Optimal defocus estimates from individual images for autofocusing a digital camera

Johannes Burge[a*] & Wilson S. Geisler[a]
[a] Center for Perceptual Systems, University of Texas at Austin, Austin, TX 78712

ABSTRACT

Recently, we developed a method for optimally estimating focus error given a set of natural scenes, a wave-optics model of the lens system, a sensor array, and a specification of measurement noise. The method is based on first principles and can be tailored to any vision system for which these properties can be characterized. Here, the method is used to estimate defocus in local areas of images (64x64 pixels) formed in a Nikon D700 digital camera fitted with a 50mm Sigma prime lens. Performance is excellent. Defocus magnitude and sign can be estimated with high precision and accuracy over a wide range. The method takes an integrative approach that accounts for natural scene statistics and capitalizes (but not does depend exclusively) on chromatic aberrations. Although chromatic aberrations are greatly reduced in achromatic lenses, we show that there are sufficient residual chromatic aberrations in a high-quality prime lens for our method to achieve good performance. Our method has the advantages of both phase-detection and contrast-measurement autofocus techniques, without their disadvantages. Like phase detection, the method provides point estimates of defocus (magnitude and sign), but unlike phase detection, it does not require specialized hardware. Like contrast measurement, the method is image-based and can operate in "Live View" mode, but unlike contrast measurement, it does not require an iterative search for best focus. The proposed approach could be used to develop improved autofocus algorithms for digital imaging and video systems.

Keywords: Defocus, natural images, optics, chromatic aberration, Bayesian statistics, autofocus, phase-detection, contrast measurement

1. INTRODUCTION

Consider a photographer who has just photographed a subject in front of a mountain landscape. Now imagine that the photographer decides to photograph the mountains alone. She will recompose the photograph and, because the camera is focused at the wrong distance, press the shutter half-way down to engage the camera's autofocus mechanism. Once the camera has cleared the focus error (i.e. autofocused on the mountains) she will fully depress the shutter and expose the photographic sensor.

The autofocus methods in most widespread use are contrast-measurement and phase-detection. These methods work well in many situations, but both suffer from serious drawbacks. Contrast measurement employs an iterative search for maximum contrast; it is slow and can be inaccurate. Phase detection requires costly specialized hardware (e.g., beam splitters, dedicated sensors) and does not work in "Live View" mode. The computer vision and engineering literatures describe many algorithms for estimating defocus from image data alone. However, these algorithms typically require simultaneous multiple images, special lens apertures, or light with known patterns projected onto the environment [1-4]. In other words, the algorithms cannot typically be used with standard camera images.

Despite the extensive body of work on autofocusing, there is still no widely accepted formal theory of defocus estimation from image data alone. Recently, we proposed such a theory by combining the principles of ideal Bayesian estimation together with a characterization of natural image statistics, a characterization of the wave-optics of the lens system, and a characterization of the spectral sensitivities, spatial sampling, and noise properties of the sensor array [5]. Here we demonstrate that the theory provides a useful alternative method for autofocusing (and potentially depth estimation) in conventional digital cameras. We apply this method to the problem of estimating defocus (magnitude and sign) from 64x64 pixel areas in images captured by a Nikon D700 SLR camera fitted with a Sigma 50mm prime lens.

[*] jburge@mail.cps.utexas.edu; 512-475-7872; 512-471-7356

Digital Photography VIII, edited by Sebastiano Battiato, Brian G. Rodricks, Nitin Sampat,
Francisco H. Imai, Feng Xiao, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8299,
82990E · © 2012 SPIE-IS&T · CCC code: 0277-786X/12 · doi: 10.1117/12.912066

Our method capitalizes on two main dimensions of information: i) the shape of the power spectrum of local image patches and ii) the differences between the local power spectra in the color channels. The difference between the color channels depends both on the correlations between the spatio-chromatic spectra of natural scenes and on the chromatic aberrations of the imaging system [5]. Although most high-quality camera lenses are 'achromatic' (i.e., many chromatic aberrations are eliminated) there are sufficient residual chromatic aberrations to estimate the sign of the defocus well above chance. In this paper we deliberately consider a high quality lens, because such lenses represent a worst-case scenario. Thus, the results presented in this paper represent a lower bound on performance; lower quality lenses (e.g. cell phone camera lenses) should generally produce better estimation performance.

By integrating the statistical structure of natural scenes, the properties of a high-quality digital SLR camera, and a Bayesian statistical analysis, we show that for each location in an individual image, it is possible to obtain accurate and precise estimates of defocus magnitude. We also show that sign estimation performance that is far better than chance. Thus, our approach provides the advantages of contrast-measurement and phase-detection autofocusing while avoiding their disadvantages.

2. METHODS & RESULTS

The defocus of a target is the difference between the lens system's focus distance and the target distance: $\Delta D = D_{focus} - D_{target}$ where ΔD is the defocus, D_{focus} is the current focus distance, and D_{target} is the target distance expressed in units of diopters (1/meters). The goal is to estimate ΔD in each local image patch.

Defocus information is determined by the statistical structure of natural scenes and the imaging system's optics, sensors, and noise characteristics. The input from a natural scene is represented by an idealized (i.e., unaffected by optics) input image, $I(\mathbf{x},\lambda)$, which gives the radiance at each location $\mathbf{x}=(x,y)$ in the sensor array for each wavelength λ. The optical system is represented by a point-spread function $psf\left(\mathbf{x},\lambda;a(\mathbf{z},\lambda),W(\mathbf{z},\lambda,\Delta D)\right)$ that gives the spatial distribution of light across the sensor array produced by a point target of wavelength λ. The point-spread function depends on the aperture function $a(\mathbf{z},\lambda)$ that specifies the shape, size, and transmittance of the pupil aperture for each wavelength. It also depends on the wavefront aberration function $W(\mathbf{z},\lambda,\Delta D)$, which depends on the position $\mathbf{z}$ in the plane of the aperture, the wavelength, and the defocus [6]. The sensor array is represented by a wavelength sensitivity function $s_c(\lambda)$ (normalized so the sensitivities sum to 1.0) and a spatial sampling function $samp_c(\mathbf{x})$ for each sensor class, *c*. Sensor noise is represented by a spatial-frequency-dependent detection threshold. Combining these factors (except for sensor noise) gives the spatial pattern of responses in each sensor class:

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

where $*$ represents two-dimensional convolution in $\mathbf{x}$. In terms of Equation (1), the goal is to estimate defocus, ΔD, at each point in an image from local sensor responses, $r_c(\mathbf{x})$, in the available sensor classes (typically, R, G, and B sensors). Thus, we need to measure the relevant natural scene statistics, as well as the optics, sensors, and noise properties of the digital imaging system of interest.

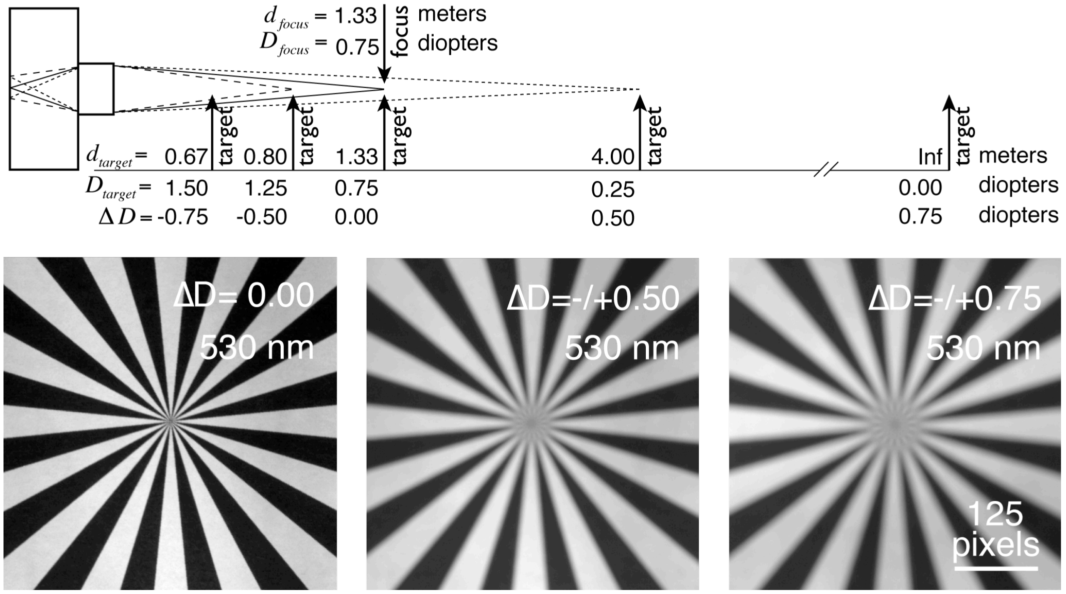
2.1 Natural Scenes

To obtain an empirical estimate of the statistical structure of idealized natural images, we collected a set of 80 well-focused three-color-channel photographs of varied scenes on and around the UT Austin campus (images available at www.cps.utexas.edu/natural_scenes) [5]. To ensure that these images were sharp, we focused the camera on optical infinity and took care that all imaged objects were at least 16 meters away. This ensured a maximum of -1/16 diopter of defocus blur in the training set. In most cases, however, the defocus blur was nearer to zero. Each input image is thus an approximation to the true idealized input image. Eight hundred 64 x 64 pixel patches were randomly selected from the photographs; four hundred for training and four hundred for testing.

Patches with RMS contrast less than 5% were excluded. This exclusion removed the small percentage of patches that were dominated by camera pixel noise (16%; 9% from blank blue sky, 7% from non-sky regions). This exclusion level is conservative. Humans typically select patches with much higher contrasts. We asked two experienced photographers, who were naïve to the purpose of the study, to select five 64x64 patches from each of 80 natural input images. They were instructed to select patches, via a mouse click, that they might try to autofocus a camera on. Neither photographer selected a single patch with contrast less than 14%.

2.2 Optics

The first step in estimating defocus is to characterize the properties of the camera lens for different defocus levels. For a fixed aperture, the primary factor in determining optical quality is the mismatch between the target distance and the lens's focus distance. As this focus error increases, the optics become progressively more low-pass. The Siemens star is good stimulus for illustrating the effect of defocus on image quality. Fig. 1 shows images (taken with our D700 camera) of a Siemens star square-wave target, illuminated by 530nm light. As focus error increases, the highest-frequency at which contrast is still visible decreases.



d_{focus} = 1.33 meters
D_{focus} = 0.75 diopters
focus
target
target
target
target
target
d_{target} = 0.67 0.80 1.33 4.00 Inf meters
D_{target} = 1.50 1.25 0.75 0.25 0.00 diopters
ΔD = -0.75 -0.50 0.00 0.50 0.75 diopters
ΔD= 0.00
530 nm
ΔD=-/+0.50
530 nm
ΔD=-/+0.75
530 nm
125 pixels

Figure 1. Distance, defocus, and the effect of defocus on image quality. The diagram shows the relationship between distance and defocus, for a camera that is focused at 1.33 meters (0.75 diopters). Image quality decreases as defocus magnitude increases. In a diffraction limited optical system, defocus levels of the same magnitude but opposite signs (-/+) produce identical defocus blur. The photographs are of a Siemens star square-wave test pattern for three different focus errors (ΔD = 0.0, 0.5, 0.75). All depicted targets were illuminated with 530nm wavelength light. Optical degradation is radially symmetric. Contrast reversals are noticeable as defocus increases.

An imaging system's optics also depends on the wavelength of light. A simple lens (e.g., the human lens) best focuses light of a specific wavelength reflected off an object at a specific distance; light of all other wavelengths is blurred. This phenomenon is known as longitudinal chromatic aberration. Longitudinal chromatic aberration reduces image quality because the images captured by different color channels are blurred differentially. In the human eye, the effect of chromatic aberration is huge: between the peak sensitivities of the long- and short-wavelength cones (570 and 445nm, respectively) [7], the optics change power by ~1 diopter [8] (Fig. 3c), the same change in power that occurs when focus distance changes from infinity to 1 meter.

Camera lens manufacturers have gone to great lengths to develop high-quality 'achromatic' lenses in which chromatic aberrations are greatly reduced (Fig. 3c). These lenses can simultaneously best focus light at two well-separated wavelengths, thereby reducing differential color blurring and improving image quality.

However, improving image quality by reducing chromatic aberrations reduces a signal that is useful for estimating the sign of focus error [5,9,10]. Therefore, one issue we address is whether sufficient residual chromatic aberrations remain in high-quality achromatic prime lenses for the sign of a focus error to be well estimated.

In the analyses that follow, we use a Sigma 50mm prime lens with a maximum aperture of f/2.8. Given its 46.8˚ horizontal field of view and the camera's 4256x2832 pixel resolution, the sampling rate of the sensor array was 91.5 samples per deg. The lens aperture was set to f/10, which corresponds to an aperture diameter of 5mm.

We characterize the optics at each wavelength and defocus level with a monochromatic point-spread function (*psf*). We estimate the monochromatic *psf* for multiple wavelengths across the visible spectrum. These monochromatic *psf*s are then combined to form three polychromatic point-spread functions, one for each color-channel. The polychromatic point-spread functions are used to simulate the effect of the camera optics on images for different levels of defocus.

To determine the effect of chromatic aberrations, we measured the camera's *psf* as a function of wavelength, at a nominal defocus of -0.5 D. First, a temporary target was positioned at 1.33 m (0.75 D) from the nodal point of the lens. Under broadband lighting, the camera's autofocus mechanism was used to focus the lens on the target. The temporary target was then removed. Next, a Siemens star square-wave target was positioned at 0.80 m (1.25 D) for a defocus (ΔD) of -0.5 D (see Fig. 1). After darkening the room, we illuminated the test pattern with a monochromatic light source and photographed the Siemens star. We repeated the procedure for wavelengths between 400 and 700 nm in 10 nm steps [11].

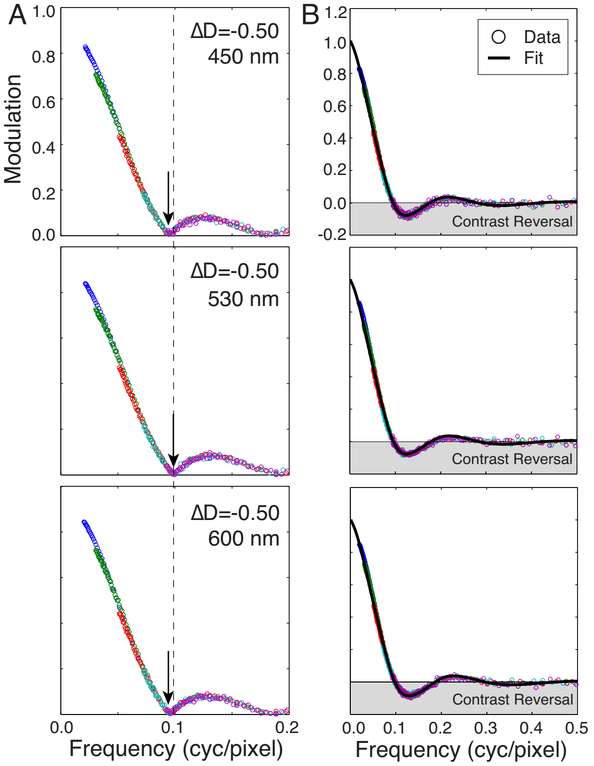


Figure 2. Estimating and fitting camera optics. **A.** Each panel shows the estimated modulation transfer function (MTF) of a target imaged with ΔD = -0.5 for three different wavelengths—450nm, 530nm, 600nm—matched to the peak wavelength sensitivity of the camera's color channels. Arrows mark the lowest frequency at which a phase reversal occurs. The lateral displacement of the arrows indicates the presence of chromatic aberrations. The MTF for 530 nm was obtained from analysis of the second photograph in Fig 1. Different colors mark MTF estimates from different harmonics of the square-wave (blue = fundamental, green = 3rd harmonic, red = 5th harmonic, etc.). **B.** OTF fits (black curves) to the data shown in A. The gray area indicates frequencies for which contrast reversals occur. Note that the axes have different scales.

We estimated the monochromatic optical transfer function (OTF)—the Fourier transform of the *psf*—by comparing the Fourier coefficients of the defocused star target to those of an idealized sharp square wave target [12]. In estimating the OTF, the optics was assumed to be radially symmetric. The procedure's accuracy was verified with simulations.

Fig 2A shows the monochromatic OTF estimates for three different wavelengths. The arrows indicate the lowest frequency at which a phase reversal occurs. (Such phase reversals are readily seen in the photograph in Fig 1. at -0.75 D of defocus.) The lateral displacement of the arrows indicates a change in the OTF with wavelength, which in turn indicates the presence of chromatic aberrations.

Then, we fit the raw OTF estimates with a model OTF generated from a single surface lens model that was as well matched to the camera lens as possible (aperture size, aperture shape, focal length, defocus). Chromatic defocus and spherical aberration were left as free parameters. The fitted OTFs matched the shape of the raw data well. Example fitted OTFs (black curves) are shown in Fig. 2b. The wavelength-dependent change in chromatic defocus and spherical aberration are shown in Fig. 3a,b.

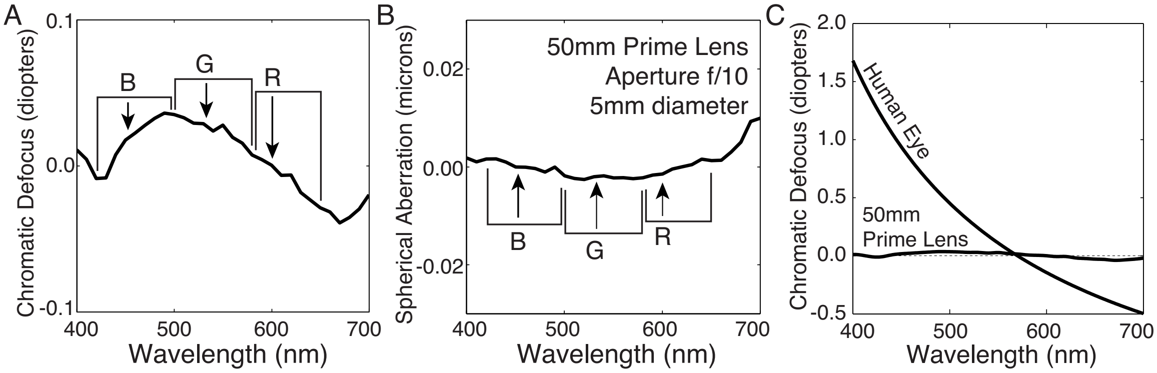


Figure 3. Chromatic aberrations in a 50mm prime lens. **A** Change in chromatic defocus with wavelength. Arrows mark the peak sensitivity of each color channel. Brackets indicate the width at half-height of each color channel's wavelength sensitivity function (see Section 2.3 & Fig. 5). **B** Change in spherical aberration with wavelength. **C** Comparison of chromatic defocus in a 50mm prime lens (same as in A) with chromatic defocus in the human lens.

Polychromatic OTFs for each color-channel were obtained by performing a weighted average of the fitted monochromatic OTFs, where the weights are given by the wavelength sensitivity functions measured via the method described in the next section. Thus, the polychromatic OTF is given by

$$OTF_c\left(\mathbf{f},\Delta D\right)=\sum_{\lambda} OTF\left(\mathbf{f},\lambda,\Delta D\right)s_c\left(\lambda\right) \tag{2}$$

where $\mathbf{f}=\left(u,v\right)$ indicates horizontal and vertically oriented frequencies.

We simulated polychromatic OTFs for 15 different defocus levels (-0.875 to 0.875 D in 1/8 D steps) by appropriately adding or subtracting defocus to the fitted monochromatic OTFs, and then performing the same weighted averaging as before. For a fixed focus distance of 1.33 m (0.75 D), this range of defocus levels corresponds to target distances ranging from 0.67 m to infinity (see Fig. 1). We assume chromatic defocus does not change with target distance (which is known to hold) and we ignore the minor changes in spherical aberration with target distance. Fig. 4 shows example polychromatic modulation transfer functions (MTFs) for select levels of defocus; the MTF is the magnitude of the OTF.

Although we have taken pains to ensure that our measurements are accurate, some inaccuracies inevitably remain. Potentially more accurate results could be obtained via direct measurement of the lens' wavefront aberrations using commercially available equipment. Nonetheless, our measurements should be more than adequate for demonstrating the value of the proposed method for estimating defocus.

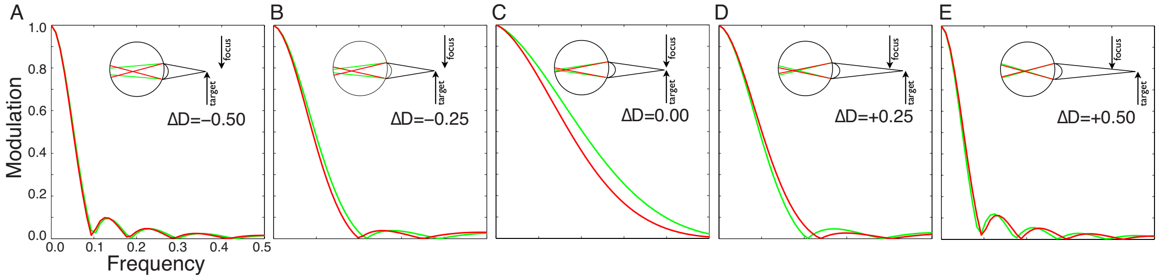


Figure 4. Polychromatic modulation transfer functions (MTFs) for five defocus levels. Red and green channel MTFs for **A** -0.50 D, **B** -0.25 D, **C** 0.00 D, **D** +0.25 D, and **E** +0.50 D. The green-channel MTFs are higher pass for negative defocus, whereas the red-channel is higher pass for positive defocus. This relationship holds for all other defocus levels. Thus, the difference between the color channels within panels provides a small but reliable signal to the sign of defocus. The difference in average shape across panels provides a reliable signal to the magnitude of defocus.

2.3 Sensor Array

Next, we determined the wavelength sensitivity and spatial sampling of the sensor array. In an otherwise dark room, a reflectance standard having a flat reflectance spectrum was illuminated with a monochromatic light source. A spectroradiometer was positioned at a 45˚ angle on one side of the reflectance standard; a camera was positioned at a 45˚ angle on the other side (Fig. 5a). The reflected spectrum was measured with the spectroradiometer and an image was captured with the camera. Care was taken not to overexpose the photos. The procedure was repeated every 10nm between 400 and 700nm. The spectral measurements, aperture, shutter speed, and average pixel value from each of the camera's color-channels were used to determine the wavelength sensitivities of the color channels [13]. The sensitivities of each color channel are shown in Fig. 5b. Spatial sampling was matched to the camera's Bayer color filter array pattern as indicated from the product specifications (Fig. 5c).

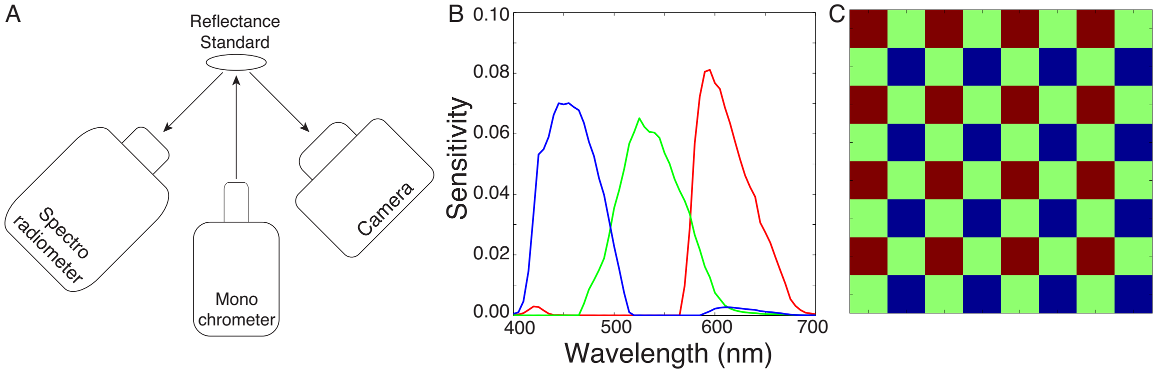


Figure 5. Characterizing the sensor array. **A** The physical setup used to measure the wavelength sensitivity of each color channel. **B** Estimated red, green, and blue pixel wavelength sensitivity functions. **C** Spatial sampling pattern of the Bayer color filter array.

2.4 Noise

The Nikon D700 has a CMOS sensor. CMOS sensors have a white spatial pixel noise component that is given by:

$$\sigma_{\bar{r}}^2 = \alpha \bar{r} + \sigma_0^2 \tag{3}$$

where $\sigma_{\bar{r}}^2$ is the variance of the sensor response, σ_0^2 is additive baseline noise variance, α is a multiplicative scalar, and $\bar{r}$ is the mean sensor response. In addition, CMOS sensors have a fixed spatial pattern of noise. Ideally some of this pattern noise could be subtracted out. However, the aim in this paper is to demonstrate the potential usefulness of our method for defocus estimation. Thus, for simplicity, we lumped the noise sources together.

The procedure for estimating the sensor noise was as follows. First, we captured images of an approximately uniformly illuminated sheet of white paper (Fig. 6, left insets) for a fixed shutter speed (1/60 sec) and a range of aperture sizes. The camera lens was defocused to increase the uniformity of the central image region. For each aperture size, we computed the mean sensor response $\bar{r}$. The variance of the sensor response $\sigma_{\bar{r}}^2$ was computed after filtering the image region with a 'derivative' kernel, $(1,-1)$, in the vertical and horizontal directions (Fig. 6, right insets). The purpose of the filtering was to remove smooth non-uniformities in gray level due to non-uniform illumination or camera optics [14]. However, the kernel also removes some noise power. We calibrated the noise power removed by the kernel by applying it to Gaussian white noise with a variance of 1.0. Multiplying the measured variance of the filtered camera images by the inverse variance of this filtered white noise corrects for the noise power removed by the kernel. The symbols in Fig. 6 represent the corrected variance in a patch as a function of mean sensor response. The solid line shows the least-squares fit of Equation (3) to the data. The procedure was repeated for each color channel. The best-fit parameters are $\alpha = 0.54$ and $\sigma_0^2 = 265$ for the red channel, $\alpha = 0.41$ and $\sigma_0^2 = 54.3$ for the green channel, and $\alpha = 0.87$ and $\sigma_0^2 = 89.3$ for the blue channel. All parameters assume a 16-bit dynamic range (i.e. maximum sensor response equals 2^{16}-1).

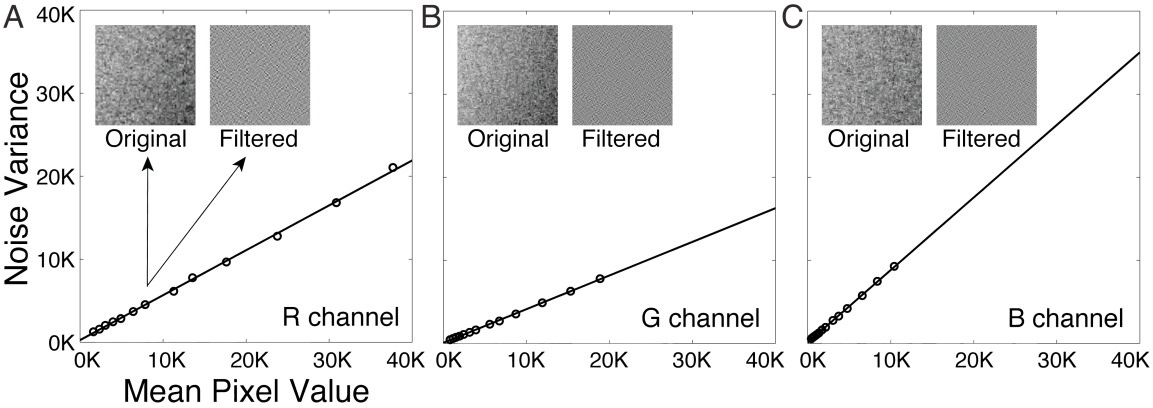


Figure 6. Characterizing sensor noise for a CMOS sensor. Noise variance is plotted as a function of the 16-bit mean pixel value in thousands for the **A** red color channel, **B** green color channel, and **C** blue color channel. Each symbol represents a different image with a different exposure. The line shows the fitted noise model (Equation 3). Best-fit parameter values are in the main text.

One potential concern is that the white-noise assumption, in conjunction with the filtering operation, may underestimate the contribution of the fixed pattern noise, because the power spectrum of the pattern noise is not perfectly flat [15]. However, we find that quadrupling the estimated noise variance in the camera has a negligible effect on the defocus estimation performance described below.

2.4 Defocus estimation

With the training set of natural inputs, and the measurements of the digital camera's optics, sensors, and noise, we investigated how well defocus magnitude and sign can be estimated from 64x64 pixel patches of natural image.

There are two steps in our method for estimating defocus: i) learn the optimal spatio-chromatic filters for defocus estimation (in the particular digital imaging system) and ii) use the filter responses to estimate defocus optimally. The first step is computationally expensive, but it must be performed only once for a given camera system. The second step is extremely fast (i.e. on the order of a millisecond) and could be implemented in the firmware of a digital imaging system.

To learn the optimal defocus filters, we need a large set of 64x64 pixel training patches. Equation 1 specifies how to determine the (noiseless) sensor responses, given an idealized hyper-spectral input $I(\mathbf{x},\lambda)$. However, as mentioned in Section 2.1, for training and testing we used well-focused three-color-channel images $\left[I_{\mathrm{R}}(\mathbf{x}), I_{\mathrm{G}}(\mathbf{x}), I_{\mathrm{B}}(\mathbf{x})\right]$ as approximations to idealized hyper-spectral images. Each channel was defocused with polychromatic point-spread functions (see Section 2.2), before being spatially sampled by the sensor array. Specifically, the sensor responses associated with each color-channel are given by

$$r_c(\mathbf{x}) = \left[I_c(\mathbf{x}) * psf_c(\mathbf{x}, \Delta D)\right] samp_c(\mathbf{x}) \tag{4}$$

where $psf_c(\mathbf{x}, \Delta D)$ is a polychromatic point-spread function, which is obtained via inverse Fourier transform of a polychromatic OTF (see Equation 2). We have previously shown that these approximations are highly accurate for simulating the sensor responses to defocused hyper-spectral images [5]. In fact, three-color-channel camera images are preferable because hyper-spectral images are often contaminated by motion blur. Gaussian white noise was added to the sensor responses of each training and test patch according to Equation 3 and the parameters in Section 2.4.

From 400 randomly sampled idealized training patches (Section 2.1) and 15 polychromatic OTFs (Section 2.2), we generated 6000 defocused training patches. We consider only the defocus information in the red (R) and green (G) sensors, because the difference in chromatic defocus is largest between R and G (see Fig. 3a). (Later we show that inclusion of the blue sensor can further improve performance.) Thus, for the present analysis, the defocus information is contained in the noisy sensor responses, $r_{\mathrm{R}}(\mathbf{x})$ and $r_{\mathrm{G}}(\mathbf{x})$. Each sensor patch was converted to a contrast patch by subtracting off and dividing by the mean. Finally, we applied a cosine window, obtained the Fourier spectra, and radially averaged the power spectra of the sensor responses (for each sensor class) for each patch. These power spectra constituted the training inputs to the algorithm.

Optimal spatio-chromatic filters were learned using a recently developed statistical technique for dimensionality reduction called Accuracy Maximization Analysis (AMA). The algorithm returns rank-ordered filters that maximize accuracy in a specific task [16]. (A Matlab implementation of AMA is available at http://jburge.cps.utexas.edu/research/Code.html.) In the present case, AMA finds the spatio-chromatic filters that maximize the accuracy for the task of defocus estimation, over the given dioptric range [5]. Here, we learned eight AMA filters. The first four filters are shown in the insets of Fig. 7.

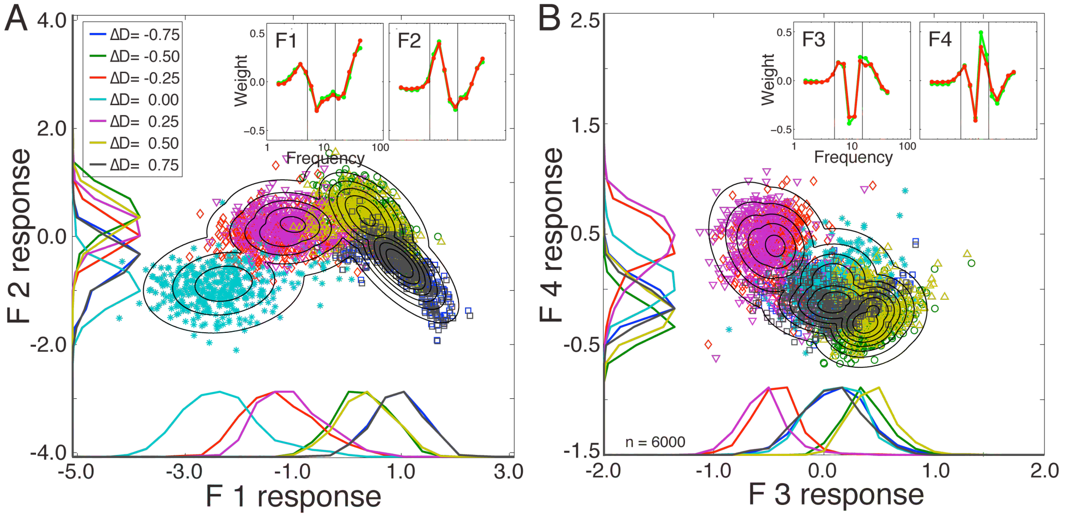


Figure 7. Defocus filters and filter response distributions (i.e., conditional likelihood distributions). **A** Joint filter response distributions and Gaussian fits for filters 1 and 2 (insets). Each symbol represents a joint response to an individual image patch. Different colors indicate different defocus levels. Filter responses cluster as a function of defocus level. Some levels are not shown for clarity. **B** Filter response distributions and Gaussian fits for filters 3 and 4 (insets). Filters 1 and 2 primarily separate the defocus levels by magnitude, whereas filters 3 and 4 do a relatively better job than filters 1 and 2 at separating defocus sign, especially for -0.50 and +0.50 D.

The red- and green-channel power spectra associated with a given patch can be described as a single combined vector, as can a particular spatio-chromatic frequency filter. Thus, the filter response *R* to a given patch is obtained by taking the dot product of the two vectors. With multiple filters, a vector of filter responses **R** is obtained for each patch in the training set. The joint responses of filters 1 and 2 and filters 3 and 4 are shown in Fig. 7a,b (the contours are iso-probability contours for fitted two-dimensional Gaussians). As can be seen, the filter responses cluster as a function of defocus level.

We characterized the joint (8 dimensional) filter response distributions for each defocus level by fitting Gaussians from the sample means and covariances, $p\left(\mathbf{R} \mid \Delta D_i\right) = gauss\left(\mathbf{R}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)$. These are the conditional likelihood functions. Bayes' rule specifies how to obtain the posterior probability of each defocus level, given the filter responses and conditional likelihood functions

$$p\left(\Delta D_i \mid \mathbf{R}\right) = \frac{p\left(\mathbf{R} \mid \Delta D_i\right) p\left(\Delta D_i\right)}{\sum_j p\left(\mathbf{R} \mid \Delta D_j\right) p\left(\Delta D_j\right)} \tag{4}$$

Fig. 8 shows the posterior probability distributions across defocus level computed for several actual defocus levels of a given example image patch, assuming a uniform prior over defocus level. Notice that there is information in the posterior distributions about both the magnitude and the sign of defocus.

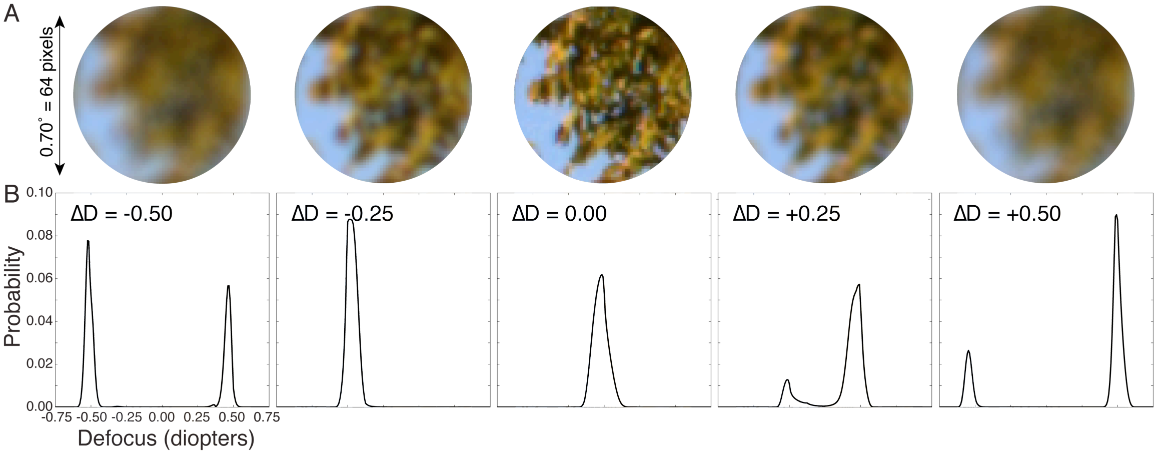


Figure 8. Defocused test patches and posterior probability distributions. **A** Example test patches for five levels of defocus (ΔD = -0.50, -0.25, 0.00, +0.25, +0.50). **B** Corresponding posterior probability distributions. Some distributions are uni-modal, others are bi-modal, and still others (not shown) are multi-modal. The locations of the peaks signal the magnitude of defocus and the ratio of the mass on either side of zero signals the sign of defocus.

To estimate of the magnitude of defocus from a posterior probability distribution, we considered two estimators, the maximum *a posteriori* (MAP) estimator (e.g., the location of the peak in each panel of Fig. 8)

$$\Delta\hat{D}_{mag} = \left| \underset{\Delta D}{\arg\max} \left[p\left(\Delta D \mid \mathbf{R}\right) \right] \right| \tag{5}$$

and the expected magnitude (EMG) estimator,

$$\Delta\hat{D}_{mag} = \sum_i \left|\Delta D\right| p\left(\Delta D_i \mid \mathbf{R}\right) \tag{6}$$

The two estimators perform similarly, but the EMG estimator is somewhat more robust.

To estimate defocus sign we compute the log posterior mass ratio (LPMR) and then take its sign:

$$LPMR = \log \frac{p(\Delta D \geq 0 \mid \mathbf{R})}{p(\Delta D \leq 0 \mid \mathbf{R})} \tag{7}$$

$$\Delta\hat{D}_{sgn} = \operatorname{sgn}(LPMR) \tag{8}$$

Note that the LPMR is the log ratio of the posterior probability mass on either side of zero (see Fig. 8).

To test the algorithm we defocused 400 randomly sampled test patches (Section 2.1) with polychromatic OTFs corresponding to each of 29 defocus levels (-0.875 to 0.875 D in 1/16 D steps) to obtain 11600 test patches. We then performed the same steps as above to generate a set of test power spectra. Thus, none of the test patches were in the training set, and only half the test defocus levels were in the training set.

Estimates of defocus magnitude are accurate and have high precision (± 0.04 D) over a wide range (Fig. 9a). Defocus sign is estimated correctly 78.4% of the time (Fig. 9b). If the above calculations are repeated for the green and blue sensors and the LPMR values for each sensor pair are added ($LPMR_{RG}$ + $LPMR_{GB}$), then sign identification performance increases to 80.2%. Performance on the test set equaled performance on the training set, indicating that the natural scenes were not under-sampled and that the filters were not over-fit.

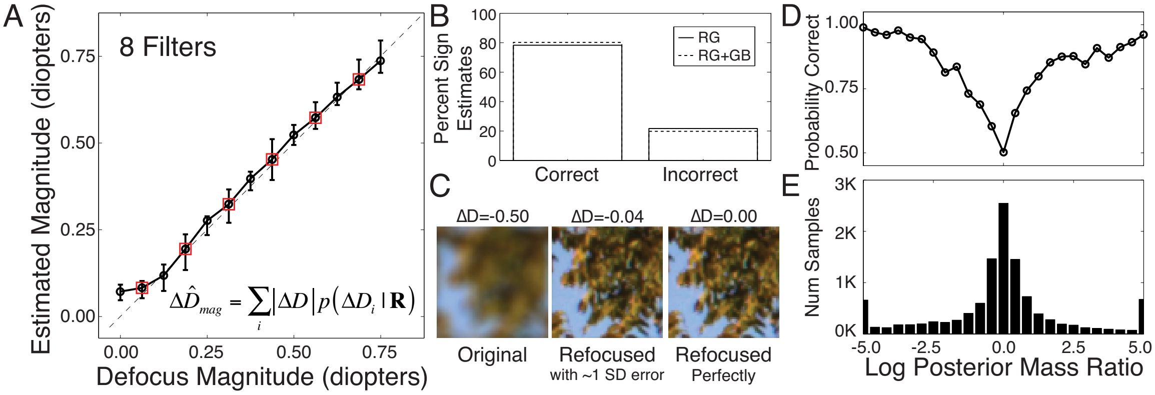


Figure 9. Estimates of defocus in test patches. **A** Median of expected magnitude (EMG) defocus estimates. Error bars are 68% confidence intervals. Red boxes indicate defocus levels not in the training set. The similarly-sized error bars at both trained and untrained levels indicate that the algorithm produces continuous estimates. **B** Accuracy of sign identification. **C** An out-of-focus image patch, a patch that has been refocused based on a typical estimation error, and a perfectly refocused patch. **D** The log posterior probability mass ratio (Equation 7) provides a patch-by-patch confidence signal that the sign estimate is correct. **E** Frequency of occurrence of different values of the log probability mass ratio.

The implications of these results are illustrated in Fig. 9c. The first panel shows an image patch that is substantially out-of-focus ($\Delta D = -0.5$). The second panel shows an image patch that has been refocused with an estimate that is in error by -0.04 D, an error that is ~1 standard deviation from the mean (i.e., half the average 68% confidence interval). The third panel shows that a patch that has been refocused with a -0.04 D error is almost indistinguishable from a perfectly focused patch.

The sign identification performance reported here (80% correct) is remarkable given the low level of chromatic aberrations in the achromatic lens (see Discussion). Indeed, the positively and negatively defocused patches in Fig. 8 are nearly indistinguishable perceptually. We have shown previously [5], that our algorithm can achieve near perfect sign identification performance in vision systems (e.g. the human visual system; see Fig. 3c) having more significant chromatic aberrations. Thus, lenses with more chromatic aberration, like the lenses on many point-and-shoot or cell-phone cameras, will almost certainly yield better performance.

Given that our estimates of defocus sign are not perfect, it might be advantageous to have a signal that is related to the probability that a given sign estimate is correct. Fig. 9d shows the probability that the sign estimate is correct as a function of the value of the LPMR. Fig. 9e shows the frequency of occurrence for different LPMR values. There is a strong relationship between the magnitude of the LPMR and percent correct. Thus, this signal could be of potential utility when designing control systems for refocusing a camera lens.

3. DISCUSSION

The results presented here demonstrate that a digital camera's sensor responses are sufficient to estimate, with high accuracy and precision, the magnitude of defocus in any given local patch of any individual natural image. Furthermore, even for a camera with a high quality achromatic prime lens, the residual chromatic aberrations are sufficient to estimate the sign of defocus with good accuracy.

Sign identification could be further improved by using data about the camera's current focus distance. For example, if the lens is focused at 2.0 m (0.5 D) and the estimated defocus magnitude is 0.75 D, the sign of the focus error must be negative (i.e., the lens must be focused behind the target), because a target cannot be beyond infinity.

An obvious question is how the algorithm is able to estimate defocus sign as well as it does given small chromatic aberrations of the Sigma prime lens (see Fig. 3c). There are two factors. The first is that there is a high spatial correlation between the images captured by the red and green channels, which effectively allows one channel to serve as a reference for the other (i.e., the large variation in power spectra across patches of natural image is largely removed by differencing the power spectra in the two channels). The second factor is that the AMA filters optimally compare details of the spectra in the two channels. In contrast, a Bayes optimal decision rule that uses only the total power in the two channels yields performance that is only a little above chance (55.3%).

Importantly, once the optimal AMA filters and combination rules have been learned for a given camera system, the computations of the optimal defocus estimates are simple and efficient. Thus, a defocus estimate corresponding to a given patch of image can be obtained in a few milliseconds on a standard laptop computer. Presumably, if these computations were built into a camera's firmware, they could be performed even faster.

Also note that the performance levels presented here probably reflect the 'worst-case' scenario for defocus estimation in digital imaging systems. The optical quality was high (low aberrations), and thus the amount of the defocus information introduced by the optics was small compared to what would be introduced by lower quality lenses. For these reasons, we speculate that our algorithm is likely to produce even more accurate estimates of magnitude and sign in point-and-shoot and cell phone cameras. In addition, it should be possible to design high quality lens systems that are optimized for estimating defocus with the method described here. Finally, we note that because our method is image-based, it will operate in live-view mode and/or in digital video cameras. When paired with an appropriate control routine for autofocusing the lens, our optimal defocus estimation method may lead to improved, more flexible autofocusing performance.

4. CONCLUSION

The two most widely used autofocus techniques, phase detection and contrast measurement, have their strengths but both also have serious weaknesses. The method proposed here combines the advantages of both phase detection and contrast measurement autofocusing without suffering their disadvantages. This work demonstrates the potential value of taking algorithmic inspiration from biological science.

ACKNOWLEDGMENTS

We thank Samuel Thurman for generously providing Matlab code. We used a slightly modified version of his code to estimate the OTFs from Siemens star targets. We thank Larry Thibos for helpful discussions on Fourier and wavefront optics. This work was supported by National Institutes of Health (NIH) Grant EY11747 and National Science Foundation (NSF) Grant IIS-1111328.

REFERENCES

[1] Pentland AP, Scherock S, Darrel T, Girod B. “Simple range cameras based on focal error”. Journal of the Optical Society of America-A, 11(11): 2925-2934 (1994).

[2] Watanabe M & Nayar SK. “Rational filters for passive depth from defocus”. International Journal of Computer Vision, 27(3): 203-225 (1997).

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

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

[5] Burge J, Geisler WS. “Optimal defocus estimation in individual natural images”. Proceedings of the National Academy of Sciences, 108 (40): 16849-16854 (2011).

[6] Goodman JW. [Introduction to Fourier Optics], McGraw-Hill, New York, 2nd Edition (1996).

[7] Stockman A, Sharpe LT. “The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype”. Vision Research, 40: 1711–1737 (2000).

[8] Thibos LN, Ye M, Zhang X, Bradley A. “The chromatic eye: A new reduced-eye model of ocular chromatic aberration in humans”. Applied Optics, 31: 3594–3600 (1992).

[9] Flitcroft DI. “A neural and computational model for the chromatic control of accommodation”. Visual Neuroscience, 5: 547–555 (1990).

[10] Wilson BJ, Decker KE, Roorda A. “Monochromatic aberrations provide an odd error cue to focus direction”. Journal of the Optical Society of America-A, 19(5): 833–839 (2002).

[11] Thibos LN, Hong X, Bradley A, Applegate RA. “Accuracy and precision of objective refraction from wavefront aberrations”. Journal of Vision, 4: 329-351 (2004).

[12] Thurman ST. “OTF Estimation Using a Siemens Star Target”, in Imaging Systems Applications, Optical Society of America Technical Digest (CD) (2011).

[13] Ing AD, Wilson JA, Geisler WS. “Region grouping in natural foliage scenes: image statistics and human performance”. Journal of Vision, 10(4): 10, 1-19 (2010).

[14] Elder JH, Zucker SW. “Local scale control for edge detection and blur estimation”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (7): 699-716 (1998).

[15] Lukas J, Fridrich J, Goljan M. “Digital Camera Identification from Sensor Noise”. IEEE Transactions on Information Security and Forensics, 1(2): 205-214 (2006).

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