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

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

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,

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

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 $\boldsymbol{\Sigma}_j$,

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

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

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

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.

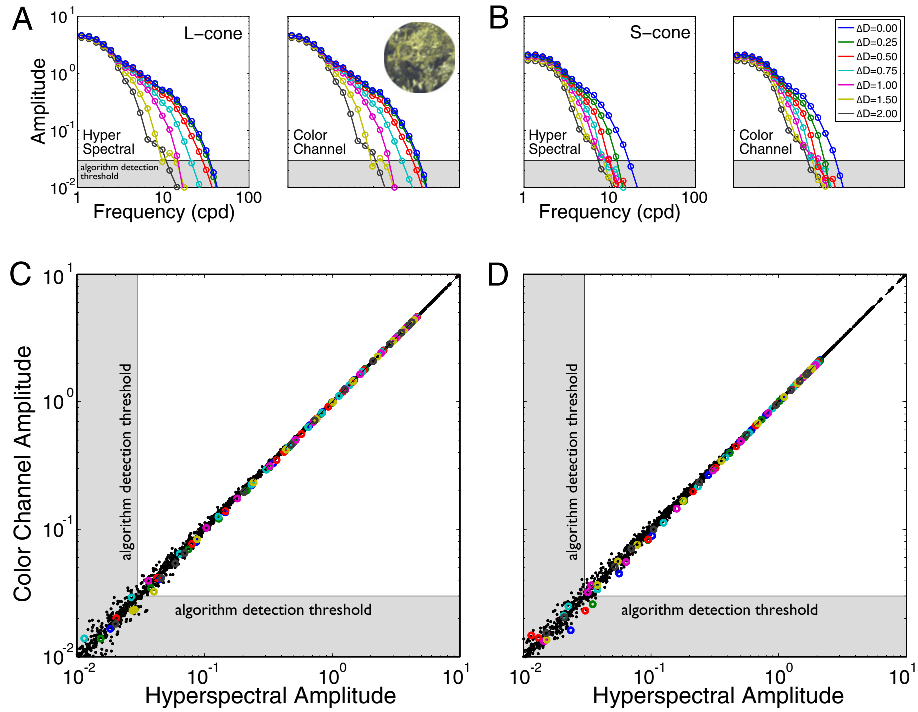

Fig. S1. Test of three-color-channel approximation to hyperspectral images. (*A*) Hyperspectral (*Left*) and color-channel (*Right*) L-cone sensor amplitude spectra for a particular patch (*Inset*). Hyperspectral sensor responses were obtained via Eq. **2** in the main text and color-channel sensor amplitude 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*.

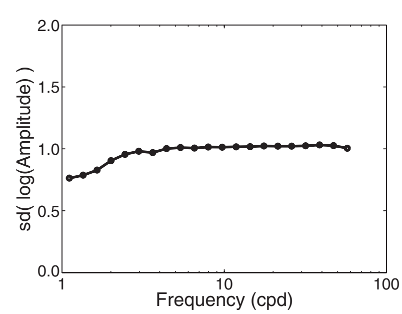

Fig. S2. Average standard deviation (SD) of logged ampliutde in each radial bin across all stimuli in the training set. The log transform nearly equalizes the SD of the amplitude within each radial bin, especially in the critical range >3 cpd.

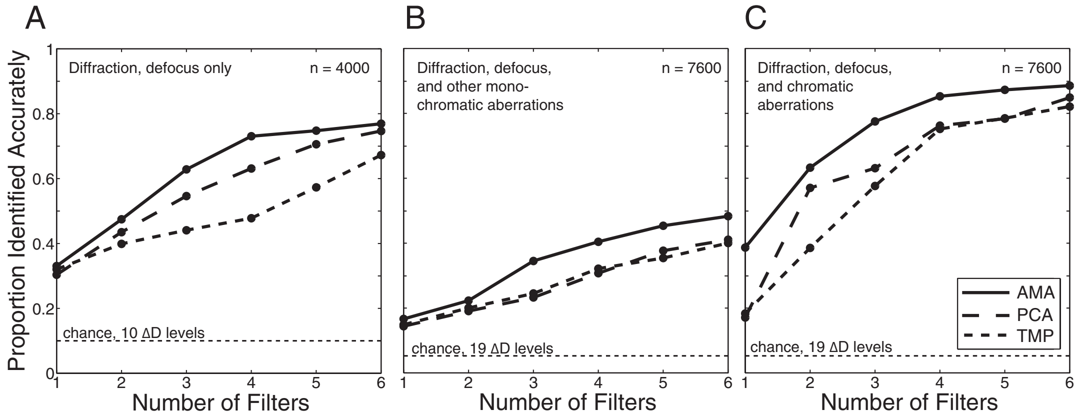

Fig. S3. Defocus filter comparison in defocus identification performance: AMA filters (solid lines) vs. PCA filters (dashed lines) and template filters (dotted lines) for the vision systems considered in the paper. Identification accuracy is plotted as a function of the number of filters. (*A*) Diffraction- 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.

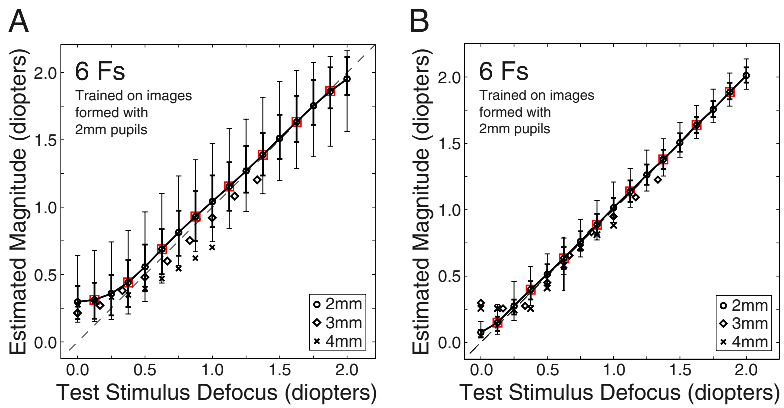

Fig. S4. Defocus magnitude estimates and filter robustness to different pupil diameters. (*A*) Results for the vision system with the monochromatic aberrations of the first author's right eye. Magnitude estimates (circles) are similar to those obtained with perfect optics (Fig. 2*D*). Although precision 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.

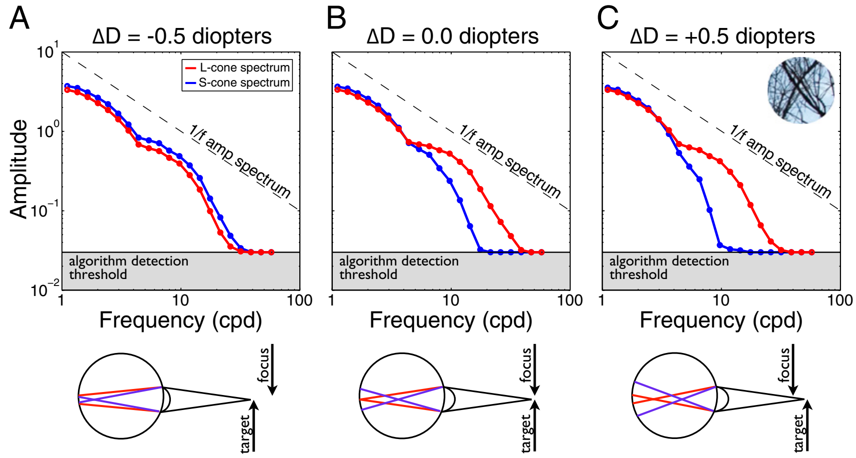

Fig. S5. Fully radially averaged L- and S-cone frequency spectra for the same patch shown in Figs. 1*C* and 3, for (*A*) −0.5, (*B*) 0.0, and (*C*) +0.5 diopters of defocus. The difference between the L- and S-cone spectra is significantly larger than the difference between the spectra in different orientation 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.

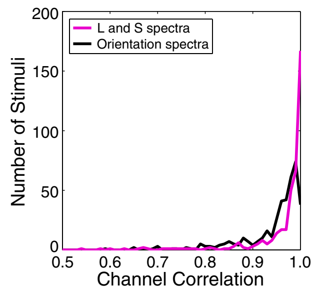

Fig. S6. Color vs. orientation channel correlation for the same collection of natural image patches. The correlation between the amplitude spectra in the color channels (L and S) is higher than the correlation between the spectra in the orientation bowties (Fig. 3*D*). This difference between the two 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).

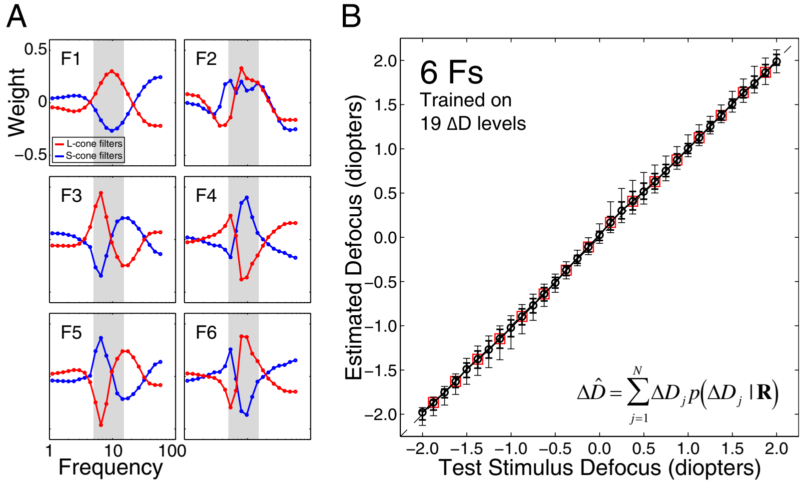

Fig. S7. Defocus filters and estimation performance for a vision system with a cone mosaic having full-resolution spatial sampling rates for both L and S cones (128 samples/degree each). The vision system was otherwise identical to the third model considered in the main text. "Training" and "test" stimuli 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.