Phase-error correction in digital holography

Samuel T. Thurman and James R. Fienup

The Institute of Optics, University of Rochester, Rochester, New York 14627, USA

Received October 29, 2007; accepted January 31, 2008;
posted February 20, 2008 (Doc. ID 89108); published March 27, 2008

The quality of images computed from digital holograms or heterodyne array imaging is degraded by phase errors in the object and/or reference beams at the time of measurement. This paper describes computer simulations used to compare the performance of digital shearing laser interferometry and various sharpness metrics for the correction of such phase errors when imaging a diffuse object. These algorithms are intended for scenarios in which multiple holograms can be recorded with independent object speckle realizations and a static phase error. Algorithm performance is explored as a function of the number of available speckle realizations and signal-to-noise ratio (SNR). The performance of various sharpness metrics is examined in detail and is shown to vary widely. Under ideal conditions with >15 speckle realizations and high SNR, phase corrections better than λ/50 root-mean-square (RMS) were obtained. Corrections better than λ/10 RMS were obtained in the high SNR regime with as few as two speckle realizations and at object beam signal levels as low as 2.5 photons/speckle with six speckle realizations. © 2008 Optical Society of America

OCIS codes: 090.1000, 090.1760, 100.3010, 100.3020, 100.3190, 040.2840.

1. INTRODUCTION

A digital hologram of an object can be recorded via the configuration shown in Fig. 1. The first beam splitter in the figure divides the collimated laser beam into a reference beam and an illumination beam, while the second beam splitter reflects the reference beam and transmits the object beam (light scattered from the object). The detector array records a digital hologram, i.e., the interference pattern between the reference and object beams. The intensity interference pattern $H(x,y)$ at the detector plane can be written as

$$H(x,y) = |R(x,y) + F(x,y)|^2, \tag{1}$$

where $R(x,y)$ and $F(x,y)$ are the optical fields associated with the reference and object beams at the detector plane, respectively. Ideally, the reference beam would be a perfect plane wave and the object beam would contain no phase errors. In such a case, the reference and object beams can be written as

$$R(x,y) = R_0 \exp[i(k_x x + k_y y)], \tag{2}$$

$$F(x,y) = \frac{1}{i\lambda z}\exp(ikz)\exp\left[i\frac{\pi}{\lambda z}(x^2+y^2)\right] \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi,\eta)\exp\left[i\frac{\pi}{\lambda z}(\xi^2+\eta^2)\right] \times \exp\left[-i\frac{2\pi}{\lambda z}(x\xi + y\eta)\right]\mathrm{d}\xi\mathrm{d}\eta, \tag{3}$$

where R_0 is the amplitude of the reference wave, k_x and k_y are the transverse components of the reference wave vector, λ is the wavelength of light, z is the longitudinal distance between the detector array and the nominal object plane, $f(\xi,\eta)$ is the optical field scattered by the object in the nominal object plane, and the Fresnel approximation [1] has been used.

Using an off-axis reference beam [2], i.e., $k_x \neq 0$ and/or $k_y \neq 0$, $F(x,y)$ can be reconstructed from $H(x,y)$. Then a coherent image of the object (with spatial resolution determined by the diameter of the detector array) can be digitally computed by inverting Eq. (3). Taking the 2D spatial Fourier transform of Eq. (1) and using the ideal form of $R(x,y)$ given in Eq. (2) yields

$$\tilde{H}(f_x,f_y) = |R_0|^2\delta(f_x,f_y) + \tilde{F}(f_x,f_y)\star\tilde{F}(f_x,f_y) + R_0^*\tilde{F}\left(\frac{k_x}{2\pi}+f_x,\frac{k_y}{2\pi}+f_y\right) + R_0\tilde{F}^*\left(\frac{k_x}{2\pi}-f_x,\frac{k_y}{2\pi}-f_y\right), \tag{4}$$

where $\tilde{H}(f_x,f_y)$ and $\tilde{F}(f_x,f_y)$ are the 2D spatial Fourier transforms of $H(x,y)$ and $F(x,y)$, respectively, (f_x,f_y) are spatial-frequency coordinates, $\delta(f_x,f_y)$ is a 2D Dirac delta function, and $\star$ denotes a cross correlation.

The first two terms on the right-hand side of Eq. (4) represent the autocorrelations of the Fourier transforms of the reference and object beams, respectively, and are centered about the dc spatial frequency. The third and fourth terms represent the Fourier transforms of the object beam and its holographic twin, respectively, which are offset from the dc spatial frequency by a distance proportional to the magnitude of the transverse reference wave vector $\sqrt{k_x^2+k_y^2}$. Provided $\sqrt{k_x^2+k_y^2}$ is sufficiently large, satisfying the holographic condition such that the third and fourth terms of Eq. (4) do not overlap the second term, the Fourier transform of the object beam can be isolated with a window function and inverse transformed to yield $F(x,y)$ to within an arbitrary multiplicative constant

1084-7529/08/040983-12/$15.00

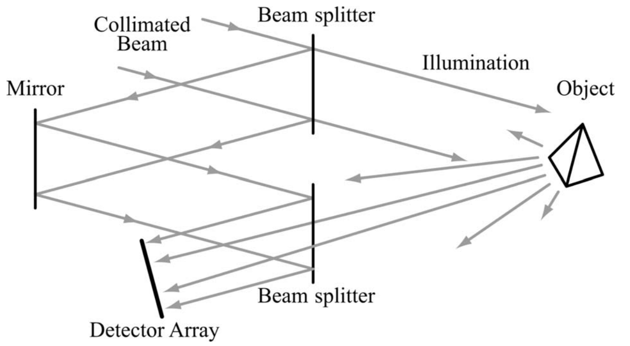

Fig. 1. Layout for recording digital holograms.

and an arbitrary piston phase (ignoring detector nonlinearities, measurement noise, and edge/window effects).

In any real system, the computed image can be degraded if the reference field is not a perfect plane wave and/or the object beam includes errors from the second beam splitter or atmospheric turbulence. In general the reference and object fields can differ from their ideal forms in both amplitude and phase; the quality of the computed image is, however, much more sensitive to phase errors than it is to amplitude errors. Thus, we will consider the case of phase errors only, such that the aberrated reference and object fields $R_\varepsilon(x,y)$ and $F_\varepsilon(x,y)$ can be written as

$$R_\varepsilon(x,y)=R(x,y)\exp[i\phi_R(x,y)], \tag{5}$$

$$F_\varepsilon(x,y)=F(x,y)\exp[i\phi_F(x,y)], \tag{6}$$

where $\phi_R(x,y)$ and $\phi_F(x,y)$ are the phase errors of the object and reference beams, respectively. Equation (6) is limited to the isoplanatic case, for which the object-beam phase errors are in a volume fairly close to the detector. Using these expressions, the hologram intensity pattern $H_\varepsilon(x,y)$ with phase errors can be written as

$$\begin{aligned} H_\varepsilon(x,y) &= |R_\varepsilon(x,y)+F_\varepsilon(x,y)|^2 \\ &= |R(x,y)+F(x,y)\exp[i\phi(x,y)]|^2, \end{aligned} \tag{7}$$

where

$$\phi(x,y)=\phi_F(x,y)-\phi_R(x,y). \tag{8}$$

When using the procedure described above for reconstructing the object field from a hologram, the result is an aberrated object field $G(x,y)$ of the form

$$G(x,y)=F(x,y)\exp[i\phi(x,y)]. \tag{9}$$

Alternatively, aberrated field measurements such as these can be obtained by heterodyne array measurements making use of a local oscillator [3].

In Section 2 below, the technical details of two approaches for correcting for unknown phase errors in $G(x,y)$ are reviewed: (i) digital shearing laser interferometry (DSLI) [4,5] and (ii) sharpness metric maximization [6–9]. Both approaches are intended for use with diffuse extended objects, but can also be used for objects having glints. DSLI requires a modest number of digital holograms with independent object speckle realizations to be recorded for a constant $\phi(x,y)$. Thus, in the case of dynamic phase errors, DSLI is limited to scenarios in which a modest number of digital holograms can be recorded with adequate signal-to-noise ratio (SNR) before $\phi(x,y)$ varies appreciably. Alternatively, the sharpness metric maximization approach can work in some cases with only a single object speckle realization, but it also benefits greatly from multiple speckle realizations. Furthermore, DSLI can be performed only when one has a noninterrupted array of measurement points, whereas the sharpness approaches can work for sparse-aperture and segmented-aperture systems. In general, the accuracy of each algorithm varies somewhat depending on the details of the object, the nature of the phase error, the number of available speckle realizations, and the SNR. Note that when maximizing sharpness for synthetic-aperture radar, as described in [7,9], one has a 1D phase error and uses multiple range lines having the same phase error to obtain the statistics needed to accurately estimate the phase error. In this paper, however, we have a 2D phase error and use multiple speckle realizations to obtain the statistics needed to accurately estimate the phase error.

Section 3 describes the details of computer simulation experiments designed to compare the relative performance of DSLI and a variety of sharpness metrics as a function of the number of available speckle realizations and the SNR for a given scene and atmospheric phase error. The residual root-mean-square (RMS) phase error (ignoring piston, tip, and tilt terms) after correction is used to quantify the performance of each algorithm. Section 4 presents results of the computer simulation experiments and Section 5 is a summary.

2. PHASE-ERROR-CORRECTION ALGORITHMS

This section describes both the DSLI and sharpness metric phase-error-correction approaches. For the development of both approaches, we assume that the object field in the detector plane has been reconstructed from digital holography or heterodyne array measurements for N independent object speckle realizations with a constant phase error $\phi(x,y)$, such that the given object field at the detector for the nth speckle realization can be written in a form analogous to Eq. (9) as

$$G_n(x,y)=F_n(x,y)\exp[i\phi(x,y)]. \tag{10}$$

A. Digital Shearing Laser Interferometry

Starting with aberrated object fields of the form given in Eq. (10), the first step in DSLI is to compute the digitally sheared quantities

$$S_x(x,y)=\frac{1}{N}\sum_{n=1}^{N}G_n(x,y)G_n^*(x-\Delta,y), \tag{11}$$

$$S_y(x,y)=\frac{1}{N}\sum_{n=1}^{N}G_n(x,y)G_n^*(x,y-\Delta), \tag{12}$$

where Δ is a shear distance, which for simplicity is shown as the same in both dimensions, but can differ in x and y. Substituting Eq. (10) into Eq. (11) yields

$$\begin{aligned} S_x(x,y) = {} & \exp\{i[\phi(x,y) - \phi(x-\Delta,y)]\} \\ & \times \frac{1}{N}\sum_{n=1}^{N} F_n(x,y)F_n^*(x-\Delta,y). \end{aligned} \tag{13}$$

Using Eq. (3), the summation in Eq. (13) can be expressed as

$$\begin{aligned} & \frac{1}{N}\sum_{n=1}^{N} F_n(x,y)F_n^*(x-\Delta,y) \\ & = \frac{1}{\lambda^2 z^2}\exp\left\{ i\frac{\pi}{\lambda z}[x^2 - (x-\Delta)^2] \right\} \\ & \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{N}\sum_{n=1}^{N} f_n(\xi,\eta) f_n^*(\xi',\eta') \\ & \times \exp\left[i\frac{\pi}{\lambda z}(\xi^2 - \xi'^2 + \eta^2 - \eta'^2) \right] \\ & \times \exp\left\{ -i\frac{2\pi}{\lambda z}[\xi'\Delta + (\xi-\xi')x + (\eta-\eta')y] \right\} \mathrm{d}\xi \mathrm{d}\eta \mathrm{d}\xi' \mathrm{d}\eta', \end{aligned} \tag{14}$$

where $f_n(\xi,\eta)$ represents the nth speckle realization of the object field in the nominal object plane. For diffuse objects, the object field is essentially delta-correlated in (ξ,η), such that the following approximation,

$$\frac{1}{N}\sum_{n=1}^{N} f_n(\xi,\eta) f_n^*(\xi',\eta') \approx \kappa I(\xi,\eta)\delta(\xi-\xi',\eta-\eta'), \tag{15}$$

becomes an equality as N approaches infinity, where κ is a constant and $I(\xi,\eta)$ is the intensity of the light reflected from the object that would be observed under spatially incoherent illumination. Inserting Eqs. (14) and (15) into Eq. (13) and integrating yields

$$\begin{aligned} S_x(x,y) = {} & \frac{\kappa}{\lambda^2 z^2}\exp\{i[\phi(x,y) - \phi(x-\Delta,y)]\} \\ & \times \exp\left[i\frac{\pi}{\lambda z}(2x\Delta - \Delta^2) \right] \\ & \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\xi,\eta)\exp\left(-i\frac{2\pi}{\lambda z}\xi\Delta \right)\mathrm{d}\xi \mathrm{d}\eta \\ = {} & \frac{\kappa}{\lambda^2 z^2}\exp\{i[\phi(x,y) - \phi(x-\Delta,y)]\} \\ & \times \exp\left[i\frac{\pi}{\lambda z}(2x\Delta - \Delta^2) \right]\tilde{I}\left(\frac{\Delta}{\lambda z}, 0 \right), \end{aligned} \tag{16}$$

where $\tilde{I}(f_x,f_y)$ is the Fourier transform of $I(\xi,\eta)$. This is analogous to the derivation of the Van Cittert–Zernike theorem [4,5,10], in which $\tilde{I}(f_x,f_y)$ represents the object mutual intensity. In a similar manner,

$$\begin{aligned} S_y(x,y) \approx {} & \frac{\kappa}{\lambda^2 z^2}\exp\{i[\phi(x,y) - \phi(x,y-\Delta)]\} \\ & \times \exp\left[i\frac{\pi}{\lambda z}(2y\Delta - \Delta^2) \right]\tilde{I}\left(0, \frac{\Delta}{\lambda z} \right). \end{aligned} \tag{17}$$

The sheared and averaged quantities $S_x(x,y)$ and $S_y(x,y)$ thus have a phase component that is a finite difference of the phase error, similar to what one gets, for example, in shearing interferometry. It is important to note that the mutual intensity contribution to the phase of $S_x(x,y)$ is limited to an arbitrary constant phase, i.e., the phase of $\tilde{I}[\Delta/(\lambda z),0]$. $S_x(x,y)$ also experiences a linear phase term $2\pi x\Delta/(\lambda z)$ associated with the quadratic phase before the integral in Eq. (3).

A number of different algorithms [11–14] exist for reconstructing wavefronts from sheared phase data. A wavefront reconstructed from the digital shear data in Eqs. (16) and (17) will ideally have a phase $\hat{\phi}(x,y)$ given by

$$\hat{\phi}(x,y) \approx \phi(x,y) + \frac{\pi}{\lambda z}(x^2+y^2) + ax + by + c, \tag{18}$$

where a and b are arbitrary tip and tilt terms associated with the piston phases of the mutual intensity terms $\tilde{I}[\Delta/(\lambda z),0]$ and $\tilde{I}[0,\Delta/(\lambda z)]$, and c is an arbitrary piston phase term. The piston, tip, and tilt phase terms do not affect image quality. An estimate of the unaberrated object field $\hat{F}_n(x,y)$ in the detector plane is then given by

$$\hat{F}_n(x,y) = G_n(x,y)\exp[-i\hat{\phi}(x,y)]. \tag{19}$$

The corresponding estimate for the field $\hat{f}_n(\xi,\eta)$ in the nominal object plane is

$$\begin{aligned} \hat{f}_n(\xi,\eta) = {} & \frac{1}{\lambda z}\exp\left[-i\frac{\pi}{\lambda z}(\xi^2+\eta^2) \right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{F}_n(x,y) \\ & \times \exp\left[i\frac{2\pi}{\lambda z}(x\xi + y\eta) \right]\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{20}$$

to within an arbitrary piston phase and (ξ,η) coordinate shift. Note that the quadratic term in Eq. (18) is already multiplying $\hat{F}_n(x,y)$, so it should not be included in this integral. Finally, a speckle-averaged intensity estimate $\hat{I}(\xi,\eta)$ of the object is computed as

$$\hat{I}(\xi,\eta) = \frac{1}{N}\sum_{n=1}^{N}|\hat{f}_n(\xi,\eta)|^2. \tag{21}$$

Since we are here concerned only with the image intensity, the quadratic phase term outside the integral in Eq. (20) can be ignored. Also, the quadratic phase term in (x^2+y^2) is canceled by the phase estimate. Thus, we can act as though both quadratic phase terms are zero and perform Fourier transforms rather than Fresnel transforms. That is, the object can be treated as though it is in the far field (Fraunhofer regime) even if it is in the near field (Fresnel regime).

B. Sharpness Metric Maximization

In the sharpness metric approach, a nonlinear optimization algorithm is used to find a phase error estimate $\hat{\phi}(x,y)$ that maximizes a quantitative measure of the sharpness of a speckle-averaged intensity estimate of the object. The functional dependence of $\hat{I}(\xi,\eta)$ on $\hat{\phi}(x,y)$ is given by Eqs. (19)–(21). We consider sharpness metrics of the following two forms

$$M=\sum_{(\xi,\eta)}\Gamma[\hat{I}(\xi,\eta)], \tag{22}$$

$$M=\sum_{(\xi,\eta)}\sum_{(\xi',\eta')\in D}\Gamma[\hat{I}(\xi,\eta)-\hat{I}(\xi-\xi',\eta-\eta')], \tag{23}$$

where the (ξ,η) summations are over a set of image samples, $\Gamma(I)$ is a nonlinear function, and D is a neighboring system of (ξ',η') shift coordinates.

Table 1 lists forms of $\Gamma(I)$ we considered for sharpness metrics for use in Eq. (22), which are called statistics-based sharpness metrics, because Eq. (22) is equivalent to an estimator for the statistical moment of $\Gamma(I)$ for the speckle-averaged image, if $\hat{I}(\xi,\eta)$ is viewed as a stochastic process. Note that metric M_1 is equivalent to the metrics S_1 and S_5 from [6] for $\alpha=2$ and $\alpha>2$, respectively, and metric M_3 yields a minimum entropy phase estimate and is equivalent to S_7 from [6].

Table 2 lists the forms of $\Gamma(\Delta I)$ we considered for use in Eq. (23), where $\Delta I=\hat{I}(\xi,\eta)-\hat{I}(\xi-\xi',\eta-\eta')$. These are called correlation metrics, since Eq. (24) is equivalent to an estimator for statistical moments between neighboring points in the speckle-averaged image. Note that M_4 with $\alpha=2$ is equivalent to a finite difference approximation of S_4 from [6]. We previously found M_5 to be useful for incoherent image restoration when there were missing areas within the spatial-frequency domain. Note that while [6] dealt with incoherent images, we are dealing here with coherent, speckled images (for $N=1$) and speckle-reduced images (for $N>1$).

It is important to understand the principle behind each metric, and how each metric differs. Note that the value of a statistics-based metric is independent of the spatial organization of an image, but is completely determined by the histogram of an image. However, the value of a correlation-based metric depends on the differences between neighboring spatial samples in an image. Due to energy conservation, the effect of maximizing a statistic-based metric is intimately related to $\Gamma''(I)$ [9], the second derivative of $\Gamma(I)$ with respect to I. As explained in [9], the third column of Table 1 indicates whether each metric concentrates more on making bright points brighter or dark points darker. Since the differences between neighboring image intensity samples are not conserved, the effect of maximizing a correlation-based metric is largely determined by $\Gamma'(\Delta I)$, the first derivative of $\Gamma(\Delta I)$ with respect to ΔI. The third column of Table 2 lists the general effect maximizing M_4 or M_5 has on the differences between neighboring image samples.

Table 1. Statistics-Based Sharpness Metrics Based on the Form of Eq. (23)

Metric	$\Gamma(I)$	Effect on Image Histogram
M_1	I^{α} for $\alpha>1$	Concentrates on making: bright points brighter for $\alpha>2$, dark points darker for $\alpha<2$
M_2	$-I^{\alpha}$ for $0<\alpha<1$	Concentrates on making dark points darker
M_3	$I\ln(I)$	Concentrates on making dark points darker

Table 2. Correlation-Based Sharpness Metrics Based on the Form of Eq. (24)

Metric	$\Gamma(\Delta I)$	Effect on Differences between Neighboring Image Points
M_4	$\lvert\Delta I\rvert^{\alpha}$	Increases differences
M_5	$\dfrac{1}{(\Delta I)^2+\alpha^2}$	Concentrates on making $\lvert\Delta I\rvert\approx\alpha$ differences smaller

3. COMPUTER SIMULATION EXPERIMENTS

This section describes the details of computer simulation experiments for exploring the relative performance of DSLI and the different sharpness metrics for phase-error correction.

The flowchart in Fig. 2 outlines the steps in simulating digital holography data. All of the simulations shown here start with the 256×256 incoherent intensity image for $I(\xi,\eta)$ shown in Fig. 3 [15]. Independent object speckle realizations were generated by

$$f_n(\xi,\eta)=\sqrt{I(\xi,\eta)}[N(0,0.5)+iN(0,0.5)] \tag{24}$$

where $N(\mu,\sigma^2)$ represents an independent random variable having a Gaussian distribution with a mean μ and variance σ^2, giving us $\langle|f_n(\xi,\eta)|^2\rangle=I(\xi,\eta)$, where the angle brackets denote the average over an ensemble of speckle

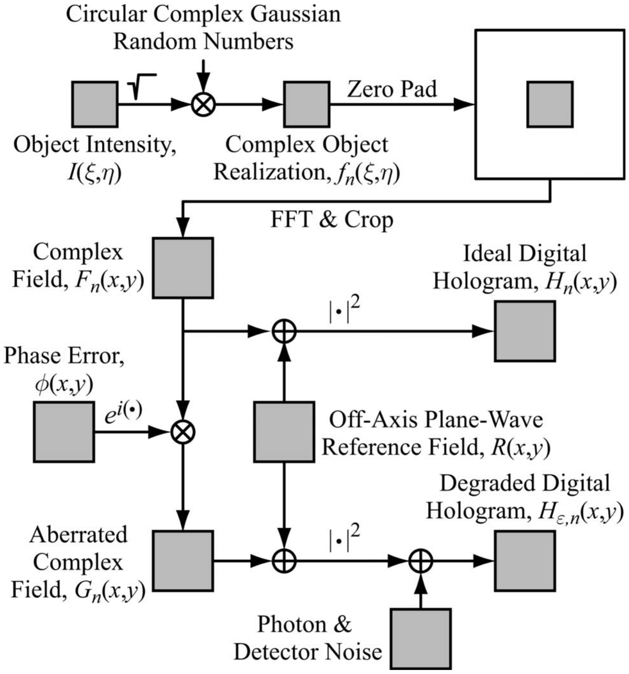

Fig. 2. Flowchart for simulating digital holography data.

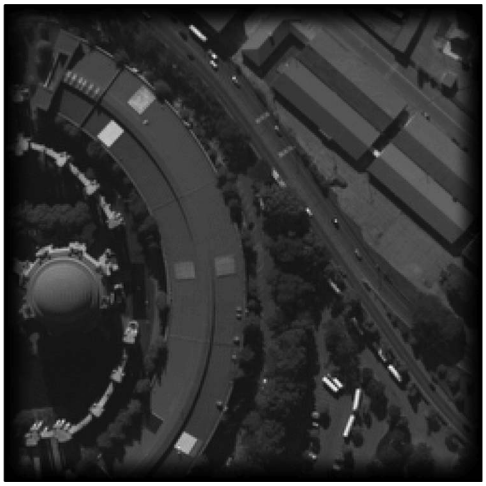

Fig. 3. Incoherent intensity image $I(\xi,\eta)$ used for simulations.

realizations. For each speckle realization, $F_n(x,y)$ was computed by zero-padding each 256×256 $f_n(\xi,\eta)$ to a size of 1024×1024, applying a fast Fourier transform (FFT), and cropping the resulting FFT to a size of 512×512. Using just an FFT to calculate $F_n(x,y)$ and neglecting the quadratic phase in Eq. (3) can be done because the solution is independent of the quadratic phase factors, as described above.

An ideal digital hologram $H_n(x,y)$ with no phase errors or noise was created by adding $F_n(x,y)$ to an ideal reference field $R(x,y)$ having the form of Eq. (2) and computing the squared modulus. A degraded digital hologram $H_{\varepsilon,n}(x,y)$ was created by: (i) including the phase error $\phi(x,y)$, shown in Fig. 4, in $F_n(x,y)$ to yield an aberrated object field, (ii) adding $R(x,y)$ to this field and computing the squared modulus, (iii) adding Poisson-distributed shot noise and Gaussian-distributed detector read noise, (iv) dividing by a detector A/D converter gain, and (v) quantizing the result to yield $H_{\varepsilon,n}(x,y)$ in digital number (DN) units of the detector.

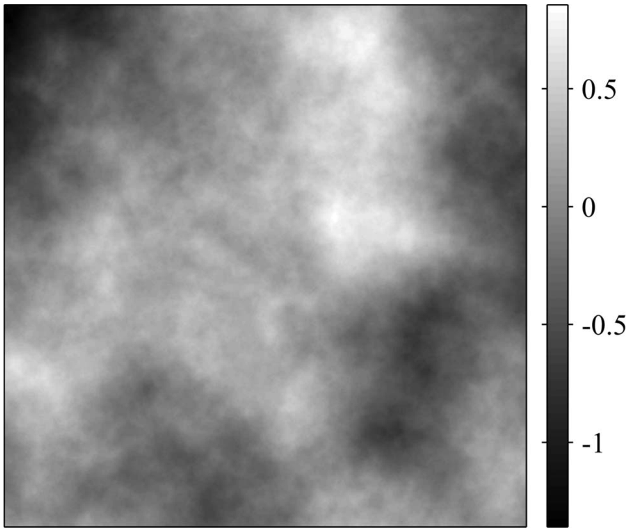

Fig. 4. Phase error $\phi(x,y)$ used for simulations in units of waves. The phase error is a random-draw atmospheric phase screen [16] (with tip and tilt subtracted) with $D/r_0=8$, where D is the width of the detector and r_0 is Fried's parameter.

The detector read noise standard deviation was 40 photoelectrons, the well depth was 5×10^4 photoelectrons, the A/D converter gain was 12.2 photoelectrons/DN, and the bit depth was 12. The shot noise statistics are determined by the measured power in units of photoelectrons. The intensity of the object beam was scaled to have a given average number of detected photoelectrons per pixel P_F. The amplitude of $R(x,y)$ was scaled to yield an average number of detected photoelectrons per pixel P_R, such that P_F+P_R was equal to 80% of the detector well depth, or 4×10^4 photoelectrons. Thus, the standard deviation of the shot noise was always ≈ 200 photoelectrons. Different SNRs were simulated by changing P_F, which in all cases was much less than P_R. The dominant noise source was shot noise. The average speckle size in the detector plane was approximately 4×4 pixels, since $f_n(x,y)$ was initially zero-padded by a factor of four along each dimension. Thus, P_F in units of average number of photoelectrons/pixel can be converted to units of object-beam photoelectrons/speckle by multiplying by 16.

The flow chart in Fig. 5 outlines the procedure for reconstructing the aberrated object field $G_n(x,y)$ from each digital hologram. $H_n(x,y)$ is multiplied by a window function, $W(x,y)$, to reduce edge effects and reduce sidelobes in the image domain, and an inverse FFT is performed on the result to yield $\tilde{H}_n(f_x,f_y)$ having the form of Eq. (4). An aberrated holographic image of the object is obtained by a coordinate shift and multiplication by another window function. Performing an FFT on the aberrated image and downsampling by a factor of two (by discarding every other sample) yields a 256×256 $G_n(x,y)$. This downsampling was done to reduce the excess memory requirements and computational burden of working with an oversampled version of $G_n(x,y)$. Both $W(x,y)$ and the holographic image window function were flattop windows with raised-cosine edges 11 pixels wide. After the 4× embedding of the object and 2× downsampling, the $G_n(x,y)$ fields were sampled at twice Nyquist and the intensity at Nyquist. Coarser sampling than this would result in reduced performance by DSLI [4,5].

For each set of $G_n(x,y)$, various phase-error estimates were generated using both DSLI and the sharpness metrics. Each algorithm yielded two polynomial-based and one point-by-point phase-error estimate $\hat{\phi}(x,y)$. The polynomial-based phase maps have the form

$$\hat{\phi}(x,y)=\sum_k C_k\psi_k(x,y), \tag{25}$$

where k is an index for the basis functions $\psi_k(x,y)$ (analogous to the Zernike polynomials) and C_k are expansion co-

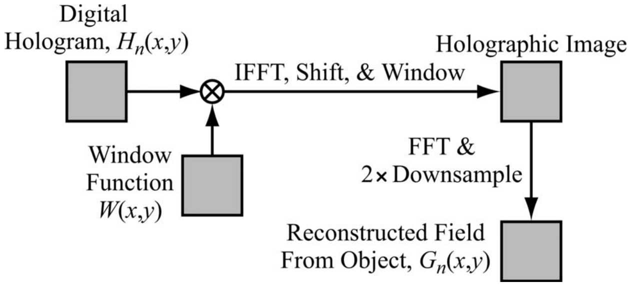

Fig. 5. Flowchart for reconstructing an object field from a digital hologram.

efficients. The first polynomial-based $\hat{\phi}(x,y)$ includes $\psi_k(x,y)$ up to 10th order (63 terms not including piston, tip, and tilt) and the second includes up to 15th order (133 terms). The 15th-order polynomial $\hat{\phi}(x,y)$ was then used as an initial guess in computing a point-by-point $\hat{\phi}(x,y)$. Starting with a polynomial-based phase estimate as the initial guess typically ensures that the algorithm is reasonably close to, and within the capture range of, the true phase-error estimate in the high-dimensional space of all point-by-point phase estimates.

For DSLI, the procedure for generating phase error estimates is:

1. Compute the quantities $S_x(x,y)$ and $S_y(x,y)$ from $G_n(x,y)$ using Eqs. (11) and (12);
2. Compute the 10th- and 15th-order polynomial-based $\hat{\phi}(x,y)$'s from $\arg[S_x(x,y)]$ and $\arg[S_y(x,y)]$ using an analytic method analogous to that of [14];
3. Starting with the 15th-order polynomial version of $\hat{\phi}(x,y)$, compute a point-by-point version of $\hat{\phi}(x,y)$ by running 100 iterations of a conjugate-gradient routine that tries to minimize the weighted mean-square error function [11]

$$E_{\mathrm{DSLI}} = \sum_{(x,y)} W(x,y)W(x-\Delta,y)\{\hat{\phi}(x,y) - \hat{\phi}(x-\Delta,y) - \arg[S_x(x,y)]\}^2 + \sum_{(x,y)} W(x,y)W(x,y-\Delta)\{\hat{\phi}(x,y) - \hat{\phi}(x,y-\Delta) - \arg[S_y(x,y)]\}^2, \tag{26}$$

where a 2× downsampled version of the window function described above is used for $W(x,y)$. $W(x,y)$ is included in Eq. (26) to reduce artifacts in $\hat{\phi}(x,y)$ arising from edge effects and the use of FFTs in computing $G_n(x,y)$.

For the sharpness metric approach, a conjugate-gradient routine was used to iteratively find the polynomial coefficients C_k or point-by-point phase values that minimize the negative of a particular sharpness metric (equivalent to maximizing the same sharpness metric). The polynomial coefficients were determined by starting with an initial guess for each coefficient of zero, performing five conjugate-gradient iterations including only up to 3rd-order terms, performing five more iterations including up to 4th-order terms, performing five more iterations including up to 5th-order terms, and so on, up to 15th order (performing five additional conjugate-gradient iterations each time an additional order of polynomials is included in the phase estimate). The point-by-point $\hat{\phi}(x,y)$ resulted from running only 25 conjugate-gradient iterations, starting from the 15th-order polynomial estimate. The polynomial estimation of the phase error mitigates, to a degree, the likelihood of oversharpening that is discussed below.

The results of the phase-correction algorithms are compared in terms of the root-mean-square (RMS) residual phase error. While the phase error introduced into the simulated data $\phi(x,y)$ is a 512×512 point-by-point phase map, the phase-error correction algorithms yield phase-error estimates $\hat{\phi}(x,y)$ having dimensions of 256×256. As a result, each $\hat{\phi}(x,y)$ was compared to a 2×2 boxcar average and 2× downsampled version of $\phi(x,y)$ having dimensions of 256×256. Piston, tip, and tilt phase differences between $\hat{\phi}(x,y)$ and $\phi(x,y)$ were then removed by finding the piston, tip, and tilt coefficients, a, b, and c, respectively, that minimize the metric

$$E = \sum_{(x,y)} W(x,y)|\exp\{i[\hat{\phi}(x,y) + a + bx + cy]\} - \exp[i\phi(x,y)]|^2. \tag{27}$$

As in Eq. (26), $W(x,y)$ is included here to reduce artifacts from edge effects. Finally the residual RMS phase error σ_ϕ was calculated as

$$\sigma_\phi^2 = \left[\sum_{(x',y')} W(x',y')\right]^{-1} \sum_{(x,y)} W(x,y) \times \{\arg[\exp\{i[\hat{\phi}(x,y) + a + bx + cy - \phi(x,y)]\}]\}^2, \tag{28}$$

where $W(x,y)$ is included again to weight down the residual phase errors at the array edges. In general, wrapped phase-error estimates were not an issue since all of the algorithms first estimated unwrapped polynomial-based phase maps and then used these results as initial guesses in forming point-by-point phase maps. Nevertheless, the particular form of Eq. (28) yields a σ_ϕ that is not affected by modulo-2π (between-π and π) differences between $\hat{\phi}(x,y)$ and $\phi(x,y)$.

4. SIMULATION RESULTS

This section presents simulation results that compare the relative performance of DSLI and various sharpness metrics for phase-error correction in two scenarios: (i) in the high-SNR regime with $P_F=640$ photoelectrons/speckle, as a function of the number of available object speckle realizations, and (ii) as a function of SNR using six object speckle realizations.

A. Performance in High Signal-to-Noise Regime

Figure 6 is a graph of the residual phase error for the DSLI algorithm versus the number of available speckle realizations. Each point in this and all the following graphs is the average σ_ϕ from five sets of $G_n(x,y)$ with independent speckle and noise realizations. For reference, the standard deviation of the phase error $\phi(x,y)$, was 0.34λ (unwrapped) and 0.28λ when computed modulo-2π, analogous to Eq. (28). Also for reference, a uniformly distributed, random, wrapped residual phase error will give $\sigma_\phi=0.29\lambda$. We point out that perfect 10th- and 15th-order polynomial estimates would yield residual phase errors of $\sigma_\phi=0.09\lambda$ and 0.06λ, respectively, for the specific $\phi(x,y)$ shown in Fig. 4. Note that these values would scale with D/r_0, i.e., for smaller D/r_0 the same levels of correction could be achieved theoretically with lower-order polynomials. As such, the 10th- and 15th-order polynomial-based phase-error estimates are expected to yield corrections no better than these limiting values.

Figure 6 indicates that DSLI yields 10th- and 15th-order polynomial phase estimates with $\sigma_\phi \approx 0.09\lambda$ and

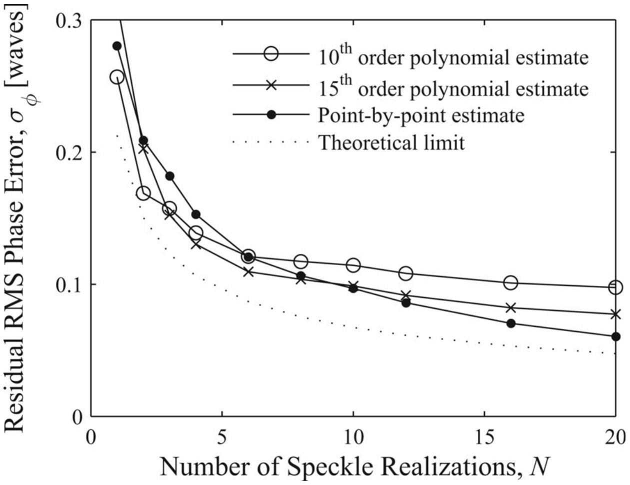

Fig. 6. DSLI performance versus number of available speckle realizations for $P_F=640$ photoelectrons/speckle. Each data point is the average of results from five trials with independent speckle and noise realizations. The dotted curve is a theoretical prediction of DSLI performance based on Eqs. (29) and (30) (ignoring photon and detector noise).

0.07λ, respectively, in the best scenario ($N=20$ speckle realizations). These are quite close to the corresponding limiting values. In this same scenario, the point-by-point estimate yielded $\sigma_\phi \approx 0.06\lambda$ (the ideal σ_ϕ value for a point-by-point phase estimate is zero). The point-by-point estimate is superior in this case because it can better match the true phase error than can the polynomial expansions. Conversely, the polynomial estimates performed better than the point-by-point estimate in the case of very few speckle realizations, because the polynomial expansion is a form of regularization that mitigates against the fitting of errors in the shear data $S_x(x,y)$ and $S_y(x,y)$ due to averaging over an insufficient number of speckle realizations. The crossover point, where the point-by-point phase estimate started to yield better corrections than the 15th-order polynomial estimate, was roughly $N=10$ speckle realizations, which also happens to be the point at which DSLI yields corrections with $\sigma_\phi \leqslant \lambda/10$.

Figure 6 also shows a theoretical prediction of σ_ϕ resulting from the speckle statistics (ignoring photon and detector noise) given by [11]

$$\sigma_\phi(N) = \sigma_{\mathrm{DSLI}}(N)\sqrt{0.6558[1+\ln(256^2)]}, \tag{29}$$

where

$$\sigma_{\mathrm{DSLI}}(N) = \sqrt{\frac{1-|\mu|^2}{2|\mu|^2 N}} \tag{30}$$

is the expected standard deviation of $\arg[S_x(x,y)]$ and $\arg[S_x(x,y)]$ due to the speckle statistics [4,5],

$$\mu(f_x,f_y) = \tilde{I}(f_x,f_y)/\tilde{I}(0,0), \tag{31}$$

and μ represents either $\mu[\Delta/(\lambda f),0]$ or $\mu[0,\Delta/(\lambda f)]$. The theoretical curve shown in Fig. 6 was calculated using $|\mu|=\mathrm{sinc}(0.5)$, since the fields were sampled at twice the Nyquist limit and the diffuse object was illuminated by a square beam. Note that Eq. (30) is valid for $N>10$. [4,5]

Figure 7 shows DSLI results for specific trials. Each column corresponds to a different number of speckle realizations N and shows an ideal speckle-averaged image (with no phase errors or noise), a speckle-averaged image degraded by both phase errors and noise, and speckle-averaged images corresponding to the 10th-order polynomial and point-by-point phase error estimates from DSLI.

The performance of the sharpness metric M_1 with $\alpha=2$ (i.e., the commonly used squared-intensity sharpness metric) is shown in Figs. 8 and 9. In all but the $N=20$ case, the point-by-point phase-error estimation failed catastrophically by converging to a $\hat{\phi}(x,y)$ that yields an oversharpened $\hat{I}(\xi,\eta)$ that contains a bright, pointlike delta function [recall that M_1 with $\alpha=2$ is maximized by stretching the histogram of $\hat{I}(\xi,\eta)$]. Further examination reveals that, in these cases, the point-by-point phase estimates contain phase vortices and yield individual speckle images, i.e., $|\hat{f}_n(\xi,\eta)|^2$ for each $n\in\{1,2,\ldots,N\}$, that contain commonly located delta functions. For $N=1$, this failure mode occurs when $\hat{\phi}(x,y)$ approximates $\phi(x,y)$ plus the phase of $F_n(x,y)$ (which typically contains numerous phase vortices). For larger N, the algorithm attempted to find a $\hat{\phi}(x,y)$ that simultaneously fits $\arg[F_n(x,y)]$ for multiple speckle realizations by matching $\arg[F_n(x,y)]$ for each n where $|F_n(x,y)|$ was large. In doing this the algorithm seemed to favor sharpening a small number of speckle realizations over the others, but each $|\hat{f}_n(\xi,\eta)|^2$ contained a commonly located bright point. For sufficiently large N, e.g., $N=20$, there were enough speckle statistics to avoid this failure mode of overemphasizing bright image points and to allow the algorithm to converge to an accurate point-by-point phase estimate with $\sigma_\phi=\lambda/20$. Conversely, the polynomial expansion provides regularization against this failure mode, i.e., a polynomial-based $\hat{\phi}(x,y)$ cannot contain the phase vortices generally required to match $\arg[F_n(x,y)]$, allowing phase-error estimates with $\sigma_\phi \leqslant \lambda/10$ with as few as $N=4$ speckle realizations.

Figure 10 shows the performance of metrics M_1 and M_2 versus the metric parameter α and of the minimum entropy metric M_3 for $N=20$ speckle realizations and $P_F=640$ photoelectrons/speckle. Note that the minimum entropy results obtained with M_3 are plotted in Fig. 10 for $\alpha=1$, since it has been shown [9] that minimizing entropy is equivalent to maximizing M_1 with $\alpha=1+\varepsilon$ for $\varepsilon \ll 1$. Using M_1 with $\alpha>2$ gives point-by-point phase estimates that yield an oversharpened $\hat{I}(\xi,\eta)$ containing a bright delta function, which results from M_1 placing increasing emphasis on making bright points brighter as α increases [9]. Thus, more than 20 speckle realizations were required to prevent this failure mode of M_1 for $\alpha>2$. For $\alpha<0.25$, we speculate that M_2 fails to converge to an accurate $\hat{\phi}(x,y)$ by placing too much emphasis on minimizing various points in the degraded image that happen to be dark due to the details of the speckle and/or noise realizations.

Figure 10 indicates that the optimum value of α for M_1 and M_2 is a value slightly larger or less than unity. This is interesting because the values of M_1 and M_2 are independent of $\hat{\phi}(x,y)$ for $\alpha=1$, since the integrated value of $\hat{I}(\xi,\eta)$ is a conserved quantity. The optimum sharpness metric is known to vary with the details of the object [9]; thus, even

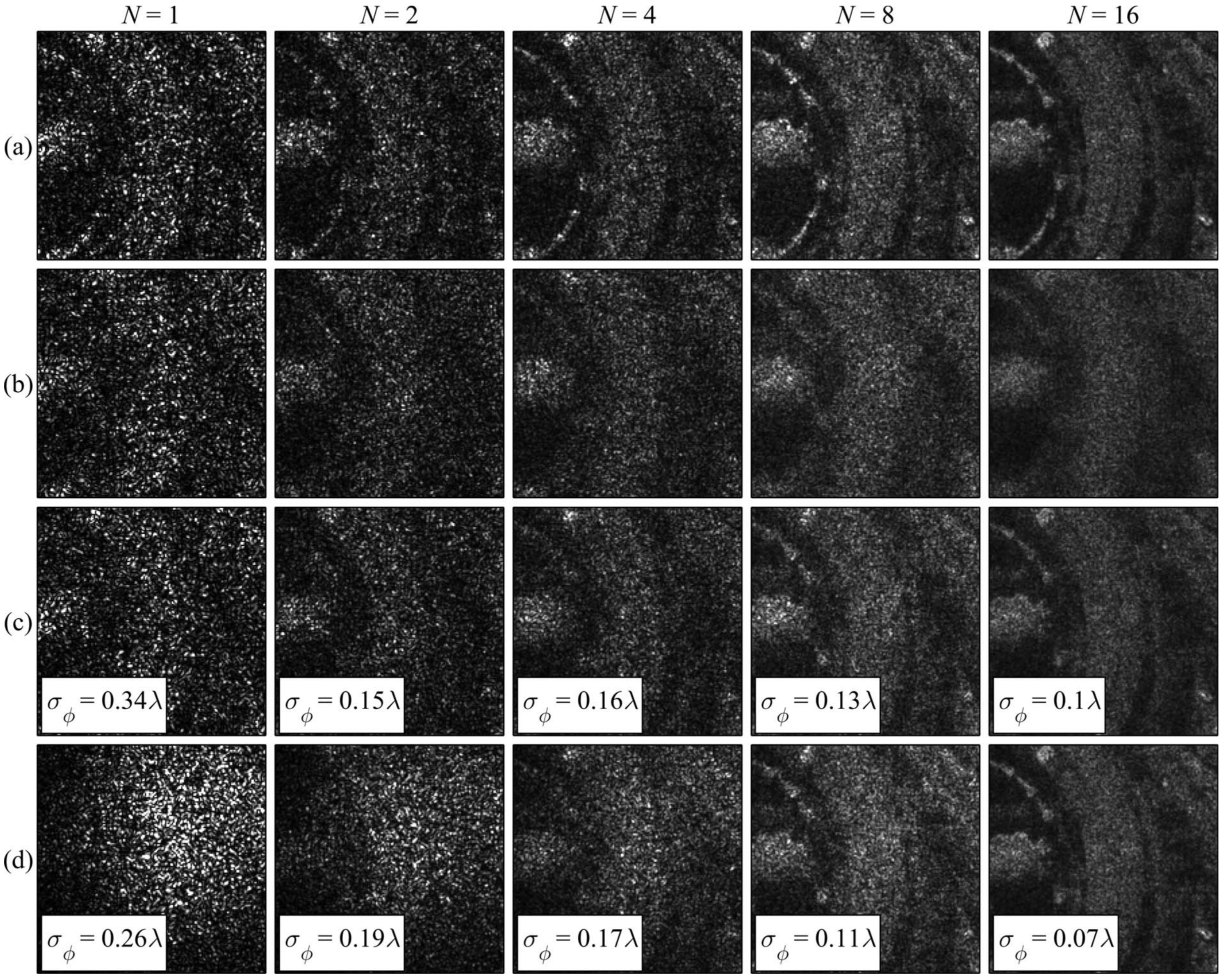

Fig. 7. DSLI results for specific trials. Each column corresponds to a different number of speckle realizations N and shows in successive rows (a) an ideal speckle-averaged image (with no phase errors or noise), (b) a speckle-averaged image degraded by both phase errors and noise, and speckle-averaged reconstructed images corresponding to (c) the 10th-order polynomial and (d) the point-by-point phase-error estimates.

though $\alpha \approx 1$ is ideal for the object used in the simulations here, the optimum value of α is expected to vary with different objects. It is worth noting that the choice of α in M_1 or M_2 can yield dramatically different results, ranging from catastrophic failure at one extreme to better than λ/50 corrections at the other extreme, with a point-by-point $\hat{\phi}(x,y)$.

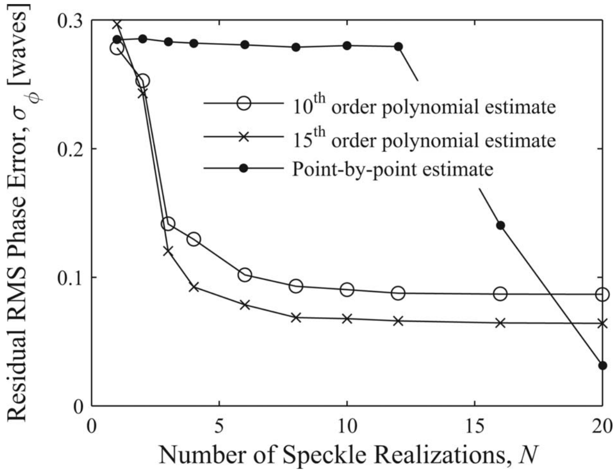

Fig. 8. Sharpness metric M_1 with $\alpha=2$ performance versus number of available speckle realizations for $P_F=640$ photoelectrons/speckle.

When using a polynomial expansion for $\hat{\phi}(x,y)$, the results obtained with M_1 and M_2 for $\alpha > 0.1$ were near the limits for perfect correction mentioned above, even for $\alpha > 2$. Figure 11 shows the results obtained using M_1 with $\alpha=1.01$ (which is near optimum for large N and large SNR) as a function of the number of available speckle realizations. The graph indicates that while the polynomial expansion was needed for regularization in the case of only one or two speckle realizations, an accurate point-by-point $\hat{\phi}(x,y)$ with $\sigma_\phi < \lambda/10$ was achieved with only three speckle realizations. A similar graph obtained by use of M_3 (negative entropy) is not shown, as the results are virtually identical to those already shown in Fig. 11.

The optimum value of α was found to vary with the number of available speckle realizations. Figure 12 shows σ_ϕ as a function of α when different numbers of speckle realizations are available. As the number of speckle realizations decreases, the optimum value of α decreases from 1.01 for $N=16$ or 20 to approximately 0.5 for $N=1$. Figure 13 shows the results of using M_2 with $\alpha=0.5$ versus the

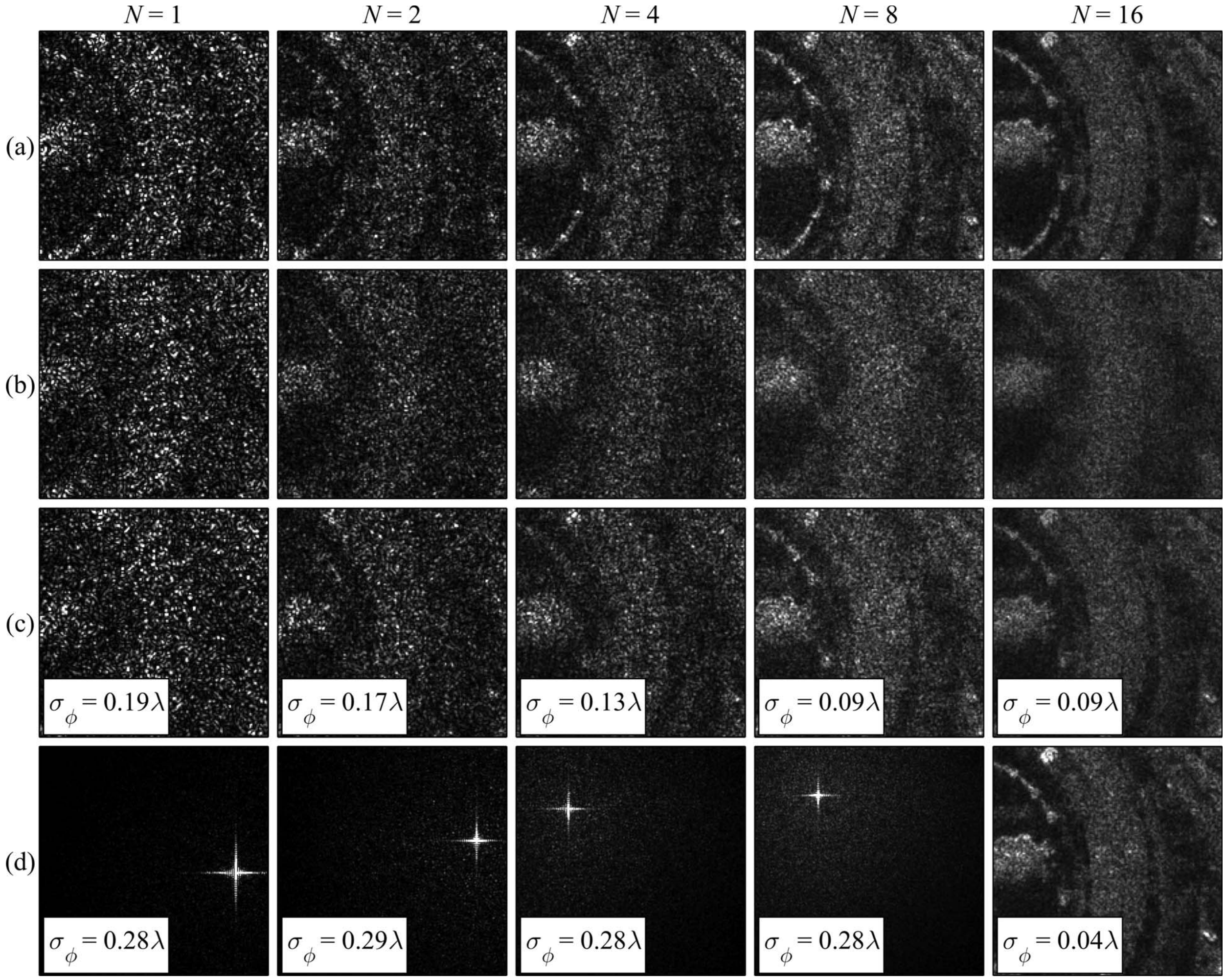

Fig. 9. Sharpness metric M_1 with $\alpha=2$ results for specific trials. Each column corresponds to a different number of speckle realizations N and shows in successive rows (a) an ideal speckle-averaged image (with no phase errors or noise), (b) a speckle-averaged image degraded by both phase errors and noise, and speckle-averaged reconstructed images corresponding to (c) the 10th-order polynomial and (d) the point-by-point phase-error estimates.

number of speckle realizations. While this value of α did not perform as well as $\alpha=1.01$ for large N, it performed better than $\alpha=1.01$ for $N=1$ and 2, as indicated by Figs. 11–13. Additionally, for $\alpha=0.5$, the polynomial expansion for $\hat{\phi}(x,y)$ was not required for regularization in the case of $N=1$ or 2 as it was for $\alpha=1.01$. Thus, in cases where

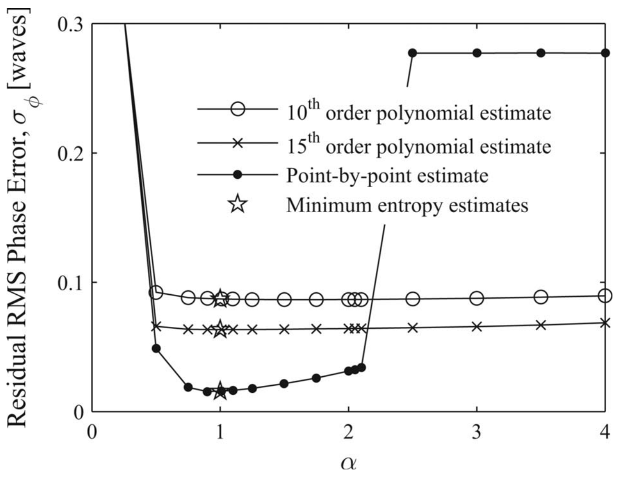

Fig. 10. Sharpness metric M_1 and M_2 performance versus α and sharpness metric M_3 performance for $P_F=640$ photoelectrons/speckle and $N=20$ speckle realizations.

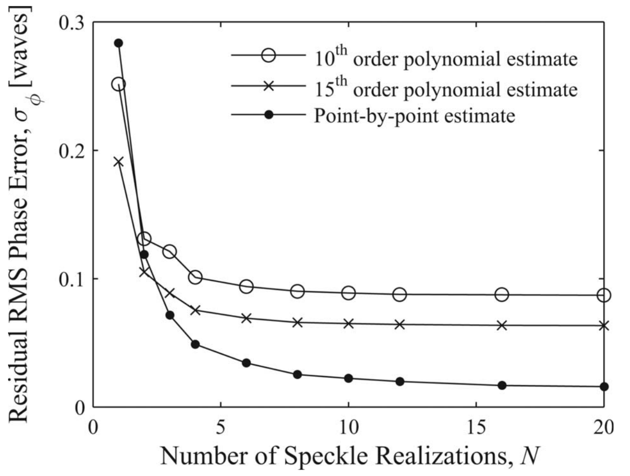

Fig. 11. Sharpness metric M_1 with $\alpha=1.01$ performance versus number of available speckle realizations for $P_F=640$ photoelectrons/speckle.

only one or two speckle realizations are available, a value of $\alpha=0.5$ is preferred over $\alpha=1.01$. Note that $\alpha=0.5$ is equivalent to maximizing the sum of the field magnitudes.

In almost all cases examined, use of sharpness metric M_4 failed catastrophically. For $\alpha<1$, M_4 concentrates on increasing the magnitude of relatively small differences between neighboring pixels in $\hat{I}(\xi,\eta)$. As a result, the algorithm failed catastrophically by converging to a $\hat{\phi}(x,y)$ that yields an image $\hat{I}(\xi,\eta)$ that looks like a correlated noise pattern. For $\alpha>1$, M_4 concentrates on increasing the magnitude of relatively large differences between neighboring image pixels. The algorithm failed catastrophically by converging to a phase estimate that yields a bright pointlike image. In a few isolated trials the polynomial-based phase estimates resulting from use of M_4 with $1.5\leqslant\alpha\leqslant2.75$ yielded $\sigma_\phi\approx\lambda/8$. However, these appear to be "lucky" cases depending on the details of the speckle and noise realizations, as no clear trends were evident. In our trials of M_4 with $\alpha\approx2$, we obtained $\sigma_\phi\approx\lambda/8$ only about a quarter of the time with data obtained under favorable conditions (high SNR and many speckle realizations).

Use of sharpness metric M_5 was extensively tested in high SNR regimes for $N=20$ speckle realizations and a wide range of α values, but failed to yield any accurate phase-error corrections. The effect of a particular α value depends on the absolute scaling of $\hat{I}(\xi,\eta)$. For large and even very modest values of α ($\geqslant0.05$ for the case considered here), M_5 concentrates on reducing large magnitude differences between neighboring pixels in the speckle-averaged image, and the algorithm failed by converging to a $\hat{\phi}(x,y)$ that yielded a severely blurred image. For very small α ($\leqslant0.01$), the algorithm did not yield any noticeable improvement in the quality of the speckle-averaged image, and we speculate that the algorithm was trying essentially to smooth out individual speckles.

B. Performance versus Signal-to-Noise Ratio

The performance of DSLI and each of the statistics-based sharpness metrics of Table 1 was explored as a function of SNR for the case of $N=6$ speckle realizations. Figure 14 shows how the performance of DSLI varies with SNR. Using a 15th-order polynomial expansion, DSLI yielded corrections with $\sigma_\phi\approx\lambda/6$ down to signal levels of $P_F=3$ photoelectrons/speckle. Figure 15 shows the performance of sharpness metric M_1 with $\alpha=1.01$ (the best performing metric in the high SNR case with $N=20$) as a function of the object beam energy. This plot shows that corrections $<\lambda/10$ were achievable down to signal levels of $P_F=2.5$ photoelectrons/speckle. The average performance of M_1 degraded sharply below this SNR.

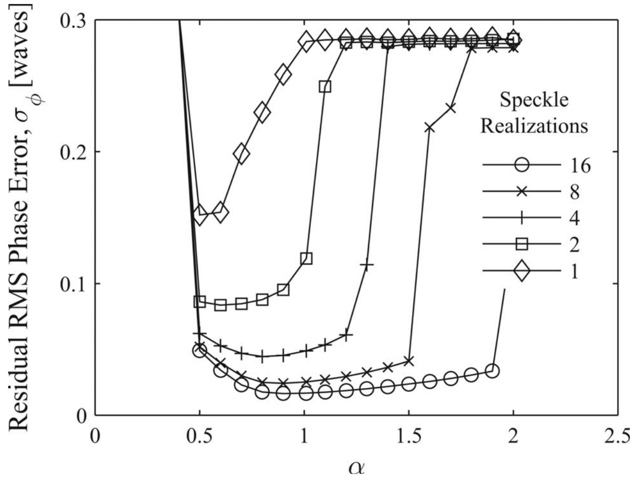

Fig. 12. Sharpness metric M_1 and M_2 performance versus α for $P_F=640$ photoelectrons/speckle with different numbers of available speckle realizations.

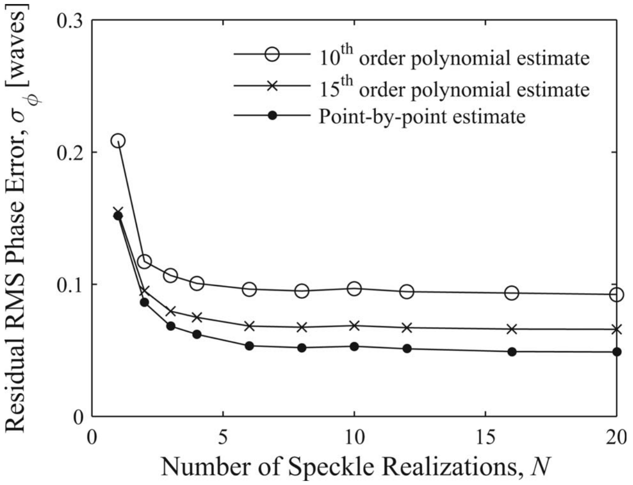

Fig. 13. Sharpness metric M_2 with $\alpha=0.5$ performance versus number of available speckle realizations for $P_F=640$ photoelectrons/speckle.

Figure 16 shows results for specific trials at different signal levels. Results (not shown) obtained with metric M_3 were nearly identical to those obtained with M_1 using $\alpha=1.01$. Figure 17 shows a plot of the performance of M_1 with $\alpha=1.25$, which was better than using $\alpha=1.01$ for $P_F<3$ photoelectrons/speckle. This suggests that the sharpness metric that performs best depends on the SNR. Figure 18 shows the performance of M_1 and M_2 over a range of different α values at different signal levels. This

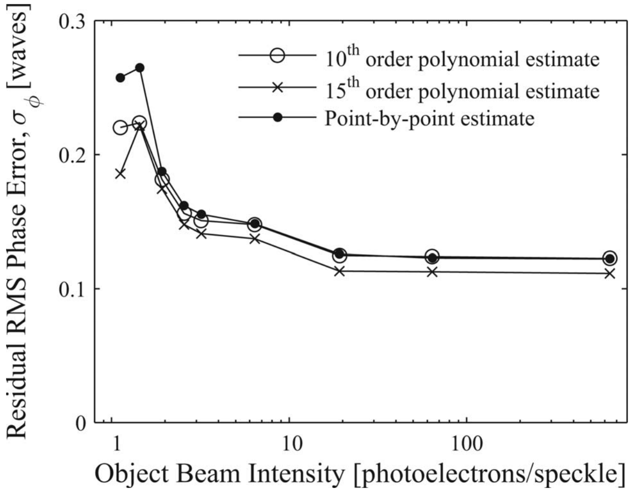

Fig. 14. DSLI performance versus object beam intensity for $N=6$ speckle realizations.

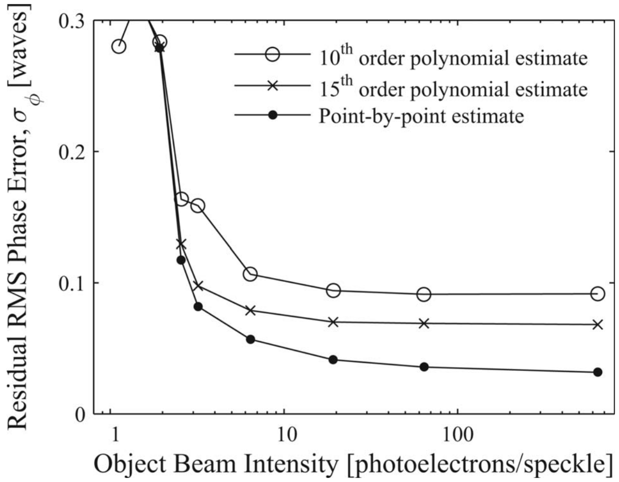

Fig. 15. Sharpness metric M_1 with $\alpha=1.01$ performance versus object beam intensity for $N=6$ speckle realizations.

graph clearly shows that the optimum value of α gradually changed from approximately 0.9 at high signal levels to about 1.25 for low signal levels for $N=6$ speckle realizations.

5. SUMMARY

We have examined the performance of digital shearing laser interferometry (DSLI) and a number of different sharpness metrics for phase-error correction for imagery computed from digital holography or heterodyne array data. Algorithm performance is expected to vary with object details, nature of the phase errors, number of available speckle realizations, and SNR. Our simulations examined algorithm performance as a function of the number of speckle realizations and SNR, while keeping the object and phase error fixed. DSLI is generally more robust than sharpness metrics, in that DSLI did not exhibit catastrophic failure modes associated with oversharpening or excessive blurring. In all cases, however, the best performing sharpness metric outperformed DSLI in terms of residual RMS phase error. We hypothesize that this happens because sharpness metrics employ *a priori* information (high-quality images typically have wider histograms than those of corresponding low-quality images) that DSLI does not.

The best performing sharpness metric was found to vary with both the number of speckle realizations and SNR. Simulation results with DSLI yielded phase-error

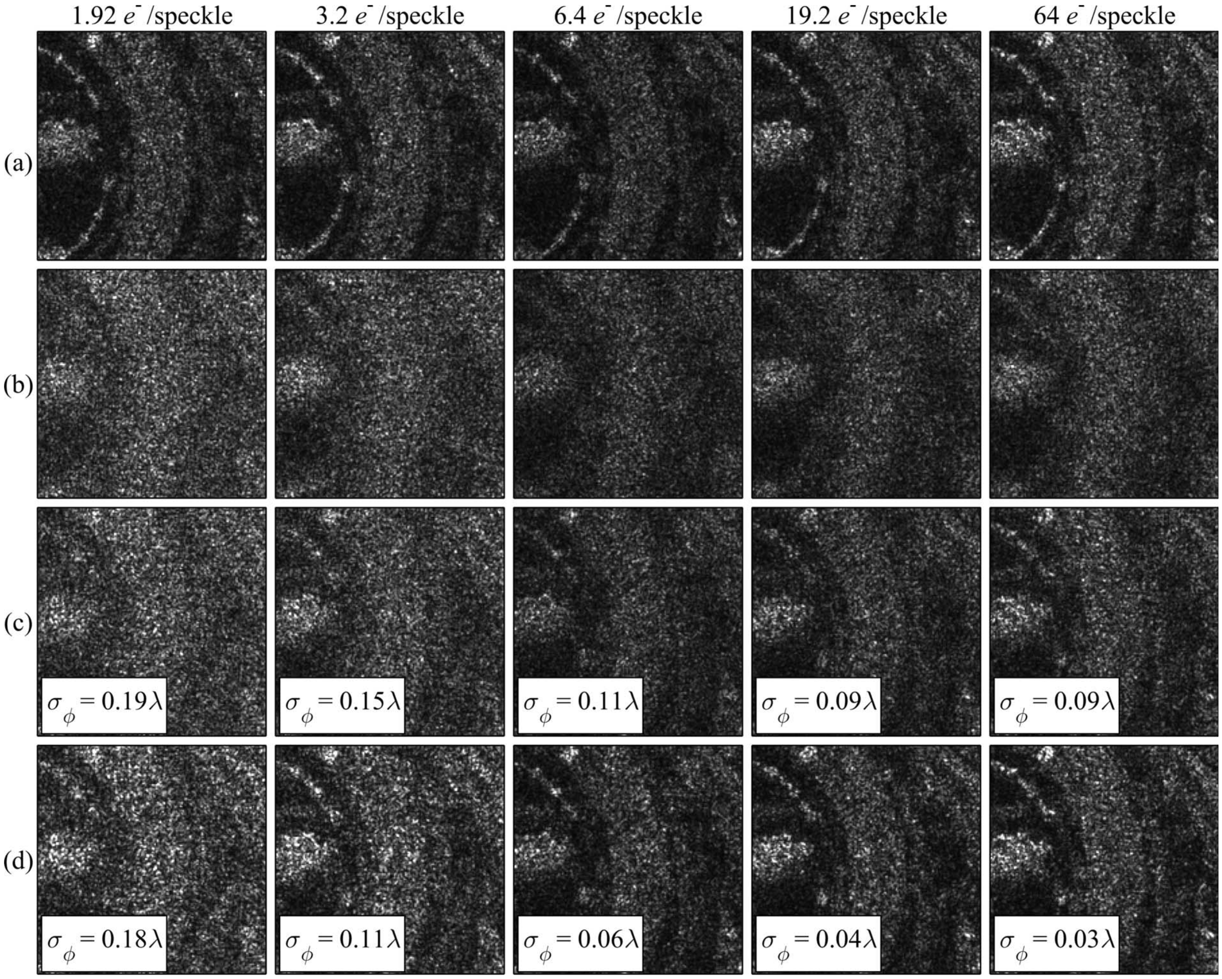

Fig. 16. Sharpness metric M_1 with $\alpha=1.01$ results for specific trials. Each column corresponds to a different object beam intensity P_F with different speckle realizations and shows in successive rows (a) an ideal speckle-averaged image (with no phase errors or noise), (b) a speckle-averaged image degraded by both phase errors and noise, and speckle-averaged reconstructed images corresponding to (c) the 10th-order polynomial and (d) the point-by-point phase-error estimates.

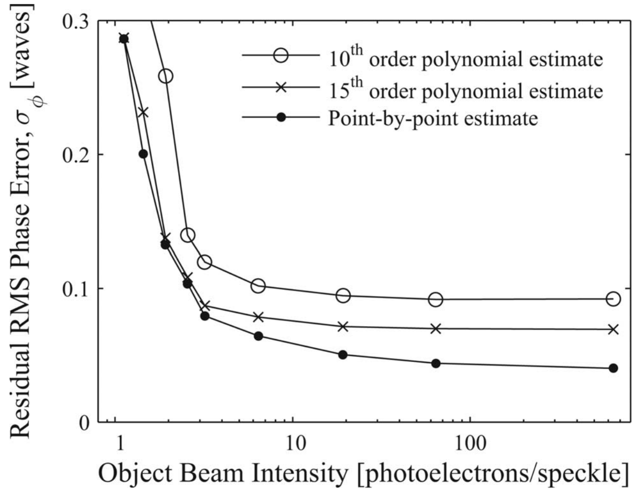

Fig. 17. Sharpness metric M_1 with $\alpha=1.25$ performance versus object beam intensity for $N=6$ speckle realizations.

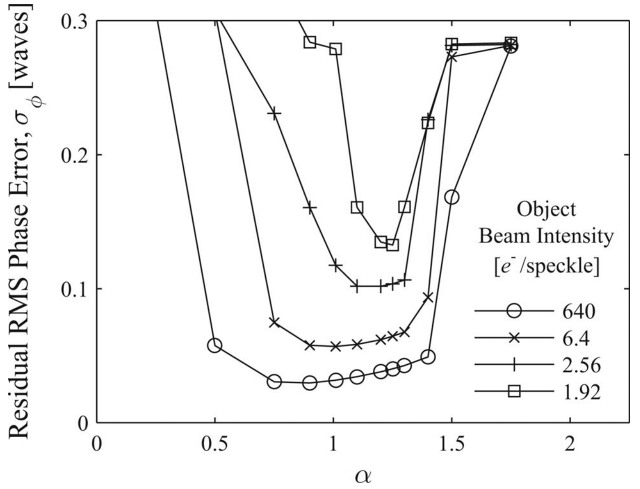

Fig. 18. Sharpness metrics M_1 and M_2 performance versus α for $N=6$ speckle realizations with different object beam intensities.

corrections with less than $\lambda/10$ RMS residual phase error in high SNR conditions (640 photons/speckle for the object beam) with ten or more speckle realizations and $\approx\lambda/6$ corrections with only ~3 photons/speckle and six speckle realizations. In the high SNR regime, the best performing sharpness metrics achieved $<\lambda/50$ corrections with >16 speckle realizations and $<\lambda/10$ corrections with as few as two speckle realizations. With only 3 photons/speckles for the object beam, the best performing sharpness metric achieved $\lambda/10$ corrections. Also worth noting is that DSLI requires samples over a contiguous aperture, but the sharpness metrics do not.

ACKNOWLEDGMENTS

This work was supported by Lockheed Martin Corporation. The authors thank Joe Marron for useful discussions and suggestions.

REFERENCES

1. J. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts, 2004).
2. E. N. Leith and J. Upatnieks, “Reconstructed wavefronts and communication theory,” J. Opt. Soc. Am. **52**, 1123–1130 (1962).
3. F. Le Clerc, L. Collot, and M. Gross, “Numerical heterodyne holography with two-dimensional photodetector arrays,” Opt. Lett. **25**, 716–718 (2000).
4. J. R. Fienup, J. N. Cederquist, J. C. Marron, T. J. Schulz, and J. H. Seldin, “Heterodyne array phasing by digital shearing laser interferometry,” in *IRIS Specialty Group on Active Systems Meeting Digest*, October 16–18, 1990.
5. J. N. Cederquist, J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, “Digital shearing laser interferometry for heterodyne array phasing,” Proc. SPIE **1416**, 266–277 (1991).
6. R. A. Muller and A. Buffington, “Real-time correction of atmospherically degraded telescope images through image sharpening,” J. Opt. Soc. Am. **64**, 1200–1210 (1974).
7. R. G. Paxman and J. C. Marron, “Aberration correction of speckled imagery with an image-sharpness criterion,” Proc. SPIE **976**, 37–47 (1988).
8. J. R. Fienup, A. M. Kowalczyk, and J. E. Van Buhler, “Phasing sparse arrays of heterodyne receivers,” Proc. SPIE **2241**, 127–131 (1994).
9. J. R. Fienup and J. J. Miller, “Aberration correction by maximizing generalized sharpness metrics,” J. Opt. Soc. Am. A **20**, 609–620 (2003).
10. J. W. Goodman, *Statistical Optics* (Wiley, 2000), Sec. 5.6, pp. 207–211.
11. D. L. Fried, “Least-squares fitting a wave-front distortion estimate to an array of phase-difference measurements,” J. Opt. Soc. Am. **67**, 370–375 (1977).
12. W. H. Southwell, “Wave-front estimation from wave-front slope measurements,” J. Opt. Soc. Am. **70**, 998–1006 (1980).
13. H. Takajo and T. Takahashi, “Least-squares phase estimation from the phase difference,” J. Opt. Soc. Am. A **5**, 416–425 (1988).
14. E. Acosta, S. Bará, M. A. Rama, and S. Rios, “Determination of phase mode components in terms of local wave-front slopes: an analytical approach,” Opt. Lett. **20**, 1083–1085 (1995).
15. Provided courtesy of Jet Propulsion Laboratories (J. B. Breckinridge).
16. R. G. Lane, A. Glindemann, and J. C. Dainty, “Simulation of a Kolmorgorov phase screen,” Waves Random Media **2**, 209–224 (1992).
17. D. L. Fried, “Optical resolution through a randomly inhomogeneous medium for very long and very short exposures,” J. Opt. Soc. Am. **56**, 1372–1379 (1966).