# Complex-wave retrieval from a single off-axis hologram

**Michael Liebling, Thierry Blu, and Michael Unser**

*Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Bâtiment de Microtechnique 4.127, CH-1015 Lausanne, Switzerland*

Received April 2, 2003; revised manuscript received November 11, 2003; accepted November 14, 2003

We present a new digital two-step reconstruction method for off-axis holograms recorded on a CCD camera. First, we retrieve the complex object wave in the acquisition plane from the hologram's samples. In a second step, if required, we propagate the wave front by using a digital Fresnel transform to achieve proper focus. This algorithm is sufficiently general to be applied to sophisticated optical setups that include a microscope objective. We characterize and evaluate the algorithm by using simulated data sets and demonstrate its applicability to real-world experimental conditions by reconstructing optically acquired holograms.
© 2004 Optical Society of America
*OCIS codes:* 090.1760, 100.2000, 100.3010, 100.5070, 180.3170

## 1. INTRODUCTION

The reconstruction of holograms by computer was first reported, some 20 years after Gabor's landmark paper,[1] by Goodman and Lawrence[2] and Yaroslavsky and co-workers.[3,4] Milestones in the evolution of the technique and algorithms have been the use of a CCD camera to acquire the hologram,[5] the reconstruction of the phase in addition to the amplitude,[6] and the measurement of polarization states.[7] A recent evolution of digital holography is its application to microscopy, which allows for truly noninvasive examination of biological samples. This promising technique has the ability to detect subwavelength changes in the morphology of living organisms.[8]

Here, we consider digital holograms that result from the interference between the wave reflected or transmitted by the object to be imaged and a plane reference wave that can be modeled with good accuracy. More specifically, we will concentrate on off-axis geometries (see Fig. 1), which permit wave front retrieval from one single, two-dimensional, real-valued intensity image. This allows one to work at high acquisition rates and makes the technique highly suitable for tracking fast biological processes.

So far, most methods that aimed at retrieving the complex wave from the hologram were directly inspired by the optical reconstruction process: The chemically processed photographic plate—the hologram—is illuminated, and the image (respectively the virtual image) is generated by the diffracted wave. The translation of this physical process into a numerical algorithm is nearly literal. Simulating the diffraction process boils down to computing the propagation of a complex wave, which can be done using several approximations.[9] The hologram may also be multiplied by an appropriate digital counterpart of a reference wave beforehand or afterwards to retrieve the proper phase.[6]

The most blatant disadvantage of approaches that imitate the physical process is that the reconstructed image is severely corrupted by interference terms: the zero-order term and the out-of-focus twin-image term. While several techniques have been proposed for removing them,[10–12] the presence of these terms still remains a determining factor that limits the quality of the reconstructed image, or at least the field of view. The present approach overcomes this disadvantage by intrinsically removing the zero-order and the twin-image terms without the need for any pre- or postprocessing.

The method we put forward has two steps. First, we estimate the amplitude and phase in the acquisition plane by applying a new algorithm that retrieves the complex wave in the CCD plane from the real-valued measurements. The key idea is to perform a nonlinear change of variables so that the reconstruction may be performed by use of a method that is reminiscent of phase-shifting techniques. The algorithm is based on a local least-squares estimation of the amplitude and phase by assuming an *a priori* model of the reference wave's phase. Once the complex object wave is recovered in the acquisition plane, we (back) propagate the wave (which contains neither zero-order nor twin-image terms) to restore a focused image using a digital implementation of the Fresnel transform. The paper concentrates on the first step, that is, the phase retrieval from a single interferogram in several configurations that fulfill the appropriate hypotheses.

The paper is organized as follows. In Section 2, we briefly review existing methods for digital, off-axis holography reconstruction and, more generally, related algorithms in interferometry. In Section 3, we present the phase-retrieval algorithm specifically targeted to the application at hand. In Section 4, we present some experimental examples and validations of the technique using both synthetic (simulated) and true measurement data.

In what follows, we use the following definition of the Fourier transform $\hat{f}(\nu)$ of a function $f(x)$

 © 2004 Optical Society of America

$$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(x)\exp(-2i\pi x\nu)\mathrm{d}x,$$

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\nu)\exp(2i\pi x\nu)\mathrm{d}\nu,$$

and the free-space propagation operator $\mathcal{R}_{(d)}$ defined for functions of two variables as the Fresnel transform

$$\mathcal{R}_{(d)}\{U\}(\mathbf{x}) = \frac{\exp(ikd)}{i\lambda d}\int\int_{-\infty}^{\infty} U(\xi,\eta)\exp\left\{\frac{i\pi}{\lambda d}[(\xi - x)^2 + (\eta - y)^2]\right\}\mathrm{d}\xi\mathrm{d}\eta, \quad (1)$$

where $\lambda$ is the wavelength of the light, $k = 2\pi/\lambda$ its wave number, and $d$ the propagation distance; also, we denote $\mathbf{x} = (x, y)$.

## 2. REVIEW OF EXISTING ALGORITHMS

Before presenting our method, we briefly review the most widely used techniques for reconstructing digital holograms, as well as a large body of related work in interferometry.

### A. Standard (Linear) Reconstruction Techniques

The hologram originates from the interference between an object and a reference wave:

$$I(\mathbf{x}) = |R(\mathbf{x}) + \Psi(\mathbf{x})|^2 = \underbrace{|R(\mathbf{x})|^2 + |\Psi(\mathbf{x})|^2}_{\text{zero-order}} + R^*(\mathbf{x})\Psi(\mathbf{x}) + \Psi^*(\mathbf{x})R(\mathbf{x}), \quad (2)$$

where $R(\mathbf{x})$ is the reference wave and $\Psi(\mathbf{x})$ the object wave evaluated in the acquisition plane. The first two intensity contributions in Eq. (2) are known as the zero-order term, while the third and fourth are the image and twin-image terms, respectively. These terms are superimposed in the hologram space, the plane in which the acquisition was made. When the hologram is acquired in an off-axis geometry {that is, $R(\mathbf{x}) = \exp[i(k_x x + k_y y)]$ where $\mathbf{k} = (k_x, k_y, k_z)$ is the wave vector}, they can be separated by taking either the hologram's Fourier transform or its Fresnel transform (see Fig. 2).

This is simply because the three terms are separated in the Fourier domain: The zero-order term is located around the origin, while the image and twin image are centered on $(-k_x/(2\pi), -k_y/(2\pi))$ (image) and $(k_x/(2\pi), k_y/(2\pi))$ (twin image), respectively. Alternatively, after application of a Fresnel transform with distance parameter $d$ and wavelength $\lambda$, the three terms will be located at the respective positions: (0, 0) (zero-order), $(-k_x/(2\pi)\lambda d, -k_y/(2\pi)\lambda d)$ (image), and $(k_x/(2\pi)\lambda d, k_y/(2\pi)\lambda d)$ (virtual image).

The most popular reconstruction methods for digital holograms are in essence linear techniques: They filter out the useless information (twin-image, zero-order) and keep the image information.[6,13,14] Instead of first transforming the hologram to another domain (Fourier domain or diffracted plane), the filtering operation can also be performed in the spatial domain by convolving the hologram with a suitable complex-valued function.[15] The frequency responses of the filters in both cases are similar.

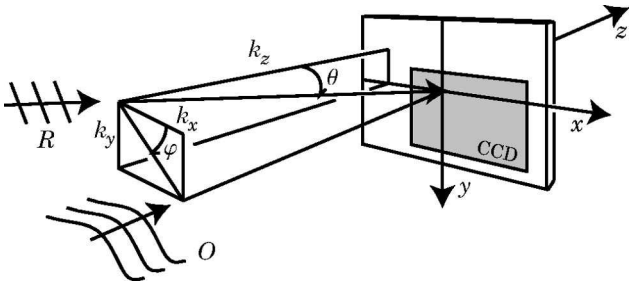


Fig. 1. Schematic view of the off-axis geometry.

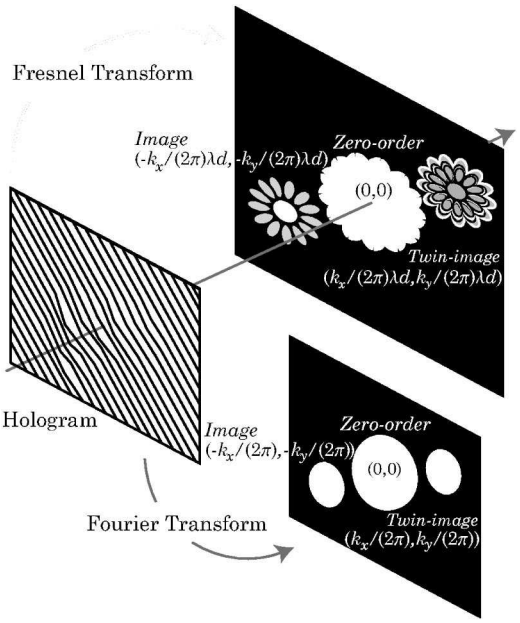


Fig. 2. Information separation of the interferogram using the Fourier, respectively Fresnel, transform.

### B. Other Related Techniques

The first part of the method that we are proposing for digital holography consists of the demodulation of a fringe pattern, a general problem for which a wide variety of algorithms have been devised. One of the most popular, the Fourier transform technique,[16,17] is essentially a linear technique that selectively filters the relevant information in the Fourier domain. In fact, it is formally equivalent to the way recent zero-order and twin-image-removal algorithms for digital holography operate. Many variants have been proposed to overcome some limitations and to optimize the technique.[14,18–20] Similar results have been achieved by using equivalent spatial convolution filters.[13,15]

Alternatively, there exist a number of methods that may be regarded as the spatial counterparts of temporal phase-shifting methods.[21,22] The latter require *several* interferograms to be recorded, each corresponding to a particular shift of the reference's phase; the phase estimation is done at each pixel by using the pixels at the same location in the other interferograms. In contrast, *spatial* phase-shifting algorithms[17,23–28] use the neighboring pix-

els (on the same line) of a single interferogram to carry out the estimations. The phase to retrieve is assumed to vary slowly and the carrier frequency to be constant. The general method has been improved in many ways, for example to compensate for nonideal acquisition devices and conditions.[29–35]

A number of spatial demodulation algorithms have been proposed that are able to demodulate interferograms with closed fringes.[36–40] Thus, they might also form the basis of a two-step procedure to reconstruct digital off-axis holograms acquired in the presence of a microscope lens, since such general setups possibly yield closed fringes. They would, however, require some adaptation to our problem, a strategy that we are not pursuing here.

Finally, a general approach for reconstructing digital holograms that inherently takes into account the statistical nature of the measured data[41,42] may also be adapted to reconstruct our type of data.

## 3. PROPOSED COMPLEX-WAVE-RETRIEVAL ALGORITHM

Our approach to reconstructing digital holograms is decomposed into two distinct and independent parts. Starting from the digitally acquired interferogram, these are:

1. Reconstruction of the object wave's amplitude and phase in the acquisition plane by using an algorithm that is reminiscent of phase-shifting methods in interferometry.
2. Numerical propagation of the reconstructed wave front to achieve proper focus in the image plane.

Here, we focus mainly on the first step. The second step, if required, can be implemented efficiently by use of a Fresnel propagation algorithm; for example, a procedure based on the Fresnelet transform.[43] From now on, we restrict ourselves to optical setups with an off-axis geometry. An illustration of such an arrangement is given in Fig. 1. We shall discuss the requirements of our method in more detail in what follows, but we can already state that they are no more demanding than those required by digital-hologram-reconstruction algorithms mentioned earlier.[6,9] The main advantages of treating the two problems independently are as follows: (1) The approach provides more flexibility for the treatment of boundary conditions, (2) there are fewer parameters to set simultaneously (decoupling effect), and (3) the zero-order and twin-image terms are implicitly suppressed—they simply do not arise because the process does not involve simulating a wave's diffraction by the hologram.

### A. Complex-Wave-Retrieval Algorithm

Our methodology requires a precise description of the reference wave's phase in the form of a parametric mathematical model. The simplest example is that of a plane reference wave. We will, however, consider a more general model for the reference wave that is also suitable for the more sophisticated optical setups that are encountered in microscopy (Subsection 4.B). Both the measurement and the calibration may be done within a single acquisition frame provided that there are regions where the object wave is constant. These regions are used for the second purpose.

The key idea of our algorithm is that we consider the measured quantity $I(\mathbf{x})$, the intensity of the hologram, and $\theta(\mathbf{x})$, the phase of the reference wave in the hologram plane $R(\mathbf{x}) = A(\mathbf{x})\exp[i\theta(\mathbf{x})]$, to vary much more rapidly with $\mathbf{x}$ than the unknown quantities $\Psi(\mathbf{x}) \in \mathbb{C}$, the complex wave to retrieve, and $A(\mathbf{x}) \in \mathbb{R}_+^*$, the amplitude of the reference wave. Although modeling the reference wave's phase has proved to be very effective, doing the same for its amplitude $A(\mathbf{x})$ is much more problematic because of the latter's highly unpredictable changes observed when repeating the experiment over longer time periods (real-time imaging). For this reason we consider it an unknown.

We regard $\Psi(\mathbf{x})$ and $A(\mathbf{x})$ to be constant (lowest-order approximation) within the neighborhood of a given point of interest $\mathbf{x}$. We shall discuss the requirements for this condition in more detail in Subsection 3.C. With this hypothesis, determining the phase and amplitude of $\Psi(\mathbf{x})$ together with $A(\mathbf{x})$ is equivalent to solving at each location $\mathbf{x}$ the following set of $M$ nonlinear equations:

$$I(\mathbf{x} + \mathbf{x}_m) = |\Psi(\mathbf{x}) + A(\mathbf{x})\exp[i\theta(\mathbf{x} + \mathbf{x}_m)]|^2, \quad (3)$$

where the $\mathbf{x} + \mathbf{x}_m$ ($m = 1,\ldots,M$) are the positions of the $M$ pixels within the considered neighborhood of $\mathbf{x}$ (see Fig. 3). We further simplify the notation by assigning an index $m$ to a variable at position $\mathbf{x} + \mathbf{x}_m$:

$$\begin{aligned} I_m &= |\Psi + A\exp(i\theta_m)|^2 \\ &= |\Psi|^2 + A^2 + 2\Re(R_m^*\Psi). \end{aligned} \quad (4)$$

We propose to retrieve the unknown parameters by solving the above nonlinear set of equations in the least-squares sense, i.e.,

$$\underset{A\in\mathbb{R}_+^*,\Psi\in\mathbb{C}}{\arg\min} \sum_m w_m |I_m - [|\Psi|^2 + A^2 + 2\Re(R_m^*\Psi)]|^2. \quad (5)$$

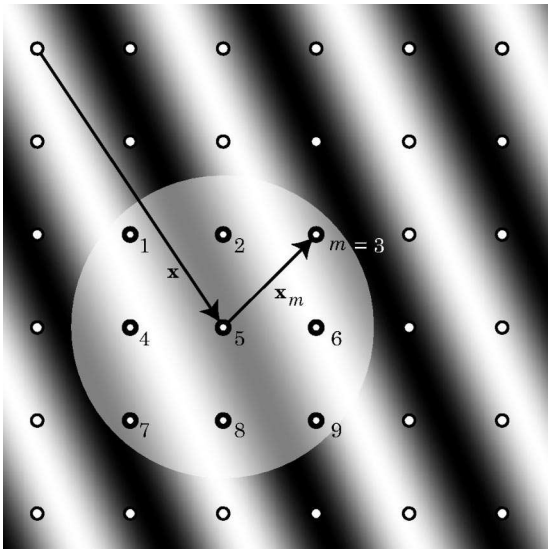


Fig. 3. Schematic hologram with a grid denoting the pixels' centers.

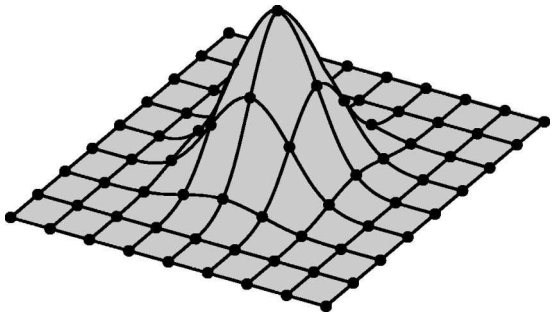

Fig. 4. Weights computed from the tensor product of two cubic B-splines and a window size $L = 9, M = 81$.

The nonnegative weights $w_m$, normalized for brevity such that $\Sigma_m w_m = 1$, ensure that the intensities that are far away from the point of interest account for less than those that are in its close vicinity, and allow us to be consistent with the hypothesis of a (nearly) constant $\Psi$ and amplitude $A$. In practice, we may use a simple indicator function or use weights that correspond to a smooth function, typically, a tensor product of B-splines of degree $n$ (e.g. $n = 3$). This is a separable, bell-shaped, Gaussian-like function that is nearly isotropic and has a finite support.[44] The weighting function can be written as

$$w(k, l) = \beta^n(k/s)\beta^n(l/s), \quad \text{with } s = (L - 1)/(n + 1), \tag{6}$$

and with $-(L - 1)/2 \leqslant k, l \leqslant (L - 1)/2$, where $L$ is the side-width of the window and where $\beta^n(x)$ denotes the central B-spline of degree $n$. An illustration of such a weighting window, for a B-spline of degree 3 (cubic B-spline) and a window of size $L^2 = 9^2 = M$ is given in Fig. 4. Other forms of windows may be used (Hanning, etc.) but the B-spline offers computational advantages in that there exist fast convolution algorithms (either waveletlike or based on iterated moving averages) with a complexity that does not depend on the window's size.[45]

We introduce the auxiliary variables $Z \in \mathbb{C}$ and $U \in \mathbb{R}_+$ defined by

$$\Psi = Z/A, \tag{7}$$

$$U = |Z|^2/A^2 + A^2. \tag{8}$$

If we now define the normalized reference wave $V_m \in \{z | z \in \mathbb{C}, |z| = 1\}$ by $V_m = R_m^*/A = \exp(-i\theta_m)$ and make the appropriate substitutions, the nonlinear problem (5) becomes

$$\arg\min_{U,Z} \sum_m w_m |I_m - U - 2\Re(V_m Z)|^2, \tag{9}$$

which may be solved by use of a linear method. The original variables $\Psi$ and $A$ are restored by the following (nonlinear) operations:

$$\Psi = Z/A, \tag{10}$$

$$A_\pm = +\left(\frac{U \pm \sqrt{U^2 - 4|Z|^2}}{2}\right)^{1/2}. \tag{11}$$

Note that these equations are well defined and consistent with the hypothesis $A \in \mathbb{R}_+^*$ since $U^2 \geqslant U^2 - 4|Z|^2 = (|Z|^2/A^2 - A^2)^2 \geqslant 0$. Furthermore, if we make the assumption that $|\Psi| < A$, i.e., that the object wave's amplitude is smaller than the reference wave's (which is experimentally advisable to yield highly contrasted fringes), there is only one possible solution for $A$. The fact that $A_- \leqslant A_+$ and the Viète relation for the roots of Eq. (8), $A_-^2 A_+^2 = |Z|^2$, imply that $A_-^2 \leqslant |Z| \leqslant A_+^2$. Thus, $A_- \leqslant |\Psi|$, which rules this case out as an acceptable solution.

The problem now boils down to finding values of $U$ and $Z$ that minimize problem (9). These must be the solutions of the normal equations:

$$\begin{aligned}
\sum_m w_m[I_m - U - 2\Re(V_m Z)] &= 0,\\
\sum_m w_m V_m[I_m - U - 2\Re(V_m Z)] &= 0,\\
\sum_m w_m V_m^*[I_m - U - 2\Re(V_m Z)] &= 0,
\end{aligned} \tag{12}$$

which are obtained by differentiating relation (9) with respect to $U$ (which is real), $Z$, and $Z^*$ (since $Z$ is complex). If we rearrange the terms, we get

$$\begin{aligned}
\sum_m w_m I_m &= U + 2\Re\left(Z \sum_m w_m V_m\right),\\
\sum_m w_m V_m I_m &= U \sum_m w_m V_m + Z \sum_m w_m V_m^2 + Z^*,\\
\sum_m w_m V_m^* I_m &= U \sum_m w_m V_m^* + Z + Z^* \sum_m w_m V_m^{*2}.
\end{aligned} \tag{13}$$

Finally, by setting $\alpha = \Sigma_m w_m V_m$ and $\beta = \Sigma_m w_m V_m^2$, we end up with a set of three linear equations to solve at every location:

$$\begin{bmatrix} 1 & \alpha & \alpha^* \\ \alpha & \beta & 1 \\ \alpha^* & 1 & \beta^* \end{bmatrix} \begin{pmatrix} U \\ Z \\ Z^* \end{pmatrix} = \begin{pmatrix} \sum_m w_m I_m \\ \sum_m w_m V_m I_m \\ \sum_m w_m V_m^* I_m \end{pmatrix}. \tag{14}$$

### B. Relation to Phase-Shifting Methods

We now briefly discuss similarities and differences between the proposed algorithm and related methods in interferometry. A central contribution of this paper is the change of variables of Eqs. (7) and (8) that transforms the nonlinear holography problem (5) into a form that may be solved with a linear algorithm. In particular, this step may be inverted uniquely, provided that the amplitude of

the reference wave is greater than that of the object wave. Once the change of variables is performed, the remaining mathematics is essentially the same as for phase-shifting algorithms: For specific choices of the reference wave (for example, a plane wave whose wave vector is horizontal) and a suitable one-dimensional weighting function, the linear part of our formulation is equivalent to previously proposed spatial phase-shifting algorithms.[17,24,25,27,31,32] One notable difference, however, is that we consider a two-dimensional spatial weighting function that allows for a much less restrictive choice of the reference wave's form (plane, parabolic, etc.) and orientation. This point is essential, since we need to deal with nonplanar reference waves in the case of microscopy. From a methodological point of view, problem (5) may also be seen as the spatial counterpart of the generalized (temporal) phase-shifting algorithm of Lai and Yatagai.[46] The analogy, however, is only formal since temporal phase shifting requires several interferograms to be measured. Our method does not include all *ad hoc* developments that have been proposed to optimize both temporal and spatial phase-shifting techniques,[47–50] except for the use of a weighting function, which has been found to have beneficial effects, albeit only in the one-dimensional case.[25,32]

We now have to investigate the appropriateness of the hypothesis of a constant phase in the vicinity of a point. This should enable us to set the right width of the weighting function (respectively the size of the neighborhood window) to consider.

### C. Sampling Considerations

The achievable resolution for holographic reconstruction is dictated mainly by two parameters: the spatial frequency of the reference wave (which influences the fringe spacing) and the sampling step of the digital acquisition system (CCD). The bounds for the minimal sampling steps can be deduced from the Shannon–Whittaker sampling theory. A thorough treatment of the sampling of digital holograms, including aspects related to the nonideal sampling by the CCD can be found in the literature.[51,52] There are also closely related discussions more specific to spatial[31,33,53] and temporal[50] phase shifting.

In the simplest case of an interference pattern measured in a plane and involving two plane waves, it is the angle that the reference wave's wave vector makes with the object wave's that specifies the spacing of the fringes. As a simple illustration we consider a plane reference wave given by

$$R(\mathbf{x}, z) = \exp[i(k_x x + k_y y + k_z z)], \tag{15}$$

whose interference with an object wave that is also a plane wave traveling perpendicularly to the acquisition device plane ($z = 0$), [i.e., $O(\mathbf{x}, z) = \exp(ikz)$, $k = (k_x + k_y + k_z)^{1/2}$] leads to a hologram of the form $I(\mathbf{x}, 0) = |\tilde{R}(\mathbf{x}, 0) + O(\mathbf{x}, 0)|^2 = 2 + 2\cos(k_x x + k_y y)$ (see Fig. 1). Let $T$ be the sampling step of the CCD. To achieve alias-free sampling of the interference pattern, one must have

$$k_x < \pi/T, \qquad k_y < \pi/T. \tag{16}$$

The maximal incidence angle $\theta_{\max}$ is defined when $k_x = k_y = \pi/T$, in which case

$$\sin(\theta_{\max}) = \frac{(k_x^2 + k_y^2)^{1/2}}{2\pi/\lambda} \leqslant \frac{\lambda}{\sqrt{2}T}. \tag{17}$$

Thus, decreasing the incidence angle ensures that the conditions for proper sampling are met. On the other hand, working with a low carrier frequency results in the expected resolution dropping. By experimenting with our algorithm, we have found that it works well if the window covers at least one period of the interference pattern. Depending on the form of the window (simple indicator function or more sophisticated weighting function), and in the case of the most rapidly varying fringes (critical sampling), the window should be at least $3 \times 3$ (respectively $7 \times 7$ in the case of a cubic B-spline). If the window size is taken too small, the linear system becomes unstable. On the other hand, taking too large a window has the consequences of increasing the computation time and decreasing the resolution of the retrieved phase.

### D. Variable-Window-Size Algorithm

So far, we have considered the weighting window to be the same for all positions at which we want to retrieve the phase, but this is not a requirement. If the reference is not a plane wave, or if the setup involves a lens, the fringes may have a different spacing depending on the $\mathbf{x}$ position; i.e., the local carrier frequency is variable. How does one choose an optimal window size in this case? In light of the discussion in Subsection 3.C, having the same window size for each position is obviously not optimal. Therefore, if we have a model for the reference wave, it makes good sense to tune and adapt the window size to it. Practically, we define a table containing the width of the weighting function to consider at each position. This table can typically be designed on the basis of the reference wave's model. Since the weighting function may have to be evaluated at every location, another reason for the choice of the B-spline is that, because it is a polynomial, its computation is reasonably fast.

### E. Computational Complexity

To compute the amplitude and phase for $N$ pixels, the presented algorithm requires $\mathcal{O}(\tilde{M}N) = \mathcal{O}(N)$ operations where $N$ is the number of retrieved phase points and $\tilde{M}$ is the average number of points covered by the window function at a given location.

The linear complexity of the present algorithm makes it attractive for large image sizes. Furthermore, the fact that we do not rely on a fast Fourier transform (FFT)—which usually imposes periodic boundary conditions—allows us to use more general boundary conditions or to work on arbitrary regions of interest without further increasing computation time. For instance, in our implementation, we used mirror boundary conditions that give a more natural way of extending the signal and have proved to be the extension of choice in many other applications. We therefore get rid of artifacts that are typically introduced by the periodic assumption. For the image sizes we used ($512 \times 512$), for a $7 \times 7$ window, and given that our implementation was designed to handle

variable window sizes (and therefore does not take advantage of possible separability and convolution properties[45]), the algorithm takes approximatively 0.51 s (respectively 1.6 s in the B-spline-weighted case) whereas an FFT-based filtering typically takes around 0.56 s (1.8 GHz PowerPC G5).

For the second part of our algorithm (i.e., propagation), which usually involves the computation of FFTs, the number of operations is $\mathcal{O}[N \log(N)]$.

## 4. RESULTS

### A. Phase-Retrieval Simulation

This first experiment consists of a simple interference simulation to test the first part of our algorithm (i.e., phase retrieval). The starting point is a test target (respectively a complex test wave) in the CCD plane. The modulus of the complex object wave is constant over the whole plane of interest while the phase follows a sinusoidal pattern depending on the polar angle. The phase varies between $-\pi/5$ and $\pi/5$. The thus-defined spoke target is shown in Fig. 5(b). To simulate the hologram, we create the interference by adding a complex reference wave. The latter wave is a plane wave that makes an angle of $\theta = 1.224°$ with the acquisition plane. It has a slowly varying, Gaussian-shaped amplitude that is typical of experimental illumination sources. The hologram is shown in Fig. 5(a). The sampling step was chosen to be $T = 10\ \mu$m (typical sampling step of commercial CCD cameras) and $\lambda = 632.8$ nm the wavelength (He-Ne laser). The resulting fringe's $k$-vector (that defines the fringe's frequency and orientation) is $\mathbf{K} = (1.5, 1.5)$, resulting in a fringe period $\Lambda = 2\pi/\|\mathbf{K}\| \approx 3$ pixels. We are now aiming at reconstructing the complex test wave from the simulated interference pattern. To reconstruct the image we used a window size of $7 \times 7$ pixels at every point with either a simple indicator function or a cubic B-spline with support of $7 \times 7$ pixels. The reconstructed phase is given in Figs. 5(c) and (d), respectively. As expected, the resolution is limited by the respective width of the weighting functions, which is smaller for the cubic B-spline than for the indicator function. This can be verified by examining the center of the reconstructed phase in Fig. 5(c), where the algorithm fails to reconstruct the high spatial frequency. In Fig. 5(d) we can see that much better results are achieved by using the cubic

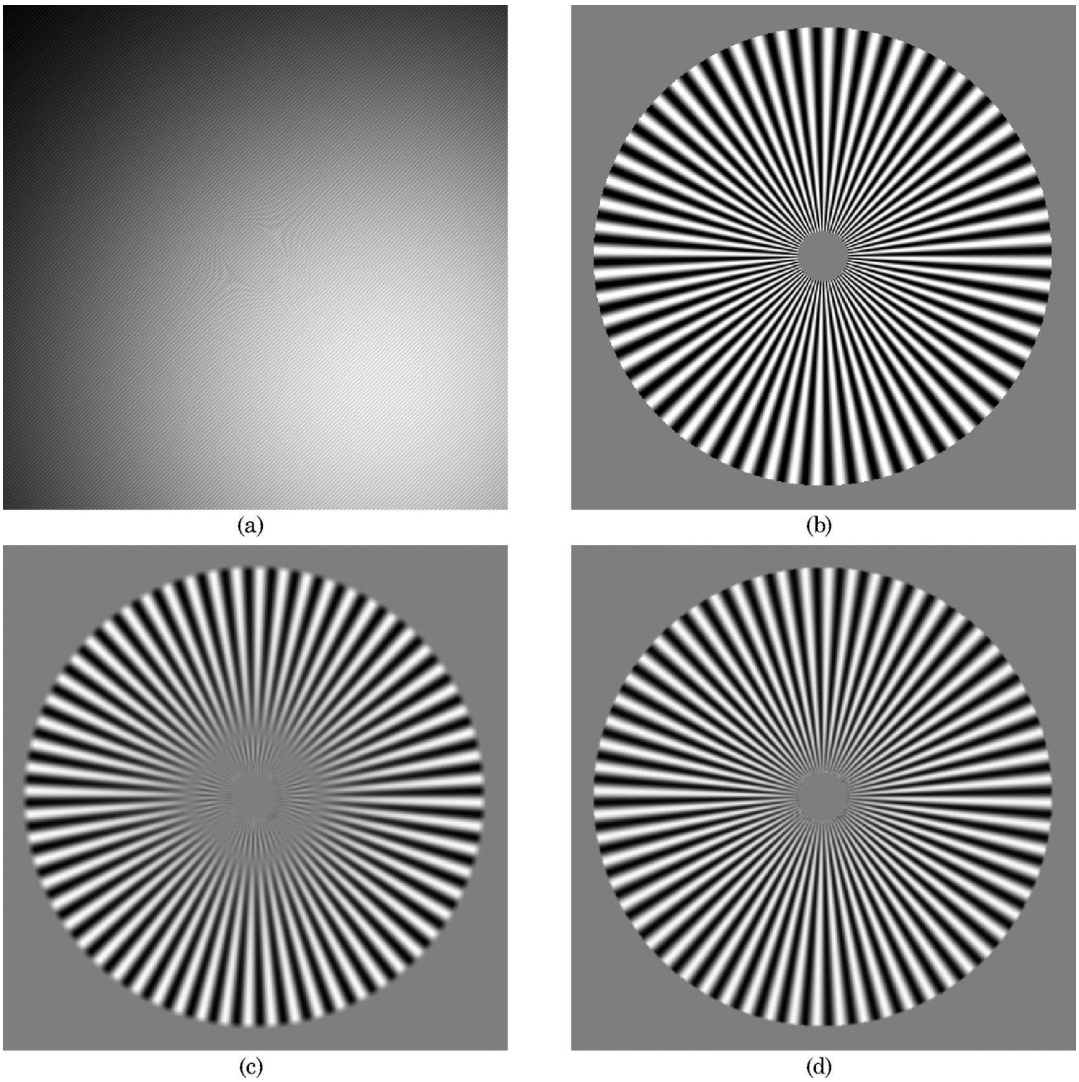


Fig. 5. (a) Simulated hologram; (b) original phase; (c) reconstructed phase; (d) reconstructed phase with weighted algorithm.

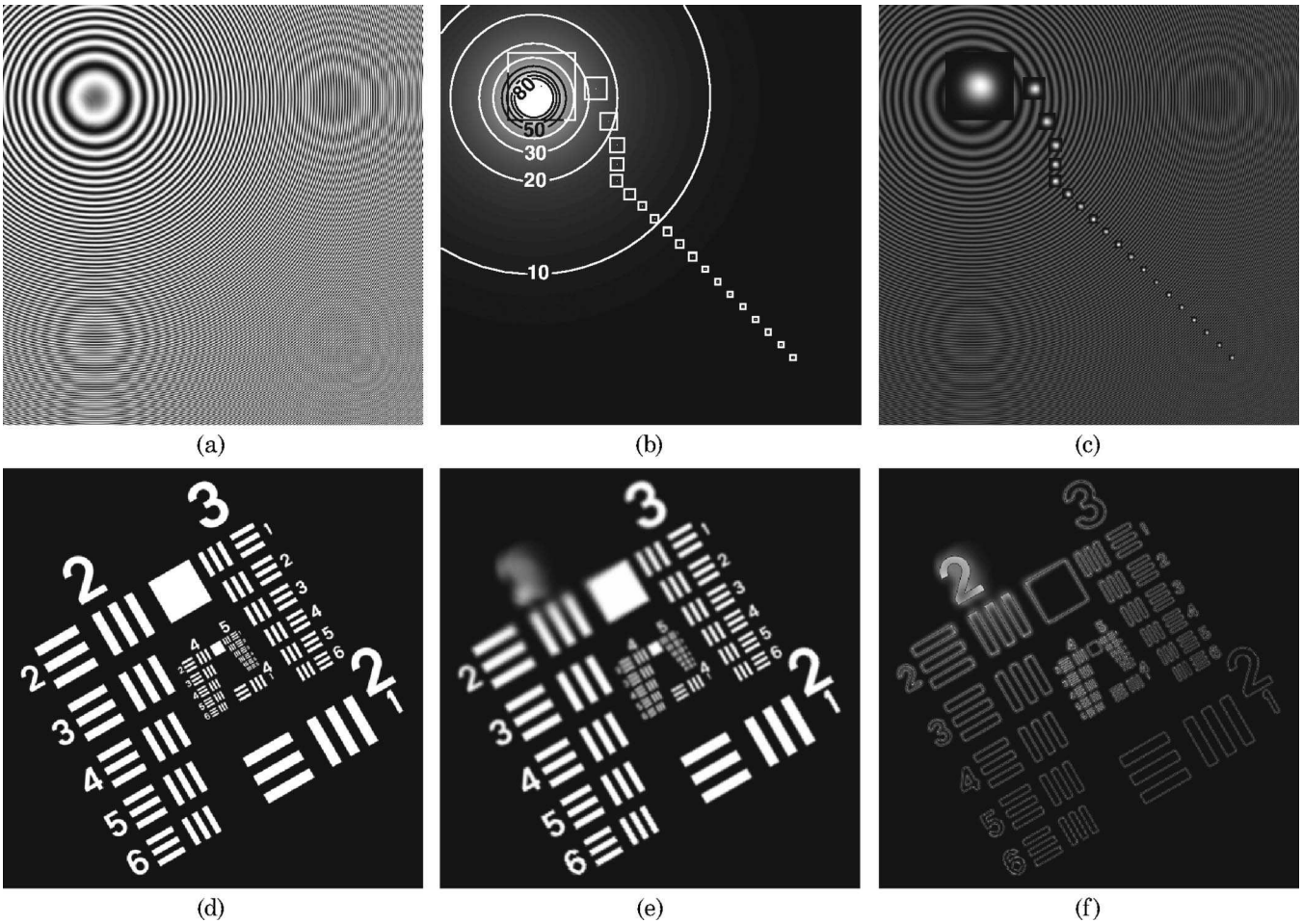


Fig. 6. (a) Simulated hologram; (b) Window width table visualization; maximum width, 80 pixels (white regions); minimum width, 7 pixels (black regions). (c) Weighted hologram. (d) Test wave's phase; in black regions it equals $\pi/30$, in bright regions $2\pi/30$. (e) Reconstructed phase. (f) Difference image: The gray scale covers the range $[0, \pi/30]$.

B-spline weighted algorithm. The support of the weighting functions and hence the computational complexity is, however, the same in both cases.

### B. Phase Retrieval for Setups Containing a Lens

We now consider the image of an object wave that has traveled through a lens:

$$\Psi_I(\mathbf{x}) = \Psi\left(\frac{\mathbf{x}}{\mathcal{M}}\right)\exp\left(i\pi\frac{\|\mathbf{x}\|^2}{\mathcal{M}\lambda f}\right), \tag{18}$$

where $\mathcal{M}$ is the magnification, $f$ the focal length of the lens, and $\lambda$ the wavelength. It is a magnified version of the object wave $\Psi(\mathbf{x})$ multiplied by a quadratic phase term. Note that, for the sake of simplicity, we have chosen to consider an infinite-aperture lens.

We considered a wavelength $\lambda = 632.8$ nm and $\pi/(\mathcal{M}\lambda f) = 1/3.001 \times 10^8\ \mathrm{m}^{-2}$. The reference wave makes an angle of $\theta = 0.6529°$ with the optical axis. We are interested in recovering $\Psi(\mathbf{x}/\mathcal{M})$ through the samples of this function. Thus we place a CCD in the image plane perpendicular to the optical axis. Its size is $500 \times 500$ pixels with a sampling step of $T = 10\ \mu$m. The measured hologram in the image plane is given by

$$I(\mathbf{x}) = \left|A(\mathbf{x})\exp[i(k_x x + k_y y)] + \Psi\left(\frac{\mathbf{x}}{\mathcal{M}}\right)\exp\left(i\pi\frac{\|\mathbf{x}\|^2}{\mathcal{M}\lambda f}\right)\right|^2, \tag{19}$$

with $A(\mathbf{x}) \in \mathbb{R}_+^*$. This expression can be written equivalently as

$$I(\mathbf{x}) = \left|A(\mathbf{x})\exp\left[i\left(\frac{1}{D}\|\mathbf{x} - \mathbf{x}_c\|^2 + C\right) + \Psi\left(\frac{\mathbf{x}}{\mathcal{M}}\right)\right]\right|^2, \tag{20}$$

where $\mathbf{x}_c = (x_c, y_c) \in \mathbb{R}^2$, $C \in \mathbb{R}$, and $D \in \mathbb{R}^*$, and it can be interpreted as the interference of the complex wave $\Psi$ with a nonplanar wave. It is thus sufficient to adjust the parameters $D$, $x_c$, and $y_c$ appropriately to specify our model. In real-world experiments, we determine them by a simple procedure based on a standard one-dimensional signal demodulation technique. We select a horizontal line segment from the measured hologram where the object wave's phase is assumed to be constant and where the measured intensity is therefore given by $I(x) = a(x) + b(x)\cos[(x - x_c)^2/D_x + C']$. The offset $a(x)$ is removed by low-pass filtering. We then compute

the corresponding analytic signal using Matlab's HILBERT function.[54] The principal value of this signal is then unwrapped and fitted to a quadratic function yielding parameters $D_x$ and $x_c$. The procedure is repeated with a vertical line segment, yielding $D_y$ and $y_c$. For an astigmatism-free system $D_x = D_y = D$. Further adjustments that may be necessary are currently done manually. In case the parameters are not set correctly, our algorithm is still operational but the phase difference between the true and the modeled wave will be added to the returned phase. One may consider other parametric models for the reference wave[55] but the procedure described above is limited to those that are separable.

Our aim in this experiment was to reconstruct a "gold standard" wave front $\Psi(\mathbf{x}/\mathcal{M})$ that has a constant intensity and whose phase is either $\pi/30$ or $2\pi/30$ according to a United States Air Force 1951 test target pattern. The phase is reproduced in Fig. 6(d). The reference wave is a plane wave with a constant amplitude that we set 10 times higher than that of the object wave. The resulting hologram is given in Fig. 6(a). The hologram consists of concentric fringes whose spacing varies in a quadratic way. We applied the variable-window-size algorithm to recover the phase. We have defined a matrix containing for each pixel of the measured hologram the size of the window to consider in order to retrieve the phase at that particular point. The maximum window size was set to $\bar{L} = 80$ and the minimum to $\underline{L} = 7$ as illustrated in Fig. 6(b). The side-width $L$ of the window was defined as the local fringe period computed by using

$$L(x, y) = \begin{cases} \underline{L} & \text{if } K(x, y) > 2\pi/\underline{L} \\ \bar{L} & \text{if } K(x, y) < 2\pi/\bar{L}, \\ 2\pi/K(x, y) & \text{otherwise} \end{cases}$$

where $K(x, y) = \|\mathbf{K}(x, y)\| = \|(x/D + 2x_c/D, y/D + 2y_c/D)\|$ is the norm of the local fringe $k$-vector. Note that we kept the minimum to be 7 pixels even though the minimal local fringe period is $\Lambda \approx 3.5$ pixels. This ensured a stable algorithm. The windows themselves were taken to be B-splines of degree 3. In Fig. 6(c) we show the hologram for which the pixels in the neighborhood of several locations [square regions shown in Fig. 6(b)] have been weighted. The reconstructed phase is given in Fig. 6(e) and the difference image between the original and reconstructed image in Fig. 6(f). We can clearly see in Fig. 6(e) that the whole field of view is available and not perturbed by any interference terms whatsoever. As we have a variable window size, the resolution is variable too. At places where the window size is large the resolution is low: The digit 2 is not recognizable anymore. On the other hand, in regions where the spatial modulation frequency is too high, the under-sampling introduces artifacts (see Group 2, element 6). Spatial frequencies of up to 32 lines per mm are resolved (Group 5, element 1) with a sampling step of the interferogram $T = 10$ $\mu$m (50 lines per mm). When we use windows with a variable width, our reconstruction time is much more substantial (6.9 s on a 1.8 GHz PowerPC G5 computer) but we believe that it can be decreased by using waveletlike techniques.

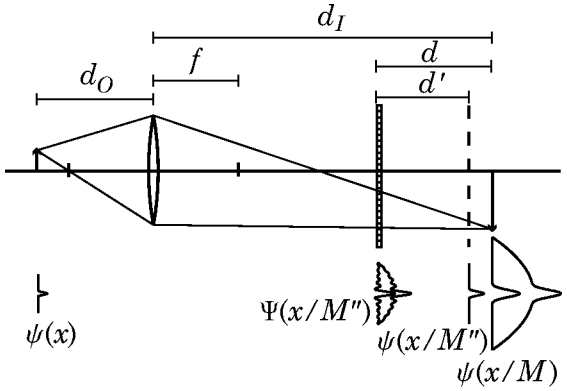


Fig. 7. Phase aberration compensation and phase retrieval in the case of a microscopy setup. The CCD is not in the image plane.

### C. Experimental Digital Holographic Microscopy

In this experiment, we have applied our method to measurements that were performed by Cuche *et al.* with a digital holographic microscope setup.[8] The specimen (a USAF test target) is illuminated by a plane wave and the reflected light is collected by a microscope objective. The CCD camera records the interference of this object wave with a reference plane wave in an off-axis geometry. The CCD is in front of the image plane at a distance $d$ (Fig. 7). This implies that the complex wave that we will recover in the camera plane will be out of focus and require a propagation step to yield the desired image of the specimen, as is the case for lensless Fresnel holography.

In the vicinity of the object, we denote the reflected wavefront by $\psi(\mathbf{x})$. The complex wave in the image plane is a magnified version of the object multiplied by a quadratic phase

$$\psi_I(\mathbf{x}) = \psi\left(\frac{\mathbf{x}}{\mathcal{M}}\right)\exp\left(i\pi\frac{\|\mathbf{x}\|^2}{\mathcal{M}\lambda f}\right). \tag{21}$$

Again, we have considered a lens of infinite aperture as our starting model. The expression of the complex wave in the CCD plane can be given by backpropagating Eq. (21) by use of the definition of the Fresnel transform

$$\tilde{\psi}(\mathbf{x}) = \mathcal{R}_{(-d)}\left\{\psi\left(\frac{\cdot}{\mathcal{M}}\right)\exp\left(i\pi\frac{\|\cdot\|^2}{\mathcal{M}\lambda f}\right)\right\}(\mathbf{x})$$

$$= \Psi(\mathbf{x})\exp\left(i\frac{\|\mathbf{x}\|^2}{D}\right), \tag{22}$$

$$\Psi(\mathbf{x}) = \mathcal{R}_{(-d')}\left\{\psi\left(\frac{\cdot}{\mathcal{M}''}\right)\right\}(\mathbf{x}), \tag{23}$$

where

$$d' = d - \frac{d^2}{\mathcal{M}f}, \tag{24}$$

$$D = \frac{\lambda(\mathcal{M}f - d)}{\pi}, \tag{25}$$

$$\mathcal{M}'' = \frac{\mathcal{M}f - d}{f}. \tag{26}$$

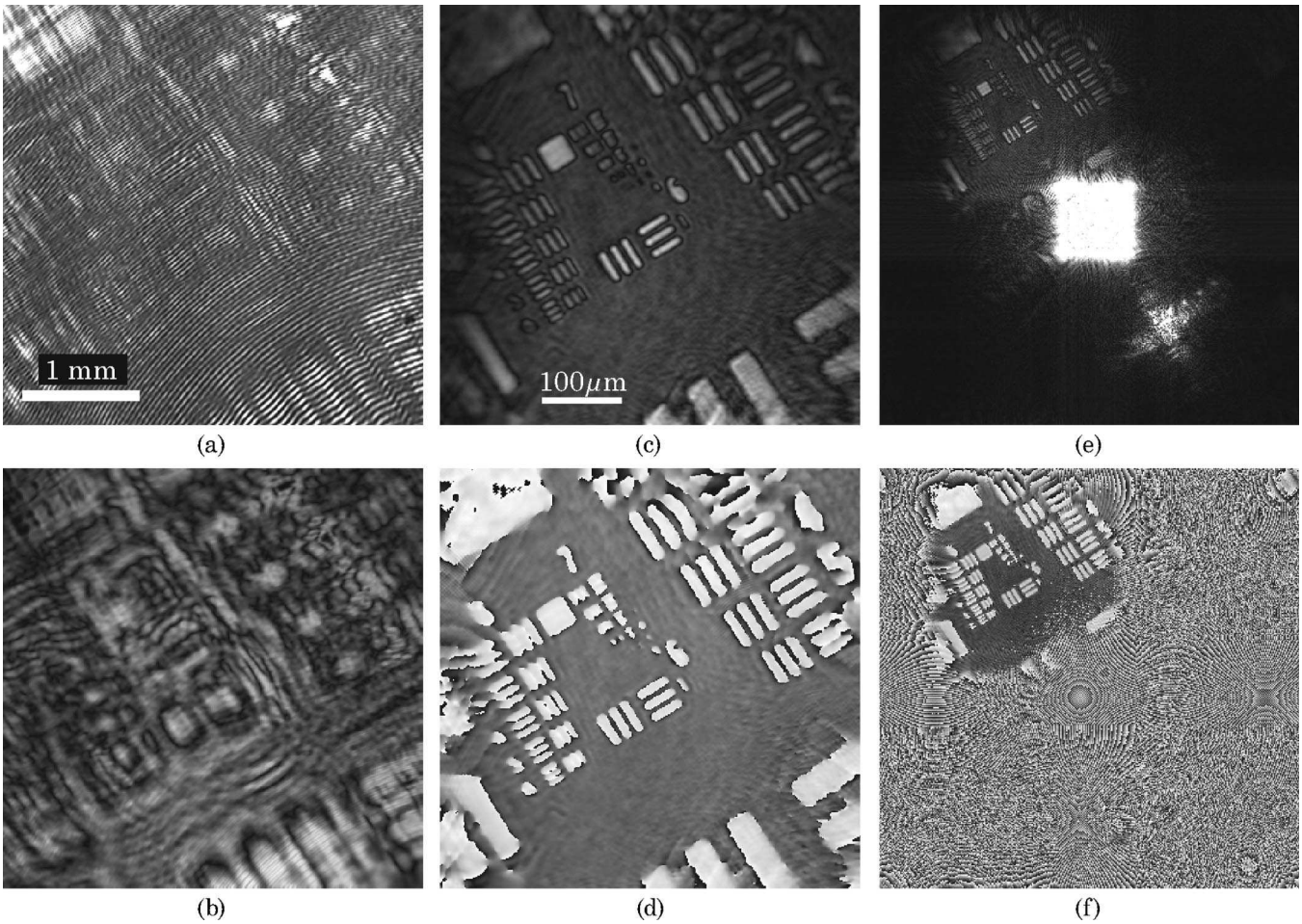


Fig. 8. (a) Measured hologram; data courtesy of E. Cuche, École Polytechnique Fédérale de Lausanne, and P. Marquet, Université de Lausanne. (b) Reconstructed amplitude in the CCD plane. Reconstructed amplitude (c) and phase (d) with adjusted focus. Reconstructed amplitude (e) and phase (f) with an alternative technique.[8] All images are 360 × 360 pixels. The phase images' gray scale covers the range $(-\pi, \pi]$.

Equation (23) is a consequence of the Fresnel transform's properties.[56]

Assuming a plane reference wave, the procedure to recover $\Psi(\mathbf{x})$ from $I(\mathbf{x})$ is the one described in subsection 4.B, as the intensity measured by the CCD camera,

$$I(\mathbf{x}) = \left| A(\mathbf{x})\exp[i(k_x x + k_y y)] + \Psi(\mathbf{x})\exp\left( i\frac{1}{D}\|\mathbf{x}\|^2 \right) \right|^2, \tag{27}$$

has the same form as Eq. (19). Once $\Psi$ is recovered, the last step is its propagation up to the proper distance $d'$ to yield

$$\psi\left( \frac{\mathbf{x}}{\mathcal{M}''} \right) = \mathcal{R}_{(d')}\{\Psi\}(\mathbf{x}). \tag{28}$$

The parameters $d'$, $D$, and $k_x$, $k_y$ are adjusted numerically. In general, these quantities cannot be obtained experimentally with the required level of accuracy. However, in the case of a reflection setup for holographic microscopy, the phase distribution over a flat background area can be assumed to be constant and the reconstruction parameters adjusted to match this constraint. The fringe period $\Lambda$ varies between 4 and 15 pixels and the window size was set accordingly in our model with the minimum window size set to 7.

In Fig. 8(a) we show the hologram, i.e., $I(\mathbf{x})$. Figure 8(b) shows the reconstructed amplitude (phase is not shown) in the CCD plane. The amplitude and phase after applying the Fresnel transform to reach proper focus is given in Figs. 8(c) and (d). In Figs. 8(e) and (f), we show the reconstructions from the same data set obtained by using an alternative (linear) technique without any zero-order or twin-image removal scheme. In the latter approach, the quadratic-phase exponential induced by the objective is compensated numerically by multiplying the diffracted wave by a numerical phase mask.[8] The zero order is the bright square portion that masks the center of the image. The image itself is on the upper left while the twin image is located in the lower right. Only the image is in focus.

Remarkably, the reconstructions using the new approach are not perturbed by the zero-order or twin-image terms. Furthermore, the field of view is also larger.

## 5. CONCLUSION

We have presented a new approach for reconstructing wavefronts from a single digital hologram. The method

works for holograms acquired in an off-axis geometry. This experimental arrangement allows for the three-dimensional information to be encoded in one single interferogram, and therefore permits work at high acquisition rates, which is required when investigating an object that undergoes fast changes. Since very low intensities can be used, it is particularly indicated for the study of biological samples. The working hypothesis for our method is that the object wave has slower local variations (low-pass spectrum) than the reference wave, a condition that tends to be met in a wide variety of applications. As the illumination amplitude often has a Gaussian amplitude, its spatial variation is also sufficiently slow to fulfill the requirement for the reference wave's amplitude.

The advantages of our method over others that rely on FFTs for the phase retrieval are manifold. Our algorithm has a linear complexity in the number of points where the amplitude and phase are to be retrieved. Since it is a local algorithm, it offers much flexibility for working in the spatial domain: variable window size, boundary conditions, etc. Also, a local defect of quality in the measurement will not deteriorate the quality of the whole reconstruction. The problem we solve is nonlinear. We take advantage of the extra information provided by the zero order and the twin image, which is discarded in linear reconstruction methods. Since the core of our algorithm is linear, the computation time remains reasonable.

To achieve reconstruction at optimal resolution in the case of a nonplanar reference wave or when microscope objectives are used, we advise use of the variable-window-size algorithm. It ensures a stable reconstruction over the whole field of view. In such circumstances, the size of the window should be adapted to the modulation frequency. To obtain a stable reconstruction, we found that the equivalent window size should be at least 3 pixels. We also recommend the use of a tensor product of cubic B-spline window functions, which are separable as well as nearly isotropic. Their finite support makes them highly suitable for a numerical implementation.

## ACKNOWLEDGMENTS

The authors thank E. Cuche and Ch. Depeursinge, École Polytechnique Fédérale de Lausanne, and P. Marquet, Université de Lausanne, for providing experimental data and for many fruitful discussions. They also thank one of the reviewers for providing many relevant pointers to the phase-shifting literature.

Corresponding author M. Liebling's e-mail address is michael.liebling@a3.epfl.ch.

## REFERENCES

1. D. Gabor, "A new microscopic principle," Nature (London) **161**, 777–778 (1948).
2. J. W. Goodman and R. W. Lawrence, "Digital image formation from electronically detected holograms," Appl. Phys. Lett. **11**, 77–79 (1967).
3. M. A. Kronrod, N. S. Merzlyakov, and L. Yaroslavskii, "Reconstruction of a hologram with a computer," Sov. Phys. Tech. Phys. **17**, 333–334 (1972).
4. L. P. Yaroslavskii and N. S. Merzlyakov, *Methods of Digital Holography* (Consultants Bureau, New York, 1980).
5. U. Schnars and W. Jüptner, "Direct recording of holograms by a CCD target and numerical reconstruction," Appl. Opt. **33**, 179–181 (1994).
6. E. Cuche, F. Bevilacqua, and Ch. Depeursinge, "Digital holography for quantitative phase-contrast imaging," Opt. Lett. **24**, 291–293 (1999).
7. T. Colomb, P. Dahlgren, D. Beghuin, E. Cuche, P. Marquet, and Ch. Depeursinge, "Polarization imaging by use of digital holography," Appl. Opt. **41**, 27–37 (2002).
8. E. Cuche, P. Marquet, and Ch. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," Appl. Opt. **38**, 6994–7001 (1999).
9. T. M. Kreis, M. Adams, and W. P. O. Jüptner, "Methods of digital holography: A comparison," in *Optical Inspection and Micromeasurements II*, C. Gorecki, ed., Proc. SPIE **3098**, 224–233 (1997).
10. E. Cuche, P. Marquet, and Ch. Depeursinge, "Spatial filtering for zero-order and twin-image elimination in digital off-axis holography," Appl. Opt. **39**, 4070–4075 (2000).
11. T. M. Kreis and W. P. O. Jüptner, "Suppression of the dc term in digital holography," Opt. Eng. **36**, 2357–2360 (1997).
12. C. Liu, Y. Li, X. Cheng, Z. Liu, F. Bo, and J. Zhu, "Elimination of zero-order diffraction in digital holography," Opt. Eng. **41**, 2434–2437 (2002).
13. T. Kreis, "Computer-aided evaluation of holographic interferograms," in *Holographic Interferometry: Principles and Methods*, P. K. Rastogi, ed. (Springer, Heidelberg, Germany, 1994), pp. 151–212.
14. T. Kreis, "Digital holographic interference-phase measurement using the Fourier-transform method," J. Opt. Soc. Am. A **3**, 847–855 (1986).
15. V. I. Vlad and D. Malacara, "Direct spatial reconstruction of optical phase from phase-modulated images," in *Progress in Optics, Vol. XXXIII*, E. Wolf, ed. (Elsevier, Amsterdam, 1994), pp. 261–317.
16. M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," J. Opt. Soc. Am. **72**, 156–160 (1982).
17. W. W. Macy, Jr., "Two-dimensional fringe-pattern analysis," Appl. Opt. **22**, 3898–3901 (1983).
18. K. A. Nugent, "Interferogram analysis using an accurate fully automatic algorithm," Appl. Opt. **24**, 3101–3105 (1985).
19. D. J. Bone, H.-A. Bachor, and R. J. Sandeman, "Fringe-pattern analysis using a 2-D Fourier transform," Appl. Opt. **25**, 1653–1660 (1986).
20. C. Roddier and F. Roddier, "Interferogram analysis using Fourier transform techniques," Appl. Opt. **26**, 1668–1673 (1987).
21. J. E. Greivenkamp and J. H. Bruning, "Phase-shifting interferometry," in *Optical Shop Testing*, 2nd ed., D. Malacara, ed. (Wiley, New York, 1992), pp. 501–598.
22. T. Kreis, *Holographic Interferometry*, Vol. 1 of Optical Metrology Series (Akademie Verlag, Berlin, 1996).
23. Y. Ichioka and M. Inuiya, "Direct phase detecting system," Appl. Opt. **11**, 1507–1514 (1972).
24. L. Mertz, "Real-time fringe-pattern analysis," Appl. Opt. **22**, 1535–1539 (1983).
25. K. H. Womack, "Interferometric phase measurement using spatial synchronous detection," Opt. Eng. **23**, 391–395 (1984).
26. P. L. Ransom and J. Kokal, "Interferogram analysis by a modified sinusoid fitting technique," Appl. Opt. **25**, 4199–4204 (1986).
27. M. Kujawińska and J. Wójciak, "Spatial phase-shifting techniques of fringe pattern analysis in photomechanics," in *Second International Conference on Photomechanics and Speckle Metrology: Moiré Techniques, Holographic Interferometry, Optical NDT, and Applications to Fluid Mechanics, Parts 1 and 2*, F.-P. Chiang, ed., Proc. SPIE **1554B**, 503–513 (1991).

28. M. Kujawińska, "Spatial phase measurement methods," in *Interferogram Analysis: Digital Fringe Pattern Measurement Techniques*, D. W. Robinson and G. T. Reid, eds. (Bristol Institute of Physics, Bristol, UK, 1993), pp. 141–193.
29. M. Servin and F. J. Cuevas, "A novel technique for spatial phase-shifting interferometry," J. Mod. Opt. **42**, 1853–1862 (1995).
30. P. H. Chan, P. J. Bryanston-Cross, and S. C. Parker, "Fringe-pattern analysis using a spatial phase-stepping method with automatic phase unwrapping," Meas. Sci. Technol. **6**, 1250–1259 (1995).
31. K. Creath and J. Schmit, "N-point spatial phase-measurement techniques for non-destructive testing," Opt. Lasers Eng. **24**, 365–379 (1996).
32. J. Schmit and K. Creath, "Window function influence on phase error in phase-shifting algorithms," Appl. Opt. **35**, 5642–5649 (1996).
33. T. Bothe, J. Burke, and H. Helmers, "Spatial phase shifting in electronic speckle pattern interferometry: minimization of phase reconstruction errors," Appl. Opt. **36**, 5310–5316 (1997).
34. R. Windecker and H. J. Tiziani, "Semispatial, robust, and accurate phase evaluation algorithm," Appl. Opt. **34**, 7321–7326 (1995).
35. A. J. Moore and F. Mendoza-Santoyo, "Phase demodulation in the space domain without a fringe carrier," Opt. Lasers Eng. **23**, 319–330 (1995).
36. M. Servin, J. L. Marroquin, and F. J. Cuevas, "Demodulation of a single interferogram by use of a two-dimensional regularized phase-tracking technique," Appl. Opt. **36**, 4540–4548 (1997).
37. J. L. Marroquin, M. Servin, and R. Rodriguez-Vera, "Adaptive quadrature filters and the recovery of phase from fringe pattern images," J. Opt. Soc. Am. A **14**, 1742–1753 (1997).
38. J. L. Marroquin, R. Rodriguez-Vera, and M. Servin, "Local phase from local orientation by solution of a sequence of linear systems," J. Opt. Soc. Am. A **15**, 1536–1544 (1998).
39. E. Yu and S. S. Cha, "Two-dimensional regression for interferometric phase extraction," Appl. Opt. **37**, 1370–1376 (1998).
40. K. G. Larkin, D. J. Bone, and M. A. Oldfield, "Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform," J. Opt. Soc. Am. A **18**, 1862–1870 (2001).
41. S. Sotthivirat and J. A. Fessler, "Relaxed ordered subsets algorithm for image restoration of confocal microscopy," in *Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI'02)* (Institute of Electrical and Electronics Engineers, New York, 2002), Vol. 3, pp. 1051–1054.
42. S. Sotthivirat and J. A. Fessler, "Penalized-likelihood image reconstruction for digital holography," J. Opt. Soc. Am. A (to be published).
43. M. Liebling, T. Blu, and M. Unser, "Fresnelets: new multiresolution wavelet bases for digital holography," IEEE Trans. Image Process. **12**, 29–43 (2003).
44. M. Unser, "Splines: a perfect fit for signal and image processing," IEEE Signal Process. Mag. **16**, 22–38 (1999).
45. M. Unser, A. Aldroubi, and S. Schiff, "Fast implementation of the continuous wavelet transform with integer scales," IEEE Trans. Signal Process. **42**, 3519–3523 (1994).
46. G. Lai and T. Yatagai, "Generalized phase-shifting interferometry," J. Opt. Soc. Am. A **8**, 822–827 (1991).
47. M. Pirga and M. Kujawińska, "Two directional, spatial-carrier, phase-shifting method for analysis of crossed closed fringe patterns," Opt. Eng. **34**, 2459–2466 (1995).
48. C. Rathjen, "Statistical properties of phase-shift algorithms," J. Opt. Soc. Am. A **12**, 1997–2008 (1995).
49. Y. Surrel, "Design of algorithms for phase measurements by the use of phase stepping," Appl. Opt. **35**, 51–60 (1996).
50. K. Hibino, B. F. Oreb, D. I. Farrant, and K. G. Larkin, "Phase-shifting algorithms for nonlinear and spatially nonuniform phase shifts," J. Opt. Soc. Am. A **14**, 918–930 (1997).
51. L. Onural, "Sampling of the diffraction field," Appl. Opt. **39**, 5929–5935 (2000).
52. T. Kreis, "Frequency analysis of digital holography," Opt. Eng. **41**, 771–778 (2002).
53. G. Paez and M. Strojnik, "Phase reconstruction from undersampled intensity patterns," J. Opt. Soc. Am. A **17**, 46–52 (2000).
54. *Signal Processing Toolbox User's Guide for Use with Matlab* (The MathWorks, Inc., Natick, Mass., 2002).
55. A. Marian, E. Cuche, and Ch. Depeursinge, "Point spread function model for microscopic image deconvolution in digital holographic microscopy," in *Novel Optical Instrumentation for Biomedical Applications*, A.-C. Boccara, ed., Proc. SPIE **5143**, 202–209 (2003).
56. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996).