \cdots, r_T) = \sum_{i=1}^{n} h_i(T)$$

$$\hat{h}_i(T) = \frac{1}{2} \log_2 \left(2\pi e \left(\left(k\varepsilon_i(T) c_i(T) \right)^2 + \sigma_0^2 \right) \right)$$

(c.f., equations 3 and 4). In these equations, $\varepsilon_i(T)$ represents the smallest eccentricity obtained so far at pixel location *i*,

$$\varepsilon_i(T) = \min_{t \le T} \varepsilon_{it}$$

and $c_i(T)$ represents the contrast observed at that eccentricity. To determine the best next fixation we need to estimate what the contrast entropy will be at every pixel, for every possible next fixation:

$$\hat{h}_i(T+1) = \frac{1}{2} \log_2 \left(2\pi e \left(\left(k\varepsilon_i(T+1) \hat{c}_i(T+1) \right)^2 + \sigma_0^2 \right) \right)$$

To evaluate this equation, we first use equation (1) to compute the MAP estimate of the unblurred image contrast at pixel location *i*, then we use equation (1) again to estimate the blurred contrast, $\hat{c}_i(T+1)$, after making a fixation to location *r*.

There are three technical matters to mention. First, the initial fixation point can be chosen arbitrarily or according to some *a priori* distribution. For our simulations, we chose the center pixel of the image as the initial fixation point. Second, we convert the differential entropy to entropy by finely sampling the Gaussian distribution to obtain a discrete probability distribution. Third, the value of σ_0 is not important as long as it is very small, but greater than zero.

4. SIMULATION RESULTS

Figures 4A and 4C show the first 9 fixations (8 saccades) obtained for two natural images that were *not* a part of the image set used to compute the contrast statistics described in section 2. These are representative of the 16 natural images we have processed. In Figure 4A we see that there are few fixations near the sky region, because there is little contrast uncertainty in low contrast regions of natural scenes, and thus little significant information to be gained by fixating those regions. We observe in Figure 4B that the algorithm fixates around the salient objects embedded in the background; i.e. the flowers. Since the entire background region of the image has similar contrast,

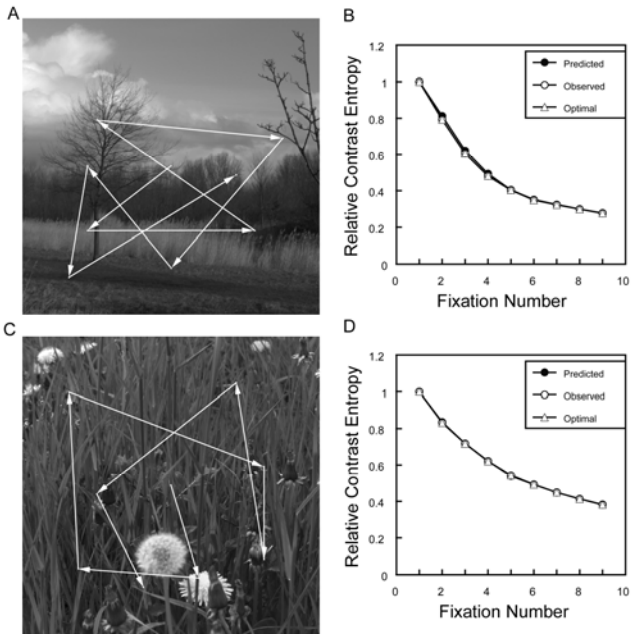

Figure 4: Fixation points selected by the principle of minimizing total contrast entropy (contrast uncertainty), using average local contrast statistics of natural images.

the fixations tend to spread out over the image. Figures 4B and 4D 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 making the fixation selected by the algorithm. As can be seen, the predicted and observed entropy 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. For the 16 test images the average ratio of optimal contrast entropy to the observed contrast entropy was 0.99. This at once demonstrates the computational efficiency of the algorithm and the utility of our contrast statistics for selecting optimal fixation points.

As a measure of how well the algorithm (which is based only upon contrast statistics) reduces the total uncertainty about the images, we computed the total mean squared error (MSE) between the original image and the image reconstructed from all the fixations up to and including the current fixation. For the 16 test images, the average ratio of the minimum possible MSE to the obtained MSE was 0.9. Thus, this simple algorithm also does a respectable job at minimizing total image uncertainty.

5. REFERENCES

[1] B.A. Wandell, *Foundations of Vision*, Sinauer Associates Inc., Publishers, 1995.

[2] D.L. Ruderman, "The statistics of natural images", Computation in Neural Systems 5, pp. 517-548, 1994.

[3] A.B. Lee, K.S. Pedersen, D. Mumford,"The Nonlinear Statistics of High-Contrast Patches in Natural Images", International Journal of Computer Vision, 54(1/2/3) pp.83-103, 2003.

[4] A. Turiel, N. Parga,"The Multi-Fractal Structure of Contrast Changes in Natural Images: from Sharp edges to Textures", Neural Computation 12, pp.763-793, 2000

[5] A.L. Yarbus, *Eye movements and vision*, Plenum Press 1967.

[6] R.P.N. Rao and D.H. Ballard, "An active vision architecture based on iconic representations," *Artificial Intelligence*, 78:461 506, 1995.

[7] W.S. Geisler and K. Chou, "Separation of low-level and high-level factors in complex tasks: Visual search," *Psychological Review*, 102, 356-378, 1995.

[8] P. Reinagel and A. M. Zador, "Natural scene statistics at the center of gaze," *Network: Computation in Neural Systems*, vol. 10, 1-10, 1999.

[9] U. Rajashekar, L. K. Cormack and A. C. Bovik, "Image features that draw fixations," IEEE International Conference on Image Processing, Vol: 3, Barcelona, Spain, September 14-17, 2003, Page(s): 313-316.

[10] E.P. Simoncelli and B.A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*,24:1193-1216, May 2001.

[11] W.S. Geisler and J.S. Perry "A real-time foveated multi-resolution system for low-bandwidth video communication" In: B. Rogowitz and T. Pappas (Eds.), *Human Vision and Electronic Imaging, SPIE Proceedings, 3299*: 294-305, 1998.

[12] S. Lee, M. S. Pattichis, and A. C. Bovik, "Rate control for foveated MPEG/H.263 video," *Proc. IEEE Int. Conf. on Image Processing*, Vol: 2, Chicago, IL, USA, October 04-07, 365-369, 1998.

[13] van Hateren database of Natural Images: http://hlab.phys.rug.nl/archive.html

[14] A.C. Bovik and R.G. Raj, "Approximating filtered scale variant signals", *IEEE Transactions on Image Processing*, Volume: 14(1), 23 – 35, 2005.

* To whom correspondence should be addressed: geisler@psy.utexas.edu

6. ACKNOWLEDGMENTS

Supported in part by NIH grant R01EY11747.