Fast and robust estimation of ophthalmic wavefront aberrations

Keith Dillon

Fast and robust estimation of ophthalmic wavefront aberrations

Keith Dillon[a,b,][*]
[a]Tulane University, Department of Biomedical Engineering, Lindy Boggs Center, Suite 500, New Orleans, Louisiana 70118, United States
[b]Formulens, LLC, P.O. Box 12232, San Diego, California 92039, United States

Abstract. Rapidly rising levels of myopia, particularly in the developing world, have led to an increased need for inexpensive and automated approaches to optometry. A simple and robust technique is provided for estimating major ophthalmic aberrations using a gradient-based wavefront sensor. The approach is based on the use of numerical calculations to produce diverse combinations of phase components, followed by Fourier transforms to calculate the coefficients. The approach does not utilize phase unwrapping nor iterative solution of inverse problems. This makes the method very fast and tolerant to image artifacts, which do not need to be detected and masked or interpolated as is needed in other techniques. These features make it a promising algorithm on which to base low-cost devices for applications that may have limited access to expert maintenance and operation. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.21.12.121511]

1 Introduction

Uncorrected refractive error is the leading cause of vision impairment worldwide.[1–3] There is a need for inexpensive solutions, particularly in the developing world, which lacks sufficient resources and specialists. Furthermore, the prevalence of myopia has been rapidly increasing in recent years in Asia,[4] putting a strain on eyecare systems. Wavefront sensors provide one potential avenue for automation to help address this problem. Unlike the phoropter,[5] which is a complex apparatus that requires operation by a trained expert to determine the best subjective correction, a wavefront sensor-based device can directly measure the optical performance of the patient's eyes. Conventional autorefractors[6] used older technologies and were generally unsuccessful at automating this evaluation.

The advantage of the wavefront sensor is that it also provides measurements of higher-order aberrations (where low order refers to defocus and astigmatic error) that have been shown to be critical for determining an accurate correction.[7,8] Given these measurements, a more accurate correction may then be performed with conventional spectacle lenses by optimizing a sharpness metric that estimates the net effect on the retinal point spread function.[9] Unfortunately, wavefront sensors still tend to be very sophisticated systems, primarily designed for expensive applications such as refractive surgery[10] when used in ophthalmic practice; in such an application, the cost of errors is very high and clearly an expert operator must be presumed to be available. For the application of determining the needed refractive correction, however, we can potentially trade off much of the complexity and cost to address a much wider population. When the goal is correction of refractive errors, an estimate of the full wavefront error is not needed. It is primarily a subset of third- and fourth-order aberrations, particularly the more central terms in the Zernike expansion such as coma and spherical aberration, which dominate the effects of aberrations on visual acuity.[7,8] Furthermore, these are the largest high-order aberrations seen in the population.[11] A better reduction in complexity would, therefore, be achieved with a method that focuses on robustly estimating just these important terms.

The most common ophthalmic wavefront sensor utilizes a Shack–Hartmann lenslet array.[10,12] This is a relatively expensive device due to fabrication costs,[13] particularly for high-density arrays, though prices might be reduced significantly given the development of high-volume fabrication techniques targeting the smartphone market.[14] Indeed, higher density arrays are desirable to improve accuracy;[15] lower density arrays suffer more from the curvature of the wavefront across each lenslet, as well as edge effects and related distortions caused by localized artifacts. A potentially cheaper high-density option[15] is grating-based sensors, such as those that utilize the Talbot effect.[16,17] These can easily use grating patterns with a very small pitch (equivalent to a high-density array), which can be produced by simply etching the grating onto a glass slide attached to the detector.

In addition to the sensor device, wavefront sensing requires a specialized algorithm to estimate the wavefront from the detected intensity pattern. Figure 1 shows an overview of common methods utilized for Shack–Hartmann and Talbot devices. With higher density lenslet arrays (and with Talbot gratings), it becomes increasingly difficult to determine the displacements for individual elements to perform the classical methods of Fig. 1(a), due to diffraction and reduced signal-to-noise ratio,[20] which suggests Fourier methods may be more attractive. As depicted in Fig. 1(b), Fourier methods typically must unwrap the phase after demodulation of the grating frequency.[21] Unfortunately phase unwrapping remains an open problem,[22] as it is ill-posed and NP-hard;[23] even in a noise-free scenario, one cannot guarantee an optimal result without testing a combinatoric number of possible combinations of 2π steps. In practical terms, phase unwrapping algorithms are particularly sensitive to localized image artifacts, as are the spot detection

---

*Address all correspondence to: Keith Dillon, E-mail: kdillon@formulens.com
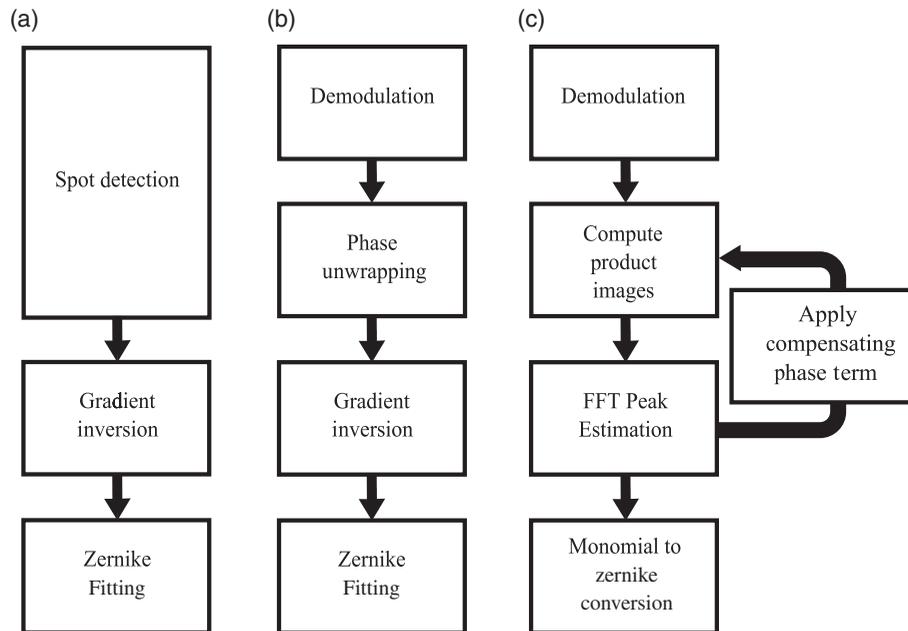
Fig. 1 (a) In classical Shack–Hartmann methods,[18] focal spot displacements are estimated to get samples of the wavefront x- and y-gradients; the wavefront is computed from these gradients via an inversion algorithm, and Zernike coefficients can be computed from the reconstructed wavefront. (b) Fourier-based methods[19] replace the spot detection step with a demodulation that directly extracts components containing the x- and y-gradients in their phase; the phase values are modulus 2π, however, so phase unwrapping is required. (c) The proposed method operates directly on the demodulated signal components, by estimating the peak locations of products of signals; no phase unwrapping or inversion is needed.

methods of Fig. 1(a), and common issues such as eyelashes or corneal scars can cause catastrophic failures in reconstruction.[24] Figure 2 shows examples of eyelashes occluding the pupil, a common artifact that can result in significant difficulties. Research on phase unwrapping continues, utilizing sophisticated nonlinear optimization techniques such as total-variation denoising,[25] unscented Kalman filters,[26] and sparse modeling.[27]

Researchers have also proposed techniques to improve or extend other components of the Fourier methods of Fig. 1(b). Demodulation can be performed by a fast filtering in the image domain.[28] Gradient inversion is itself a nontrivial inverse problem, and a variety of techniques have been developed[29] including fast noniterative approaches.[30] If the wavefront phase is small, it may also be extracted directly after demodulation in the spatial frequency domain, by taking the anti-hermitian component after centering,[31] which can then be directly used in Fourier-based inversion methods. However, this approach requires symmetry of the signal aperture, and the small-angle approximation leads to increasing errors for larger aberrations.

As noted earlier, there are only a limited number of major aberration components of interest for refractive correction, so a more specialized approach may sidestep the major difficulties of these conventional reconstruction algorithms. This paper formulates a direct approach to estimate major aberrations, generally described in Fig. 1(c), which does not need phase unwrapping or iterative inversion. The approach uses a numerical technique mathematically similar to a spatial heterodyne, to create a diversity of images with different aberrations. At each polynomial order, the algorithm estimates each aberration coefficient by finding the peak frequency component. This technique is employed in a successive fashion to compute and correct for successively lower aberrations until only the low-order terms remain to be calculated much more accurately. Furthermore, the fact that this approach is not reliant on high-quality estimates of spot locations or pixel phase allows us to take advantage of higher pitch gratings, which reduces the relative effect of edges and localized artifacts, and also allows for a very compact and easily produced sensor. We will demonstrate the performance of
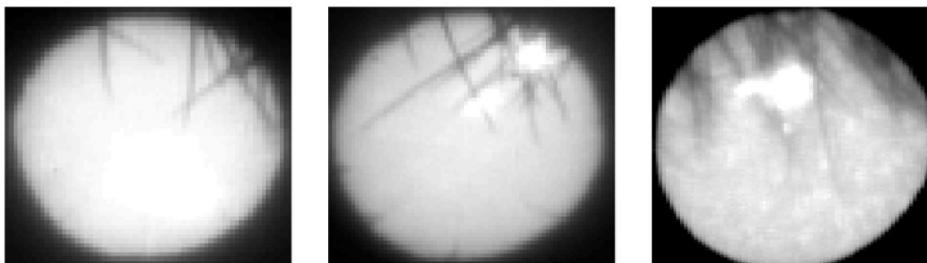
Fig. 2 Examples of eyelash shadows partly occluding pupil signal; the bright regions are reflections of the laser source from the cornea and subsequent structures.
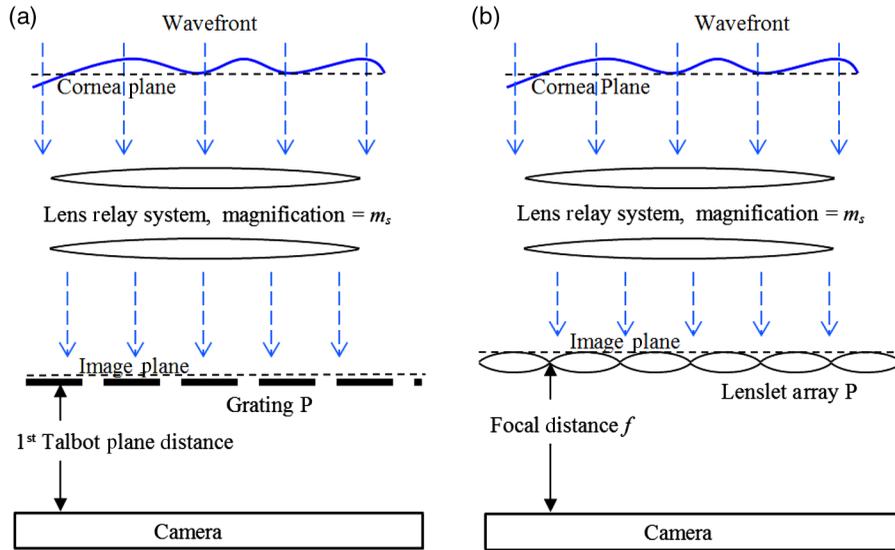
Fig. 3 Wavefront sensor: (a) employing grating with pitch P at appropriate distance from a camera to utilize Talbot effect and (b) employing lenslet array with pitch P at appropriate distance from camera to produce focused spot pattern.

the method with simulations depicting the robustness against severe artifacts, as well as over a range of realistic aberration magnitudes.

2 Method

An ophthalmic wavefront sensor typically utilizes a monochromatic laser source scattered from a point on the subject's retina. Viewing this as a point source, the scattered light passes through the eye's optics in reverse, and deviations of the resulting signal from a planar wavefront provide an estimate of the eye's optical aberrations. Most commonly, the returning signal is reimaged from the eye onto a wavefront sensor, such as in Fig. 3, which depicts a sensor utilizing the Talbot effect (a) and a Shack–Hartmann array (b). The resulting image at the camera may be described as the intensity of a interference pattern of the following equation:

$$s(x,y) = \sum_{n,m} C_{n,m} \cos\left\{ m\frac{2\pi}{P}[x - \alpha(x,y)] + n\frac{2\pi}{P}[y - \beta(x,y)] \right\},$$
(1)

where $C_{n,m}$ are the amplitudes for the different spectral components; P is the pitch of the device array (e.g., the lenslet spacing); and $\alpha(x,y)$ and $\beta(x,y)$ are proportional to the horizontal and vertical gradients, respectively, of the wavefront incident on the sensor, $w(x,y)$

$$\alpha(x,y) = d\frac{\partial}{\partial x}w(x,y),$$
(2)

$$\beta(x,y) = d\frac{\partial}{\partial y}w(x,y),$$
(3)

where d is a constant that depends on the type of sensor. Neglecting magnification for simplicity for a Shack–Hartmann sensor,[21] we have $d = f$, the focal length of the lenslets. For a Talbot sensor,[17] $d = z_T = 2P^2/\lambda$, the Talbot distance for the sensor, where λ is the wavelength of the light (assuming

the detector is at the first Talbot plane, otherwise incorporate an appropriate integer for the plane used). Figure 4 shows a simulated example depicting the result of a Talbot sensor with a sinusoidal grating pattern, as well as a Shack–Hartmann sensor with an equal pitch. The primary difference is the scalars $C_{n,m}$ for the harmonics.

The two-dimensional Fourier transform of the detected image is

$$\tilde{s}(k_x, k_y) = \iint \exp\{-i2\pi(k_x x + k_y y)\} s(x,y) \mathrm{d}x \mathrm{d}y.$$
(4)

The frequency content of the image is depicted in Fig. 5. The important information can be retained with just the first-linear harmonic terms where $(n,m) = (1,0)$ and $(0,1)$ denoted by the boxes in Fig. 5. For a plane wave, the peaks of the first terms will be located at $(0, \pm k_0)$ and $(\pm k_0, 0)$, where $k_0 = (1/P)$. Assuming the information is sufficiently concentrated near these points, we can extract the subimages depicted in Fig. 6, where the origins correspond to the $(0, k_0)$ and $(k_0, 0)$ points, respectively, from the original spatial frequency image. By taking the inverse Fourier transform of these subimages, we get complex images modulated by the phase gradients

$$s_\alpha(x,y) = C_{1,0} \exp\left[-\frac{2\pi i}{P}\alpha(x,y)\right],$$

$$s_\beta(x,y) = C_{0,1} \exp\left[-\frac{2\pi i}{P}\beta(x,y)\right].$$
(5)

2.1 Low-Order Aberrations

In the absence of high-order aberrations, we can immediately calculate the defocus and astigmatism based on the peak locations, via a relatively simple and well-known Fourier transform technique. A purely quadratic surface can be written using Zernike basis functions as
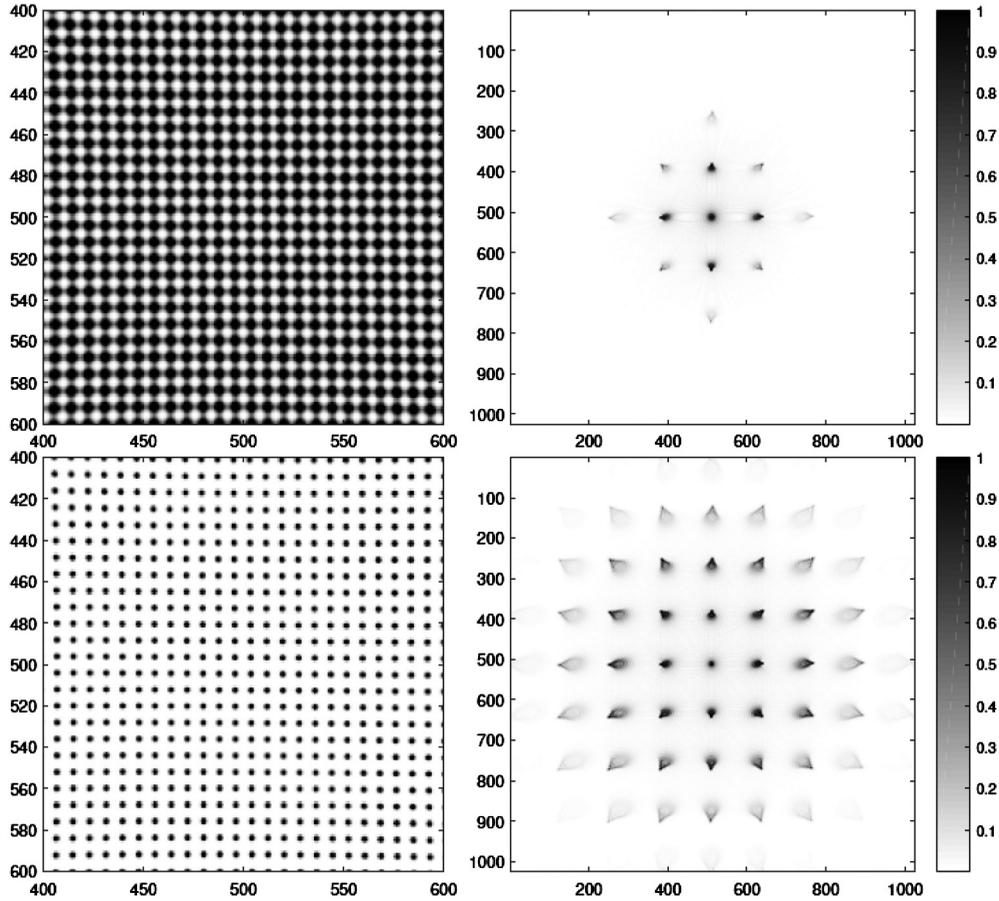
Fig. 4 Simulated grating (top row) and Shack–Hartmann (bottom row) result; detector image (left) and spatial-frequency image (right). Detector images are 1024×1024 pixels covering 10 mm, with a pitch of 42 μm and a magnification of 2.0; here images are zoomed to central 200×200 pixel region for visibility of pattern.
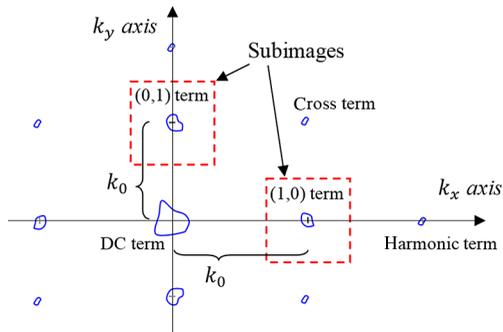
Fig. 5 Example of $\tilde{s}(k_x, k_y)$, two-dimensional spatial Fourier transform of sensor image, depicting locations of subimage selection about harmonic terms of interest.

$$w(x,y) = c_3 2\sqrt{6}xy + c_4\sqrt{3}(2x^2 + 2y^2 - 1) + c_5\sqrt{6}(x^2 - y^2).$$
(6)

The derivatives of this wavefront form linear functions, yielding

$$\alpha(x,y) = d(c_3 2\sqrt{6}y + c_4 4\sqrt{3}x + c_5 2\sqrt{6}x),$$
$$\beta(x,y) = d(c_3 2\sqrt{6}x + c_4 4\sqrt{3}y - c_5 2\sqrt{6}y),$$
(7)

and so the Fourier transforms of Eq. (5) have the form of delta functions (presuming infinite continuous functions for clarity)

$$\tilde{s}_\alpha(k_x, k_y) = \delta[2\pi(x - X_\alpha), 2\pi(y - Y_\alpha)],$$
$$\tilde{s}_\beta(k_x, k_y) = \delta[2\pi(x - X_\beta), 2\pi(y - Y_\beta)],$$
(8)

with peak locations as labeled in Fig. 6. Solving for the Zernike coefficients based on the peak locations gives the following estimate:

$$c3 = -\frac{P}{4\sqrt{6}d}(X_\beta + Y_\alpha),$$

$$c4 = -\frac{P}{8\sqrt{3}d}(X_\alpha + Y_\beta),$$

$$c5 = -\frac{P}{4\sqrt{6}d}(X_\alpha - Y_\beta).$$
(9)

For some applications, such as system calibration or determining the amount of compensation to use, this simple estimate may suffice. In the presence of high-order aberrations, however, this approach will become increasingly inaccurate.

2.2 High-Order Aberrations

To address high-order aberrations, we will take advantage of the fact that ophthalmic wavefronts are dominated by a small
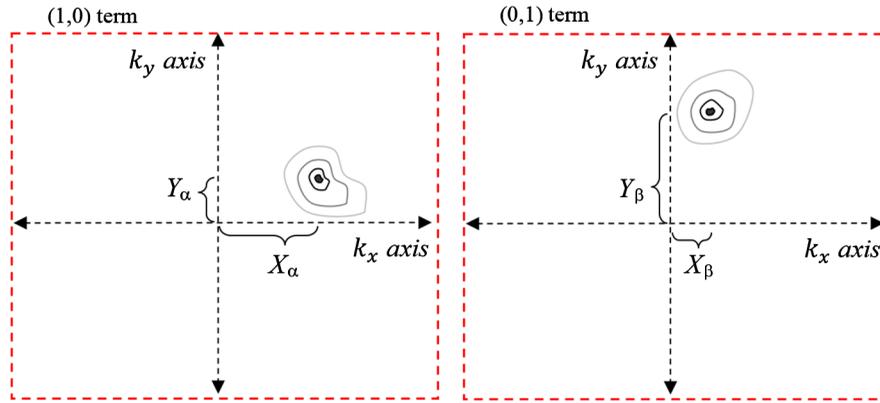
Fig. 6 Subimages s_α and s_β extracted from Fig. 5, depicting peak locations resulting from aberrations.

number of Zernike terms, while retaining the robustness and simplicity of Fourier-transform-based methods such as in Sec. 2.1. We address high-order terms by numerically computing shifted products of the demodulated images and their conjugates, and thereby convert higher-order terms to lower orders by a process similar to differentiation. This will allow us to subsequently utilize a low-order estimation method like that of Sec. 2.1. While the approach appears quite involved mathematically, the derivation only requires manipulations of small matrices, and the implementation consists of phase adjustments and fast Fourier transforms (FFTs) of the relatively small subimages around the peaks.

The first step is to process the subimages as follows, for both the s_α and s_β subimages,

$$
\begin{aligned}
s_\alpha^{(x)}(x,y) &= s_\alpha(x-\Delta,y)s_\alpha(x+\Delta,y)^*, \\
s_\alpha^{(xx)}(x,y) &= s_\alpha^{(x)}(x-\Delta,y)s_\alpha^{(x)}s_\alpha(x+\Delta,y)^*, \\
s_\alpha^{(y)}(x,y) &= s_\alpha(x,y-\Delta)s_\alpha(x,y+\Delta)^*, \\
s_\alpha^{(yy)}(x,y) &= s_\alpha^{(y)}(x,y-\Delta)s_\alpha^{(y)}(x,y+\Delta)^*, \\
s_\alpha^{(xy)}(x,y) &= s_\alpha^{(y)}(x-\Delta,y)s_\alpha^{(y)}(x+\Delta,y)^*.
\end{aligned}
\tag{10}
$$

Recall that the demodulated subimages are complex, unlike the original detected image. We will label the phase of these subimages with α for those derived from the $(1,0)$ term and β for those derived from the $(0,1)$ terms. For example, the phase of $s_\alpha^{(x)}$ is $\alpha^{(x)}$, and the phase of $s_\beta^{(x)}$ is $\beta^{(x)}$. The parameter Δ is similar conceptually to the distance used in a finite-difference differentiation estimate (instead of first and second derivatives, we refer to them as first and second products). The precise choice is not critical, except that a large Δ discards more information (as depicted in Fig. 7), while a smaller Δ will result in a smaller frequency shift for our subsequent calculation, making the result more susceptible to noise and quantization errors. Section 3 gives a simulation demonstrating that the output is relatively insensitive to Δ except for extreme values.

The result of such combinations is most easily understood in terms of transformations of monomials. Recall that a vector $\mathbf{z}$ of Zernike coefficients describing all aberrations up to some order can be described equivalently by a vector $\mathbf{p}$ of coefficients for monomial terms.[32] The vector $\mathbf{p}$ is related to $\mathbf{z}$ by a simple linear transformation (i.e., $\mathbf{z} = \mathbf{M}_{m2z}\mathbf{p}$). A fourth-order wavefront can be described as the monomial expansion
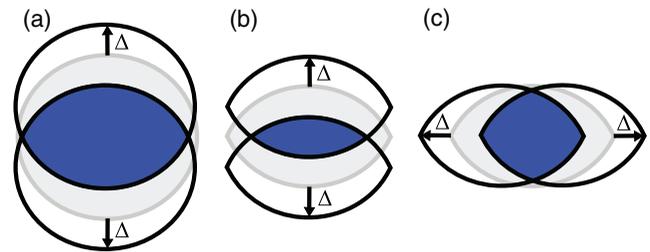
Fig. 7 An excessively large Δ parameter reduces the amount of pupil information utilized (a) by 2Δ for the first product $s^{(y)}$ [and similar for $s^{(x)}$], (b) by 4Δ for the second products $s^{(yy)}$ [and similar for $s^{(xx)}$], and (c) by 2Δ in both directions for the remaining second product $s^{(xy)}$. So, a smaller Δ is desirable to retain the most pupil area.

$$
w(x,y) = \sum_{i=0}^{4}\sum_{j=0}^{4-i} p_{i,j}x^i y^j,
\tag{11}
$$

for which the coefficients may be easily visualized in the matrix form (using base-zero indexing)

$$
\mathbf{P} = \begin{pmatrix}
p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} & p_{0,4} \\
p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & 0 \\
p_{2,0} & p_{2,1} & p_{2,2} & 0 & 0 \\
p_{3,0} & p_{3,1} & 0 & 0 & 0 \\
p_{4,0} & 0 & 0 & 0 & 0
\end{pmatrix},
\tag{12}
$$

with $\mathbf{p} = \mathrm{vec}(\mathbf{P})$ using the vectorization operator, an operator that reforms matrix elements into a vector (purely for convenience of notation). In matrix form, separable linear operations may be performed using a matrix $\mathbf{M}$, such as $\mathbf{M}^T\mathbf{P}$ to transform the x-coordinate and $\mathbf{PM}$ to transform the y-coordinate. See Ref. 32 for a more thorough introduction.

The coefficients of the gradient wavefronts $\alpha(x,y)$ and $\beta(x,y)$ can be computed using a differentiation matrix $\mathbf{M}_D$, to get

$$
\mathbf{P}_\alpha = \mathbf{M}_D^T\mathbf{P} = \begin{pmatrix}
p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & 0 \\
2p_{2,0} & 2p_{2,1} & 2p_{2,2} & 0 & 0 \\
3p_{3,0} & 3p_{3,1} & 0 & 0 & 0 \\
4p_{4,0} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix},
\tag{13}
$$

$$\mathbf{P}_\beta = \mathbf{PM}_D = \begin{pmatrix} p_{0,1} & 2p_{0,2} & 3p_{0,3} & 4p_{0,4} & 0 \\ p_{1,1} & 2p_{1,2} & 3p_{1,3} & 0 & 0 \\ p_{2,1} & 2p_{2,2} & 0 & 0 & 0 \\ p_{3,1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (14)$$

In this section, we will assume the device constants $P = d = 1$ to simplify the notation; the scale factor P/d may be applied when done, the same as it appears in Sec. 2.1 for low-order coefficient estimates. Note that the locations (row and column, counting from zero) in the matrix $\mathbf{P}_\alpha$ and $\mathbf{P}_\beta$ give the power of the new term via an expansion similar to Eq. (11) for $\alpha(x, y)$ and $\beta(x, y)$, respectively. Hence we can easily see that the wavefronts are now third order, as would be expected from differentiation, since the highest row or column index (when counting from zero) for a nonzero element is three.

To determine the effect of the shifts by Δ, we use translation matrices of the form

$$\mathbf{M}_\Delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \Delta & 1 & 0 & 0 \\ \Delta^2 & 2\Delta & 1 & 0 \\ \Delta^3 & 3\Delta^2 & 3\Delta & 1 \end{pmatrix}. \quad (15)$$

So, e.g., to determine the monomial coefficients for $\alpha^{(x)}(x, y) = \alpha(x - \Delta, y) - \alpha(x + \Delta, y)$ given $\mathbf{P}_\alpha$ we would simply compute $\mathbf{M}_{-\Delta}^T \mathbf{P}_\alpha - \mathbf{M}_\Delta^T \mathbf{P}_\alpha$. The resulting coefficients are

$$\mathbf{P}_{\alpha^{(x)}} = \begin{pmatrix} 4\Delta p_{2,0} + 8\Delta^3 p_{4,0} & 4\Delta p_{2,1} & 4\Delta p_{2,2} & 0 \\ 12\Delta p_{3,0} & 12\Delta p_{3,1} & 0 & 0 \\ 24\Delta p_{4,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (16)$$

A similar computation for $\alpha^{(xx)}(x, y) = \alpha^{(x)}(x + \Delta, y) - \alpha^{(x)}(x - \Delta, y)$ gives

$$\mathbf{P}_{\alpha^{(xx)}} = \begin{pmatrix} 24\Delta^2 p_{3,0} & 24\Delta^2 p_{3,1} & 0 & 0 \\ 96\Delta^2 p_{4,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (17)$$

The wavefront described by these coefficients is linear. Writing out the polynomial gives

$$\alpha^{(xx)}(x, y) = \sum_{i=0}^{4} \sum_{j=0}^{4-i} [\mathbf{P}_{\alpha^{(xx)}}]_{i,j} x^i y^j$$
$$= 24\Delta^2 p_{3,0} + 96\Delta^2 p_{4,0} x + 24\Delta^2 p_{3,1} y. \quad (18)$$

Hence, the processed image will be of the form (where, again, we assume $P = d = 1$)

$$s_\alpha^{(xx)}(x, y) = C_{1,0} \exp[-2\pi i \alpha^{(xx)}(x, y)]$$
$$= C_{1,0} \exp[-2\pi i (96\Delta^2 p_{4,0} x + 24\Delta^2 p_{3,1} y)], \quad (19)$$

and its two-dimensional Fourier transform will have a peak at a location

$$(k_x, k_y) = (96\Delta^2 p_{4,0}, 24\Delta^2 p_{3,1}). \quad (20)$$

So by finding the location of this peak, we can estimate the $p_{4,0}$ and $p_{3,1}$ coefficients of $w(x, y)$. With similar logic applied to the other processed images, we are able to estimate all fourth-order terms, to get

$$p_{4,0} = \frac{1}{96\Delta^2} X[s_\alpha^{(xx)}],$$
$$p_{3,1} = \frac{1}{72\Delta^2} \{Y[s_\alpha^{(xx)}] + X[s_\alpha^{(xy)}] + X[s_\beta^{(xx)}]\},$$
$$p_{2,2} = \frac{1}{64\Delta^2} \{Y[s_\alpha^{(xy)}] + X[s_\alpha^{(yy)}] + Y[s_\beta^{(xx)}] + X[s_\beta^{(xy)}]\},$$
$$p_{1,3} = \frac{1}{72\Delta^2} \{Y[s_\alpha^{(yy)}] + Y[s_\beta^{(xy)}] + X[s_\beta^{(yy)}]\},$$
$$p_{0,4} = \frac{1}{96\Delta^2} Y[s_\beta^{(yy)}], \quad (21)$$

where the terms that appear in multiple images are averaged, and the coordinates for the peak of the Fourier transform of an image s are denoted as $[X(s), Y(s)]$.

The next step is to remove these components from the lower order images, which we achieve by applying a phase adjustment based on the quantity of the term that exists in each (see Appendix). For example, from Eq. (16), we see that there are quadratic components resulting from the fourth-order terms of w, namely, $24\Delta p_{4,0}$ in the (2,0) entry, $12\Delta p_{3,1}$ in the (1,1) entry, and $4\Delta p_{2,2}$ in the (0,2) entry. As a result we would apply

$$\delta_\alpha^{(x)}(x, y) = \exp[2\pi i \Delta (24 p_{4,0} x^2 + 12 p_{3,1} xy + 4 p_{2,2} y^2)] \quad (22)$$

to $s_\alpha^{(x)}(x, y)$. Similarly for the other three quadratic images, we would apply

$$\delta_\alpha^{(y)}(x, y) = \exp[2\pi i \Delta (6 p_{3,1} x^2 + 8 p_{2,2} xy + 6 p_{1,3} y^2)],$$
$$\delta_\beta^{(x)}(x, y) = \exp[2\pi i \Delta (6 p_{3,1} x^2 + 8 p_{2,2} xy + 6 p_{1,3} y^2)],$$
$$\delta_\beta^{(y)}(x, y) = \exp[2\pi i \Delta (4 p_{2,2} x^2 + 12 p_{1,3} xy + 24 p_{0,4} y^2)]. \quad (23)$$

Now we can calculate the cubic terms from $w(x, y)$ using the peaks of the Fourier transform of these corrected images, as

$$p_{3,0} = \frac{1}{12\Delta} X[\delta_\alpha^{(x)} s_\alpha^{(x)}],$$

$$p_{2,1} = \frac{1}{12\Delta} \{Y[\delta_\alpha^{(x)} s_\alpha^{(x)}] + X[\delta_\alpha^{(y)} s_\alpha^{(y)}] + X[\delta_\beta^{(x)} s_\beta^{(x)}]\},$$

$$p_{1,2} = \frac{1}{12\Delta} \{Y[\delta_\beta^{(x)} s_\beta^{(x)}] + X[\delta_\beta^{(y)} s_\beta^{(y)}] + Y[\delta_\alpha^{(y)} s_\alpha^{(y)}]\},$$

$$p_{0,3} = \frac{1}{12\Delta} Y[\delta_\beta^{(y)} s_\beta^{(y)}]. \tag{24}$$

Finally, we correct the third- and fourth-order terms in the original gradient subimages, to accurately compute the second-order aberrations, by forming

$$\delta_\alpha(x,y) = \exp[2\pi i\Delta(4p_{4,0}x^3 + 3p_{3,1}x^2y + 2p_{2,2}xy^2 + p_{1,3}y^3 + 3p_{3,0}x^2 + 2p_{2,1}xy + p_{1,2}y^2)],$$

$$\delta_\beta(x,y) = \exp[2\pi i\Delta(p_{3,1}x^3 + 2p_{2,2}x^2y + 3p_{1,3}xy^2 + 4p_{0,4}y^3 + p_{2,1}x^2 + 2p_{1,2}xy + 3p_{0,3}y^2)]. \tag{25}$$

With these, we can calculate

$$p_{2,0} = \frac{1}{2} X(\delta_\alpha s_\alpha),$$

$$p_{1,1} = \frac{1}{2} [Y(\delta_\alpha s_\alpha) + X(\delta_\beta s_\beta)],$$

$$p_{0,2} = \frac{1}{2} Y(\delta_\beta s_\beta). \tag{26}$$

Finally, we form the vector of monomial coefficients that can be transformed to Zernike coefficients for application

$$\mathbf{p} =$$
$$(p_{2,0}, p_{1,1}, p_{0,2}, p_{3,0}, p_{2,1}, p_{1,2}, p_{0,3}, p_{4,0}, p_{3,1}, p_{2,2}, p_{1,3}, p_{0,4})^T. \tag{27}$$

To summarize, we have the following algorithm:

1. Demodulate detected image, producing complex subimages: s_α and s_β.

2. Compute first product images: $s_\alpha^{(x)}, s_\alpha^{(y)}, s_\beta^{(x)}, s_\beta^{(y)}$.

3. From first product images, compute second product images: $s_\alpha^{(xx)}, s_\alpha^{(xy)}, s_\alpha^{(yy)}, s_\beta^{(xx)}, s_\beta^{(xy)}, s_\beta^{(yy)}$.

4. From FFT peaks of second product images, compute fourth-order coefficients: $p_{4,0}, p_{3,1}, p_{2,2}, p_{1,3}, p_{0,4}$.

5. Using fourth-order coefficients, generate phase corrections: $\delta_\alpha^{(x)}, \delta_\alpha^{(y)}, \delta_\beta^{(x)}, \delta_\beta^{(y)}$.

6. From FFT peaks of first product images with phase corrections applied, compute third-order coefficients: $p_{3,0}, p_{2,1}, p_{1,2}, p_{0,3}$.

7. Using third-order coefficients, generate phase corrections: $\delta_\alpha, \delta_\beta$.

8. From FFT peaks of original subimages with phase corrections applied, compute second-order coefficients: $p_{2,0}, p_{1,1}, p_{0,2}$.

9. Transform monomial coefficients to Zernike coefficients: $\mathbf{z} = \mathbf{M}_{m2z}\mathbf{p}$.
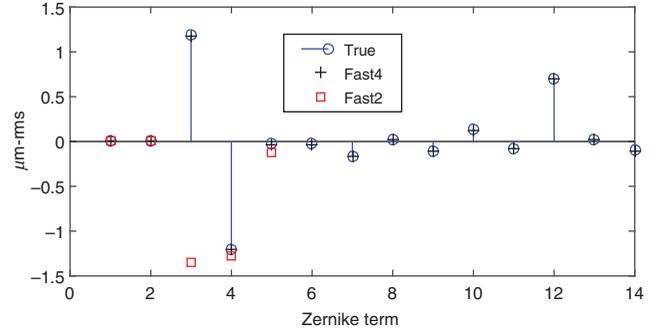

Fig. 8 True Zernike coefficients used in simulation, and result of algorithm to estimate fourth-order Zernikes (Fast4), as well as a version that stops after estimating second-order coefficients (Fast2). Net rms error (the root of the sum of the errors for each coefficient squared) is 0.043 μm-rms for Fast4, and 2.65 μm-rms for Fast2.
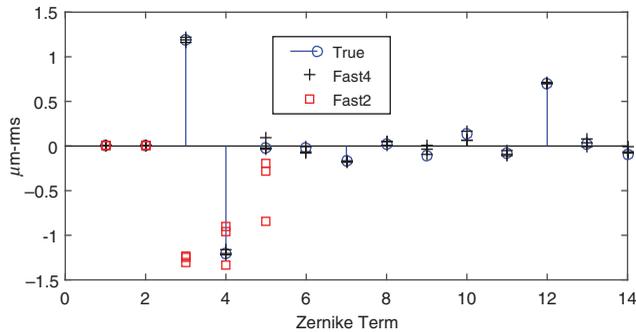
3 Results

Next, the method was tested with a variety of simulations of a Talbot sensor. The simulations used a pupil size of 5 mm with a detected image size of 10 mm for 1024×1024 pixels. The grating pitch was 42 μm, the wavelength was 850 nm, and the magnification was $m_s = 2.0$ (this simply required that a scaling be applied to the final terms based on their order). Figure 8 shows the estimates versus true coefficients for the example from Fig. 4. Here, the result of the simple second-order estimation of Eq. (9) is also demonstrated, which does not use the subsequent high-order estimation and their correction. Figure 9 shows the Fourier transforms of the subimages calculated in the method, for one of the two linear harmonic terms used [specifically for just the $(n, m) = (1,0)$ term]. First note that the peak in (a), the original subimage, exhibits a large amount of high order, as do the subimages in (b) and (c), computed from a single-shifted product. However, the second-shifted product
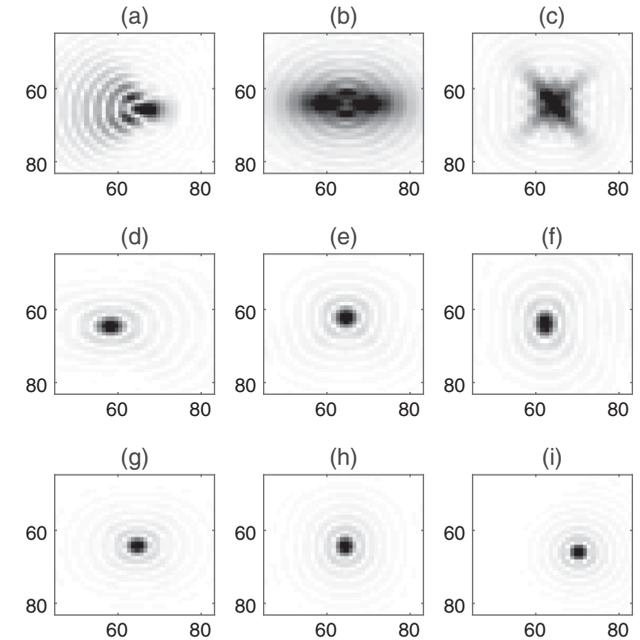

Fig. 9 Example of Fourier transforms of subimages used in processing, showing aberrations of peak: (a) $\tilde{s}_\alpha$, (b) $s_\alpha^{(x)}$, (c) $s_\alpha^{(y)}$, (d) $s_\alpha^{(xx)}$, (e) $s_\alpha^{(xy)}$, (f) $s_\alpha^{(yy)}$, (g) $\delta_\alpha^{(x)} s_\alpha^{(x)}$, (h) $\delta_\alpha^{(y)} s_\alpha^{(x)}$, and (i) $\delta_\alpha s_\alpha$.
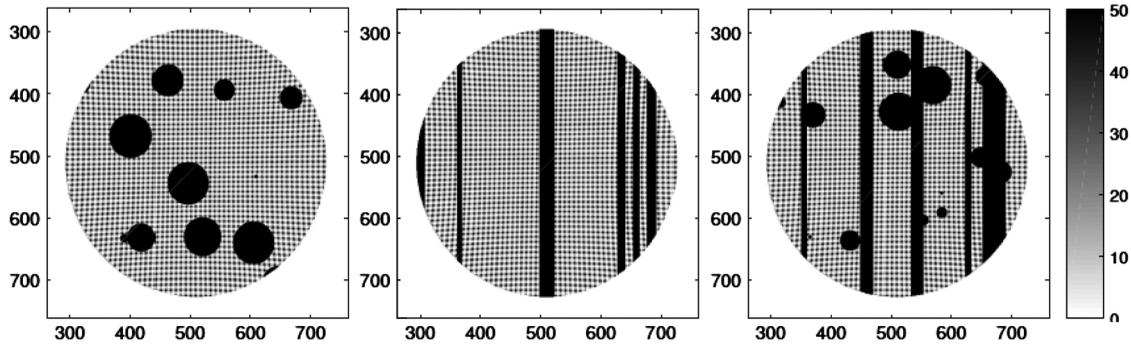
Fig. 10 Image artifact examples, with random spots or lines to demonstrate algorithm robustness.

Fig. 11 Resulting estimates for the three artifact examples. Net rms errors (the root of the sum of the errors for each coefficient squared) for Fast4 are 0.056, 0.18, and 0.19 μm-rms. As compared to an error of 0.043 μm-rms for the example with no artifacts. The net errors for Fast2 are 2.58, 2.56, and 2.74 μm-rms for the three cases, comparable to the 2.65 μm-rms error in the artifact-free case.
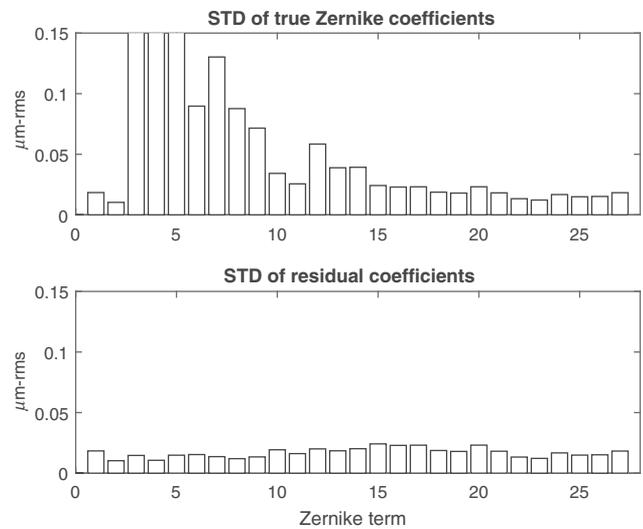
Fig. 12 Standard deviations of coefficients for each Zernike term in simulated data (top) and in residual error after estimate (bottom). Standard deviations of true Zernike coefficients 3, 4, and 5, are 0.27, 1.40, and 0.4274, respectively.

images in (d), (e), and (f) result in Airy disks, as there is no remaining high order at this point. Similarly, the correction of the previous subimages, depicted in (g), (h), and (i), result in Airy disks in the Fourier transforms, as the high order has been accurately estimated and removed. Figures 10 and 11 show results for the same case with severe artifacts intentionally added to the image; again, the high-order coefficients are accurately estimated.

The previous examples simulate a purely fourth-order wavefront, meaning no fifth- or higher-order aberrations are present. This presumption is the basis for the simplifications we were able to make. To test the effect of realistic levels of higher-order aberrations, a large set of images were simulated based on a secondary analysis of real sixth-order Zernike coefficients for measurements from 1500 subjects collected at multiple clinics in North America and Asia utilizing multiple different aberrometers. Figure 12 shows the standard deviations for each Zernike coefficient at 5 mm diameter versus the residual (true minus estimate), which generally agrees with the statistics found in other studies.[11] We see that the magnitude of the terms drops off quickly after fourth order (corresponding to coefficients above 14). Typically, values in this region are <0.02 μm-rms, which is comparable to the repeatability of the measurement due to biological variation. In Fig. 12, we see that the error after correction is reduced to this range as well.

Finally, this large simulation was used to test the parameter choice Δ. This parameter directly results in the shift of the peaks within the processed subimages, hence larger Δ makes this

estimate more accurate. Though note also that interpolation of the Fourier transform, peak locations will generally be needed to accurately measure small coefficients. The choice of larger Δ also comes at a price, however, of utilizing less data from the subimages as described in Fig. 7. In our initial testing, a moderate choice of seven pixels was sufficient, and the result was not very sensitive to this choice until extremes were reached. This agrees with the simulation results shown in Fig. 13, where error was found to be small and relatively constant between the extremes of small shifts and a maximum of about 13.

3.1 Discussion

This paper presented an approach to estimate wavefront aberrations that takes advantage of the limited order of aberrations seen in ophthalmic applications, in order to provide a measurement that is simple and robust, advantages that are critical for low-cost applications. The drawbacks of more sophisticated methods include the need for an expert operator to achieve the best measurement possible and visually reject images containing problematic artifacts to avoid capturing images that result in catastrophic phase-unwrapping failures. Consider the risks of using phase unwrapping to estimate the frequency of
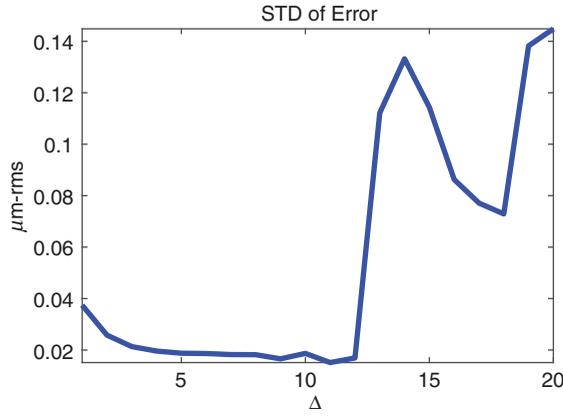
Fig. 13 Average wavefront error versus shift parameter Δ for large population of simulated wavefronts. For this pupil size and sample spacing, the result is relatively insensitive to choices of Δ in the range from three to twelve.

a noisy signal versus simply using a Fourier transform and finding the peak. Sophisticated methods also require additional expert algorithm development to further detect problematic artifacts that cannot be avoided by the operator, such as scars and reflections. These expert algorithms generally require careful tuning as with any detection and classification algorithm, and this tends to limit the range of applicable subjects, including certain population groups. The key to the approach presented here is the ability of transform methods to exploit the redundancy from many grating periods, which also complements the choice of accurate and inexpensive grating-based sensor elements such as the Zernike-based sensor used in the examples. It should also be noted that the same approach works for a Shack–Hartmann sensor, as the subimages will be mathematically equivalent, as depicted in Fig. 5.

A fourth-order representation was used because this captures the majority of aberrations in most eyes, as we demonstrated with the population data analysis, but it is straightforward to extend the method to higher orders. The most obvious way would be to simply continue the process of computing additional shifted products. Another tactic might be to use a hybrid technique that takes advantage of the reductions in high order in the shifted-product subimages to subsequently utilize a simpler phase estimate, perhaps which no longer needs to unwrap the phase, such as fast Fourier demodulation.[31] At the very least, it would be simple to test the quality of the Fourier transform peak to determine when the method has succeeded.

Appendix: Coefficient Matrices

We list here all the coefficient matrices resulting from the shifted-product operations used

$$\mathbf{P}_{\alpha^{(x)}} = \begin{pmatrix} 4\Delta p_{2,0} + 8\Delta^3 p_{4,0} & 4\Delta p_{2,1} & 4\Delta p_{2,2} & 0 \\ 12\Delta p_{3,0} & 12\Delta p_{3,1} & 0 & 0 \\ 24\Delta p_{4,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(28)

$$\mathbf{P}_{\alpha^{(y)}} = \begin{pmatrix} 2\Delta p_{1,1} + 2\Delta^3 p_{1,3} & 4\Delta p_{1,2} & 6\Delta p_{1,3} & 0 \\ 4\Delta p_{2,1} & 8\Delta p_{2,2} & 0 & 0 \\ 6\Delta p_{3,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(29)

$$\mathbf{P}_{\alpha^{(xx)}} = \begin{pmatrix} 24\Delta^2 p_{3,0} & 24\Delta^2 p_{3,1} & 0 & 0 \\ 96\Delta^2 p_{4,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(30)

$$\mathbf{P}_{\alpha^{(xy)}} = \begin{pmatrix} 8\Delta^2 p_{2,1} & 16\Delta^2 p_{2,2} & 0 & 0 \\ 24\Delta^2 p_{3,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(31)

$$\mathbf{P}_{\alpha^{(yy)}} = \begin{pmatrix} 8\Delta^2 p_{1,2} & 24\Delta^2 p_{1,3} & 0 & 0 \\ 16\Delta^2 p_{2,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(32)

$$\mathbf{P}_{\beta^{(x)}} = \begin{pmatrix} 2\Delta p_{1,1} + 2\Delta^3 p_{3,1} & 4\Delta p_{1,2} & 6\Delta p_{1,3} & 0 \\ 4\Delta p_{2,1} & 8\Delta p_{2,2} & 0 & 0 \\ 6\Delta p_{3,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(33)

$$\mathbf{P}_{\beta^{(y)}} = \begin{pmatrix} 4\Delta p_{0,2} + 8\Delta^3 p_{0,4} & 12\Delta p_{0,3} & 24\Delta p_{0,4} & 0 \\ 4\Delta p_{1,2} & 12\Delta p_{1,3} & 0 & 0 \\ 4\Delta p_{2,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(34)

$$\mathbf{P}_{\beta^{(xx)}} = \begin{pmatrix} 8\Delta^2 p_{2,1} & 16\Delta^2 p_{2,2} & 0 & 0 \\ 24\Delta^2 p_{3,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(35)

$$\mathbf{P}_{\beta^{(xy)}} = \begin{pmatrix} 8\Delta^2 p_{1,2} & 24\Delta^2 p_{1,3} & 0 & 0 \\ 16\Delta^2 p_{2,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

(36)

$$\mathbf{P}_{\beta^{(yy)}} = \begin{pmatrix} 24\Delta^2 p_{0,3} & 96\Delta^2 p_{0,4} & 0 & 0 \\ 24\Delta^2 p_{1,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

(37)

References

1. N. G. Congdon, D. S. Friedman, and T. Lietman, "Important causes of visual impairment in the world today," *J. Am. Med. Assoc.* **290**, 2057–2060 (2003).
2. C. A. McCarty, "Uncorrected refractive error," *Br. J. Ophthalmol.* **90**, 521–522 (2006).
3. L. Pizzarello et al., "Vision 2020: the right to sight: a global initiative to eliminate avoidable blindness," *Arch. Ophthalmol.* **122**, 615–620 (2004).
4. I. G. Morgan, K. Ohno-Matsui, and S.-M. Saw, "Myopia," *Lancet* **379**, 1739–1748 (2012).
5. D. B. Elliott, *Clinical Procedures in Primary Eye Care*, Elsevier Health Sciences, Philadelphia, Pennsylvania (2013).
6. I. C. J. Wood, "A review of autorefractors," *Eye* **1**, 529–535 (1987).
7. L. N. Thibos et al., "Accuracy and precision of objective refraction from wavefront aberrations," *J. Vision* **4**(4), 9–9 (2004).
8. J. D. Marsack, L. N. Thibos, and R. A. Applegate, "Metrics of optical quality derived from wave aberrations predict visual performance," *J. Vision* **4**(4), 8–328 (2004).
9. D. R. Williams, "Sharpness metric for vision quality," U.S. Patent No. 7077522 B2 (2006).
10. N. Maeda, "Clinical applications of wavefront aberrometry—a review," *Clin. Exp. Ophthalmol.* **37**, 118–129 (2009).
11. J. Porter et al., "Monochromatic aberrations of the human eye in a large population," *J. Opt. Soc. Am. A* **18**, 1793 (2001).
12. L. A. Carvalho, "A simple and effective algorithm for detection of arbitrary Hartmann–Shack patterns," *J. Biomed. Inf.* **37**, 1–9 (2004).
13. M. Lombardo and G. Lombardo, "New methods and techniques for sensing the wave aberrations of human eyes," *Clin. Exp. Optom.* **92**, 176–186 (2009).
14. R. Voelkel, "Wafer-scale micro-optics fabrication," *Adv. Opt. Technol.* **1**(3), 135–150 (2012).
15. M. Lombardo et al., "Adaptive optics technology for high-resolution retinal imaging," *Sensors* **13**, 334–366 (2012).
16. N. H. Salama et al., "Wavefront sensor using the Talbot effect," *Opt. Laser Technol.* **31**, 269–272 (1999).
17. Y. Liu et al., "Z-View diffractive wavefront sensor: principle and applications," *Proc. SPIE* **6018**, 601809 (2005).
18. C. Canovas and E. N. Ribak, "Comparison of Hartmann analysis methods," *Appl. Opt.* **46**, 1830–1835 (2007).
19. Y. Carmon and E. N. Ribak, "Phase retrieval by demodulation of a Hartmann–Shack sensor," *Opt. Commun.* **215**, 285–288 (2003).
20. D. Podanchuk et al., "Bottlenecks of the wavefront sensor based on the Talbot effect," *Appl. Opt.* **53**, B223 (2014).
21. E. J. Sarver, J. Schwiegerling, and R. A. Applegate, "Extracting wavefront error from Shack–Hartmann images using spatial demodulation," *J. Refract. Surg.* **22**, 949–953 (2006).
22. J. Ma, J. B. Son, and J. D. Hazle, "An improved region growing algorithm for phase correction in MRI," *Magn. Reson. Med.* **76**, 519–529 (2016).
23. C. W. Chen and H. A. Zebker, "Network approaches to two-dimensional phase unwrapping: intractability and two new algorithms," *J. Opt. Soc. Am. A* **17**, 401 (2000).
24. V. Akondi et al., "Phase unwrapping with a virtual Hartmann–Shack wavefront sensor," *Opt. Express* **23**, 25425 (2015).
25. H. Y. H. Huang et al., "Path-independent phase unwrapping using phase gradient and total-variation (TV) denoising," *Opt. Express* **20**, 14075 (2012).
26. Z. Cheng et al., "Practical phase unwrapping of interferometric fringes based on unscented Kalman filter technique," *Opt. Express* **23**, 32337 (2015).
27. V. Katkovnik and J. Bioucas-Dias, "Wavefront reconstruction in phase-shifting interferometry via sparse coding of amplitude and absolute phase," *J. Opt. Soc. Am. A* **31**, 1801 (2014).
28. A. Talmi and E. N. Ribak, "Direct demodulation of Hartmann–Shack patterns," *J. Opt. Soc. Am. A* **21**, 632 (2004).
29. L. Huang et al., "Comparison of two-dimensional integration methods for shape reconstruction from gradient data," *Opt. Lasers Eng.* **64**, 1–11 (2015).
30. A. Talmi and E. N. Ribak, "Wavefront reconstruction from its gradients," *J. Opt. Soc. Am. A* **23**, 288 (2006).
31. Y. Carmon and E. N. Ribak, "Fast Fourier demodulation," *Appl. Phys. Lett.* **84**, 4656–4657 (2004).
32. K. Dillon, "Bilinear wavefront transformation," *J. Opt. Soc. Am. A* **26**(8), 1839–1846 (2009).

Keith Dillon is cofounder of Formulens, LLC, which applies advanced numerical methods to optical system optimization. He received his PhD degree from the University of California, San Diego, his MS degree from Rice University, and his BS from Tulane University. His research interests include computational imaging and vision, particularly neuroimaging, and entrepreneurship. He is also currently a postdoctoral fellow at Tulane University.