Accurate Image-Based Estimates of Focus Error in the Human Eye and in a Smartphone Camera

Estimation of focus error is a key consideration in the design of any advanced image-capture system. Today's contrast-based auto-focus algorithms in digital cameras perform more slowly and less accurately than the human eye. New methods for estimating focus error can close this gap. By making use of optical imperfections, like chromatic aberration, these new methods could significantly improve the performance of digital auto-focusing techniques.

by Johannes Burge

THE visual systems of humans and other animals perform powerful computations that exploit information in retinal images that is useful for critical sensory-perceptual tasks. The information in retinal images is determined by the statistical structure of natural scenes, projection geometry, and the properties of the optical system and the retina itself. Task performance is determined by the quality of the information available in retinal images and by how well that information is exploited. To characterize the theoretical limits of performance in a specific natural task, all these factors must be accounted for.

Nearly all sighted mammals have lens-based imaging systems (eyes) that focus and defocus light on the retinal photoreceptors. The estimation of focus error (*i.e.*, defocus) is one particularly important natural task. Focus information is useful for a wide range of tasks, including depth estimation, eye-growth regulation, and accommodation control.[6,8,15] Typical lenses focus light from only one distance at a time, but natural scenes contain objects and surfaces at many distances. Most regions in images of depth-varying scenes are therefore out-of-focus and blurry under normal observing situations. The amount of image blur caused by a given focus error depends on the lens optics and the size and shape of the lens aperture.

For tasks that depend on high-resolution images, image blur can be a significant impediment. To sharply image an out-of-focus target, the lens must be refocused so that the focus distance equals the target distance. It has been estimated that humans refocus their eyes more than 100,000 times per day.[10,12] Perhaps because of all this practice, human accommodation (biological autofocusing) is fast, accurate, and precise. Two- to three-hundred milliseconds after presentation of a defocused target, the human lens refocuses ballistically with (approximately) the correct magnitude in the correct direction nearly 100% of the time.[7]

Consumers are often frustrated by the slow speed and inaccuracy of image-based smartphone autofocus routines. Achieving the speed of human accommodation would be a great improvement. The most popular image-based autofocus routine is contrast detection. This is a "guess-and-check" procedure that employs an iterative search for maximum contrast. The procedure is non-optimal for at least two reasons: (1) Contrast-detection autofocus does not provide information about focus error sign; when simple detection algorithms start the search for best focus, the direction of the initial response (closer *vs.* farther) is random. (2) Contrast-detection autofocus does not provide estimates of focus error magnitude; in the search for best focus, the focus adjustment often crosses the point of best focus and then must turn around and come back.

Here, we describe recent advances in our ability to estimate focus error from small patches of individual images. We show that precise unbiased estimates of focus error can be obtained for both the human visual system and for a popular smartphone camera. Chromatic aberrations that are introduced by

Johannes Burge *is currently an Assistant Professor at the University of Pennsylvania where he is a member of the Department of Psychology and the Neuroscience and Bioengineering Graduate Groups. He can be reached at jburge@psych.upenn.edu.*

0362-0972/1/2017-018$1.00 + .00 © SID 2017

the lenses of these vision systems can be used to resolve the sign ambiguity. Thus, the approach has the potential to significantly improve image-based autofocus routines in smartphone cameras, medical devices for assistive vision, and other electronic imaging devices.

Background

Focus-error estimation suffers from an inverse-optics problem; from image information alone, it is impossible to determine with certainty whether a given image pattern is due to focus error (blur) or some feature of the scene (*e.g.,* shadows). Focus-error estimation is also said to suffer from a sign ambiguity; under certain conditions, focus errors of the same magnitude but different signs produce identical images. These issues may make it seem that accurate focus-error estimation from individual images is impossible. However, in many vision systems, the optical properties of the lens and the sensing properties of the photosensor array, together with the statistical properties of natural images, make a solution possible. We now discuss these factors.

Statistical Properties of Natural Images

Natural images are remarkably varied. In natural viewing conditions, the eye images a staggering variety of object colors, shapes, sizes, and textures [Fig 1(a)]. In spite of this variation, there is one property of natural images that is relatively stable: the shape of the amplitude spectrum. Most well-focused natural-image patches have amplitude spectra with a $1/f$ fall-off; *i.e.*, in a typical patch, there is 10× less contrast at 10 cpd (cycles per degree) and 30× less at 30 cpd than at 1 cpd. Of course, the shape of the amplitude spectrum varies somewhat with patch content, and variability increases as patch size decreases. Nevertheless, the shape of the natural amplitude spectrum is stable enough. To obtain an empirical estimate of the statistical structure of natural images, we collected a large database of well-focused images of natural scenes.[2.]

Optical Properties of Lenses

Focus-error changes the shape of the amplitude spectrum. Small focus errors attenuate the spectrum (*i.e.*, power) at high frequencies; intermediate focus errors attenuate the spectrum at intermediate frequencies, and so on [Fig. 1(b)]. These shape changes provide information about focus-error magnitude [Fig. 1(c)]. However, under certain conditions, lenses provide no information about the sign of the error (focus too close *vs.* too far). For example, in an ideal optical system with monochromatic light, image quality is degraded by focus error (*i.e.*, defocus) and diffraction alone. Focus errors of the same magnitude but opposite signs thus yield identical point-spread functions (PSFs) and corresponding modulation-transfer functions [MTFs; Fig. 1(b)]. The effect of this type of focus error on the amplitude spectrum of a representative natural image patch is shown in Fig. 1(c).

In real optical systems with broadband light, image quality is degraded not just by defocus and diffraction, but also by chromatic and monochromatic aberrations other than defocus (*e.g.*, astigmatism). Although these aberrations reduce best-possible image quality, they introduce information into retinal images than can be used to precisely estimate the magnitude and sign of focus error.[2,3,17] Here, we focus on the usefulness of chromatic aberration in the human visual system[14] and smartphone cameras.

Sensing Properties of Photosensors

For chromatic aberrations to be useful, the vision system must be able to sense them. The human visual system and most cameras have arrays of sensors that are differentially sensitive to long-, medium-, and short-wavelength light. In human vision, the sensitivities of the long- (L), medium- (M), and short- (S) wavelength cones peak at 570, 530, and 445 nm, respectively[13] In the human eye, the change in chromatic defocus between the peak sensitivities of the L and S cones is approximately 1 diopter (D).[1] In many cameras, the sensitivity of the red, green, and blue sensors

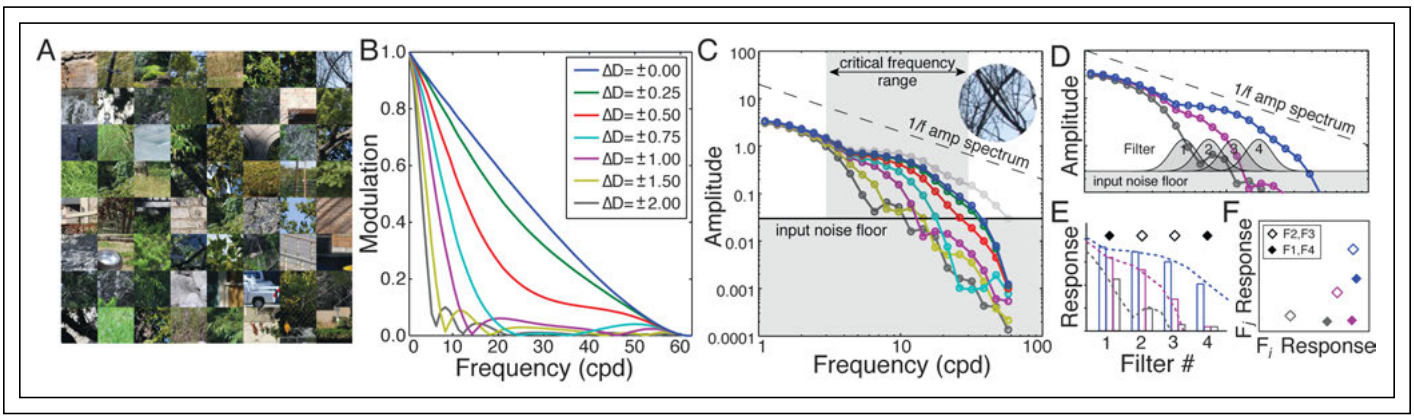

Fig. 1: Signals for focus-error estimation: (a) Natural image variation is substantial. (b) Monochromatic modulation transfer function (MTF) in a diffraction limited lens for a range of focus errors (colors). The MTF is the modulus of the Fourier transform of the point-spread function (PSF). (c) The amplitude spectrum of a particular local patch (1°, inset) changes shape systematically with focus error (colors matched to b). (d) Spatial-frequency filters (Gaussian bumps labeled 1–4) tiling the critical band of the spatial-frequency domain. (e) Each filter responds according to power in the spectrum in its passband. The responses provide a digital approximation to the shape of the amplitude spectrum. (f) Joint filter responses. Filter 2 and 3 responses (open symbols) to spectra with different focus errors are significantly further apart than filter 1 and 4 responses (closed symbols). Hence, filters 2 and 3 provide more useful information for classifying focus error in this patch.

peak at 590, 530, and 460 nm. In most cameras, chromatic defocus is markedly less than in the human eye. But even in high-quality achromatic prime lenses, measureable chromatic defocus occurs between the R and B sensors.[3]

General Principle of Estimation

The first job of a good estimator is to determine the signal features that carry good information about the task-relevant variable. Figure 1(d) shows the amplitude spectra of four generic filters (shaded Gaussian bumps), along with spectra for three amounts of focus error. Each filter increases its response according to the local power in the amplitude spectrum (above the noise floor) at the spatial frequencies to which each filter is sensitive. This set of spatial-frequency filters [Fig. 1(d)] provides a digital approximation of amplitude spectra [Fig. 1(e)], much like a bass equalizer on a car stereo provides a digital approximation of the amplitude spectra of sound waves. Figure 1(f) plots the responses of the filters against each other. Filters 2 and 3 are more useful than 1 and 4 for discriminating the three focus errors in the patch.

The problem of estimating focus error in a particular image patch is trivial compared to the task of estimating focus error in a random image patch. Natural-image variation introduces task-irrelevant changes in the typical $1/f$ shape of the amplitude spectrum that makes the problem difficult. But focus error can be estimated because it introduces shape changes that are more dramatic than those introduced by image variation. In general, if a measureable signal varies more due to the task-relevant variable than to task-irrelevant image variation, then the accurate estimation of the task-relevant variable is possible.[4,5] For the current task of focus error estimation in human and smartphone camera lenses, this condition holds.

Figure 2 demonstrates that this condition holds in the human visual system. Figure 2(a) shows examples from a training set; focus error varies down the rows, image content varies across the columns [Fig 2(a)]. Image variation introduces task-irrelevant variability in the shape of the spectrum [Fig. 2(b)], but focus error introduces much larger changes [Fig. 2(c)]. The most useful changes due to focus error occur within a critical spatial-frequency band. Natural images, because of their $1/f$ spectra, rarely have power exceeding the noise floor at high spatial frequencies. Focus error has little effect on low spatial frequencies. Thus, intermediate spatial frequencies carry the most useful information about focus error. This is the critical frequency band.

Human chromatic aberration [Figs. 2(b) and 2(c), (insets)] causes systematic differences between the spectra in two (or more) color channels that provide useful information about the sign of focus error. For negative errors (*i.e.*, focus too far), the short-wavelength sensor image is in better focus than the long-wavelength sensor image. For positive errors (focus too close), the long-wavelength

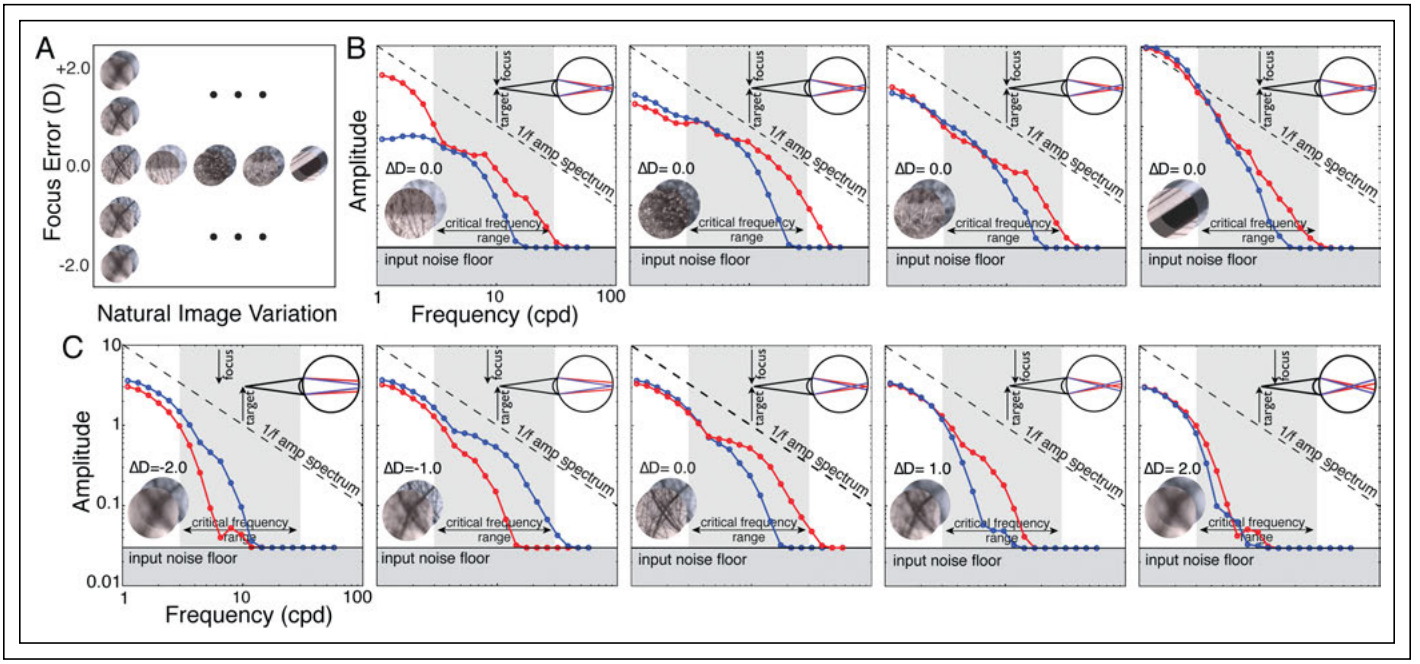

Fig. 2: *Impact of natural-image variability and focus error on shapes of amplitude spectra. Results shown for a lens with human chromatic aberration for the L- and S-cone images and for a 2-mm pupil. (a) Training set of natural image patches with different focus errors (8400 patches = 21 focus errors x 400 patches per error). (b) Amplitude spectra of the L-cone image (red) and S-cone image (blue) for four different well-focused image patches. (c) Amplitude spectra for the same patch with five different focus errors. The eyeball icon indicates focus error geometry: Negative and positive focus errors correspond to when the lens is focused behind and in front of the target, respectively. The shape of the amplitude spectrum varies randomly with the image patch and changes systematically with the focus error. The amplitude spectrum shape provides good information about focus-error magnitude. The L-cone or S-cone spectrum with more energy at higher frequencies provides good information about focus-error sign.*

sensor image is in better focus. Chromatic aberration thus introduces a useful signal for determining the sign of a focus error.

Results

We developed an algorithm for estimating focus error based on the principles and observations described above.[2] We next describe its performance for the human visual system and for a popular smartphone camera: the Samsung Galaxy S4. For the human visual system, we assumed a 2-mm pupil (typical for daylight), optics with human chromatic aberration, sensors with the wavelength sensitivities of the L and S cones, and a plausible input noise level.[16] For the Galaxy S4, we assumed a fixed 1.7-mm aperture and measured its optics, wavelength sensitivity, and noise in the R and B sensors.[3] (Two of the three available sensors are used for computational simplicity. Similar performance is obtained with all three sensors together.) Note that image blur due to focus error decreases as aperture size decreases. Vision systems with larger apertures and comparable optics will, in general, yield more accurate results than those presented here.

Next, in each vision system we found the spatial-frequency filters that are most useful for estimating focus error from –2.5 to +2.5D using Accuracy Maximization Analysis, a recently developed task-specific method for dimensionality reduction. Assuming a focus distance of 40 cm, this range of focus errors corresponds to distances of 20 cm to infinity. For the human visual system, the filters operate on the amplitude spectra of the L- and S-cone sensor images. For the Galaxy S4 smartphone, the filters operate on the amplitude spectra of the R- and B-sensor images.

The four most useful filters for estimating focus error in the human eye are shown in Fig. 3(b). These filters find the spectral features that provide the best possible information about focus error, given the variability of natural images and the effect of focus error in each color channel on the captured images' amplitude spectra. The filters concentrate in and near the frequency range known to drive

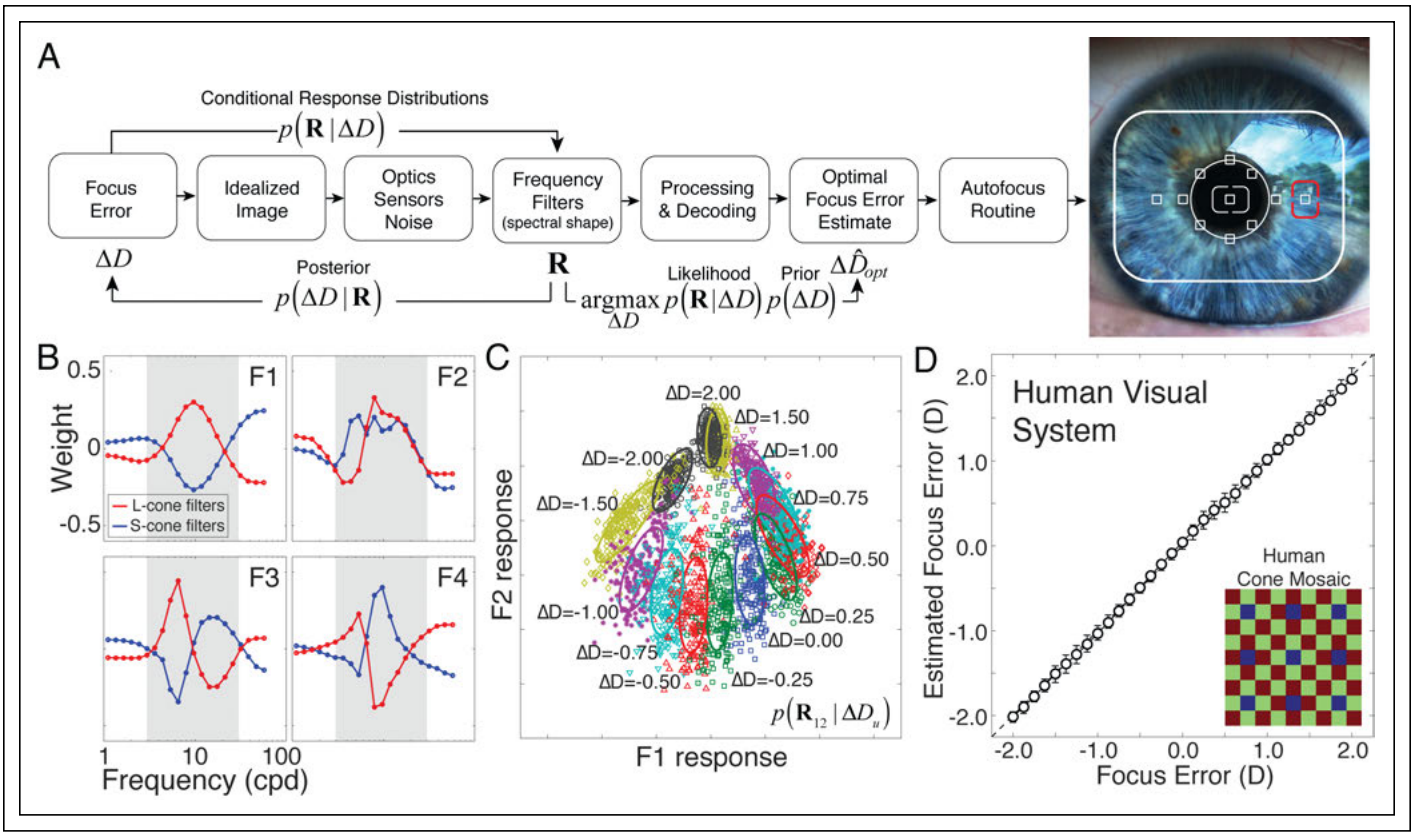

Fig. 3: *Focus-error estimation in the human visual system. (a) Schematic of optimal focus-error estimation and how it can be used to eliminate focus error as part of an autofocusing routine. The estimate of focus error can be used as input to an autofocus routine to null focus error. (b) Spatial-frequency filters that extract the most useful information for estimating focus error in the human visual system. The filters weight and sum of the amplitude spectra of captured L-cone and S-cone images. The first filter is selective for differences in the shapes of the L- and S-cone amplitude spectra and is most useful for discriminating focus-error sign. The second filter is less selective for differences between the color channels. The filters apply more weight to an intermediate frequency band because this band carries the most useful information. (c) Filters 1 and 2 responses to different retinal images (symbols) with different focus errors (colors). The conditional filter responses cluster as a function of focus error and can be approximated by a Gaussian distribution. (d) Optimal focus-error estimates across thousands of test images. Error bars represent 68% confidence intervals. Inset shows the rectangular approximation of the human-cone mosaic used to sample the images.*

human accommodation.[9] These filters also have properties that are similar to chromatic double-opponent cells in early visual cortex,[11] which have primarily been studied in the context of color processing.

The responses of the two most useful filters to thousands of randomly sampled natural-image patches with different amounts of focus error are shown in Fig. 3(c). Each symbol represents the filter responses to a particular individual image patch. Each color represents a different focus error. The fact that the responses cluster by focus error indicates that the filters extract good information about focus error from the shape of the amplitude spectrum. Next, we characterized the joint filter responses by fitting Gaussians $gauss(\mathbf{R};\mu_u,\Sigma_u) = p(\mathbf{R}|\Delta D_u)$ to each response cluster, where μ_u and Σ_u are the sample means and covariance [colored ellipses, Fig. 3(b)]. Figure 3(d) shows focus-error estimation performance in the human visual system for thousands of randomly sampled image patches. In humans, high-precision ($\pm 1/16D$) unbiased estimates of focus error are obtainable from small patches from the L- and S-cone sensor images of natural scenes.

The human visual system has much more chromatic aberration than the lenses in typical DSLR and smartphone cameras. How well do these same methods work in DSLRs and smartphones? We have previously examined the performance attainable in a DSLR camera.[3] Here, we determine focus-error estimation performance in the Galaxy S4. We measured the R, G, and B sensor wavelength sensitivities and the optics of the Galaxy S4 over a range of 5D and then used our methods to estimate focus error.

Estimation results are shown in Fig. 4. Figure 4(a) shows focus-error estimates for each of four randomly sampled patches across the range of focus errors. In each subpanel, the inset shows the posterior probability distribution over focus error for the condition circled in red. For reference, the full-size image from which the four patches were sampled is shown in Fig. 4(b). Performance is

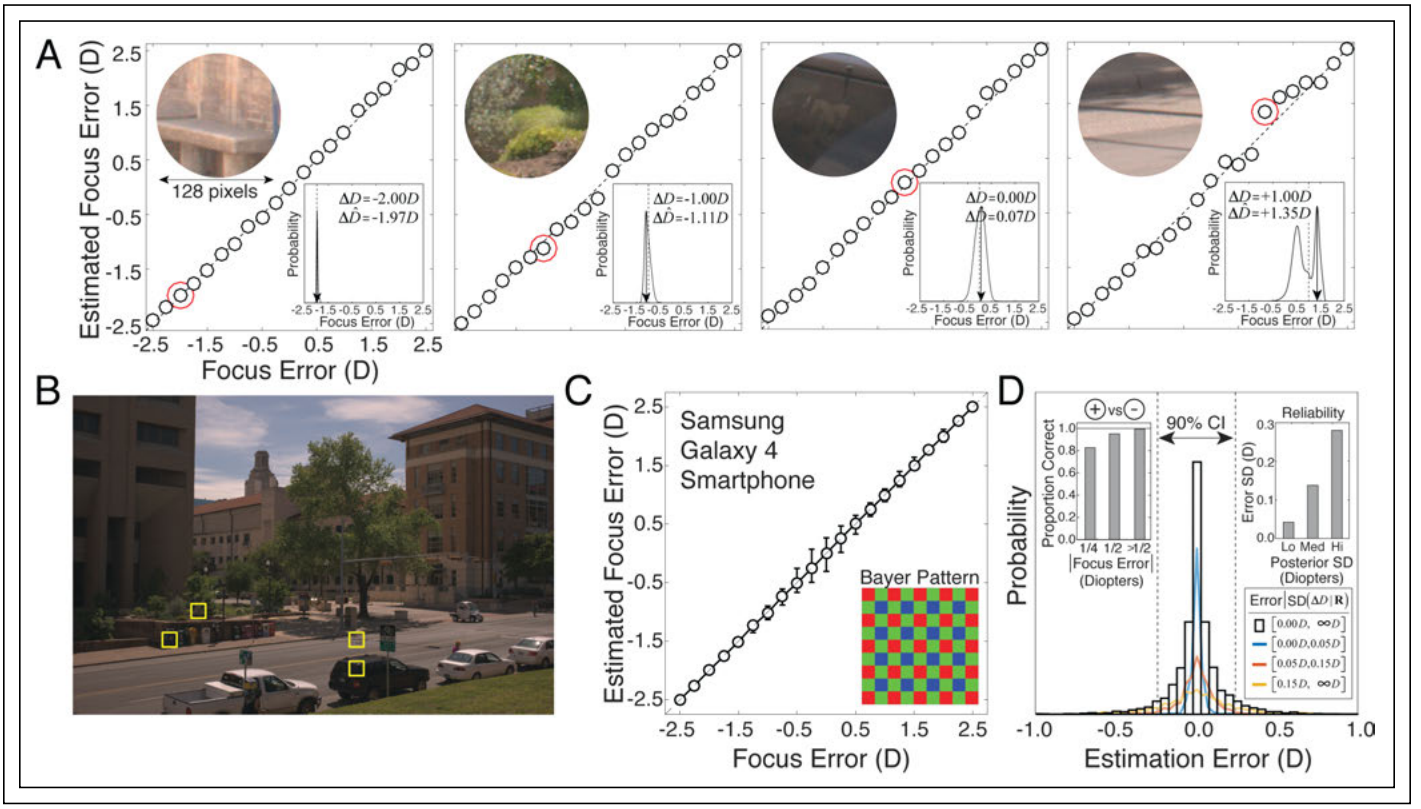

Fig. 4: *Focus-error estimation with Samsung Galaxy S4 smartphone optics and sensors. (a) Focus-error estimation for four randomly sampled natural image patches (128 x 128 pixels) over –2.5 to +2.5D. Insets show the particular image patch (without blur) and the posterior probability over focus error for one particular groundtruth focus error (red circle). Dashed vertical line indicates the true focus error. The variance (width) of the posterior can be used as a measure of estimate reliability. Performance is nearly identical with 64 x 64 pixel patches. (b) Original image from which the patches were sampled. (c) Average estimation performance as a function of focus error across 8400 test patches (21 focus errors x 400 patches). Error bars are 68% confidence intervals. Inset shows the sensor pattern that was used to sample the images. (d) Grand histogram of estimation errors. 90% of estimates are accurate to +0.25D (approximately the human blur detection threshold).[10] Colored lines show error histogram conditioned on the standard deviation of the posterior: low (SD = 0.00–0.05D; blue), medium (SD = 0.05–0.15D; red), high (SD > 0.15D; orange). Upper right inset shows that the standard deviation of the estimation error increases with the standard deviation of the posterior probability distribution. Upper left inset shows the proportion of the time focus-error sign is estimated correctly as a function of the true focus error. For focus errors 0.5D or larger, the sign is estimated correctly 99% of the time.*

good for each patch, but it is not perfect, and some patches produce more accurate estimates than others. For example, estimates for the patch in the rightmost subpanel of Fig. 4(a) are the least accurate on average. The shadows against the street curb make the sharp patch (inset) look somewhat blurry. Some of the same features that confuse humans seem to confuse the algorithm. Also, a featureless surface carries no information about focus error, and therefore yields highly inaccurate estimates. This variability in accuracy across patches is an unavoidable aspect of estimation performance with natural stimuli.[10]

It would therefore be advantageous for an autofocus routine to have not just an estimate of focus error but of each estimate's reliability. The standard deviation (width) of the posterior probability distribution predicts the reliability of each patch-by-patch estimate. This signal could therefore have utility in the design of a control system for autofocusing a smartphone camera.

Estimation performance in the Samsung Galaxy S4, averaged across thousands of patches, is shown in Figs. 4(c) and 4(d). None of the test patches were in the training set, indicating that the estimation algorithm should generalize well to arbitrary images. The grand histogram of estimate errors is shown in Fig. 4(d). Errors are generally quite small. 90% of the estimates are within +0.25D of the correct value. Given the 1.7-mm aperture and 4.2-mm focal length of the Galaxy S4 (*f*-stop of *f*/2.4), errors of ~0.25D will be within the depth of field. Sign estimation was also accurate.

The colored lines in Fig. 4(d) show error histograms conditioned on the standard deviation of the posterior probability distribution. When the posterior probability distribution has a low standard deviation [*e.g.*, Fig 4(a), left panel] errors are very small. When the posterior probability distribution has a high standard deviation [*e.g.*, Fig 4(a), right panel], errors tend to be larger. These results show that, in both humans and a popular smartphone camera, accurate estimates of focus error (including sign) can be obtained from small patches of individual images.

Applications

The method described here provides highly accurate estimates of focus error, given the optics and sensors in a popular smartphone camera, and it has the potential to significantly improve the autofocus routines in smartphone cameras and other digital-imaging devices. It has the advantages of both contrast-measurement and phase-detection autofocus techniques, without their disadvantages. Like phase detection, the method provides estimates of focus error (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. And because the method is image based and can be implemented exclusively in software, it has the potential to improve performance without increasing manufacturing cost.

This same method for estimating focus error may also be useful for improving certain medical technologies. A number of different assistive vision devices have hit the market in recent years. These devices act, essentially, as digital magnifying glasses. If these devices could benefit from improved autofocusing, our method could apply there as well.

References

[1]F. Autrusseau, L. Thibos, and S. K. Shevell, "Chromatic and wavefront aberrations: L-, M-, and S-cone stimulation with typical and extreme retinal image quality," *Vision Research* **51**(21–22), 2282–2294 (2011); http://doi.org/10.1016/j.visres.2011.08.020

[2]J. Burge and W. S. Geisler, "Optimal defocus estimation in individual natural images," *Proceedings of the National Academy of Sciences of the United States of America* **108**(40), 16849–16854 (2011); http://doi.org/10.1073/pnas.1108491108

[3]J. Burge and W. S. Geisler, "Optimal defocus estimates from individual images for autofocusing a digital camera," *Proc. IS&T/SPIE 47th Annual Meeting, Proc. SPIE* (2012); http:// doi.org/10.1117/12.912066

[4]J. Burge and W. S. Geisler, "Optimal disparity estimation in natural stereo images," *J. Vision* **14**(2) (2014); http://doi.org/10.1167/ 14.2.1

[5]J. Burge and W. S. Geisler, "Optimal speed estimation in natural image movies predicts human performance," *Nature Communications* **6**, 7900 (2015); http://doi.org/10.1038/ncomms8900

[6]R. T. Held, E. A. Cooper, J. F. O'Brien, and M. S. Banks, "Using Blur to Affect Perceived Distance and Size," *ACM Transactions on Graphics* **29**(2), 19:1–19:16 (2010); http://doi.org/10.1145/1731047.1731057

[7]S. Kasthurirangan, A. S. Vilupuru, and A. Glasser, "Amplitude dependent accommodative dynamics in humans," *Vision Research* **43**(27), 2945–2956 (2003).

[8]P. B. Kruger, P. B., Mathews, S. M. Katz, K. R. Aggarwala, and S. Nowbotsing, "Accommodation without feedback suggests directional signals specify ocular focus," *Vision Research* **37**(18), 2511–2526 (1997).

[9]K. J. MacKenzie, D. M. Hoffman, and S. J. Watt, "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control," *Journal of Vision* **10**(8), 22 (2010); http://doi.org/10.1167/10.8.22

[10]S. Sebastian, J. Burge, and W. S. Geisler, "Defocus blur discrimination in natural images with natural optics," *Journal of Vision* **15**(5), 16 (2015); http://doi.org/10.1167/15.5.16

[11]R. Shapley and M. J. Hawken, "Color in the cortex: single- and double-opponent cells," *Vision Research* **51**(7), 701–717 (2011); http://doi.org/10.1016/j.visres.2011.02.012

[12]W. W. Sprague, E. A. Cooper, S. Reissier, B. Yellapragada, and M. S. Banks, "The natural statistics of blur," *Journal of Vision* **16**(10), 23 (2016); http://doi.org/10.1167/16.10.23

[13]A. Stockman and L. T. Sharpe, "The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype," *Vision Research* **40**(13), 1711–1737 (2000).

[14]L. N. Thibos, M. Ye, X. Zhang, and A. Bradley, "The chromatic eye: a new reduced-eye model of ocular chromatic aberration in humans," *Applied Optics* **31**(19), 3594–3600 (1992).

[15]C. F. Wildsoet and K. L. Schmid, "Emmetropization in chicks uses optical vergence and relative distance cues to decode defocus," *Vision Research* **41**(24), 3197–3204 (2001).

[16]D. R. Williams, "Visibility of interference fringes near the resolution limit," *J. Opt. Soc. Am. A* **2**(7), 1087–1093 (1985).

[17]B. J. Wilson, K. E. Decker, and A. Roorda, "Monochromatic aberrations provide an odd-error cue to focus direction," *J. Opt. Soc. Am. A*, *Optics, Image Science, and Vision*, **19**(5), 833–839 (2002). ■