Optimal Imaged-based Defocus Estimates from Individual Natural Images

Johannes Burge and Wilson S. Geisler
Center for Perceptual Systems, 1 University Station, University of Texas at Austin, Austin, TX 78712
jburge@mail.cps.utexas.edu
geisler@psy.utexas.edu

Abstract: We present a general method for estimating defocus blur from first principles, given a set of natural scenes and properties of the vision system. Local, high-precision, signed estimates are obtained for a model human visual system.

OCTS Codes: (100.0100) Image Processing; (110.0110) Imaging Systems; (330.0330) Vision, color, and visual optics

1. Introduction

Biological visual systems perform powerful computations that exploit information in retinal images useful to the perceptual, behavioral, and biological tasks that organisms perform. The information in retinal images is determined by the statistical structure of natural scenes, and by properties of the organism's optical systems and photosensor arrays. Performance is jointly determined by the quality of the available information and by the efficiency with which that information is processed. To characterize the theoretical limits of performance in a natural task, one must account for all these factors [1]. The 3D structure of the environment is one of the most important environmental attributes that organisms estimate, given organisms' needs to interact with the environment in their search for shelter, food, and mates. Defocus may be the most widely available depth cue in the animal kingdom.

Vision begins with lens systems that focus and defocus light on the retinal photoreceptors. Lenses focus light perfectly from only one distance, and natural scenes contain objects at many distances. Thus, defocus is generally present in images of natural scenes. Although defocus degrades image quality, it plays an important role in depth estimation, accommodation control, eye growth regulation, and the predatory behavior of many small animals [2-5]. Defocus is also central to engineering and clinical eye-care applications (e.g. digital cameras, myopia prevention, refractive corrections, Lasik). Despite these facts, it is unknown how biological systems estimate defocus [4]. This is a significant theoretical gap. The computer vision and engineering literatures describe algorithms for defocus estimation. But they typically require simultaneous multiple images, special lens apertures, or light with known patterns projected onto the environment [6-8]. Mammalian visual systems usually lack these advantages. These algorithms therefore cannot serve as plausible models of defocus estimation in many biological visual systems.

Here, we describe a principled approach for estimating defocus in small regions of individual images, given a training set of natural images, a wave-optics model of the lens system, a photosensor array, and a specification of measurement noise. We show for the human visual system that high-precision, unbiased estimates are obtainable under natural viewing conditions, and that chromatic aberrations fully resolve the sign ambiguity. And we show that simple receptive fields, similar to those in retina and early visual cortex, suffice to extract the information optimally. The approach can be tailored to any environment-vision system pairing: natural or man-made, animal or machine. Thus, it establishes a framework, based on basic physical principles and established Bayesian statistics, for analyzing the psychophysics and neurophysiology of defocus estimation in species across the animal kingdom, and for creating optimal image-based defocus and depth estimation algorithms for computational vision systems.

2. Methods & Results

There are two broad steps in our approach to defocus estimation: i) characterize defocus information available for processing in the vision system of interest, ii) discover spatial frequency filters that are optimally diagnostic of defocus and use them to estimate defocus. The defocus of a target region is the difference between the lens system's current power and the power required to focus the target region: $\Delta D = D_{focus} - D_{target}$ where ΔD is the defocus, D_{focus} is the current power, and D_{target} is the power required to image the target sharply (target distance), expressed in units of diopters (1/meters). The goal is to estimate ΔD in each local region of an image. We start with a formal description of factors that determine the defocus information available for processing. Defocus information is jointly determined by the properties of natural scenes, the optical system, and the photosensor array of the vision system.

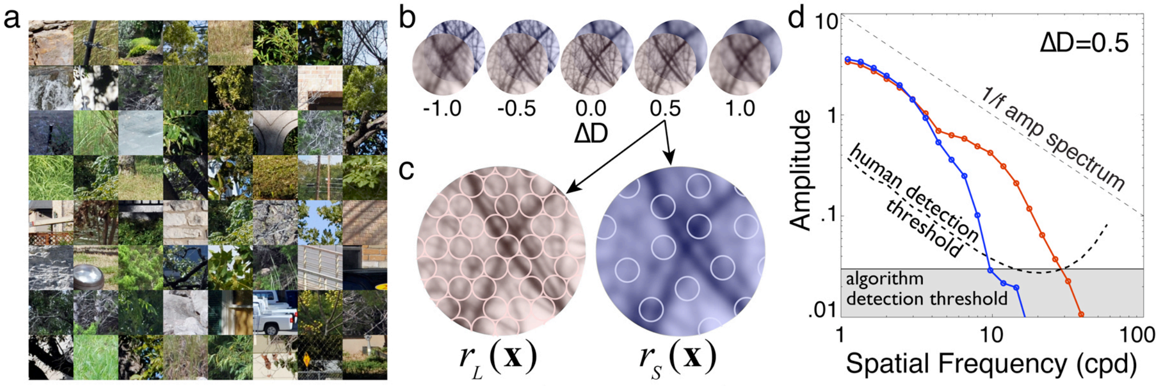

Fig. 1: Factors determining defocus information. a. *Natural scenes*. Natural scene inputs were approximated with well-focused photographs. The camera lens was focused on infinity and imaged objects were minimum 16m from the camera. **b.** *Optics*. L- and S-cone retinal images for five different levels of defocus. Chromatic aberration causes the S-cone image to be less sharp than the L-cone image for myopic defocus and more sharp for hyperopic defocus. **c.** *Sensor sampling*. L- and S-cone sensor sampling is shown for an image with 0.5 diopters of defocus. Note the smaller number of S-cone samples. **d.** *Neural noise and inefficiency*. Radially-averaged amplitude spectra of the L- and S-cone images for an image with 0.5 diopters defocus. The dashed black curve shows the human neural detection threshold. The solid black line shows the threshold imposed on the algorithm.

The input from a natural scene (Fig. 1a) is represented by an idealized (i.e. unaffected by optics) image, $I(\mathbf{x},\lambda)$, which gives the radiance at each location $\mathbf{x}=(x,y)$ in the plane of the sensor array for each wavelength λ. The optical system is represented by a point-spread function, $psf\left(\mathbf{x},\lambda;a(\mathbf{z}),W(\mathbf{z},\lambda,\Delta D)\right)$, which gives the spatial distribution of light across the sensor array produced by a point target of wavelength λ. The form of the point-spread function depends on the aperture function, $a(\mathbf{z})$, which specifies the shape, size, and transmittance of the pupil aperture. It also depends on the wavefront aberration function, which depends on the position $\mathbf{z}$ in the plane of the aperture, the wavelength of light λ, and defocus. The aperture function determines the effect of diffraction on image quality. The wave aberration function determines degradations in image quality not attributable to diffraction (Fig. 1b). The sensor array is represented by a wavelength sensitivity function $s_c(\lambda)$ and a spatial sampling function $samp_c(\mathbf{x})$ for each sensor class, c (Fig. 1c). Neural noise and other processing inefficiencies are represented by a spatial-frequency-dependent detection threshold (Fig. 1d). Combining these factors (except for neural inefficiencies) gives the spatial responses in a given sensor class:

$$r_c(\mathbf{x})=\left(\sum_{\lambda}\left[I(\mathbf{x},\lambda)*psf\left(\mathbf{x},\lambda;a(\mathbf{z}),W(\mathbf{z},\lambda,\Delta D)\right)\right]s_c(\lambda)\right)samp_c(\mathbf{x}) \qquad (1)$$

where $*$ represents two-dimensional convolution in $\mathbf{x}$. The goal is to estimate defocus, ΔD, at each point in an image from local sensor responses (Equation 1) in the available sensor classes.

We model a vision system with a 2mm pupil, human chromatic aberrations, sensors with the wavelength and spatial sampling of human long-wavelength (L) and short-wavelength (S) cones, and detection thresholds determined from human psychophysical data [9]. Thus, in this case, the defocus information is contained in the sensor responses, $r_L(\mathbf{x})$ and $r_S(\mathbf{x})$. To discover filters that are optimally diagnostic of defocus given the variation in natural images, we sampled hundreds of 1 deg patches from natural sensor images that had been defocused by different amounts (-2 to 2 diopters in ¼ diopter steps). Patches used for 'training' were not used for 'testing'. Next, we performed a fast-Fourier transform to obtain their Fourier spectra. Then, we use a recently developed technique for dimensionality reduction—Accuracy Maximization Analysis [10] (AMA)—to find the Bayes-optimal, rank-ordered, spatial-frequency defocus filters that, for a fixed number of filters, maximize defocus estimation accuracy in the least-squared sense over a given dioptric range. In this case, AMA operates on the radially-averaged power spectra of each sensor class (Fig 1d).

The optimal defocus filters are similar to chromatic double-opponent cells in early visual cortex [11] (Fig. 2a). Also, the filters concentrate their energy near the frequency range known to drive human accommodation [12] (5-15 cpd). The final step is to use the filter responses to estimate defocus. To do this, we estimated the joint filter response distributions for each defocus level, $p\left(\mathbf{R}\mid\Delta D_j\right)$, by fitting Gaussians to the sample means and covariances (Fig. 2b). Given a sufficient number of defocus levels, continuous estimates are obtained with the standard formula for minimum mean-squared error [13]: $\Delta\hat{D} = \sum_{j=1}^{N} \Delta D_j \, p\left(\Delta D_j \mid \mathbf{R}\right)$, where ΔD_j is one of N defocus levels and $p\left(\Delta D_j \mid \mathbf{R}\right)$ is the posterior probability of that defocus level (by Bayes' rule) given the observed filter response vector, **R**.

High-precision (+1/16 diopter), unbiased defocus estimates are obtainable from natural images in vision systems with human chromatic aberrations, L- and S-cone sensors, and typical human neural contrast detection thresholds (Fig. 2c).

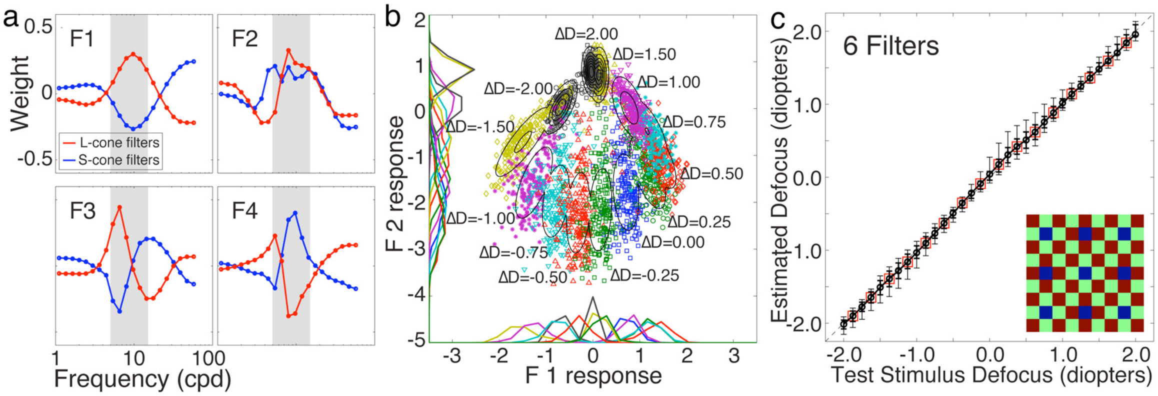

Fig. 2. Estimating defocus. a Optimal spatial-frequency defocus filters, **b** Filter responses and Gaussian fits to the filter response distributions. Different colors indicate different defocus levels. Training patches were used to determine the filters and the response distributions, **c.** Defocus estimates from test patches. Error bars show 68% (thick bars) and 90% (thin bars) confidence intervals on the estimates. Boxes indicate defocus levels that were not in the training set. The inset shows the pattern of L (red), M (green), and S (blue) cones in the rectangular mosaic that was used to sample the retinal images (57, 57, and 14 samples per deg, respectively).

3. Conclusion

Our work has at least four major benefits. First, it prescribes how to characterize the information in captured images relevant for estimating defocus, thus enabling the rigorous study of the perception, behavior, and neurophysiology of defocus estimation using natural stimuli. Second, it specifies how to determine the theoretical limits of performance by any biological or machine vision system given its constraints. Third, it determines the optimal filters (i.e. bases) for defocus estimation, which can in turn be used to make principled predictions about neurophysiological receptive fields involved in performing relevant computations. Fourth, it prescribes the design of algorithms that make optimal use of the available defocus information. In light of the recent trend in the machine vision community of taking algorithmic inspiration from biological science, this work may have broad practical applications.

4. Acknowledgments

Supported by NIH grant EY11747

5. References

[1] Geisler WS, & Ringach, D (2009) “Natural Systems Analysis”. *Visual Neuroscience*, 26, 1-3.

[2] Held RT, Cooper EA, O’Brien JF, Banks MS (2010). “Using blur to affect perceived distance and size”. *ACM Transactions on Graphics*, 29(2): 19.1-19.16.

[3] Kruger PB, Mathews S, Aggarwala KR, & Sanchez N (1993). “Chromatic aberration and ocular focus: Fincham revisited”. *Vision Research*, 33(10): 1397-1411.

[4] Wallman J, & Winawer (2004). “Homeostasis of eye growth and the question of myopia”. *Neuron*, 43: 447-468.

[5] Harkness L (1977). “Chameleons use accommodation cues to judge distance”. *Nature*, 267(26): 346-349.

[6] Pentland AP, Scherock S, Darrel T, Girod B (1994). “Simple range cameras based on focal error”. *JOSA A*, 11(11): 2925-2934.

[7] Zhou C, Lin S, Nayar S (2009). “Coded aperture pairs for depth from defocus”. In IEEE International Conference in Computer Vision (ICCV), October.

[8] Levin A, Fergus R, Durand F, Freeman W (2007). “Image and depth from a conventional camera with a coded aperture”. *ACM Transactions On Graphics*, 26(3): 70.1-70.9.

[9] Williams DR (1985). “Visibility of interference fringes near the resolution limit”. *JOSA A* 2(7): 1087-1093.

[10] Geisler WS, Najemnik J, Ing, AD (2009). “Optimal stimulus encoders for natural tasks.” *Journal of Vision*, 9(13): 17, 1–16.

[11] Johnson EN, Hawken MJ, Shapley R (2001). “The spatial transformation of color in the primary visual cortex of the macaque monkey”. *Nature Neuroscience*, 4(4): 409-416.

[12] Mathews S, & Kruger PB (1994). “The spatiotemporal transfer function of human accommodation”. *Vision Research*, 34(15): 1965-1980.

[13] Hastie T, Tibshirani R, Friedman J (2009). *The elements of statistical learning 2nd edition*, New York: Springer.