Optimal defocus estimation in individual natural images

Johannes Burge[1] and Wilson S. Geisler

Center for Perceptual Systems, University of Texas at Austin, Austin, TX 78712

Edited by Tony Movshon, New York University, New York, NY, and approved August 22, 2011 (received for review May 31, 2011)

Defocus blur is nearly always present in natural images: Objects at only one distance can be perfectly focused. Images of objects at other distances are blurred by an amount depending on pupil diameter and lens properties. Despite the fact that defocus is of great behavioral, perceptual, and biological importance, it is unknown how biological systems estimate defocus. Given a set of natural scenes and the properties of the vision system, we show from first principles how to optimally estimate defocus at each location in any individual image. We show for the human visual system that high-precision, unbiased estimates are obtainable under natural viewing conditions for patches with detectable contrast. The high quality of the estimates is surprising given the heterogeneity of natural images. Additionally, we quantify the degree to which the sign ambiguity often attributed to defocus is resolved by monochromatic aberrations (other than defocus) and chromatic aberrations; chromatic aberrations fully resolve the sign ambiguity. Finally, we show that simple spatial and spatio-chromatic receptive fields extract the information optimally. The approach can be tailored to any environment–vision system pairing: natural or man-made, animal or machine. Thus, it provides a principled general framework for analyzing the psychophysics and neurophysiology of defocus estimation in species across the animal kingdom and for developing optimal image-based defocus and depth estimation algorithms for computational vision systems.

optics | sensor sampling | Bayesian statistics | depth perception | auto-focus

In a vast number of animals, 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 information is nearly always present in images of natural scenes. Defocus information is vital for many natural tasks: depth and scale estimation (1, 2), accommodation control (3, 4), and eye growth regulation (5, 6). However, little is known about the computations visual systems use to estimate defocus in images of natural scenes. The computer vision and engineering literatures describe algorithms for defocus estimation (7, 8). However, they typically require simultaneous multiple images (9–11), special lens apertures (11, 12), or light with known patterns projected onto the environment (9). Mammalian visual systems usually lack these advantages. Thus, these methods cannot serve as plausible models of defocus estimation in many visual systems.

Although defocus estimation is but one problem faced by vision systems, few estimation problems have broader scope. Vision scientists have developed models for how defocus is used as a cue to depth and have identified stimulus factors that drive accommodation (biological autofocusing). Neurobiologists have identified defocus cues and transcription factors that trigger eye growth, a significant contributor to the development of near-sightedness (5). Comparative physiologists have established accommodation's role in predatory behavior across the animal kingdom, in species as diverse as the chameleon (13) and the cuttlefish (14). Engineers have developed methods for autofocusing camera lenses and automatically estimating depth from defocus across an image. However, there is no widely accepted formal theory for how defocus information should be extracted from individual natural images. The absence of such a theory constitutes a significant theoretical gap.

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 sensor array, and a specification of noise and processing inefficiencies. We begin by considering a vision system with diffraction- and defocus-limited optics, a sensor array sensitive only to one wavelength of light, and sensor noise consistent with human detection thresholds. We then consider more realistic vision systems that include human monochromatic aberrations, human chromatic aberrations, and sensors similar to human photoreceptors.

The defocus of a target region is defined as the difference between the lens system's current power and the power required to bring the target region into focus,

$$\Delta D = D_{\text{focus}} - D_{\text{target}}, \qquad [1]$$

where ΔD is the defocus, D_{focus} is the current power, and D_{target} is the power required to image the target sharply, expressed in diopters (1/m). The goal is to estimate ΔD in each local region of an image.

Estimating defocus, like many visual estimation tasks, suffers from the "inverse optics" problem: It is impossible to determine with certainty, from the image alone, whether image blur is due to defocus or some property of the scene (e.g., fog). Defocus estimation also suffers from a sign ambiguity: Under certain conditions, point targets at the same dioptric distances nearer or farther than the focus distance are imaged identically. However, numerous constraints exist that can make a solution possible. Previous work has not taken an integrative approach that capitalizes on all of these constraints.

Results

The information for defocus estimation is jointly determined by the properties of natural scenes, the optical system, the sensor array, and sensor noise. The input from a natural scene can be represented by an idealized image $I(\mathbf{x}, \lambda)$ that gives the radiance at each location $\mathbf{x} = (x, y)$ in the plane of the sensor array, for each wavelength λ. An optical system degrades the idealized image and can be represented by a point-spread function $psf(\mathbf{x}, \lambda, \Delta D)$, which gives the spatial distribution of light across the sensor array produced by a point target of wavelength λ and defocus ΔD. The sensor array is represented by a wavelength sensitivity function $s_c(\lambda)$ and a spatial sampling function $samp_c(\mathbf{x})$ for each class of sensor c. Combining these factors gives the spatial pattern of responses in a given class of sensor,

$$r_c(\mathbf{x}) = \left(\sum_{\lambda} [I(\mathbf{x}, \lambda) * psf(\mathbf{x}, \lambda, \Delta D)] s_c(\lambda) \right) samp_c(\mathbf{x}), \qquad [2]$$

where $*$ represents 2D convolution in $\mathbf{x}$. Noise and processing inefficiencies then corrupt these sensor responses. The goal is to

Author contributions: J.B. and W.S.G. designed research; J.B. performed research; J.B. analyzed data; and J.B. and W.S.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. E-mail: jburge@mail.cps.utexas.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1108491108/-/DCSupplemental.

estimate defocus from the noisy sensor responses in the available sensor classes.

The first factor determining defocus information is the statistical structure of the input images from natural scenes. These statistics must be determined by empirical measurement. The most accurate method would be to measure full radiance functions $I(\mathbf{x}, \lambda)$ with a hyperspectral camera. However, well-focused, calibrated digital photographs were used as approximations to full radiance functions. This approach is sufficiently accurate for the present purposes (*SI Methods* and Fig. S1); in fact, it is preferred because hyperspectral images are often contaminated by motion blur. Eight hundred 128 × 128 pixel input patches were randomly sampled from 80 natural scenes containing trees, shrubs, grass, clouds, buildings, roads, cars, etc.; 400 were used for training and the other 400 for testing (Fig. 1*A*).

The next factor is the optical system, which is characterized by the point-spread function. The term for the point-spread function in Eq. **2** can be expanded to make the factors determining its form more explicit,

$$psf(\mathbf{x}, \lambda; a(\mathbf{z}, \lambda), W(\mathbf{z}, \lambda, \Delta D)), \qquad [3]$$

where $a(\mathbf{z}, \lambda)$ specifies the shape, size, and transmittance of the pupil aperture, and $W(\mathbf{z}, \lambda, \Delta D)$ is a wave aberration function, which depends on the position $\mathbf{z}$ in the plane of the aperture, the wavelength of light, the defocus level, and other aberrations introduced by the lens system (15). The aperture function determines the effect of diffraction on the image quality. The wave aberration function determines degradations in image quality not attributable to diffraction. A perfect lens system (i.e., limited only by diffraction and defocus) converts light originating from a point on a target object to a converging spherical wavefront. The wave aberration function describes how the actual converging wavefront differs from a perfect spherical wavefront at each point in the pupil aperture. The human pupil is circular and its minimum diameter under bright daylight conditions is ∼2 mm (16); this pupil shape and size are assumed throughout the paper. Note that a 2-mm pupil is conservative because defocus information increases (depth-of-focus decreases) as pupil size increases.

The next factor is the sensor array. Two sensor arrays were considered: an array having a single sensor class sensitive only to 570 nm and an array having two sensor classes with the spatial sampling and wavelength sensitivities of the human long-wavelength (L) and short-wavelength (S) cones (17). (A system sensitive only to one wavelength will be insensitive to chromatic aberrations.) Human foveal cones sample the retinal image at ∼128 samples/degree; this rate is assumed throughout the paper. Thus, the 128 × 128 pixel input patches have a visual angle of 1 degree.

The last factor determining defocus information is the combined effect of photon noise, system noise, and processing inefficiencies. We represent this factor in our algorithm by applying a threshold determined from human psychophysical detection thresholds (18). (For the analyses that follow, we found that applying a threshold has a nearly identical effect to adding noise.)

The proposed computational approach is based on the well-known fact that defocus affects the spatial Fourier spectrum of sensor responses. Here, we consider only amplitude spectra (19), although the approach can be generalized to include phase spectra. There are two steps to the approach: (*i*) Discover the spatial frequency filters that are most diagnostic of defocus and (*ii*) determine how to use the filter responses to obtain continuous defocus estimates. A perfect lens system attenuates the amplitude spectrum of the input image equally in all orientations. Hence, to estimate the spatial frequency filters it is reasonable, although not required, to average across orientation. Fig. 1*B* shows how spatial frequency amplitudes are modulated by different magnitudes of defocus (i.e., modulation transfer functions). Fig. 1*C* shows the effect of defocus on the amplitude spectrum of a sampled retinal image patch; higher spatial frequencies become more attenuated as defocus magnitude increases. The gray boundary represents the detection threshold imposed on our algorithm. For any given image patch, the shape of the spectrum above the threshold would make it easy to estimate the magnitude of defocus of that patch. However, the substantial variation of local amplitude spectra in natural images makes the task difficult. Hence, we seek filters tuned to spatial frequency features that are optimally diagnostic of the level of defocus, given the variation in natural image patches.

To discover these filters, we use a recently developed statistical learning algorithm called accuracy maximization analysis (AMA) (20). As long as the algorithm does not get stuck in local minima, it finds the Bayes-optimal feature dimensions (in rank order) for maximizing accuracy in a given identification task (see http://jburge.cps.utexas.edu/research/Code.html for Matlab implementation of AMA). We applied this algorithm to the task of identifying the defocus level, from a discrete set of levels, of sampled retinal image patches. The number of discrete levels was picked to allow accurate continuous estimation (*SI Methods*). Specifically, a random set of natural input patches was passed through a model lens system at defocus levels between 0 and 2.25 diopters, in 0.25-diopter steps, and then sampled by the sensor array. Each sampled image patch was then converted to a contrast image by subtracting

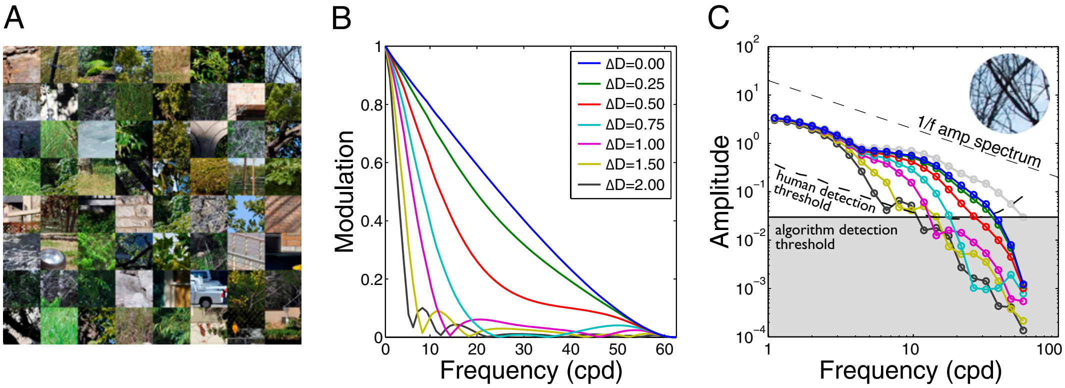

Fig. 1. Natural scene inputs and the effect of defocus in a diffraction- and defocus-limited vision system. (*A*) Examples of natural inputs. (*B*) Optical effect of defocus. Curves show one-dimensional modulation transfer functions (MTFs), the radially averaged Fourier amplitude spectra of the point-spread functions. (*C*) Radially averaged amplitude spectra of the top-rightmost patch in *A*. Circles indicate the mean amplitude in each radial bin. Light gray circles show the spectrum of the idealized natural input. The dashed black curve shows the human neural detection threshold.

and dividing by the mean. Next, the contrast image was windowed by a raised cosine (0.5 degrees at half height) and fast-Fourier transformed. Finally, the log of its radially averaged squared amplitude (power) spectrum was computed and normalized to a mean of zero and vector magnitude of 1.0. [The log transform was used because it nearly equalizes the standard deviation (SD) of the amplitude in each radial bin (Fig. S2). Other transforms that equalize variability, such as frequency-dependent gain control, yield comparable performance.] Four thousand normalized amplitude spectra (400 natural inputs × 10 defocus levels) constituted the training set for AMA.

Fig. 2*A* shows the six most useful defocus filters for a vision system having diffraction- and defocus-limited optics and sensors that are sensitive only to 570 nm light. The filters have several interesting features. First, they capture most of the relevant information; additional filters add little to overall accuracy. Second, they provide better performance than filters based on principal components analysis or matched templates (Fig. S3). Third, they are relatively smooth and hence could be implemented by combining a few simple, center-surround receptive fields like those found in retina or primary visual cortex. Fourth, the filter energy is concentrated in the 5–15 cycles per degree (cpd) frequency range, which is similar to the range known to drive human accommodation (4–8 cpd) (21–23).

The AMA filters encode information in local amplitude spectra relevant for estimating defocus. However, the Bayesian decoder built into the AMA algorithm can be used only with the training stimuli, because that decoder needs access to the mean and variance of each filter's response to each stimulus (20). In other words, AMA finds only the optimal filters.

The next step is to combine (pool) the filter responses to estimate defocus in arbitrary image patches, having arbitrary defocus. We take a standard approach. First, the joint probability distribution of filter responses to natural image patches is estimated for the defocus levels in the training set. For each defocus level, the filter responses are fit with a Gaussian by calculating the sample mean and covariance matrix. Fig. 2*B* shows the joint distribution of the first two AMA filter responses for several levels of defocus. Fig. 2*C* shows contour plots of the fitted Gaussians. Second, given the joint distributions (which are six dimensional, one dimension for each filter), defocus estimates are obtained with a weighted summation formula

$$\Delta\hat{D} = \sum_{j=1}^{N} \Delta D_j p(\Delta D_j|\mathbf{R}), \qquad [4]$$

where ΔD_j is one of the N trained defocus levels, and $p(\Delta D_j|\mathbf{R})$ is the posterior probability of that defocus level given the observed vector of filter responses **R**. The response vector is given by the dot product of each filter with the normalized, logged amplitude spectrum. The posterior probabilities are obtained by applying Bayes' rule to the fitted Gaussian probability distributions (*SI Methods*). Eq. **4** gives the Bayes optimal estimate when the goal is to minimize the mean-squared error of the estimates and when N is sufficiently large, which it is in our case (*SI Methods*).

Defocus estimates for the test patches are plotted as a function of defocus in Fig. 2*D* for our initial case of a vision system having perfect optics and a single class of sensor. None of the test patches were in the training set. Performance is quite good. Precision is high and bias is low once defocus exceeds ~0.25 diopters, roughly the defocus detection threshold in humans (21, 24). Precision decreases at low levels of defocus because a modest change in defocus (e.g., 0.25 diopters) does not change the amplitude spectra significantly when the base defocus is zero; more substantial changes occur when the base defocus is nonzero (24, 25) (Fig. 1*C*). The bias near zero occurs because in vision systems having perfect optics and sensors sensitive only to a single wavelength, positive and negative defocus levels of identical magnitude yield identical amplitude spectra. Thus, the bias is due to a boundary effect: Estimation errors can be made above but not below zero.

Now, consider a biologically realistic lens system having monochromatic aberrations (e.g., astigmatic and spherical). Although such lens systems reduce the quality of the best-focused image, they can introduce information useful for recovering defocus sign (26). To examine this possibility, we changed the optical model to include the monochromatic aberrations of human eyes. Aberration maps for two defocus levels are shown for the first author's right eye (Fig. 3*A*). At the time the first author's optics were measured, he had 20/20 acuity and 0.17 diopters of astigmatism, and his higher-order aberrations were about equal in magnitude to his astigmatism (Table S1). Spatial frequency attenuation due to the lens optics now differs as a function of the defocus sign. When focused behind the target (negative defocus), the eye's 2D modulation transfer function (MTF) is oriented near the positive oblique; when focused in front (positive defocus), the MTF has

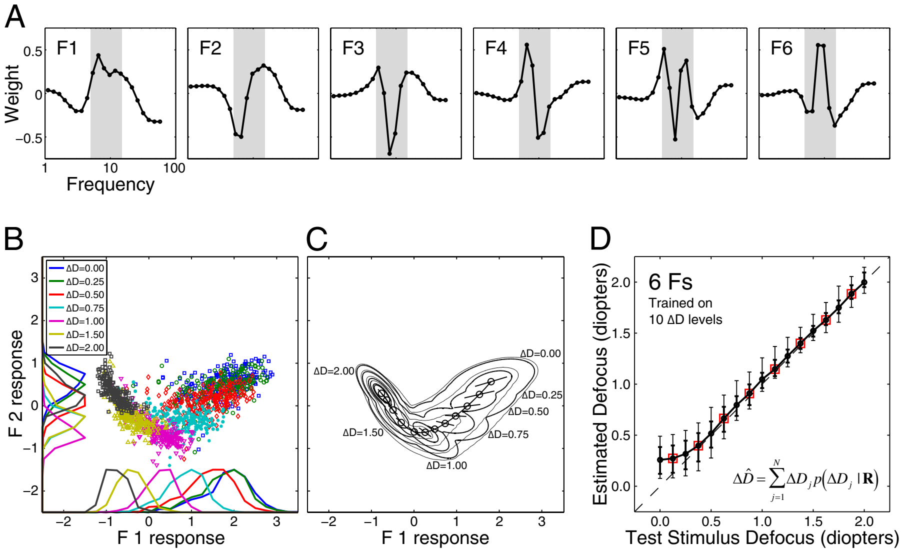

Fig. 2. Optimal filters and defocus estimation. (*A*) The first six AMA filters. Filter energy is concentrated in a limited frequency range (shaded area). (*B*) Filter responses to amplitude spectra in the training set (1.25, 1.75, and 2.25 diopters not plotted). Symbols represent joint responses from the two most informative filters. Marginal distributions are shown on each axis. (*C*) Gaussian fits to filter responses. Thick lines are iso-likelihood contours on the maximum-likelihood surface determined from fits to the response distributions at trained defocus levels. Thin lines are iso-likelihood contours on interpolated response distributions (*SI Methods*). Circles indicate interpolated means separated by a *d′* (i.e., Mahalanobis distance) of 1. Line segments show the direction of principle variance and ±1 SD. (*D*) Defocus estimates for test stimuli. Circles represent the mean defocus estimate for each defocus level. Error bars represent 68% (thick bars) and 90% (thin bars) confidence intervals. Boxes indicate defocus levels not in the training set. The equal-sized error bars at both trained and untrained levels indicates that the algorithm outputs continuous estimates.

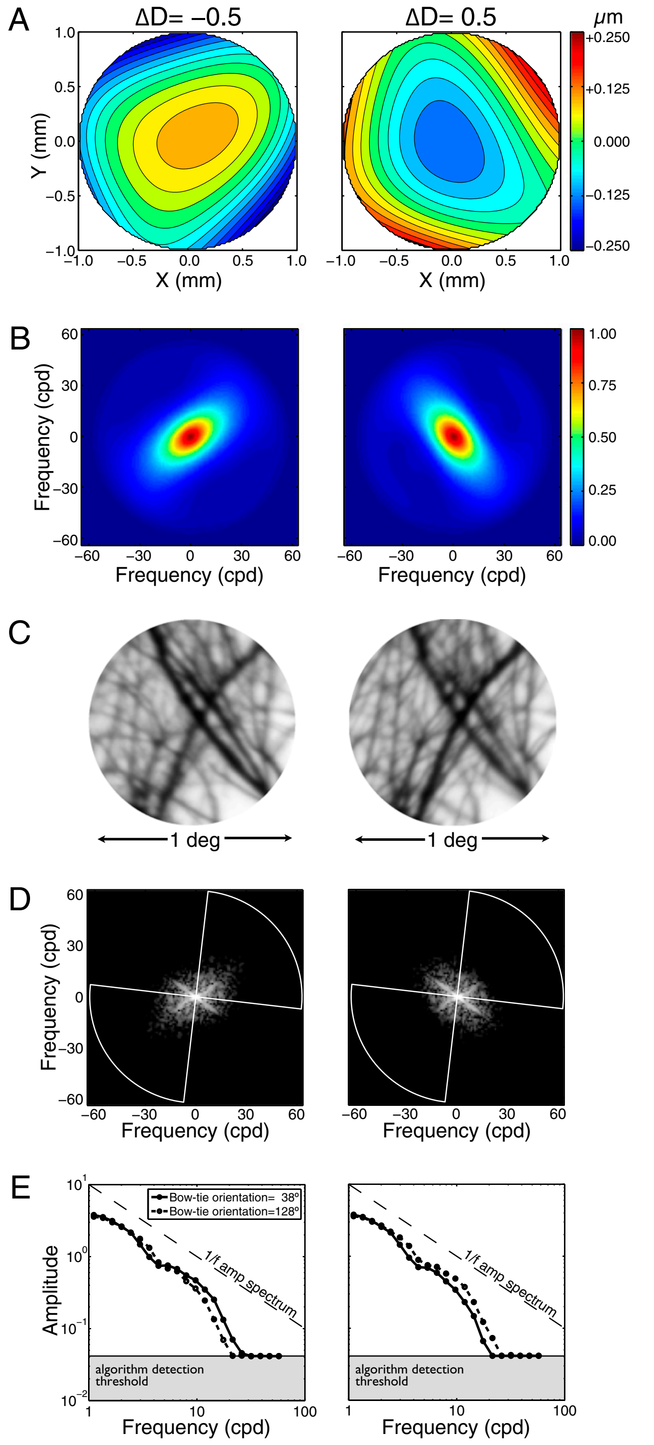

Fig. 3. Effect of defocus sign in a vision system with human monochromatic aberrations. (*A*) Wavefront aberration functions of the first author's right eye for −0.5 and +0.5 diopters of defocus (*x* and *y* represent location in the pupil aperture). Color indicates wavefront errors in micrometers. (*B*) Corresponding 2D MTFs. Orientation differences are due primarily to astigmatism. Color indicates transfer magnitude. (*C*) Image patch defocused by −0.5 and +0.5 diopters. Relative sharpness of differently oriented image features changes as a function of defocus sign. (*D*) Logged 2D-sampled retinal image amplitude spectra. The spectra were radially averaged within two "bowties" (one shown, white lines) that were centered on the dominant orientations of the negatively and positively defocused MTFs (*SI Methods*). (*E*) Thresholded bowtie amplitude spectra. Curves show the bowtie amplitude spectra at the dominant orientations of the negatively and positively defocused MTFs (solid and dashed curves, respectively).

the opposite orientation (Fig. 3*B*). Image features oriented orthogonally to the MTF's dominant orientation are imaged more sharply. This effect is seen in the sampled retinal image patches (Fig. 3*C*) and in their corresponding 2D amplitude spectra (Fig. 3*D*).

Many monochromatic aberrations in human optics contribute to the effect of defocus sign on the MTF, but astigmatism—the difference in lens power along different lens meridians—is the primary contributor (27). Interestingly, astigmatism is deliberately added to the lenses in compact disc players to aid their autofocus devices.

To examine whether orientation differences can be exploited to recover defocus sign, optimal AMA filters were relearned for vision systems having the optics of specific eyes and the same single-sensor array as before. There were two procedural differences: (*i*) Instead of averaging radially across all orientations, the spectra were radially averaged in two orthogonal "bowties" (Fig. 3*D*) centered on the MTF's dominant orientation (*SI Methods*) for each sign of defocus (Fig. 3*E*). (*ii*) The same training natural inputs were passed through the optics at defocus levels ranging from −2.25 to 2.25 diopters in 0.25-diopter steps, yielding 7,600 thresholded spectra (400 natural inputs × 19 defocus levels).

The filters for the first author's right eye (Fig. 4*A*) yield estimates of defocus magnitude similar in accuracy to those in Fig. 2*D* (Fig. S4*A*). Importantly, the filters now extract information about defocus sign. Fig. 4*B* (black curve) shows the proportion of test stimuli where the sign of the defocus estimate was correct. Although performance was well above chance, a number of errors occurred. Similar performance was obtained with "standard observer" optics (28); better performance was obtained with the first author's left eye, which has more astigmatism. Thus, a vision system with human monochromatic aberrations and a single sensor class can estimate both the magnitude and the sign of defocus with reasonable accuracy.

Finally, consider a vision system with two sensor classes, each with a different wavelength sensitivity function. In this vision system, chromatic aberrations can be exploited. It has long been recognized that chromatic aberrations provide a signed cue to defocus (29, 30). The human eye's refractive power changes by ~1 diopter between 570 and 445 nm (31), the peak sensitivities of the L and S cones. Typically, humans focus the 570-nm wavelength of broadband targets most sharply (32). Therefore, when the eye is focused on or in front of a target, the L-cone image is sharper than the S-cone image; the opposite is true when the lens is focused sufficiently behind the target. Chromatic aberration thus introduces sign information in a manner similar to astigmatism. Whereas astigmatism introduces a sign-dependent statistical tendency for amplitudes at some orientations to be greater than others, chromatic aberration introduces a sign-dependent tendency for one sensor class to have greater amplitudes than the other.

Optimal AMA filters were learned again, this time for a vision system with diffraction, defocus, chromatic aberrations, and sensors with spatial sampling and wavelength sensitivities similar to human cones. In humans, S cones have ~1/4 the sampling rate of L and medium wavelength (M) cones (33). We sampled the retinal image with a rectangular cone mosaic similar to the human cone mosaic. For simplicity, M-cone responses were not used in the analysis. The amplitude spectra from L and S sensors were radially averaged because the optics are again radially symmetric. Optimal filters are shown in Fig. 4*C*. Cells with similar properties (i.e., double chromatically opponent, spatial-frequency bandpass receptive fields tuned to the same frequency) have been reported in primate early visual cortex (34, 35). Such cells would be well suited to estimating defocus (30).

A vision system sensitive to chromatic aberration yields unbiased defocus estimates with high precision (~ ±1/16 diopters) over a wide range (Fig. 4*D*). Sensitivity to chromatic aberrations also allows the sign of defocus to be identified with near 100% accuracy (Fig. 4*B*, magenta curve). The usefulness of chromatic aberrations

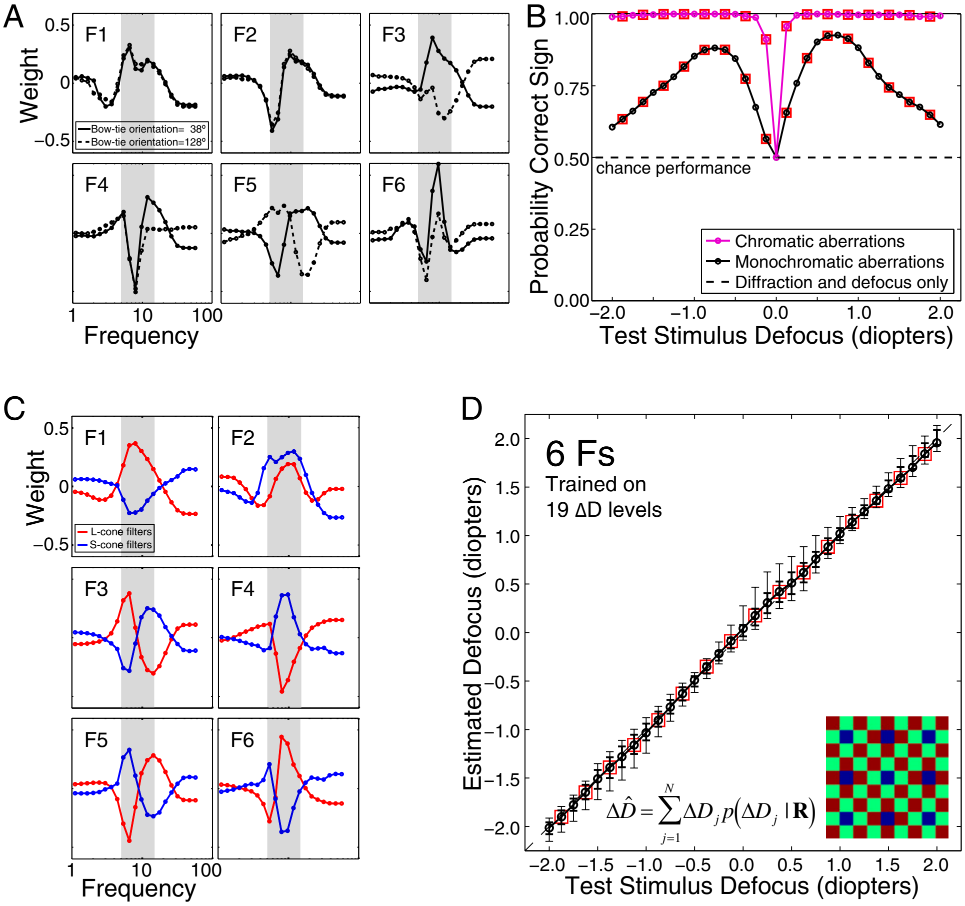

Fig. 4. Optimal filters and defocus estimates for vision systems with human monochromatic or chromatic aberrations. (*A*) Optimal filters for a vision system with the optics of the first author's right eye and a sensor array sensitive only to 570 nm light. Solid lines show filter sensitivity to orientations in the "bowtie" centered on the dominant orientation of the negatively defocused MTF (Fig. 3 *D* and *E*). Dotted lines show filter sensitivities to the other orientations. (*B*) Defocus sign identification. The black curve shows performance for a vision system with the first author's monochromatic aberrations. The magenta curve shows performance for a system sensitive to chromatic aberration. (*C*) Optimal filters for the system sensitive to chromatic aberrations. Red curves show L-cone filters. Blue curves show S-cone filters. *Inset* in *D* shows the rectangular mosaic of L (red), M (green), and S (blue) cones used to sample the retinal images (57, 57, and 14 samples/degree, respectively). M-cone responses were not used in the analysis. (*D*) Defocus estimates using the filters in *C*. Error bars represent the 68% (thick bars) and 90% (thin bars) confidence intervals on the estimates. Boxes mark defocus levels not in the training set. Error bars at untrained levels are as small as at trained levels, indicating that the algorithm makes continuous estimates.

is due to at least three factors. First, the ∼1-diopter defocus difference between the L- and S-cone images produces a larger signal than the difference due to the monochromatic aberrations in the analyzed eyes (Fig. S5; compare with Fig. 3*E*). Second, natural L- and S-cone input spectra are more correlated than the spectra in the orientation bowties (Fig. S6); the greater the correlation between spectra is, the more robust the filter responses are to variability in the shape of input spectra. Third, small defocus changes are easier to discriminate in images that are already somewhat defocused (21, 24). Thus, when the L-cone image is perfectly focused, S-cone filters are more sensitive to changes in defocus, and vice versa. In other words, chromatic aberrations ensure that at least one sensor will always be in its "sweet spot".

How sensitive are these results to the assumptions about the spatial sampling of L and S cones? To find out, we changed our third model vision system so that both L and S cones had full resolution (i.e., 128 samples/degree each). We found similar filters and only a small performance benefit (Fig. S7). Thus, defocus estimation performance is robust to large variations in the spatial sampling of human cones.

Some assumptions implicit in our analysis were not fully consistent with natural scene statistics. One assumption was that the statistical structure of natural scenes is invariant with viewing distance (36). Another was that there is no depth variation within image patches, which is not true of many locations in natural scenes. Rather, defocus information was consistent with planar fronto-parallel surfaces displaced from the focus distance. Note, however, that the smaller the patch is (in our case, 0.5 degrees at half height), the less the effect of depth variation. Nonetheless, an important next step is to analyze a database of luminance-range images so that the effect of within-patch depth variation can be accounted for. Other aspects of our analysis were inconsistent with the human visual system. For instance, we used a fixed 2-mm pupil diameter. Human pupil diameter increases as light level decreases; it fluctuates slightly even under steady illumination. We tested how well the filters in Fig. 4 can be used to estimate defocus in images obtained with other pupil diameters. The filters are robust to changes in pupil diameter (Fig. S4 *A* and *B*). Importantly, none of these details affect the qualitative findings or main conclusions.

We stress that our aim has been to show how to characterize and extract defocus information from natural images, not to provide an explicit model of human defocus estimation. That problem is for future work.

Our results have several implications. First, they demonstrate that excellent defocus information (including sign) is available in natural images captured by the human visual system. Second, they suggest principled hypotheses (local filters and filter response pooling rules) for how the human visual system should encode and decode defocus information. Third, they provide a rigorous benchmark against which to evaluate human performance in tasks involving defocus estimation. Fourth, they demonstrate the potential value of this approach for any organism with a visual system. Finally, they demonstrate that it should be possible to design useful defocus estimation algorithms for digital imaging systems without the need for specialized hardware. For example, incorporating the optics, sensors, and noise of digital cameras into our framework could lead to improved methods for autofocusing.

Defocus information is even more widely available in the animal kingdom than binocular disparity. Only some sighted animals have visual fields with substantial binocular overlap, but nearly all have lens systems that image light on their photoreceptors. Our results show that sufficient signed defocus information exists in individual natural images for defocus to function as an absolute depth cue once pupil diameter and focus distance are known. In this respect, defocus is similar to binocular disparity, which functions as an absolute depth cue once pupil separation and fixation distance are known. Defocus becomes a higher precision depth cue as focus distance decreases. Perhaps this is why many smaller animals, especially those without consistent binocular overlap, use defocus as their primary depth cue in predatory behavior (13, 14). Thus, the theoretical framework described here could guide behavioral and neurophysiological studies of defocus and depth estimation in many organisms.

In conclusion, we have developed a method for rigorously characterizing the defocus information available to a vision system by combining a model of the system's wave optics, sensor sampling, and noise with a Bayesian statistical analysis of the sensor responses to natural images. This approach should be widely applicable to other vision systems and other estimation problems, and it illustrates the value of natural scene statistics and statistical decision theory for the analysis of sensory and perceptual systems.

Methods

Natural Scenes. Natural scenes were photographed with a tripod-mounted Nikon D700 14-bit SLR camera (4,256 × 2,836 pixels) fitted with a Sigma 50-mm prime lens. Scenes were those commonly viewed by researchers at the University of Texas at Austin. Details on camera parameters (aperture, shutter speed, ISO), on camera calibration, and on our rationale for excluding very low contrast patches from the analysis are in *SI Methods*.

Optics. All three wave-optics models assumed a focus distance of 40 cm (2.5 diopters), a single refracting surface, and the Fraunhoffer approximation, which implies that at or near the focal plane the optical transfer function (OTF) is given by the cross-correlation of the generalized pupil function with its complex conjugate (15). The wavefront aberration functions of the first author's eyes were measured with a Shack–Hartman wavefront sensor and expressed as 66 coefficients on the Zernike polynomial series (Table S1). The coefficients were scaled to the 2-mm pupil diameter used in the analysis from the 5-mm diameter used during wavefront aberration measurement (37).

A refractive defocus correction was applied to each model vision system before analysis began to ensure 0-diopter targets were focused best. Details on this process, and on how the dominant MTF orientations in Fig. 3 were determined, are in *SI Methods*.

Sensor Array Responses. To account for the effect of chromatic aberration on the L- and S-cone sensor responses in the third vision system, we created polychromatic point-spread functions for each sensor class. See *SI Methods* for details.

Noise. To account for the effects of sensor noise and subsequent processing inefficiencies, a detection threshold was applied at each frequency (e.g., Fig. 1*C*); amplitudes below the threshold were set equal to the threshold amplitude. The threshold was based on interferometric measurements that bypass the optics of the eye (18) under the assumption that the limiting noise determining the detection threshold is introduced after the image is encoded by the photoreceptors.

Accuracy Maximization Analysis. AMA was used to estimate optimal filters for defocus estimation. See *SI Methods* for details on the logic of AMA.

Estimating Defocus. Given an observed filter response vector **R**, a continuous defocus estimate was obtained by computing the expected value of the posterior probability distribution over a set of discrete defocus levels (Eq. **4**). Details of this computation, of likelihood distribution estimation, and of likelihood distribution interpolation are in *SI Methods*.

ACKNOWLEDGMENTS. We thank Austin Roorda for measuring the first author's monochromatic aberrations, and Larry Thibos for helpful discussions on physiological optics. We thank David Brainard and Larry Thibos for comments on an earlier version of the manuscript. This work was supported by National Institutes of Health (NIH) Grant EY11747 to WSG.

1. Held RT, Cooper EA, O'Brien JF, Banks MS (2010) Using blur to affect perceived distance and size. *ACM Trans Graph* 29(2):19.1–19.16.
2. Vishwanath D, Blaser E (2010) Retinal blur and the perception of egocentric distance. *J Vis* 10:26, 1–16.
3. Kruger PB, Mathews S, Aggarwala KR, Sanchez N (1993) Chromatic aberration and ocular focus: Fincham revisited. *Vision Res* 33:1397–1411.
4. Kruger PB, Mathews S, Katz M, Aggarwala KR, Nowbotsing S (1997) Accommodation without feedback suggests directional signals specify ocular focus. *Vision Res* 37:2511–2526.
5. Wallman J, Winawer J (2004) Homeostasis of eye growth and the question of myopia. *Neuron* 43:447–468.
6. Diether S, Wildsoet CF (2005) Stimulus requirements for the decoding of myopic and hyperopic defocus under single and competing defocus conditions in the chicken. *Invest Ophthalmol Vis Sci* 46:2242–2252.
7. Pentland AP (1987) A new sense for depth of field. *IEEE Trans Patt Anal Mach Intell* 9:523–531.
8. Wandell BA, El Gamal A, Girod B (2002) Common principles of image acquisition systems and biological vision. *Proc IEEE* 90(1):5–17.
9. Pentland AP, Scherock S, Darrel T, Girod B (1994) Simple range cameras based on focal error. *J Opt Soc Am A* 11:2925–2934.
10. Watanabe M, Nayar SK (1997) Rational filters for passive depth from defocus. *Int J Comput Vis* 27:203–225.
11. Zhou C, Lin S, Nayar S (2011) Coded aperture pairs for depth from defocus and defocus blurring. *Int J Comput Vis* 93(1):53–69.
12. Levin A, Fergus R, Durand F, Freeman W (2007) Image and depth from a conventional camera with a coded aperture. *ACM Trans Graph* 26(3):70.1–70.9.
13. Harkness L (1977) Chameleons use accommodation cues to judge distance. *Nature* 267:346–349.
14. Schaeffel F, Murphy CJ, Howland HC (1999) Accommodation in the cuttlefish (Sepia officinalis). *J Exp Biol* 202:3127–3134.
15. Goodman JW (1996) *Introduction to Fourier Optics* (McGraw-Hill, New York), 2nd Ed.
16. Wyszecki G, Stiles WS (1982) *Color Science: Concepts and Methods, Quantitative Data and Formulas* (Wiley, New York).
17. Stockman A, Sharpe LT (2000) The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Res* 40:1711–1737.
18. Williams DR (1985) Visibility of interference fringes near the resolution limit. *J Opt Soc Am A* 2:1087–1093.
19. Field DJ, Brady N (1997) Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Res* 37:3367–3383.
20. Geisler WS, Najemnik J, Ing AD (2009) Optimal stimulus encoders for natural tasks. *J Vis* 9(13):17, 1–16.
21. Walsh G, Charman WN (1988) Visual sensitivity to temporal change in focus and its relevance to the accommodation response. *Vision Res* 28:1207–1221.
22. Mathews S, Kruger PB (1994) Spatiotemporal transfer function of human accommodation. *Vision Res* 34:1965–1980.
23. Mackenzie KJ, Hoffman DM, Watt SJ (2010) Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodative control. *J Vis* 10(8):22, 1–20.
24. Wang B, Ciuffreda KJ (2005) Foveal blur discrimination of the human eye. *Ophthalmic Physiol Opt* 25:45–51.
25. Charman WN, Tucker J (1978) Accommodation and color. *J Opt Soc Am* 68:459–471.
26. Wilson BJ, Decker KE, Roorda A (2002) Monochromatic aberrations provide an odd-error cue to focus direction. *J Opt Soc Am A Opt Image Sci Vis* 19(5):833–839.
27. Porter J, Guirao A, Cox IG, Williams DR (2001) Monochromatic aberrations of the human eye in a large population. *J Opt Soc Am A Opt Image Sci Vis* 18:1793–1803.
28. Autrusseau F, Thibos LN, Shevell S (2011) Chromatic and wavefront aberrations: L-, M-, and S-cone stimulation with typical and extreme retinal image quality. *Vision Res*, in press.
29. Fincham EF (1951) The accommodation reflex and its stimulus. *Br J Ophthalmol* 35:381–393.
30. Flitcroft DI (1990) A neural and computational model for the chromatic control of accommodation. *Vis Neurosci* 5:547–555.
31. Thibos LN, Ye M, Zhang X, Bradley A (1992) The chromatic eye: A new reduced-eye model of ocular chromatic aberration in humans. *Appl Opt* 31:3594–3600.
32. Thibos LN, Bradley A (1999) Modeling the refractive and neuro-sensor systems of the eye. *Visual Instrumentation: Optical Design and Engineering Principle*, ed Mouroulis P (McGraw-Hill, New York), pp 101–159.
33. Packer O, Williams DR (2003) Light, the retinal image, and photoreceptors. *The Science of Color*, ed Shevell SK (Elsevier, Amsterdam), 2nd Ed, pp 41–102.
34. Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215–243.
35. Johnson EN, Hawken MJ, Shapley R (2001) The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat Neurosci* 4:409–416.
36. Ruderman DL (1994) The statistics of natural images. *Network* 5:517–548.
37. Campbell CE (2003) Matrix method to find a new set of Zernike coefficients from an original set when the aperture radius is changed. *J Opt Soc Am A Opt Image Sci Vis* 20:209–217.

Supporting Information

Burge and Geisler 10.1073/pnas.1108491108

SI Methods

Natural Scenes. Camera aperture diameter was set to 5 mm (f/10). Maximum shutter duration was 1/100 s. ISO was set to 200. To ensure well-focused photographs, the lens was focused on optical infinity, and care was taken that imaged objects were at least 16 m from the camera (i.e., maximum defocus in any local image patch was 1/16 diopter). Ten 128 × 128-pixel patches were randomly selected from each of 80 photographs; half were used for training and half for testing. RAW photographs were calibrated via a previously published procedure and were converted either to 14-bit luminance or long, medium, and short wavelength (LMS) cone responses, depending on which type of sensor array was being modeled (1). We excluded all natural input patches that had <5% root-mean–squared (rms) contrast *before* they were passed through a model eye's optics. This exclusion removed the small percentage of patches that were dominated by camera pixel noise and that would largely fall below the human detection threshold (16%; 7% from non-sky regions and 9% from blank blue sky). Defocus estimates from these patches are (unsurprisingly) of low quality. However, vision systems have access to local contrast and hence could disregard defocus estimates from image locations with very low contrast. Including these patches in the analysis has no discernable effect on the estimated filters and only a minor effect on overall estimation performance.

Optics. Patches were defocused by simulated optical systems. Before analysis began, a refractive defocus correction was applied to each model vision system so that 0-diopter targets were focused best. We applied the correction that maximized the volume under the MTF scaled by the neural contrast sensitivity function (2). This metric accurately predicts the refractive correction that humans judge best (3).

When the optical model included monochromatic aberrations other than defocus, the dominant orientation of the MTF changed with the sign of defocus. To estimate the dominant orientation for each sign, the MTF was computed for each of 65 evenly spaced negative defocus levels between −0.75 and −0.25 diopters and 65 positive defocus levels between +0.25 and +0.75 diopters. Each MTF was convolved with a bowtie function and the result was fitted with a Von Mises function (circular Gaussian). The function peak was the estimated orientation for that defocus level. We then found the two orientations that were best centered in the estimated orientation distributions for the positive and negative defocus levels, with the constraint that these two orientations differed by 90 degrees. Forcing dominant orientations to be perpendicular is justified when astigmatism is the primary aberration that changes with defocus sign, because then the principal directions of lens surface curvature are always perpendicular.

Sensor Array Responses. To account for chromatic aberration and its effect on L- and S-cone sensor responses, single-wavelength point-spread functions (PSFs) were computed every 5 nm between 400 and 700 nm (3). The wavelength-dependent change in refractive power of the human eye was taken from the literature (4). Separate polychromatic PSFs were obtained for each cone class by weighting the single-wavelength PSFs by the L- and S-cone sensitivity functions (5) and by the D65 daylight illumination spectrum and then summing

$$psf_c(\mathbf{x}, \Delta D) = \frac{1}{K}\sum_{\lambda} psf(\mathbf{x}, \lambda, \Delta D) s_c(\lambda) \mathrm{D65}(\lambda), \qquad \textbf{[S1]}$$

where K is a normalizing constant that sets the PSF volume to 1.0. Retinal images were obtained by transforming the RGB values of the input photographs to LMS values and then by convolving the L- and S-cone input channels with the polychromatic PSFs (6). This procedure was repeated for each defocus level under consideration.

To implement the reduced spatial sampling rates of the L and S cones, we sampled the retinal images using the rectangular array shown in Fig. 4*D*, *Inset*. Then, we linearly interpolated back to full resolution. Linear interpolation is justified because it cannot add useful information into the image.

Accuracy Maximization Analysis (AMA). The logic of AMA is as follows. Consider encoding each training stimulus with a small population of filters that each apply a linear weighting function with a specified response noise (here, a small amount of Gaussian noise). Suppose that the linear weighting functions are known. In that case, it is easy to compute the mean and variance of each filter's response to each training sample. If these means and variances are known, then a closed-form expression can be derived for the approximate accuracy of the Bayesian optimal decoder with access to the means and variances (7). Finally, this closed-form expression can be used to search the space of linear weighting functions to find the functions (filters) that give the most accurate performance. We searched for these functions using gradient descent after initializing each weighting function with random values. Different random initializations yielded the same final estimated filters. A Matlab implementation of AMA and a short discussion of how to apply it are available at http://jburge.cps.utexas.edu/research/Code.html.

AMA is a form of dimensionality reduction similar to principal components analysis (PCA) with one critically important difference: AMA finds the training set components (feature dimensions) that are optimal for a particular task whereas PCA finds the components that account for the highest proportion of variance in the training set, without regard to task. The fact that PCA and AMA filters differ indicates (unsurprisingly) that retinal amplitude spectra variability exists that is not due to defocus. Another difference is that PCA is required to find orthogonal components, whereas AMA has no such requirement. Like PCA, AMA components are found sequentially: The first component is selected to maximize accuracy then the second component is selected to maximize accuracy when used in conjunction with the first component, and so on.

Estimating Defocus. Bayes' rule gives the posterior probability of each specific defocus level ΔD_j,

$$p(\Delta D_j|\mathbf{R}) = \frac{p(\mathbf{R}|\Delta D_j)p(\Delta D_j)}{\sum\limits_{k=1}^{N} p(\mathbf{R}|\Delta D_k)p(\Delta D_k)}, \qquad \textbf{[S2]}$$

where $p(\mathbf{R}|\Delta D_j)$ is the likelihood of the observed filter response vector given that defocus level, and $p(\Delta D_j)$ is the prior probability of that defocus level. We assumed that the likelihood for each defocus level is a multidimensional Gaussian (one dimension per filter) with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$,

$$p\left(\mathbf{R}|\Delta D_j\right) = gauss\left(\mathbf{R}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right), \qquad \text{[S3]}$$

where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ were set to the sample mean and covariance matrix of the raw filter responses (e.g., Fig. 2 *B* and *C*). In our test set, the prior probabilities of the defocus levels were equal. Thus, the prior probabilities factor out of Eq. **S2**.

Increasing the number of discrete defocus levels in the training set increases the accuracy of the continuous estimates. (Identification of discrete defocus levels becomes equivalent to continuous estimation as the number of levels increases.) However, increasing the number of discrete defocus levels increases the training set size and the computational complexity of learning filters via AMA. In practice, we found that excellent continuous estimates are obtained using 0.25-diopter steps for training, followed by interpolation to estimate Gaussian distributions between steps. Interpolated distributions were obtained by fitting a cubic spline through the response distribution means and linearly interpolating the response distribution covariance matrices. Interpolated distributions were added until the maximum d' (i.e., Mahalanobis distance) between neighboring distributions was ≤ 0.5.

To prevent boundary condition effects, we trained on defocus levels that were 0.25 diopters more out of focus than the largest defocus level for which we present estimation performance.

Testing the Three-Color-Channel Approximation of Full Radiance Functions. Idealized hyperspectral radiance functions $I(\mathbf{x}, \lambda)$ contain the radiance at each location $\mathbf{x}$ in the plane of the sensor array for each wavelength λ, as would occur in a hypothetical optical system that does not degrade image quality at all. Throughout the paper we used well-focused calibrated three-color-channel digital photographs $I_c(\mathbf{x})$ as approximations to idealized hyperspectral radiance functions. To test whether this approximation was justified, we obtained a set of hyperspectral reflectance images (8), multiplied them by the D65 irradiance spectrum (to obtain radiance images), and then processed them according to two workflows. (The actual measured irradiance spectra were flatter than the D65 spectrum, making the following test more stringent.)

In the first workflow, hyperspectral images were processed exactly as specified by Eq. **2** in the main text. The idealized image $I(\mathbf{x}, \lambda)$ was convolved with wavelength-specific point-spread functions and weighted by the wavelength sensitivity of each sensor class, before being spatially sampled by each sensor class. We refer to the sensor responses resulting from this workflow as "hyperspectral" sensor responses.

In the second workflow, hyperspectral images were converted to three-channel LMS images and were defocused with polychromatic point-spread functions (*Methods*), before being spatially sampled by the sensor array. Specifically, each class of sensor response was given by

$$r_c(\mathbf{x}) = [I_c(\mathbf{x})^{*} psf_c(\mathbf{x}, \Delta D)] samp_c(\mathbf{x}), \qquad \text{[S4]}$$

where each image channel was obtained by taking the dot product of the wavelength distribution at each pixel with the sensor wavelength sensitivity: $I_c(\mathbf{x}) = \sum_\lambda I(\mathbf{x}, \lambda) s_c(\lambda)$. We refer to the sensor response resulting from this workflow as the "color-channel" sensor responses.

Finally, we fast-Fourier transformed both the hyperspectral and color-channel sensor responses and compared their amplitude spectra (Fig. S1). The analysis shows that for the present purposes, it is justified to approximate sensor responses by using polychromatic point-spread functions to defocus three-channel color images.

Defocus Filter Comparison (AMA vs. PCA vs. Templates). We compared defocus-level identification performance of the AMA defocus filters to the performance of defocus filters that were obtained via suboptimal methods. AMA filters substantially outperform filters determined via PCA and template matching. Template filters were created by multiplying the average natural input spectrum with the modulation transfer function for each defocus level (i.e., the template filters were the average retinal amplitude spectra for each defocus level). The test stimuli from the main text were projected onto each set of filters to obtain the filter response distributions. Each filter response distribution was fit with a Gaussian. A quadratic classifier was used to determine the classification boundaries. The proportion correctly identified was computed as a function of the number of filters (Fig. S3).

1. Ing AD, Wilson JA, Geisler WS (2010) Region grouping in natural foliage scenes: Image statistics and human performance. *J Vis* 10(4):10, 1–19.
2. Williams DR (1985) Visibility of interference fringes near the resolution limit. *J Opt Soc Am A* 2:1087–1093.
3. Thibos LN, Hong X, Bradley A, Applegate RA (2004) Accuracy and precision of objective refraction from wavefront aberrations. *J Vis* 4:329–351.
4. Thibos LN, Ye M, Zhang X, Bradley A (1992) The chromatic eye: A new reduced-eye model of ocular chromatic aberration in humans. *Appl Opt* 31:3594–3600.
5. Stockman A, Sharpe LT (2000) The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Res* 40:1711–1737.
6. Ravikumar S, Thibos LN, Bradley A (2008) Calculation of retinal image quality for polychromatic light. *J Opt Soc Am A Opt Image Sci Vis* 25:2395–2407.
7. Geisler WS, Najemnik J, Ing AD (2009) Optimal stimulus encoders for natural tasks. *J Vis* 9(13):17, 1–16.
8. Foster DH, Nascimento SMC, Amano K (2004) Information limits on neural identification of colored surfaces in natural scenes.. *Vis Neurosci* 21:331–336.

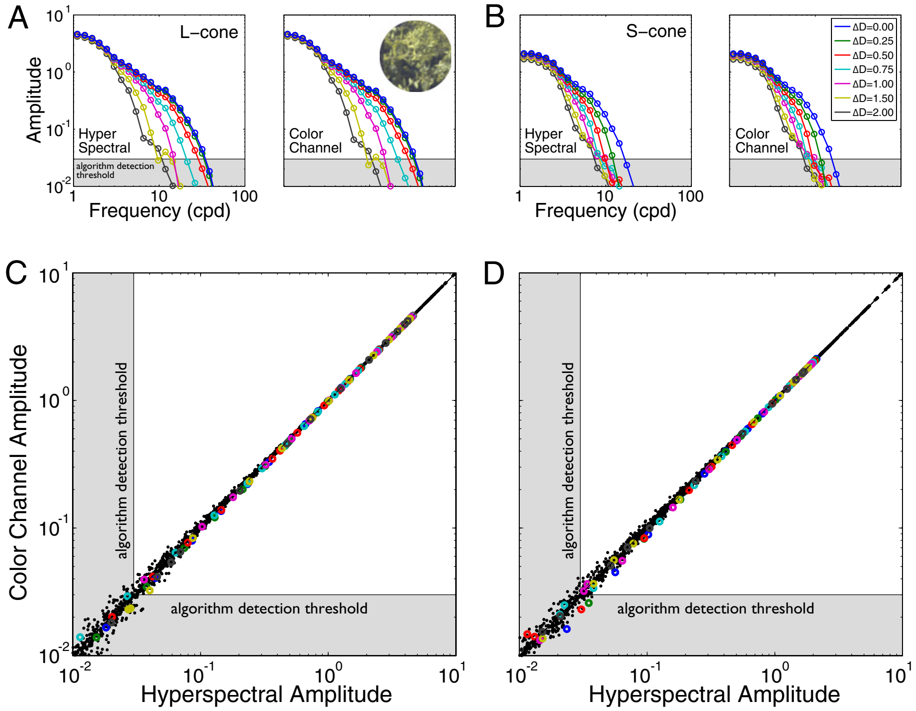

Fig. S1. Test of three-color-channel approximation to hyperspectral images. (*A*) Hyperspectral (*Left*) and color-channel (*Right*) L-cone sensor amplitude spectra for a particular patch (*Inset*). Hyperspectral sensor responses were obtained via Eq. **2** in the main text and color-channel sensor amplitude spectra were obtained via Eq. **S4**, the approximation that was used throughout the paper. Different colors indicate different defocus levels. The gray area shows the threshold below which amplitudes were not used in the analysis. (*B*) Hyperspectral (*Left*) and color-channel (*Right*) S-cone sensor amplitude spectra of the same patch (*Inset* in *A*). (*C*) Hyperspectral vs. color-channel amplitudes in the L-cone channel for 20 patches randomly selected from the hyperspectral image database (8). The approximation (Eq. **S4**) is perfect if all points fall on the unity line. Colored circles show the correspondence between the amplitudes from the particular patch shown in *A*. Black dots show the correspondence for amplitudes in the other 19 test patches. (*D*) Hyperspectral vs. color-channel amplitudes in the S-cone channel for the same 20 patches. Colored circles show the correspondence between the amplitudes shown in *B*.

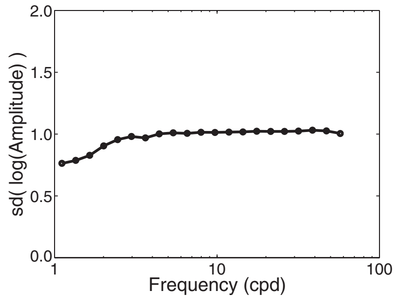

Fig. S2. Average standard deviation (SD) of logged ampliutde in each radial bin across all stimuli in the training set. The log transform nearly equalizes the SD of the amplitude within each radial bin, especially in the critical range >3 cpd.

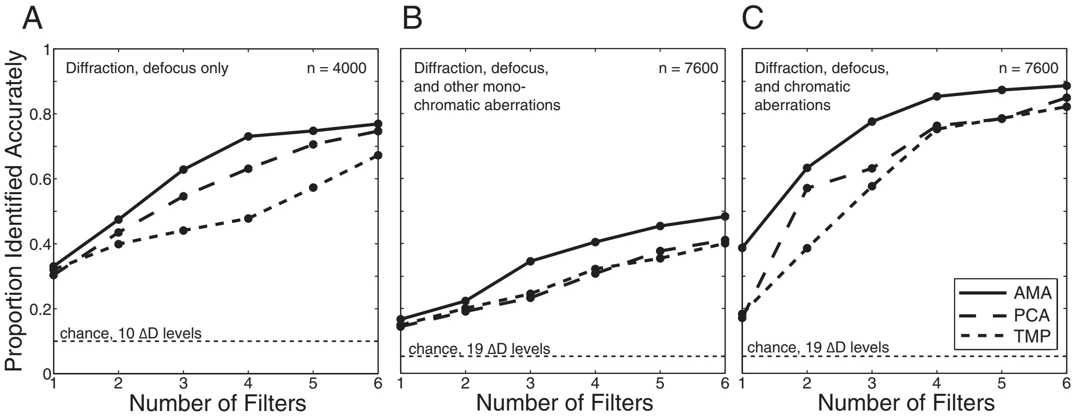

Fig. S3. Defocus filter comparison in defocus identification performance: AMA filters (solid lines) vs. PCA filters (dashed lines) and template filters (dotted lines) for the vision systems considered in the paper. Identification accuracy is plotted as a function of the number of filters. (*A*) Diffraction- and defocus-limited vision system with a sensor array sensitive only to 570 nm light. (*B*) Vision system limited by the monochromatic aberrations of the first author's right eye. (*C*) Vision system with diffraction, defocus, and chromatic aberration and with a sensor array composed of two sensors with wavelength sensitivities similar to the human L and S cones. Note that chance performance is higher in *A* than in *B* and *C* by nearly a factor of 2 because there were more defocus levels used in *B* and *C* than in *A* (19 vs. 10). To directly compare identification performance in *A* to that in *B* and *C*, multiply the identification performance in *A* by 10/19.

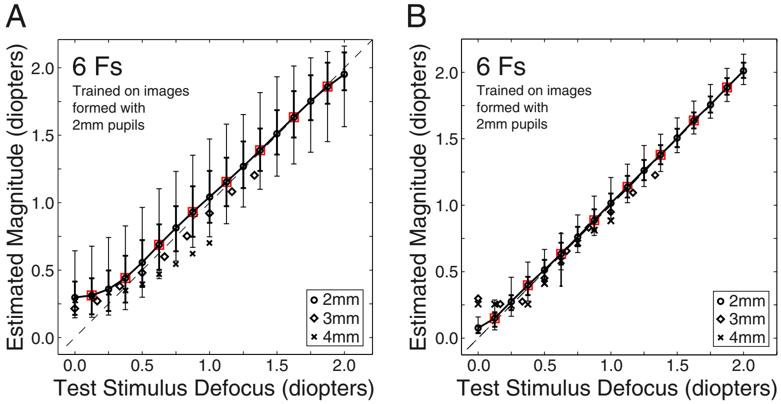

Fig. S4. Defocus magnitude estimates and filter robustness to different pupil diameters. (*A*) Results for the vision system with the monochromatic aberrations of the first author's right eye. Magnitude estimates (circles) are similar to those obtained with perfect optics (Fig. 2*D*). Although precision is somewhat reduced, the monochromatic aberrations introduce the benefit of enabling decent estimates of defocus sign (Fig. 4*B*). Diamonds and crosses show defocus estimates for images formed with 3- and 4-mm pupils, respectively, instead of the 2-mm pupil images upon which the filters were trained. (*B*) Results for the vision system sensitive to chromatic aberrations having sensors like human L and S cones. Defocus estimates are robust to changes in pupil diameter. The robustness of the estimates means that filters determined for one pupil diameter can generalize well for other pupil diameters. The correct pupil diameter was assumed in all cases. If incorrect pupil diameters are assumed, defocus estimates are scaled by the ratio of the correct and assumed diameters. Note that under the geometric optics approximation, 2-mm pupils with 2.0 diopters of defocus produce the same defocus blur (i.e., blur circle diameter) as 3- and 4-mm pupils with 1.33 and 1.0 diopters of defocus, respectively.

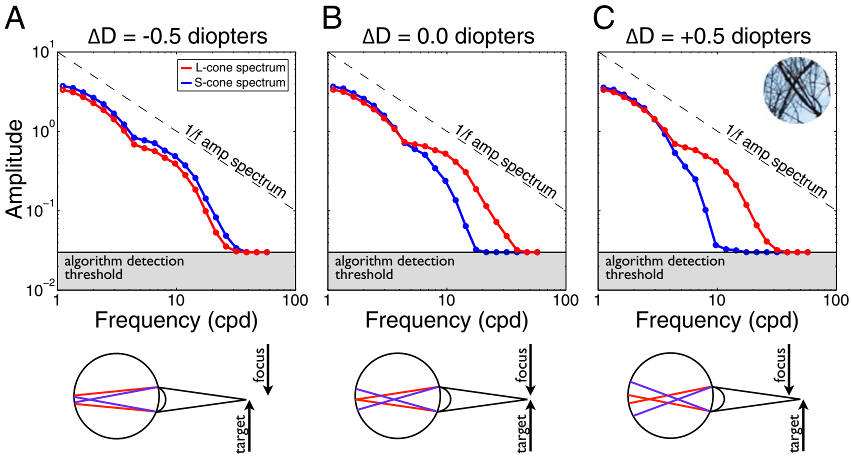

Fig. S5. Fully radially averaged L- and S-cone frequency spectra for the same patch shown in Figs. 1*C* and 3, for (*A*) −0.5, (*B*) 0.0, and (*C*) +0.5 diopters of defocus. The difference between the L- and S-cone spectra is significantly larger than the difference between the spectra in different orientation bands introduced by the monochromatic aberrations of the first author's right eye (Fig. 3*E*). In other words, the signal introduced by the optics is larger for chromatic than for the monochromatic aberrations in the analyzed eyes.

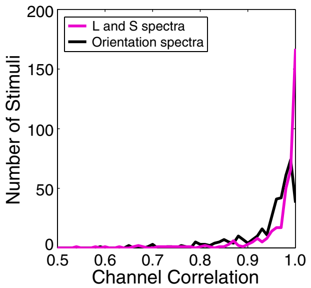

Fig. S6. Color vs. orientation channel correlation for the same collection of natural image patches. The correlation between the amplitude spectra in the color channels (L and S) is higher than the correlation between the spectra in the orientation bowties (Fig. 3*D*). This difference between the two correlations was to be expected. Wavelength illumination and reflectance functions are broadband, suggesting that color channels should be highly correlated. On the other hand, the amplitude at different orientations varies considerably with image content (e.g., an obliquely oriented edge).

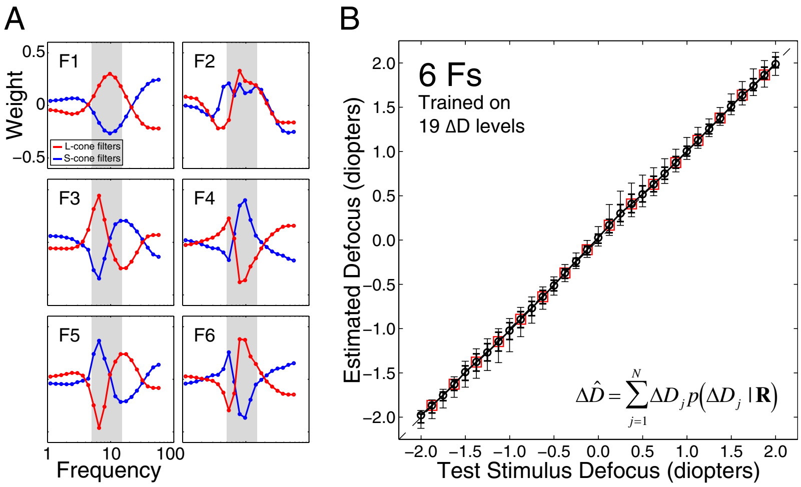

Fig. S7. Defocus filters and estimation performance for a vision system with a cone mosaic having full-resolution spatial sampling rates for both L and S cones (128 samples/degree each). The vision system was otherwise identical to the third model considered in the main text. "Training" and "test" stimuli from the main text were used to train filters and test estimation performance. (*A*) Optimal defocus filters are comparable to the filters shown in Fig. 4*C*. As expected, in these filters spatial frequency selectivity is slightly higher than in the main text, because the L- and S-cone image undersampling does not occur in this system. (*B*) Defocus estimates. Performance is comparable to that shown in Fig. 4*D*, although precision is slightly increased. Thus, the sampling rates of human cones do not significantly reduce defocus estimation performance.

Table S1. Johannes Burge, right eye, Zernike coefficients, 2-mm pupil diameter

j	*n*	*m*	Zernike coefficient, μm	Zernike term
1	0	0	0	Piston
2	1	−1	0	Tilt
3	1	1	0	Tilt
4	2	−2	0.033296604	Astigmatism
5	2	0	−0.000785912	Defocus
6	2	2	0.007868414	Astigmatism
7	3	−3	0.021247462	Trefoil
8	3	−1	−0.002652952	Coma
9	3	1	−0.004069984	Coma
10	3	3	−0.001117291	Trefoil
11	4	−4	−0.003315845	
12	4	−2	0.000470568	Secondary astigmatism
13	4	0	−0.002159882	Spherical
14	4	2	−0.003245562	Secondary astigmatism
15	4	4	0.000722913	
16	5	−5	0.000152741	
17	5	−3	−0.000338946	
18	5	−1	0.000409569	Secondary coma
19	5	1	0.000433756	Secondary coma
20	5	3	−0.000141623	
21	5	5	−0.000425779	
22	6	−6	−2.19851*E*-05	
23	6	−4	0.00011365	
24	6	−2	−8.65552*E*-06	
25	6	0	0.000103126	Secondary spherical
26	6	2	7.40655*E*-05	
27	6	4	9.48473*E*-07	
28	6	6	4.66819*E*-05	
29	7	−7	5.89112*E*-06	
30	7	−5	1.73869*E*-07	
31	7	−3	2.9185*E*-06	
32	7	−1	−8.47174*E*-06	
33	7	1	−7.90212*E*-06	
34	7	3	2.59235*E*-06	
35	7	5	7.59019*E*-06	
36	7	7	−3.07495*E*-06	
37	8	−8	2.43143*E*-06	
38	8	−6	1.77089*E*-07	
39	8	−4	−1.30228*E*-06	
40	8	−2	−3.92712*E*-07	
41	8	0	−1.59687*E*-06	
42	8	2	−9.91955*E*-07	
43	8	4	1.00225*E*-07	
44	8	6	−7.46211*E*-07	
45	8	8	−2.76361*E*-06	
46	9	−9	−1.60158*E*-08	
47	9	−7	−2.31327*E*-08	
48	9	−5	−1.97329*E*-08	
49	9	−3	−3.49865*E*-09	
50	9	−1	4.11879*E*-08	
51	9	1	4.64632*E*-08	
52	9	3	−1.72462*E*-08	
53	9	5	−4.16899*E*-08	
54	9	7	4.61718*E*-09	
55	9	9	7.37214*E*-08	
56	10	−10	3.85138*E*-08	
57	10	−8	−1.07015*E*-08	
58	10	−6	−1.00234*E*-09	
59	10	−4	4.98049*E*-09	
60	10	−2	4.99783*E*-09	
61	10	0	9.41298*E*-09	
62	10	2	5.92213*E*-09	

Table S1. Cont.

j	*n*	*m*	Zernike coefficient, μm	Zernike term
63	10	4	−1.47403*E*-09	
64	10	6	5.24061*E*-09	
65	10	8	1.78739*E*-08	
66	10	10	−8.1141*E*-09	

Astigmatism: RMS wavefront error, 0.03421 μm. Higher-order aberrations: RMS wavefront error, 0.02245 μm.