Principles and techniques of digital holographic microscopy

Myung K. Kim
Principles and techniques of digital holographic microscopy

Myung K. Kim
University of South Florida, Department of Physics, 4202 E. Fowler Avenue, Tampa, Florida 33620
mkkim@cas.usf.edu

Abstract. Digital holography is an emerging field of new paradigm in general imaging applications. We present a review of a subset of the research and development activities in digital holography, with emphasis on microscopy techniques and applications. First, the basic results from the general theory of holography, based on the scalar diffraction theory, are summarized, and a general description of the digital holographic microscopy process is given, including quantitative phase microscopy. Several numerical diffraction methods are described and compared, and a number of representative configurations used in digital holography are described, including off-axis Fresnel, Fourier, image plane, in-line, Gabor, and phase-shifting digital holographies. Then we survey numerical techniques that give rise to unique capabilities of digital holography, including suppression of dc and twin image terms, pixel resolution control, optical phase unwrapping, aberration compensation, and others. A survey is also given of representative application areas, including biomedical microscopy, particle field holography, micrometrology, and holographic tomography, as well as some of the special techniques, such as holography of total internal reflection, optical scanning holography, digital interference holography, and heterodyne holography. The review is intended for students and new researchers interested in developing new techniques and exploring new applications of digital holography. © 2010 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/6.0000006]

Keywords: digital holography; phase microscopy; three-dimensional microscopy; numerical diffraction; biomedical imaging.

Paper SR090109 received Dec. 5, 2009; accepted for publication Mar. 17, 2010; published online May 14, 2010.

1 Introduction

Digital holography (DH) is an emerging technology of new paradigm in general imaging applications. By replacing the photochemical procedures of conventional holography with electronic imaging, a door opens to a wide range of new capabilities. Although many of the remarkable properties of holography have been known for decades, their practical applications have been constrained because of the cumbersome procedures and stringent requirements on equipment. A real-time process is not feasible, except for photorefractives and other special materials and effects. In digital holography, the holographic interference pattern is optically generated by superposition of object and reference beams, which is digitally sampled by a charge-coupled device (CCD) camera and transferred to a computer as an array of numbers. The propagation of optical fields is completely and accurately described by diffraction theory, which allows numerical reconstruction of the image as an array of complex numbers representing the amplitude and phase of the optical field. Digital holography offers a number of significant advantages, such as the ability to acquire holograms rapidly, availability of complete amplitude and phase information of the optical field, and versatility of the interferometric and image processing techniques. Indeed, digital holography by numerical diffraction of optical fields allows imaging and image processing techniques that are difficult or not feasible in real-space holography.

1946-3251/2010/$25.00 © 2010 SPIE
Holography was invented in 1948 by Dennis Gabor in an effort to improve the resolution of the electron microscope, where the correction of electron lens aberrations posed increasing technical difficulty. Instead of attempting to perfect the electron imaging lens, Gabor dispensed it altogether and realized that the diffraction pattern of the electron beam contained complete information regarding the amplitude and phase of the electron wave. The record of the electron wave diffraction is then used to optically synthesize the object field. This allowed the use of the optics of visible light for image formation, and was much easier and more developed compared to electron optics. He named the new imaging principle holography, for its ability to record the whole optical field [1–3].

The holography principle was immediately applied to recording and imaging by visible light [4]. But it had to wait for two critical inventions before its full potential was to be realized. One was the powerful coherent source of light in lasers to provide high quality interference contrast. The other, due to Leith and Upatnieks [5–7], was off-axis illumination with a separate reference wave, thus eliminating the problem of the zero-order and twin images of the Gabor’s on-axis configuration. Soon many new techniques and applications of holography began to develop. Holography is now a mature field, and an excellent survey is given, for example, in Ref. 8. It was also realized early on that the use of nonplanar structures of the reference beam can lead to various manipulations of the holographic output, ranging from image magnification to more complex image processing, such as encryption, pattern recognition, associative memory, and neural networks [9]. Instead of photographic plates or films, real-time holography is possible with photorefractives and other nonlinear optics materials. In fact, we now understand much of nonlinear optics as generalizations of the holography principle, including phase conjugation, four-wave mixing, spectral hole burning, and photon echo [10]. The holography process is being developed for other regions of the electromagnetic spectrum, especially x-ray holography, with its prospect of atomic resolution [11], as well as for microholography of living organisms [12].

For many application areas, real-time operation is of critical importance but is difficult with conventional holography. Even photorefractives and other nonlinear optics systems require substantial equipment and technical care to implement them, and have seen only limited practical applications. Digital holography replaces physical and chemical recording processes with electronic ones, and the optical reconstruction process with numerical computation. The propagation of optical fields is completely and accurately described by diffraction theory, and in 1967, Goodman and Lawrence demonstrated the feasibility of numerical reconstruction of an image from a Fourier hologram detected by a vidicon camera [13]. Schnars and Jueptner, in 1994, were the first to use a CCD camera directly connected to a computer as the input, and compute the image in a Fresnel holography setup [14]. In what is now called digital holography (DH), holographic interference is produced by optical processes in real space, while reconstruction is by numerical computation. Conversely, in computer-generated holography (CGH), the hologram can be produced by numerical computation inside a computer, followed by printing or other outputs to real space [15]. Reconstruction is then carried out by optical means. CGH has many interesting properties and applications, such as the ability to arbitrarily prescribe desired amplitude and phase properties of the output optical field starting from fictitious objects. CGH is not a subject of this review [16].

By direct electronic recording of holographic interference, and because of the increasing speed of holographic computation, real-time holographic imaging is now possible, and more importantly, the complete and accurate representation of the optical field as an array of complex numbers allows many imaging and processing capabilities that are difficult or infeasible in real-space holography [17]. Various useful and special techniques have been developed to enhance the capabilities and extend the range of applications. In digital holographic microscopy, a single hologram is used to numerically focus on the holographic image at any distance [18,19]. Direct access to the phase information leads to quantitative phase microscopy with nanometer sensitivity of transparent or reflective phase objects, [20,21] and allows further manipulations such as aberration correction [22]. Multiwavelength optical phase unwrapping is a fast and robust method for removing 2π-discontinuities compared to software algorithm-based methods [23].
A significant constraint of digital holography is the pixel count and resolution of the imaging devices. Suppression of the zero-order and twin images by phase-shifting digital holography allows efficient use of the pixel array [24]. Digital Gabor holography, without separate reference beams, is useful for particle imaging applications by providing 4-D space-time records of particle fields [25]. Digital holography naturally evolved from the effort to utilize electronic imaging in interferometry, such as in electronic speckle pattern interferometry (ESPI) [26]. Metrology of deformations and vibrations is a major application area of digital holography [27]. Optical processing, such as pattern recognition and encryption, by digital holography also offers new capabilities [28].

Basic principles of diffraction and general holography are outlined in Sec. 2, and a general description of digital holographic microscopy (DHM) is given in Sec. 3, with an emphasis on the quantitative phase microscopy by DHM. Methods of numerical calculation of diffraction are described and compared in Sec. 4, and a number of main types of interferometer configurations used in digital holography experiments are given in Sec. 5. There are many numerical techniques that lead to the unique and powerful capabilities of digital holography, described in Sec. 6. Then in Sec. 7, a survey is given of the application areas of DHM, as well as special techniques that expand the capabilities and applications of digital holography. This review has an emphasis on microscopy applications of digital holography, and therefore omits some major areas of digital holography development, such as metrology of macroscopic systems and image processing of holographic data [17].

2 Basic Theory of Holography

2.1 Scalar Diffraction Theory

First, some of the main results of scalar diffraction theory are recalled and applied to the description of basic holographic image formation. We start by writing down the Fresnel-Kirchoff diffraction formula for the general problem of diffraction from a 2-D aperture depicted in Fig. 1 [29].

2.1.1 Huygens convolution

Given the optical field $E_0(x_0, y_0)$ over the input plane $\Sigma_0$ at $z = 0$, the field $E(x, y)$ over the output plane $\Sigma$ at $z$ is, with the wavelength $\lambda = 2\pi/k$,

$$E(x, y; z) = -\frac{i k}{2 \pi z} \int \int d x_0 d y_0 E_0(x_0, y_0) \exp \left[ i k \sqrt{(x - x_0)^2 + (y - y_0)^2 + z^2} \right].$$

(1)

This is a convolution integral:

$$E(x, y; z) = E_0 \ast S_H,$$

(2)
where the point spread function (PSF) is

\[
S_H(x, y; z) = -\frac{ik}{2\pi z} \exp\left[i k \sqrt{x^2 + y^2 + z^2}\right],
\]

representing the Huygens spherical wavelet.

### 2.1.2 Fresnel transform

For paraxial approximation, valid for \(z^3 \gg \frac{k^2}{\lambda} \left( (x - x_0)^2 + (y - y_0)^2 \right)_{\text{max}}^2 \), the Fresnel PSF is

\[
S_F(x, y; z) = -\frac{ik}{2\pi z} \exp\left[i kz + \frac{ik}{2z}(x^2 + y^2)\right].
\]

For example, for \(\lambda = 0.6 \mu m\) and \(\left[\sqrt{(x - x_0)^2 + (y - y_0)^2}\right]_{\text{max}} = 5 \text{ mm}\), one needs \(z > z_{\text{min}} = 93 \text{ mm}\). Then the diffraction is described with a single Fourier transform, as

\[
E(x, y; z) = (2\pi) \exp\left[\frac{ik}{2z}(x^2 + y^2)\right] \mathcal{F}\{E_0(x_0, y_0)S_F(x_0, y_0; z)\}[k_x, k_y].
\]

We denote the Fourier transform of a function \(f(x, y)\) with respect to spatial frequencies \((k_x, k_y)\) as

\[
\mathcal{F}\{f(x, y)\}[k_x, k_y] = \frac{1}{2\pi} \int \int_{\Sigma_0} dxdy f(x, y) \exp[-i(k_x x + k_y y)] = \hat{f}(k_x, k_y).
\]

The spatial frequencies in Eq. (5) are

\[
k_x = k_x^0 \frac{x}{z}; \quad k_y = k_y^0 \frac{y}{z}.
\]

### 2.1.3 Angular spectrum

An alternative approach to describe diffraction is by analysis of the angular spectrum. Given the field \(E_0(x_0, y_0)\) at the input plane \(\Sigma_0(z = 0)\), its angular spectrum is defined as the Fourier transform

\[
A_0(k_x, k_y) = \mathcal{F}\{E_0\} = \frac{1}{2\pi} \int \int_{\Sigma_0} dxdy_0 E_0(x_0, y_0) \exp[-i(k_x x_0 + k_y y_0)].
\]

Then of course, the input field \(E_0(x_0, y_0)\) is the inverse Fourier transform

\[
E_0(x_0, y_0) = \mathcal{F}^{-1}\{A_0\} = \frac{1}{2\pi} \int \int_{\Sigma_0} dk_x dk_y A_0(k_x, k_y) \exp[i(k_x x_0 + k_y y_0)].
\]

The exponential phase factor is the \((x_0, y_0)\) projection of a plane wave with a wave vector \(k = (k_x, k_y, k_z)\), where \(k_z = \sqrt{k^2 - k_x^2 - k_y^2}\). After propagation over a distance \(z\), the plane wave acquires an additional phase factor \(\exp[ik_z z]\), so that the field \(E(x, y)\) at \(\Sigma(z)\) is
\[ E(x, y; z) = \frac{1}{2\pi} \int \int_{\Sigma_0} dk_x dk_y A_0(k_x, k_y) \exp \left[ i \left( k_x x + k_y y + \sqrt{k_x^2 - k_y^2} z \right) \right] \text{circ} \left( \frac{\sqrt{k_x^2 + k_y^2}}{k} \right) \]

\[ = F^{-1} \left\{ A_0(k_x, k_y) \exp \left[ i \sqrt{k_x^2 - k_y^2} \text{circ} \left( \frac{\sqrt{k_x^2 + k_y^2}}{k} \right) \right] \right\} [x, y]. \quad (10) \]

The circle function \( \text{circ} \), whose value is one where the argument is less than one and is zero otherwise, is necessary to restrict \( k_z \) to be real. Ordinarily, \( k^2 \geq k_x^2 + k_y^2 \), and the circle function can be dropped. We can also express Eq. (10) as a convolution. To save space, all \((x, y)\) terms are abbreviated with \((x)\). Implied \((y)\) terms should be clear from the context.

\[ E(x, y; z) = \frac{1}{(2\pi)^2} \int_{\Sigma_0} \int dx_0 E_0(x_0) \int_{\Sigma_0} \int dk_x \exp i[k_x(x - x_0)] \exp \left( i \sqrt{k_x^2 - k_y^2} z \right) \]

\[ = \frac{1}{2\pi} \int_{\Sigma_0} \int dx_0 E_0(x_0) F^{-1} \left\{ \exp \left( i \sqrt{k_x^2 - k_y^2} z \right) \right\} [x - x_0] \]

\[ = E_0 \odot S_A, \]

\[ S_A(x, y; z) = \frac{1}{2\pi} F^{-1} \left\{ \exp \left( i \sqrt{k_x^2 - k_y^2} z \right) \right\} [x, y]. \quad (12) \]

Note that the Fresnel PSF can be expressed as

\[ S_F(x, y; z) = \frac{1}{2\pi} \left\{ \exp \left[ i k z - \frac{i z}{2k} (k_x^2 + k_y^2) \right] \right\}. \quad (13) \]

Under paraxial approximation, the Fresnel transform and angular spectrum methods are equivalent.

### 2.2 Holography of Point Sources

It is useful to consider holographic imaging by point sources [30,31]. Referring to Fig. 2, suppose two point sources \( E_1 \delta(x - x_1, y - y_1, z - z_1) \) and \( E_2 \delta(x - x_2, y - y_2, z - z_2) \) emit spherical waves toward the hologram plane \( \Sigma_0(x_0, y_0) \) at \( z = 0 \). Again, \((x, y)\) pairs of expressions are
mostly abbreviated with \((x)\) only. Using Fresnel approximation, the fields at \(z = 0\) are

\[
E_1(x_0, y_0) = E_1 \exp \left[ -ikz_1 - \frac{ik}{2z_1}(x_0 - x_1)^2 \right] \\
E_2(x_0, y_0) = E_2 \exp \left[ -ikz_2 - \frac{ik}{2z_2}(x_0 - x_2)^2 \right].
\]

(14)

The intensity on the hologram plane is

\[
I_{12} = |E_1 + E_2|^2 = |E_1|^2 + |E_2|^2 + 2E_1E_2 \cos \left[ k(z_1 - z_2) + \frac{k}{2z_{12}}(x_0 - x_{12})^2 + k\zeta_{12} \right],
\]

where

\[
\frac{1}{\zeta_{12}} = \frac{1}{z_1} - \frac{1}{z_2}; \quad \frac{x_{12}}{\zeta_{12}} = \frac{x_1 - x_2}{z_2}; \quad \zeta_{12} = \frac{1}{2} \frac{(x_1 - x_2)^2}{z_1 - z_2}.
\]

(15)

This is a Fresnel zone pattern of a point source located at \((x_{12}, y_{12}, z_{12})\). Now illuminate the hologram with a third spherical wave of a possibly different wavelength \(\lambda' = 2\pi / k' (\mu \equiv k' / k)\) from the point source \(E_3\delta(x - x_3, y - y_3, z - z_3)\):

\[
E_3(x_0, y_0) = E_3 \exp \left[ -ik'z_3 - \frac{ik'}{2z_3}(x_0 - x_3)^2 \right].
\]

(17)

The optical field at another plane \(\Sigma(x, y)\) at an arbitrary \(z\) is calculated using the Fresnel diffraction formula. We calculate only the twin image terms arising from

\[
I_{12}^{\pm} = E_1E_2 \exp \left[ \pm ik(z_1 - z_2) \pm \frac{ik}{2\zeta_{12}}(x_0 - x_{12})^2 \pm ik\zeta_{12} \right].
\]

(18)

Then,

\[
E^{\pm}(x, y; z) = -\frac{ik'}{2\pi z} \exp(ik'z) \int d\Sigma_0 I_{12}^{\pm} E_3 \exp \left[ \frac{ik'}{2z}(x - x_0)^2 \right] \\
= -\frac{ik'}{2\pi z} E_1E_2E_3 \exp \left[ \pm ik(z_1 - z_2) - ik'(z_3 - z) \pm ik\zeta_{12} \right] \\
\times \int d\Sigma_0 \exp \left[ -\frac{ik'}{2} \left( \frac{1}{Z^{\pm} - \frac{1}{z}} \right) x_0^2 + ik' \left( \frac{x^{\pm}}{Z^{\pm} - \frac{x}{z}} \right) x_0 + \frac{ik}{2} \left( \frac{x_0^2}{\zeta_{12}} - \frac{\mu x_0^2}{z_3} + \frac{\mu x_0^2}{z} \right) \right].
\]

(19)

After some algebraic effort, we obtain

\[
E^{\pm}(x, y; z) = \alpha^{\pm}E_1E_2E_3 \exp \left[ \pm ik(z_1 - z_2) - ik'(z_3 - z) \right] \exp \left[ \frac{ik'}{2} \left( \frac{(x - X^{\pm})^2}{z - Z^{\pm}} \right) + i\Phi^{\pm} \right],
\]

(20)
where

\[
\frac{1}{Z^\pm} = \mp \frac{1}{\mu z_2} + \frac{1}{z_3} = \mp \frac{1}{\mu z_1} + \frac{1}{z_3},
\]

\[
\frac{X^\pm}{Z^\pm} = \mp \frac{x_{12}}{\mu z_2} + \frac{x_3}{z_3} = \mp \frac{x_1}{\mu z_1} \pm \frac{x_2}{\mu z_2} + \frac{x_3}{z_3},
\]

\[
\Phi^\pm = \frac{k x_1^2}{2z_1} \left[ (x_1 - x_3)^2 \frac{1}{z_1 z_3} \pm \frac{(x_2 - x_1)^2}{z_2 z_3} \right] Z^\pm
\]

and \(\alpha^\pm = (1 - \frac{Z^\pm}{z})^{-1}\).

The results show that the fields \(E^\pm(x, y; z)\) are spherical waves centered at \((X^\pm, Y^\pm, Z^\pm)\). Various magnifications can be calculated. The lateral, longitudinal, and angular magnifications are, respectively,

\[
M^\pm_x = \frac{\partial X^\pm}{\partial x_1} = \mp \frac{Z^\pm}{\mu z_1} = \mp \frac{1}{\mu z_1} \pm \frac{z_1}{z_2} + \frac{z_1}{z_3};
\]

\[
M^\pm_z = \frac{\partial Z^\pm}{\partial z_1} = \mp \mu M^\pm_x; \tag{22}
\]

\[
M^\pm_{x/z} = \frac{\partial (X^\pm/Z^\pm)}{\partial (x_1/z_1)} = \mp \frac{1}{\mu}.
\]

For the case of equal wavelengths, \(\mu = 1\), and the reference sources on the optical axis, \((x_2, y_2) = (x_3, y_3) = (0, 0)\), the prior expressions simplify to

\[
\frac{1}{Z^\pm} = \mp \frac{1}{z_1} \pm \frac{1}{z_2} + \frac{1}{z_3}; \quad X^\pm = \mp x_1 \frac{Z^\pm}{z_1},
\]

\[
\Phi^\pm = \frac{k x_1^2}{2z_1} \left[ \frac{1}{z_2} \pm \frac{1}{z_3} \right] Z^\pm, \quad \alpha^\pm = \frac{1}{1 - \frac{z}{Z^\pm}}. \tag{23}
\]

These results are based on the quadratic (Fresnel) approximation in Eq. (4). If higher order terms are included, then one obtains the third-order aberration terms: spherical aberration, coma, astigmatism, field curvature, and distortion [30,31]. With the higher order terms, magnification or wavelength mismatch can introduce aberrations. For the most part, the following theoretical descriptions will be within the Fresnel approximation. If the finite size of the hologram is taken into account, the image point has a finite spread determined by the numerical aperture of the hologram [32].

### 3 Digital Holographic Microscopy

#### 3.1 General Description of Digital Holographic Microscopy

A basic digital holographic microscopy (DHM) setup consists of an illumination source, an interferometer, a digitizing camera, and a computer with necessary programs. Most often a laser is used for illumination with the necessary coherence to produce interference. All different types of lasers have been used, from ubiquitous HeNe lasers and diode lasers, to diode-pumped and doubled YAG lasers (often referred to simply as a solid-state laser), argon lasers, as well as tunable dye lasers and Ti:sapphire femtosecond lasers. For multiwavelength techniques, two or more different lasers can be coupled into the interferometer, or a tunable laser can be employed. There are also low-coherence techniques for the purpose of reducing speckle and spurious
interference noise, or generating contour or tomographic images. A short-pulse (picosecond or femtosecond) laser can be used, or a tunable laser can be turned into a broadband source by removing the tuning element. Even an LED typically has $10 \, \mu m$ or so coherence length, which can be sufficient for holographic microscopy. DHM using a 10.6-$\mu m$ CO$_2$ infrared laser [33], deep UV (193 nm), [34] and 32-nm soft x-ray [35] has been demonstrated.

Two main types of interferometers, the Michelson interferometer for reflective objects and the Mach-Zehnder interferometer for transmissive objects, are depicted in Fig. 3. In each diagram, the light-green beams are the input from the laser, the light blue is the reference beam path, and the light red depicts image formation of an object point. In both designs, the object is illuminated with a plane wave, and the reference arrives at the CCD plane with the same wavefront curvature as the object wave, except for an offset in the angle of incidence for off-axis holography. The Mach-Zehnder types require more components but offer more flexibility in alignment, especially when microscopic imaging optics are used. Interferometers can also include various apertures, attenuators, and polarization optics to control the reference and object intensity ratio. Polarization optics can also be used for the explicit purpose of birefringence imaging. There can also be various types of modulators such as piezo-mounted optics, liquid crystal phase modulators, acousto-optic, or electro-optic modulators to establish modulated signals. Techniques such as the lensless Fourier holography [36] configuration can be used for magnification, but in practice achievable magnification is limited and explicit magnification by microscope objective lenses is preferred and necessary. Another lens can be used in the reference arm to match the curvatures of the object and reference wavefronts.

Fig. 3 (a) Michelson interferometer for digital holographic microscopy of reflective specimen. (b) Mach-Zehnder interferometer for digital holographic microscopy of transmissive specimen. BS: beamsplitters; L: lenses.
There are many versatile techniques in digital holography that compensate for various types of aberrations and imperfections of the optical system (see Sec. 6.4), and therefore, in comparison with conventional holography, the optical and mechanical requirements can be significantly less stringent.

Typically a CCD, or more recently CMOS, cameras are used to capture and digitize a holographic interference pattern. The pixel size of these devices is several microns with pixel counts of 1000² or so. These parameters are the main limiting factors in DHM resolution and prescribes the range of applications, but one would expect them to continue to improve in the coming years. The captured hologram pattern is digitized by the camera, or a frame grabber, and input to the computer as a 2-D array of integers with 8-bit or higher grayscale resolution. The main task of the computer is to carry out the numerical diffraction to compute the holographic image as an array of 2-D complex numbers. In addition, the computer program handles a number of other tasks, such as pre- and postprocessing of the images, rendering and storage of images, as well as timing and other necessary controls of the apparatus.

An example of the DHM process is shown in Fig. 4 using a resolution target with field of view (FOV) = 200 × 150 μm². Figure 4(a) is the hologram, with detail shown in the inset, where the interference fringes are visible. Figure 4(b) is the angular spectrum (Fourier transform), showing the zero-order and twin image peaks. One of the twin terms is selected with a numerical bandpass filter (yellow circle). The filtered hologram is then used for numerical diffraction over an appropriate distance, which results in the reconstructed holographic image as a 2-D array of complex numbers. The amplitude and phase images in 4(c), 4(d), respectively, are obtained by taking the absolute magnitude and phase of the complex array. The phase image represents a phase profile of the optical field reflected from the object surface or transmitted through a thickness of a transparent object. The phase profile has the precision of a fraction of optical wavelength, and therefore reveals nanometric variations of the surface or the optical thickness of the specimen. In Fig. 4(d), minute smudges of some kind are visible, apparently some fraction of wavelength thickness, which the amplitude image completely misses. The phase image color scale ranges $2\pi$ from blue to red. The object surface is slightly tilted with respect to the reference wavefront, and such tilt and other aberrations can readily be compensated by numerical techniques described later.

A well-known distinctive feature of holography is the 3-D content of the image information. In DHM, a single hologram is used to reconstruct the optical field at any distance from the hologram, within the limitation of the approximation method used. For example, Fig. 5(a) shows a hologram of a paramecium. From the single hologram, the image is calculated at various distances, which are then assembled into a video clip in 5(b). It shows the paramecium image going through a best focus, precended and followed by more defocused images, emulating the turning of a focusing knob on a conventional microscope. In Fig. 6(a), a series of holograms of a live paramecium and several euglenas are captured. In the scene, the paramecium and the euglenas swim not only in lateral directions but also in varying depths. In conventional video microscopy, the focal plane would be fixed and whatever happened to be in that plane would be recorded, but information on objects not in the focal plane would be permanently
3.2 Quantitative Phase Microscopy

Many microscopic biological specimens, such as living cells and their intracellular constituents, are mostly transparent and therefore problematic for conventional bright-field microscopy. A number of techniques have been developed for rendering transparent phase objects visible [37] that have played very important roles in the development of modern biology and medicine. For example, in dark-field microscopy, only the scattering centers and boundaries contribute to the image signal against a zero background. In the Zernike phase contrast microscope, the phase variation is converted into amplitude variation by use of a phase plate and spatial filtering. In differential interference contrast (DIC) microscopy, the interference of two sheared polarization components results in images that have a shadow effect and thus gives a 3-D perception of the object. Interference microscopy, using a Michelson objective for example, produces fringes of equal thickness of a transparent object. Although these techniques are effective in making transparent objects visible, the phase-to-amplitude conversion is nonlinear, and there are significant artifacts in the images such as the halo in Zernike phase contrast and the
Kim: Principles and techniques.

Fig. 7  Digital holographic microscopy process (onion cells) (FOV = 100 × 100 μm, 416 × 416 pixels): (a) hologram, (b) angular spectrum, (c) amplitude image, (d) phase image, (e) unwrapped phase image, and (f) phase image in pseudo-3-D view.

disappearance of contrast along the direction perpendicular to shear in DIC. These techniques do not produce quantitative phase images.

The optical phase of the light transmitted through transparent objects can convey quantitative information about the object, such as its physical thickness and index of refraction [38], which in turn are functions of physical density or chemical concentration properties. High precision measurements of optical phase can thus reveal subtle changes in these parameters that accompany cellular processes. To obtain quantitative phase images, one can perform an interferometric measurement of a focused beam of light on an object, and scan the beam over the object in a raster fashion. Optical profilers based on scanning interferometers are especially well suited for imaging applications in materials science, as in MEMS and nanofabrication, because of the high precision obtainable and the static nature of the objects being imaged [39,40]. On the other hand, the speed constraint and mechanical complexity of scanning interferometers can significantly restrict the range of applications in biomedical imaging [41], where one needs to make observations of dynamic processes under widely varying environments. There have been some recent developments in 2-D quantitative phase microscopy. In phase-shifting interference microscopy [42,43], the quantitative phase image is obtained from a combination of three or more interferograms. There is also a noninterferometric method to extract quantitative phase images from differential focusing properties of bright-field intensity images alone [44,45].

Digital holography is a very effective process for achieving high-precision quantitative phase microscopy. The phase image is immediately and directly available as soon as the 2-D complex array of the holographic image is calculated. A single hologram exposure is required. It does not involve raster scanning. Most importantly, the phase image is a quantitative representation of the object profile with nanometer, and even subnanometer, precision [20,21,46–48]. An example of DHM imaging of a layer of onion cells is shown in Fig. 7, where Fig. 7(a) is the hologram and Fig. 7(b) is its angular spectrum. Because of the structure of the specimen, the spectral peaks are more diffuse compared to Fig. 4. Figure 7(c) is the amplitude image, analogous to what one would see through a conventional microscope, and Fig. 7(d) is the phase image. The onion cells apparently have thicknesses of several microns, and therefore the phase profile varies by several cycles of 2π radians. A public-domain phase unwrapping algorithm is used to remove the 2π discontinuities in Fig. 7(e), and it is rendered in pseudocolor pseudo-3-D perspective in Fig. 7(f). Figure 7(f) is pseudo-3-D in the sense that the apparent height profile is the profile of optical thickness that includes both physical thickness and index variation, and one needs to use caution in interpreting such an image. Figure 8 displays a few more examples of quantitative phase microscopy (QPM) images by DHM. Figure 8(a) is one group of three bars on a resolution target. The noise level in the flat area of the image is measured to be 3 nm and the thickness of the chromium film is measured to be about 50 nm, consistent with the manufacturer’s estimate. Figures 8(b), 8(c) are fixed SKOV-3 ovarian cancer cells, where one can discern several intracellular components such as the nuclear membrane and chromosomes. Figure 8(d) shows several red blood cells, while in Fig. 8(e) one can notice a fold of the cheek epithelial cell, as well as its nucleus and mitochondria. Figure 8(f) is an image of a small quartz crystal in common sand.

SPIE Reviews 018005-11 Vol. 1, 2010
3.3 Comparisons of Analog and Digital Holographic Microscopy

There are a number of significant distinctions between analog (AH) and digital (DH) holographies. Most obviously, DH does not involve photochemical processing. Therefore, DH is orders of magnitude faster and can be performed at video rates. Additional hardware required in DH is the CCD camera and a computer, while the need for dark room facilities and a supply of chemicals is unnecessary. Furthermore, because of the high sensitivity of CCD compared to photographic emulsion, the exposure time is reduced by orders of magnitude. For example, a CCD pixel area of 100 $\mu$m$^2$ can detect as few as several photons, whereas a similar area of a high-sensitivity photographic plate requires many millions of photons. Short exposure time in turn implies much reduced requirement on the mechanical stability of the apparatus. Heavy optical tables with vibration isolation are often not critical. On the other hand, the main issue of DH is low resolution. A typical CCD pixel is several microns across, while the grains on a photographic emulsion may be 2 orders of magnitude finer. This limits the spatial frequency of the fringes and therefore the angular size of the object to a few degrees for DH, while a full 180 deg is possible for AH. The familiar parallax effect of display holograms of AH is currently not feasible in DH [49]. The real strength of DH, however, is the whole range of powerful numerical techniques that can be applied once the hologram is input into a computer. One simple but significant example relates to Fig. 3, where a lens is used to magnify the hologram FOV to match the CCD size. Once the computer reads the hologram into an array, one only needs to specify the dimension of the FOV and the wavelength, and proceed to compute the numerical diffraction. In AH, however, to properly read out the magnified or demagnified hologram, the wavelength also needs to be scaled proportionately, a task that is highly cumbersome at the least and infeasible in most cases. Another example is holographic interferometry using multiple wavelengths. In AH interferometry, multiple holograms are produced and repositioned exactly, and ideally each hologram needs to be illuminated with a different wavelength, which can be physically impossible. Most often the superposed holograms are illuminated with a single wavelength, and the resulting aberrations are unavoidable. In DH, however, the superposition simply consists of an addition of several numerical arrays. There is no limitation on the number

Fig. 8 Examples of quantitative phase microscopy by digital holography: (a) resolution target (25 × 25 $\mu$m, 452 × 452 pixels); (b) SKOV-3 ovarian cancer cells (60 × 60 $\mu$m, 404 × 404 pixels); (c) SKOV-3 ovarian cancer cell (60 × 60 $\mu$m, 404 × 404 pixels); (d) red blood cells (50 × 50 $\mu$m, 404 × 404 pixels); (e) cheek epithelial cell (60 × 60 $\mu$m, 404 × 404 pixels); and (f) quartz crystal of sand (60 × 60 $\mu$m, 404 × 404 pixels).
of arrays, and furthermore, there are ways to preprocess the arrays to compensate for chromatic and other aberrations if present. More examples of the power of numerical processing in DH will become evident in the following discussions.

Because of its sensitivity and technical versatility, quantitative phase microscopy is a very important and active area of research and applications in digital holography [20,21,46]. Aberrations or other deformations of wavefronts can easily be compensated by using a matching reference wave [22,50,51]. Multiwavelength optical phase unwrapping (see Sec. 6.3) allows nanometric-precision phase imaging over a range of many micrometers without many problems associated with common software-based unwrapping methods [23,52]. Biomedical microscopy application is an area that can benefit significantly from the new capabilities of digital holography by providing label-free, minimally invasive, and highly sensitive methods of imaging subtle changes in the physical and physiological states of cells and tissues [19,47,53–55]. Materials and MEMS technologies can also utilize digital holography in characterization and testing of various structures [56–61].

4 Numerical Diffraction

4.1 Methods of Numerical Diffraction

There are a number of different methods developed for numerically calculating the diffraction field. Precursors of digital holography [13,62] involved Fourier holography and therefore required a single Fourier transform for reconstruction. This was significant because of the limitations on computational power. More common approaches in digital holography have been the Fresnel transform and Huygens convolution methods, as well as, more recently, the angular spectrum method [27,63]. These methods are closely related to each other, and the terminology is not uniform. The Fresnel transform method is a direct digital translation of the familiar Fresnel diffraction formula of Eq. (5), and is therefore amenable to direct comparison of digital results with analytical expressions, which are mostly developed using the Fresnel approximation. The Huygens convolution and angular spectrum methods appear very similar analytically, but their digital implementation leads to significant differences, as described next. Among the three methods, only the angular spectrum method requires no minimum reconstruction distance and has other advantages. Other methods of numerical diffraction include wavelet or Fresnelet approach [64,65], fractional Fourier transform for highly astigmatic systems [66–69], and windowed Fourier analysis [70]. The Wigner space diagram is utilized for space-bandwidth analysis of holographic image formation [71–73]. 3-D optical transfer function (OTF) for DHM is given in Ref. 74.

4.1.1 Fresnel transform

Under Fresnel approximation, the propagation of optical field from a plane $\Sigma_0$ at $z = 0$ to another plane $\Sigma$ along the z axis is given by Eq. (5):

$$ E(x, y; z) = (2\pi) \exp \left[ \frac{ik}{2z} (x^2 + y^2) \right] F\{E_0(x_0, y_0) S_F(x_0, y_0; z)\} [k_x, k_y]. \tag{24} $$

$$ S_F(x, y; z) = -\frac{ik}{2\pi z} \exp \left[ ikz + \frac{ik}{2z} (x^2 + y^2) \right]. \tag{25} $$

The Fresnel transform method therefore involves a single Fourier transform, and can be efficiently implemented by the use of the fast Fourier transform (FFT) [14,27,46,75–80]. For numerical calculation, a $X_0 \times Y_0$ area of the input field $E_0(x_0, y_0)$ is sampled on an $N_x \times N_y$ array with pixel size $\delta x_0 \times \delta y_0$. The output field is then another $N_x \times N_y$ array of area $X \times Y$ and pixel size $\delta x \times \delta y$. An example of numerical diffraction by the Fresnel transform method...
Fig. 9 Numerical diffraction by Fresnel transform method. Each frame is $100 \times 100 \, \mu m^2$ with $256 \times 256$ pixels. Top row is amplitude images and the bottom is phase images, at distances 10, 30, 50, 100, and 200 $\mu m$, and $\lambda = 0.633 \, \mu m$.

Fig. 10 Comparison of the numerical diffraction methods for propagation over a range of $z = 0 \sim 250 \, \mu m$ : (a) Fresnel transform, (b) Huygens convolution, and (c) angular spectrum methods. Upper row is amplitude profiles and lower row is phase profiles.

is shown in Fig. 9, where the input pattern at $z = 0$ is the letter “F” in an opaque screen of $100 \times 100 \, \mu m^2$ area with $256 \times 256$ pixels, and the wavelength is assumed to be $\lambda = 0.633 \, \mu m$. The top row shows the amplitude of the diffracted field at distances of $z = 10, 30, 50, 100, \text{and } 200 \, \mu m$. The bottom row shows corresponding phase patterns. In the phase images, the factor $\exp(-ikz)$ has been multiplied to remove the rapid phase variation due to overall $z$-propagation.

One pixel in $E(x, y; z)$ corresponds to $\delta k_x = \frac{1}{2} \delta x$, which in the $\Sigma$-plane corresponds to

$$
\delta x = \frac{z}{k} \delta k_x = \frac{2\pi z}{kX_0} = \frac{\lambda}{X_0} z,
$$

where $X_0$ is the size of the $\Sigma_0$ plane. That is, the pixel size and the $\Sigma$-plane size grow linearly with the distance, as is evident in Fig. 9. In fact, the field of view at the five distances are 16, 48, 81, 160, and 320 $\mu m$. There are methods to address the problem of nonconstant pixel resolution, which is discussed in Sec. 6.2. Also evident is the presence of a minimum distance to obtain a valid diffraction pattern. To avoid aliasing [81], the output plane $\Sigma$ needs to be at least as large as the input plane $\Sigma_0$, which leads to $z_{\text{min}} = \frac{X_0^2}{4\lambda}$. Figure 10(a) shows a $yz$-cross section of the propagation over a range of $z = 0 \sim 250 \, \mu m$ along a vertical line indicated by the blue arrow in Fig. 9. The location of $z_{\text{min}} = 62 \, \mu m$ is indicated with a red arrow in Fig. 10(a), and at shorter distances one observes wrapping and aliasing of the calculated output image.
4.1.2 Huygens convolution

The diffraction can also be calculated as a convolution integral [Eq. (2)],

$$E(x, y; z) = E_0 \otimes S_H = F^{-1} \{ F[E_0] \cdot F\{S_H\} \},$$  \hspace{1cm} (27)

$$S_H(x, y; z) = -\frac{ik}{2\pi z} \exp[i k \sqrt{x^2 + y^2 + z^2}].$$  \hspace{1cm} (28)

Thus, the Huygens convolution is computed with three Fourier transforms [18,27,76,82,83]. Numerical diffraction by Huygens convolution using the same set of parameters as in Fig. 9 is shown in Figs. 10(b) and 11. The short-distance behavior is significantly better than the Fresnel method. Also note that one can use the Fresnel PSF $S_F$ as an approximation to $S_H$ and obtain results with similar qualities. Such approximation, however, is really not necessary.

4.1.3 Angular spectrum

The diffraction is calculated by the angular spectrum method according to

$$E_A(x, y; z) = F^{-1} \left\{ F[E_0][k_x, k_y] \exp \left[i \sqrt{k_x^2 - k_y^2 - k_z^2} \right] \text{circ} \left( \frac{\sqrt{k_x^2 + k_y^2}}{k} \right) \right\} [x, y],$$  \hspace{1cm} (29)

which requires two Fourier transforms [19,63,84–89]. For image plane holographic microscopy, where the input field $E_0$ may be actually a magnified image of a microscopic object, the effective pixel size may be smaller than the wavelength, and the argument of the square root in Eq. (29) may become negative in parts of the spectral domain, and the circle function is then required. Numerical diffraction by angular spectrum using the same set of parameters as in Fig. 9 is shown in Figs. 10(c) and 12. This method yields valid diffraction images at short distances, down to zero.

4.2 Comparison of Methods

Referring to Fig. 13, the angular spectrum method is based on propagation of plane waves. Sampling of the plane waves by discrete pixels of the CCD does not vary with distance, and therefore the angular spectrum method does not have any distance limitations. On the other hand, the Fresnel transform method is based on the propagation of spherical (or parabolic
approximation) wavefronts. When the center of curvature is too close to the CCD array, local fringe frequency on the CCD plane may be higher than the Nyquist frequency. This occurs when the distance is smaller than $z_{\text{min}} = X_0 \sqrt{2 \delta x \lambda / \lambda} - 1$, which in our example is 72 $\mu$m. The behavior of diffraction field calculated by the three methods over a range of $z = 0 \sim 250$ $\mu$m is shown in Fig. 10. Moreover, there also is a maximum distance that yields correct diffraction patterns for the Fresnel transform method. For large enough distances, the fringe period becomes larger than the entire CCD array, failing to record any diffraction information. This occurs when the distance is larger than $z_{\text{max}} = X_0^2 / (2 \lambda^2)$, which, for our numerical example, is 7900 $\mu$m. The angular spectrum does not have this limitation either. It should be noted that the terminology and relevant approximations of the various methods in the literature are not uniform, and comparison of results, such as the minimum and maximum distances, must be made with some care [90].

4.3 Digital Sampling of Hologram

The finite size of and discrete sampling by the CCD array modifies the holographic imaging properties that we obtained in Sec. 2.2 [73,76,80,91–94]. Suppose the size of the CCD array is $X_0 \times Y_0$ with $N_x \times N_y$ pixels, so that the pixel pitch is $\delta x_0 \times \delta y_0 = (X_0 / N_x) \times (Y_0 / N_y)$. Further suppose that the sensitive area of a CCD pixel is $\gamma_x \delta x_0 \times \gamma_y \delta y_0$, where $\gamma_x$ and $\gamma_y$ are...
the fill factors. The CCD sampling function is then written as [93,95–97]

\[
P(x_0, y_0) = \text{rect}\left(\frac{x_0}{X_0}\right) \cdot \left[\text{comb}\left(\frac{x_0}{\delta x_0}\right) \ominus \text{rect}\left(\frac{x_0}{\gamma/\delta x_0}\right)\right],
\]

(30)

where again we abbreviate the \( y \)-terms and also drop the subscript from \( \gamma \). The comb function \( \text{comb}(x/\delta x) \) is a series of delta functions with \( \delta x \) intervals, and the rectangle function \( \text{rect}(x/a, y/b) \) has a value of one within the rectangle of size \( a \times b \) and zero outside. We consider the holographic image formation by point sources as in Sec. 2.2, but in Eq. (19), we multiply the sampling function \( P(x_0, y_0) \) inside the Fresnel diffraction integral, so that

\[
E^\pm = \frac{ik'}{2\pi z} E_1 E_2 E_3 \exp\left[\pm ik(z_1 - z_2) - ik'(z_3 - z) \pm ik\xi_{12}\right] \\
\times \int_{X_0} d_{x_0} P(x_0) \exp\left[-\frac{i k'}{2} \left(\frac{1}{Z^\pm} - \frac{1}{z}\right) x_0^2 + \frac{i k'}{2} \left(\frac{X_0^\pm}{Z^\pm} - \frac{x}{z}\right) x_0 \right. \\
\left. + \frac{i k'}{2} \left(\frac{1}{Z^\pm} - \frac{1}{z}\right) x_0^2 \right].
\]

(31)

The effect of the sampling function on the integral is, for a function \( f(x_0) \),

\[
\int_{X_0} d_{x_0} P(x_0) \cdot f(x_0) = \sum_{m=-X_0/2\delta x_0}^{X_0/2\delta x_0} \int_{(m-\gamma/2)\delta x_0}^{(m+\gamma/2)\delta x_0} d_{x_0} f(x_0).
\]

(32)

Also note that at the position of the holographic image \( z = Z^\pm \) and the integral and summation are easily carried out.

\[
E^\pm(x; z) = \frac{-ik'}{2\pi z} E_1 E_2 E_3 \exp[\pm ik(z_1 - z_2) - ik'(z_3 - z) \pm ik\xi_{12}] \exp\left[\frac{i k}{2} \left(\frac{1}{Z^\pm} - \frac{1}{z}\right) \frac{x_0^2}{\delta x_0^2} \right. \\
\left. + \frac{1}{Z^\pm} \left(\frac{x}{x_0} - 1\right) \frac{Z^\pm}{x_0} \right] \sin\left[\frac{k'}{2} \left(\frac{x-x_0^\pm}{\delta x_0}\right) \frac{Z^\pm}{x_0} \right] \sin\left[\frac{k'}{2} \left(\frac{x-x_0^\pm}{\delta x_0}\right) \frac{Z^\pm}{x_0}\right].
\]

(33)

The lower line of Eq. (33) shows the effect of digital sampling [76]. The sinc function has a width of \( \frac{4\pi Z^\pm}{k' \delta x_0} \) between the first zeroes. For a small fill factor \( \gamma \to 0 \), the image is uniform in amplitude throughout the image plane. On the other hand, integration over the pixel area with a larger fill factor results in vignetting, especially for shorter image distance \( Z^\pm \). The sinc-over-sine factor of Eq. (33) gives the lateral resolution of a point image to be \( \frac{2\pi Z^\pm}{k' \delta x_0} \), which shows the expected dependence on the numerical aperture of the camera array [32]. Also note that the point image repeats over a period of \( \frac{4\pi Z^\pm}{k' \delta x_0} \). This period has to be larger than the size of the camera array [81], which leads to the minimum distance \( z_{\text{min}} = \frac{X_0^2}{2k' \delta x_0^2} \), as seen in Sec. 4.1. To estimate the axial resolution, suppose all three source points are on the \( z \) axis. Then the reconstructed field along the \( z \) axis is given by

\[
E^\pm(0; z) \propto \sum_{m=-X_0/2\delta x_0}^{X_0/2\delta x_0} \int_{(m-\gamma/2)\delta x_0}^{(m+\gamma/2)\delta x_0} d_{x_0} \exp\left[-\frac{i k'}{2} \left(\frac{1}{Z^\pm} - \frac{1}{z}\right) \frac{x_0^2}{\delta x_0^2} \right].
\]

(34)
The integral sum is close to zero unless the phase of the exponential varies slowly within the range $X_0$, which leads the axial resolution to be $\frac{2\lambda'Z^2}{X_0^2}$.

For a holographic microscopy setup with an objective lens, as depicted in Fig. 3, the source point $E_1, \delta(x - x_1, y - y_1, z - z_1)$ may in fact refer to an object point imaged and magnified by the lens. In that case, $E_1$ is a point within the PSF of the image, and the description follows as before. Alternatively, and equivalently, $E_1$ refers to an actual object point and the CCD array refers to its image projected into the object volume through the objective lens. In that case, the CCD array is demagnified by the lens and an object point is spread by a demagnified PSF as well.

5 Digital Holography Configurations

5.1 Off-Axis Fresnel Holography

An overview of the main types of interferometer configurations used in digital holography is given. First, Fresnel holography refers to a configuration where the object is a finite distance from the hologram plane and the reference is usually a plane wave. Then the images form at the object position and its mirror position with respect to the hologram, with unit magnification, as shown in Fig. 14. To avoid the overlap of the reference and the images, the reference and object waves are offset by an angle, as was originally done by Leith and Upatnieks [6]. Image position and magnification can be manipulated by using references other than a plane wave, according to Eqs. (21) and (22). Placement at a large enough distance and the use of Fresnel transform allow imaging of an object larger than the CCD array size, such as in macroscopic metrology applications [14,77,98,99]. Also, an imaging lens can be used to form a magnified [46] or demagnified image of the object, which then propagates to the hologram plane. For example, in Ref. 77, a negative lens is used to form a demagnified image of a large object, thus reducing spatial frequency bandwidth.

5.2 Fourier Holography

As is well known, the field at the back focal plane is the Fourier transform of the object field at the front focal plane. The hologram is recorded with a plane reference wave [Fig. 15(a)]. In reconstruction, a plane reference wave and a lens produce the inverse transform, which is the image [Fig. 15(c)]. Alternatively, lensless Fourier holography is possible by placing a point source reference at the object plane [Fig. 15(b)], and the reconstruction proceeds in the same manner as before [36,94]. But in digital holography, the reconstruction is especially simple, for it requires only a single Fourier transform of the recorded hologram [13,62]. In the first digital Fourier microholography experiment of Ref. 62, a drop of glycerol placed next to the specimen on a microscope slide acted as the focusing lens to form the point source reference in front of the Fourier lens. There a numerical lens was also introduced to focus the image at varying distances. High resolution microscopy is possible using a relatively simple setup [100,101],

Fig. 14 Off-axis Fresnel holography. Blue represents reference beams and red represents propagation from a point on the object: (a) recording and (b) reconstruction.
Fig. 15 Fourier and lensless Fourier holography: (a) Fourier hologram recording using a lens, (b) lensless Fourier hologram recording, and (c) reconstruction by Fourier transform, represented with the Fourier lens.

and a detailed analysis of image formation in lensless digital Fourier holography is given in Ref. 91. In lensless Fourier microholography, the object can be placed close to the sensor, which increases the numerical aperture and improves resolution [102]. But this introduces aberration in the reconstruction because of violation of the Nyquist frequency requirement. The hologram is expanded and interpolated, followed by multiplication of a transfer function, before Fourier transform, to obtain an aberration-free high resolution image. Ref. 98 makes a comparison of various interferometer configurations in the context of macroscopic metrology applications.

5.3 Image Plane Holography

The object can be placed close to the hologram plane, so that $z_1 \approx 0$. Then $Z^\pm \approx z_1$ and $X^\pm = -x_1$, so that the image forms near the hologram plane at the object position. In real-space holography, this is useful for creating holograms that can be viewed with low coherence light, because the image distance is not very large compared to the coherence length. In microscopic holography, this would not be useful because there is no magnification, unless one uses a microscope to view the hologram. On the other hand, an objective lens can be used to form near the hologram plane a magnified image of the object, and the holographic image would reconstruct the magnified image [Fig. 16(a)]. The image will correctly reproduce the amplitude profile of the object, but not the phase profile. Alternatively, the objective lens can be used to form a magnified image of the holographic interference, including both the object and write reference [Fig. 16(b)]. In real-space holography, a reconstruction using the same wavelength from the magnified hologram would result in an unmagnified image of the original object size. To maintain the magnification, one would have to use a wavelength longer by the same magnification factor, but such a long wavelength would be in the far infrared or microwave.

In digital holographic microscopy, this configuration is particularly flexible because the acquired hologram can be numerically scaled according to the physical dimensions of the object, regardless of the image size on the CCD. In effect, a demagnified CCD array is placed...
Fig. 16 Recording of image plane holograms by projection of a magnified image of the object on the hologram plane, in superposition with a (a) plane reference wave or (b) wavefront-matched reference wave.

near the object position. Once the magnified holographic interference is input into the computer, it is only a matter of assigning the original microscopic frame size and the correct wavelength that was used for creating the holographic interference. The numerically reconstructed image will then correctly represent the amplitude and phase of the original microscopic object space.

Image plane digital holography has been useful in a number of different application areas, including biological microscopy [47,103], where it is important to be able to monitor the live specimen being imaged. It is also advantageous for improving light collection efficiency in particle velocimetry [104]. For DHM using low coherence light, one necessarily operates near image plane configuration [83,105,106].

5.4 In-Line Holography

Off-axis holography is necessary to avoid the overlap of the zero-order and holographic images. But this in effect reduces the information content of the hologram to one quarter of the pixel count, which is at a premium in digital holography. With in-line holography, the object field is in general alignment with the reference beam, and the entire hologram pixel count is utilized, which also leads to shorter minimum distances for Fresnel reconstruction and higher resolution of the resultant image [107] (Fig. 17). A number of methods have been proposed and demonstrated to reduce or eliminate the effect of the zero-order and twin image, as is described later (see Sec. 6.1). The zero-order (or dc) term can be partially suppressed simply by subtracting the average intensity of the entire hologram, or alternatively by taking the Fourier transform of the hologram and applying a high-pass filter near the zero frequency. The effectiveness of high-pass filtering depends on the spectral content of the object. Separate exposures of the reference and object beams and their subtraction from the hologram before zero-order filtering improves the result. The twin image removal is less straightforward [79,108] and requires special techniques such as the phase-shifting multiexposure method. On the other hand, the high-pass filtering method can be effective when the dynamic nature of the object precludes the phase-shifting method.
Fig. 17 In-line holography: (a) in-line superposition of object and reference beams, and (b) reconstruction of superposed zero-order and twin images.

Fig. 18 Gabor holography: (a) recording by superposition of the reference and its scattered component from a point object, and (b) reconstruction of a point image and its defocused twin.

5.5 Gabor Holography

In Gabor holography, the object is illuminated with a single beam of light and there is no separate reference wave (Fig. 18). The part of the light that is scattered by the object is the object wave, and the remainder that does not undergo scattering acts as the reference wave. The method is more effective the smaller the object is, so that the reference is not excessively disturbed. Because of this constraint and because of the simplicity of the optical setup, Gabor holography is particularly useful for particle image analysis [109–111] as well as for thin fibers [112]. Digital Gabor holography (DGH) can provide new capabilities for wider applications of this simple configuration [113–115]. For example, in Ref. 25, the difference of two consecutive holograms in a holographic movie completely subtracts the background while revealing movement of particles or microbes. For microscopic particles, the twin image problem is often negligible, because even at relatively short distances the Fraunhofer condition is satisfied and the twin image is completely defocused. In Ref. 116, simple inspection of intensity profiles through a particle allowed an estimate to be made of the particle’s 3-D position coordinates within an accuracy of a few hundred nanometers. When the derivative of a second-order polynomial fitted to the intensity profiles was taken, the X, Y, and Z position coordinates of particles could be determined within 50 nm. An underwater device utilized DGH to investigate particles, bacteria, paramecium, and other swimming microbes [117,118], and the detailed flow field of a feeding copepod was measured using DGH [119]. The DGH configuration is also in-line, but in this review in-line holography refers to one with a separate reference wave. With DGH, the object must fill only a small portion of the field, while in-line holography with references does not have such a constraint. With the in-line configuration, the subtraction of the zero-order and twin terms is important, while with DGH those terms presumably have negligible effect.

5.6 Phase-Shifting Digital Holography

The in-line configuration makes use of the full pixel count in forming the holographic image, but the zero-order and the twin image terms are superposed on the image. A very effective method of removing these terms was introduced by Yamaguchi and Zhang [24], where the complex field at the hologram is obtained by phase-shifting interferometry. From the complex field at the
Kim: Principles and techniques...

**Fig. 19** Phase-shifting digital holography. PZT: piezomounted mirror for modulation of reference phase.

hologram plane, including the amplitude and phase information, the optical field at any other plane can be obtained by one of the numerical diffraction methods.

For simplicity, assume that the reference is a plane wave normally incident on the hologram plane: $E_R(x, y) = E_R \exp(i\psi)$. The object wave has the amplitude $E_O(x, y)$ and phase $\phi(x, y)$ distributions, so that

$$E_O(x, y) = E_O(x, y) \exp[i\phi(x, y)] = E_O(x, y) [\cos \phi + i \sin \phi]. \quad (35)$$

Then the interference intensity is

$$I(\psi, x, y) = |E_O + E_R|^2 = E_R^2 + E_O^2(x, y) + 2E_RE_O(x, y) \cos[\phi(x, y) + \psi]. \quad (36)$$

In four-step phase-shifting digital holography (PSDH), four holograms with phase shifts $\psi = 0, \pi/2, \pi, 3\pi/2$ are acquired, for example by a piezo-mounted reference mirror (Fig. 19):

$$I_0 = E_R^2 + E_O^2 + 2E_RE_O \cos \phi$$

$$I_{\pi/2} = E_R^2 + E_O^2 - 2E_RE_O \sin \phi$$

$$I_{\pi} = E_R^2 + E_O^2 - 2E_RE_O \cos \phi$$

$$I_{3\pi/2} = E_R^2 + E_O^2 + 2E_RE_O \sin \phi, \quad (37)$$

which are then numerically combined to extract the amplitude and phase profiles, so that

$$E_O(x, y) = \frac{1}{4E_R} [(I_0 - I_{\pi}) + i(I_{3\pi/2} - I_{\pi/2})]$$

$$\phi(x, y) = \tan^{-1} \left[ \frac{I_{3\pi/2} - I_{\pi/2}}{I_0 - I_{\pi}} \right]. \quad (38)$$

This completely defines the complex optical field $E_O(x, y; 0)$ of the object at the hologram plane, and the diffraction theory can be used to calculate the optical field $E_O(x, y; z)$ at any distance $z$ from the hologram. A similar procedure is given in Ref. 120. These procedures completely remove the contributions from the zero-order and twin image terms. The number
of required hologram exposures is reduced by one in three-step PSDH [76] with phase shifts \( \psi = 0, \pi / 2, \pi \):

\[
E_0(x, y) = \frac{1 + i}{4E_R} [(I_0 - I_{\pi / 2}) + i(I_{\pi / 2} - I_{\pi / 2})]
\]

\[
\varphi(x, y) = \tan^{-1}\left[\frac{I_{\pi / 2} - I_{\pi / 2}}{I_0 - I_{\pi / 2}}\right].
\]

A two-step method is also possible [121,122]

\[
E_O(x, y) = \frac{(I_0 - E_O^2 - E_R^2) - \exp (i\psi)(I_{\psi} - E_O^2 - E_R^2)}{E_R [1 - \exp (2i\psi)]},
\]

which requires two phase-shift exposures plus separate exposures of the object and reference intensities. Recently, a two-step-only method without the need for separate intensity exposure of references or objects was introduced [123]. Methods are also available for phase extraction with unknown or random phase shifts [124,125].

The general phase-shifting principle is applied in other various configurations. For example, in Ref. 126, the phase shift is introduced on a portion of the Fourier spectrum by projecting an apertured phase modulator onto the focal plane of the imaging lens, reminiscent of Zernike phase contrast microscopy. The object is positioned on the focal plane and the CCD acquires three or four Fourier holograms with appropriate phase shifts, which eliminates the zero-order and twin spectral components. There is another method of extracting the phase profile from a single interferogram by fitting sinusoids over the slowly varying interference intensity. The method is called the spatial carrier phase shift method or sinusoid fitting method [98,127], which requires a single interferogram exposure but at the expense of the spatial resolution. It operates on the interference fringes at the image plane and does not require Fourier transform. It is valid if the phase changes slowly over several pixels, so that

\[
\varphi(x, y) = -\tan^{-1}\left[\frac{I(x + \Delta x, y) - I(x - \Delta x, y)}{I(x + \Delta x, y) - 2I(x, y) + I(x - \Delta x, y)} \tan \frac{k_0\Delta x}{2}\right],
\]

where \( I(x, y) \) is the interference pattern, \( k_0 \) is the carrier fringe frequency, and \( \Delta x \) is the pixel pitch. A more general method that works for curved reference waves as well as plane waves was introduced in Ref. 128. Reference 104 describes a few different phase-shifting methods.

Error in phase shift in reconstruction relative to the phase shift in recording can lead to incomplete cancellation of the dc and conjugate terms [129]. One way to estimate and correct the error is by minimizing the error in the reconstructed amplitude image compared to the object [121]. More general methods are given in Refs. 130, 131, and 132, where the phase shift between two exposures is calculated based on a statistical consideration of the interferogram, and its correction for intensity instability is also given [133]. In Ref. 134, a difference frequency between the object and reference beams was set up using acousto-optic modulators such that the phase shift between consecutive CCD frames could be precisely controlled, and achieved ultimate sensitivity of one photoelectron noise per pixel – also called heterodyne digital holography. An analysis of noise and sensitivity of PSDH has been made in Ref. 135.

The phase-shifting concept for conventional holography was originally introduced by Gabor and Goss [136], but the complexity of the optomechanical system was substantial, whereas with digital implementation, algebraic manipulation of the various terms presents no such difficulty. Hence the technique finds applications in many different areas, including microscopy [137], surface shape measurement [138], and color holography [139]. PSDH can be applied with Fresnel configurations for macroscopic imaging [24] or image plane configuration for microscopy [98].
6 Numerical Techniques for Digital Holographic Microscopy

6.1 Suppression of DC and Twin Image Terms

The DC term, which includes reference and object field intensities, can be reduced by subtracting
the average value from the hologram array. Since the object as well as the reference fields have
spatial variations, the DC component has a finite amount of spectral spread around the zero
frequency. A high-pass filtering effect was obtained by blanking a 3×3 pixel neighborhood of
the zero frequency from the spectrum [78]. A more effective approach is to apply a numerical
filter or mask directly on the Fourier (or angular) spectrum of the off-axis hologram, as first
demonstrated by Cuche, Marquet, and Depeursinge [140]. The method can be used not only to
suppress the DC term, but also to select one of the twin first order terms as well as eliminate
spurious spectral components due to parasitic reflections and interference, thus improving the
quality of the reconstructed image. A similar effect can be achieved physically by placing an
appropriate mask in the Fourier plane of a 4f configuration of lenses, but the numerical method
affords significant flexibility and versatility. For example, the numerical spectral mask can be
easily set up with a smoothed window function to reduce fringing in the reconstructed image.
Separate acquisition and subtraction of reference and object intensities from the hologram can
be useful [107], even with off-axis numerical filtering, by allowing larger bandwidth of the
filter. Note that the hologram filtered numerically can have positive and negative pixel values
with the average near zero, a feature not feasible in a real-space hologram.

In Ref. 108, the DC component is filtered out by applying a high-pass filter in the Fourier
spectrum of an in-line hologram. The twin image term is reduced, but not eliminated, by
numerically imaging the aperture and suppressing the out-of-focus conjugate noise outside the
aperture. Alternatively, with an aperture at a Fourier plane, the aperture image is suppressed and
the conjugate part is used to reconstruct the conjugate image of the object. Combinations of two
or more holograms with stochastic change in the object speckles between them resulted in the
removal of the dc and twin terms [79,141]. Phase-shifting digital holography very efficiently
eliminates the dc and twin image terms, as described before.

6.2 Pixel Resolution Control in the Fresnel Transform Method

In the Fresnel transform method of numerical diffraction, the pixel size varies linearly with
distance, as \( \delta x = \frac{\lambda z}{X_0} \). Therefore, the field of view of the image plane increases with distance,
and therefore objects larger than the CCD array can be imaged using the Fresnel transform
if the object is at a large enough distance. On the other hand, this can cause problems when
multiple holographic images with different image distances need to be combined for holographic
interferometry. Pixel resolution also varies with wavelength, which can cause problems for
multiwavelength holography methods, such as color holography or optical phase unwrapping.
One way to address the issue is by interpolation of the pixels. The amplitude images can
interpolate correctly, but the phase images likely would not because of wrapped phase at
multiples of \( 2\pi \). A better solution is provided by zero padding [142], where the hologram
is padded with zero-valued pixels around the boundary, so that the effective hologram size
\( X_0' = N \delta x \) increases linearly with the reconstruction distance \( z \), and the pixel resolution at the
image plane \( \delta x' = \frac{\lambda z}{X_0} \) remains constant.

Another method is provided in a two-step Fresnel transform over distances \( z_1 \) and \( z_2 \), such
that \( z = z_1 + z_2 \) is the total distance [143] (Fig. 20). Given the frame size of the hologram \( X_0 \),
first the Fresnel transform is calculated at an intermediate plane at \( z_1 \). The frame size there
is \( X_1 = \frac{Nz_1}{X_0} \). Then another Fresnel transform is calculated over the rest of the distance \( z_2 \),
so that the final image frame size is \( X = \frac{Nz_2}{X_1} = \frac{z_2}{z_1} X_0 \). Therefore, the final frame size can be
adjusted by proper choice of the ratio \( \frac{z_2}{z_1} \). A variation was introduced in Ref. 144 using a two-step
process. The first diffraction to an intermediate plane is calculated using an angular spectrum
method, whose pixel resolution is constant over any distance. The second diffraction to the final
image plane then has variable resolution. This method can handle small hologram-to-image
distances, whereas in the two-step Fresnel method, if the distance is very small compared to the
minimum reconstruction distance $|z| \ll z_{\text{min}}$, then the ratio $z_2/z_1$ cannot have much variation. With
two-step Fresnel transform methods, there can be a loss of information if the Fresnel frame size
$X_1 = N \lambda z_1$ is less than the actual optical diffraction size, which depends on the spectral content
of the object [75]. To avoid distortion of the final image due to the loss of diffraction field, one
of the Fresnel steps is replaced with an angular spectrum or Rayleigh-Sommerfeld calculation,
which covers a larger frame than the Fresnel frame. In Ref. 145, the shifted Fresnel transform
and tiling approach was used to control pixel resolution as well as array size, in the context of
CGH.

6.3 Optical Phase Unwrapping

Phase images generated by digital holography, as well as most other phase imaging techniques,
suffer from modulo $2\pi$ ambiguities. An object whose optical thickness variation exceeds the
wavelength produces wrapped phase images, with discontinuities at every $2\pi$ of the phase
profile. Numerous phase unwrapping algorithms have been developed [146,147], but it remains
challenging to find solutions that can efficiently address all different types of phase topologies.
This is because most of the unwrapping procedures are based on different strategies to find
the phase discontinuities and to make judgments on how to stitch the discontinuous regions.
Most often the algorithms are computationally demanding and have difficulty handling different
types of phase topologies. Optical phase unwrapping (OPU) based on multiwavelength digital
holography offers a method that is fast, efficient, and deterministic.

For example, two holograms of the same object are acquired using two different wavelengths
$\lambda_1$ and $\lambda_2$, and the phase images $\varphi_1(x, y)$ and $\varphi_2(x, y)$ are produced from them. Each of these
phase profiles range in phase from 0 to $2\pi$, and the corresponding optical thickness profiles
range up to $\lambda_1$ and $\lambda_2$, respectively. Now, subtracting the two $\Delta \varphi = \varphi_1 - \varphi_2$, followed by
adding $2\pi$ wherever $\Delta \varphi < 0$, results in a new phase image $\Phi_{12}$ that ranges from 0 to $2\pi$, whose
effective wavelength, or synthetic wavelength, is given by

$$\Lambda_{12} = \frac{\lambda_1 \lambda_2}{|\lambda_1 - \lambda_2|}.$$  \hspace{1cm} (42)

The new phase image handles optical thickness variations up to $\Lambda_{12}$, which can be made
large enough to cover the object’s maximum thickness variation by choosing small enough
wavelength differences. If, however, the original phase images have certain amounts of noise,
say $\varepsilon \cdot 2\pi$, then the new phase image contains the same amount of phase noise, which translates
to noise in the optical thickness profile, $\varepsilon \Lambda_{12}$ instead of $\varepsilon \lambda_1$, magnified by the same factor as the synthetic wavelength. The noise can be reduced back to the original level by using the new phase map as a guide to decide how to unwrap the $\varphi_1$ phase map. That is, the new phase map is given by

$$\Phi_{12} = \varphi_1 + \text{int} \left[ \frac{\Phi'_1 \Lambda_{12}}{\lambda_1} \right] \cdot 2\pi, \quad (43)$$

where int stands for integer quotient. This scheme works if the amplified noise does not exceed the original wavelength $\varepsilon \Lambda_{12} < \lambda_1$, which sets the minimum difference wavelength, and therefore the maximum synthetic wavelength:

$$|\lambda_1 - \lambda_2| > \varepsilon \lambda_2; \quad \Lambda_{12} < \frac{\lambda_1}{\varepsilon}. \quad (44)$$

If the noise is more excessive, or a larger synthetic wavelength is needed, one can continue with a hierarchical method using three or more wavelengths as follows [148]. Start from $\lambda_1$ and choose $\lambda_2 > \lambda_1$ so that $\Lambda_{12} = \frac{\lambda_1}{\varepsilon}$, i.e.

$$\lambda_2 = \frac{1}{\frac{1}{\lambda_1} - \Lambda_{12}}. \quad (45)$$

The noise associated with the new phase map of $\Lambda_{12}$ is $\varepsilon_{12} = \frac{\varepsilon}{\lambda_1}$. Next, choose $\lambda_3 > \lambda_2$ to form

$$\Lambda_{13} = \frac{\lambda_1 \lambda_3}{\lambda_3 - \lambda_1} < \Lambda_{12}, \quad (46)$$

which automatically satisfies Eq. (44). Now use $\Lambda_{12}$ and $\Lambda_{13}$ as the two new phase maps to form a new combination

$$\Lambda_{23} = \frac{\Lambda_{12} \Lambda_{13}}{\Lambda_{12} - \Lambda_{13}}, \quad (47)$$

such that $\Lambda_{23} = \frac{\lambda_1}{\varepsilon}$. The noise associated with $\Lambda_{23}$ is $\varepsilon_{23} = \frac{\varepsilon}{\lambda_2}$. In general, the $n$’th wavelength $\lambda_n > \lambda_{n-1}$ is chosen to form

$$\Lambda_{n-1,n} = \frac{\Lambda_{1,n-1} \Lambda_{1,n}}{\Lambda_{1,n-1} - \Lambda_{1,n}} = \frac{\lambda_{n-1} \lambda_n}{\lambda_n - \lambda_{n-1}}, \quad (48)$$

such that $\Lambda_{n-1,n} = \frac{\lambda_{n-2,n-1}}{\varepsilon_{n-2,n-1}}$, where $\varepsilon_{n-2,n-1} = \frac{\varepsilon}{\Lambda_{n-2,n-1}}$. The process continues until $\Lambda_{n-1,n}$ is large enough for the $z$-range of the object.

The optical phase unwrapping method was applied to quantitative phase microscopy in Refs. 23 and 149. Two-wavelength OPU can be achieved with single exposure digital holography by angular multiplexing [21,52,54,59], as shown in Fig. 21. The double interferometer consists of one common object arm and two separate reference arms illuminated by two different wavelength lasers. The reference arms are aligned so that the fringes of the two wavelengths are perpendicular to each other. On the angular spectrum, the peaks appear as two distinct pairs, so that the holographic images can be processed separately by selecting the appropriate peak for each wavelength. Figure 22(d) is an example of a DHM phase image of the surface of a polished coal sample, unwrapped by OPU, while Figs. 22(e), 22(f) are unwrapped by an available software algorithm. Evidently, the software-based method has difficulty handling isolated areas of phase profile, while the OPU generates correct phase profiles, regardless of the topology. The OPU consists of only several algebraic and Boolean operations, and therefore is
Kim: Principles and techniques...

Fig. 21  Two-wavelength holographic interferometer. The ND filters and polarizers P1 and P2 are used to control the intensity of the laser beams [59].

very fast and computational demand is low. The method is entirely deterministic and does not depend on any estimations of topology of a pixel's neighborhood.

For macroscopic measurements with thicknesses in the range of centimeters, the wavelength difference needs to be $\sim 10^{-5} \lambda$ or several gigahertz in frequency difference, which can be produced by laser cavity modulation [150] or electrooptic modulation. For microscopic imaging of $< 50 - \mu m$-thick biological cells, the wavelength difference needs to be $\Delta \lambda > 5$ nm, which would require separate lasers or a tunable laser. For larger wavelength differences, the image formation can be affected by chromatic aberration of the system. The optical elements or the object itself may have chromatic aberration, or in the case of the Fresnel transform method for numerical diffraction, the reconstructed image size depends on the wavelength. In DH, it is a simple matter to compensate for the aberration by adjustment of the reconstruction distance [151] or by subtracting reference holograms without the object [51]. The OPU was also applied to phase-shifting interference microscopy using three LEDs as light sources [152,153]. The method is fast enough for real-time vibration analysis [70].

6.4 Aberration Compensation

6.4.1 Wavefront compensation

Because of the direct numerical access to the phase profile of the wavefront, with digital holography it is possible to manipulate the phase profiles with flexibility and versatility unmatched by any other imaging methods [18,46]. For example, in DHM with microscopic magnification, the use of a curvature-matching objective lens is advantageous to reduce the fringe frequency, but it is tedious and unnecessary to align the lenses exactly. Instead, as in Fig. 23, any residual curvature can be compensated for by using a curved numerical reference wave [22,54].

Compensation of spherical aberration was demonstrated in Ref. 154 by multiplying the aberration function to the hologram and fine adjusting the parameters for best PSF. More generally, a numerical parametric lens is used for shifting, magnification, and complete aberration compensation in Ref. 155, and in Ref. 50, the Zernike polynomial fit is used to subtract curvature and aberrations from the background of the DHM image. In Ref. 156, astigmatism is corrected by effectively applying two different reconstruction distances for $x$ and $y$ axes. Correction of anamorphism is demonstrated in Ref. 157. In lensless Fourier microholography, the object can be placed close to the sensor, which increases the numerical aperture and improves resolution [102,158]. But this introduces aberration in the reconstruction because of violation of the Nyquist frequency requirement. The hologram is expanded and interpolated, followed by multiplication of a transfer function, before Fourier transform. An aberration-free high
resolution image is obtained. In Ref. 159, several methods of aberration compensation are described and compared.

### 6.4.2 Chromatic aberration

Chromatic aberration can affect multiwavelength digital holography techniques, with its main effect being the variation of the position and magnification of the image. Aberration correction in conventional microscopy entails complex and expensive compound lens design, optimized only for a limited range of parameters. In digital holography, on the other hand, images of different wavelengths can be calculated at different appropriate distances and

Fig. 22 Two-wavelength optical phase unwrapping on images of a porous coal sample: (a) amplitude image; single-wavelength phase maps reconstructed at (b) $\lambda_1 = 532$ nm and (c) $\lambda_2 = 633$ nm; (d) 3-D rendering of the dual-wavelength phase map; software unwrapped phase maps reconstructed at (e) $\lambda_1 = 532$ nm and (f) $\lambda_2 = 633$ nm for comparison. All image sizes are $98 \times 98 \mu m^2$ [59].
magrnifications adjusted for best alignment before numerically superposing the reconstructed images [151]. In Ref. 51, for two-wavelength OPU, the wavefront aberration is imaged separately without the object of interest, and subtracted from the unwrapped phase image of the object.

6.5 Diffraction with Inclined Planes

In some techniques and applications of digital holography, situations arise where one needs to consider diffraction between planes that are inclined with respect to each other. An example is for the examination of particle fields in an arbitrarily oriented plane. Another example is total internal reflection holographic microscopy [160], where the geometry of the optical system dictates holographic reconstruction on a plane at a large angle with respect to the optical axis (see Sec. 7.6). Development of integrated optics is another area where the pertinent optical plane can change in arbitrary directions.

A number of techniques have been developed, with the main strategy being avoiding direct integration of the diffraction integral so that one can continue to use the fast Fourier transform. In Ref. 161, the diffraction is described with a full Rayleigh-Sommerfeld integral equation, where the authors make the point that as far as numerical computation is concerned, there is no need for Fresnel or Fraunhofer approximation. The propagation between tilted planes is incorporated through coordinate transformation. The transformation is a rotation in the Fourier domain (Fig. 24). Using the angular spectrum method, the input optical field is Fourier transformed to obtain its angular spectrum. The wave vector of a plane wave transforms according to

\[
k'_x = k_x \cos \beta + k_z \sin \beta
\]

\[
k'_z = -k_x \sin \beta + k_z \cos \beta.
\]

(49)
Because of the rotation, the uniform intervals of frequency samples become nonuniform, and its inverse Fourier transform causes numerical problem. In Refs. 162 and 163, the spectrum is interpolated into uniform intervals. The final image, obtained by the inverse Fourier transform of the product of the angular spectrum and propagation phase factor, contains some error due to the interpolation.

A digital hologram contains full 3-D information, which leads to the possibly surprising fact that, in principle, the image can be reconstructed on a plane tilted by an arbitrary angle, up to 90 deg. The problem is that the diffraction integral is not directly amenable to the use of FFT. In Ref. 87, three coordinate transforms are utilized to achieve image reconstruction without loss of the frequency contents, and with the same pixel sizes as the hologram. The first two transformations are the rotation and tilt of the image plane, and the last is for change of the viewing angle. It incorporates a combination of the interpolation and nonuniform discrete Fourier transform (NDFT) and achieves reduced computation time. In Ref. 164, the diffraction integral is carried out using FFT along the nontilted axis, while the tilted axis is handled by direct summation of the integrand.

A conceptually straightforward but less efficient method is to build up a 3-D volume of reconstructed image pixels (voxels), and to compose the image along an arbitrarily oriented plane by collecting the appropriate pixels [113]. Reconstruction of a holographic image with variable viewing angles is described in Refs. 49 and 165. A tomographic image over an inclined plane in the image volume of digital interference holography was demonstrated in Refs. 166 and 167.

6.6 Other Numerical Techniques

6.6.1 Apodization

The sharp boundaries of hologram aperture can cause spurious fringes in reconstructed images, especially phase images. The “ringing” can be reduced by apodization of the boundary of the hologram with a smoothly attenuating function [168].

6.6.2 Extended focus imaging

The depth of focus of a microscope narrows as the square of the lateral resolution, and therefore only a thin section, often a fraction of one micron, is visible in clear focus. A DHM hologram, on the other hand, does contain information of the full 3-D space of the specimen. This property is brought out to image the full volume of an object in extended focus imaging [169]. From a single hologram, a stack of amplitude images is calculated over a range of distances. A phase image is also obtained from the hologram, which is then unwrapped to generate a height profile of the object’s surface. The unwrapped phase image is then used as a guide to extract parts of the amplitude images that are at the right distance to be in focus. These parts are stitched together to form a complete single amplitude image with all the object points in focus.

6.6.3 Focus detection

In classical imaging systems, it is difficult or impossible to recover the focused image from a defocused one. With digital holography, the image can be calculated at any distance from the hologram. To determine if an image is in focus, one can apply subjective judgment to the sharpness of the image, or such judgment can be quantified using a sharpness metric. A more deterministic method for finding the focus distance, if the specimen is either a pure amplitude or pure phase object, is given in Ref. 170, where a theorem is proven that the sum of absolute values of the complex image array is an extremum at the focal distance. For an amplitude object, \( \int \int dx \, dy \, |E(x, y; z)| \) is minimum when \( z \) is at the focus, and therefore \( E \) is real, whereas at other distances \( E \) would be complex. For a phase object, the integral is maximum at the focus, where \( E \) has constant magnitude across the frame.
6.6.4 Super-resolution

The resolution of the DH image is determined by the pixel count and the size of the CCD array, and is the main limitation of the DHM. There have been a number of methods developed to overcome the limitation, mainly by synthesizing larger apertures of the imaging system, for example by translating the CCD camera to cover a larger area [171–173]. In Ref. 174, a vertical cavity surface-emitting laser (VCSEL) array was used as the laser source, and a number of laser elements were turned on in sequence to illuminate the optical system with varying tilt angles. The combined hologram then contains larger spectral bandwidth and results in improved resolution. A grating can be used to redirect diffraction components that otherwise would leave the imaging system, and thus increase the resolution of the digital hologram [175,176].

6.6.5 Ultrahigh sensitivity

In Refs. 177 and 178, the influence of the camera shot noise on the accuracy of QPM is investigated, and it is found that a phase image with a signal-to-noise ratio (SNR) above 10 can be obtained with a mean intensity lower than 10 object photons per pixel. The quantum limit of sensitivity at less than one photon per pixel has been demonstrated in Ref. 179.

7 Applications and Special Techniques of Digital Holographic Microscopy

7.1 Biomedical Microscopy

Digital holography offers a number of novel capabilities for biomedical microscopy. Quantitative phase microscopy by digital holography (DH-QPM) has been applied to image various cell types, including SKOV-3 ovarian cancer cells [47,54], fibroblast cells [19], testate amoeba [180], diatom skeletons [181], and red blood cells [53]. It is also used to investigate various cellular dynamics such as drug-induced changes in pancreas tumor cells [55]. In Ref. 182, DHM is used to monitor laser microsurgery on red blood cells (RBCs), rat kangaroo kidney epithelial cells (PTK2), and goldfish retinal rod cells for quantitative evaluation of the damage and repair of cells and cellular organelles in real time. In Ref. 53, live erythrocyte (red blood cell) membrane fluctuations of 37 nm are measured and compared with ethanol-fixed cells, which had much smaller 5-nm fluctuation. The contributions to optical thickness from physical thickness and index of refraction are decoupled by phase mapping with two perfusion solutions of different refractive indices. The RBC refractive index is measured to be $n = 1.394 \pm 0.008$. Deformation of phospholipid vesicles in a microchannel flow was measured by DH-QPM, as a model of red blood cells in capillary flow [58].

Time-varying cellular and subcellular features are imaged with submicron, diffraction-limited resolution [19]. Movies of holographic amplitude and phase images of living microbes and cells, such as paramecium swimming among other microbes and fibroblast cells in the process of migration, are created from a series of holograms and reconstructed with numerically adjustable focus.

Another notable technique for quantitative phase microscopy is the Fourier phase microscope (FPM), which has a configuration that resembles the Zernike phase microscope but with the phase filter replaced with a spatial light modulator to allow phase-shift image acquisition in a common path interferometer [43], as depicted in Fig. 25.

Digital Gabor holography is excellent for imaging 3-D distribution of particulate microbes [25,109]. Taking the difference between consecutive frames of a time series is especially useful for tracking the motion of the microbes, which also efficiently subtracts the background noise. Underwater instruments have been constructed for monitoring marine planktons and particulates [117,118]. DGH was used to investigate the flow field generated by the appendages of a copepod [119]. A sample volume of water containing copepods is seeded with tracer particles and illuminated with an expanded and collimated HeNe laser. The 3-D positions
of the copepod and tracer particles can be determined from the numerical reconstruction of
the hologram, but the axial resolution (∼500 μm) is much poorer than the lateral resolution
(∼7.4 μm). The disparity in resolution is overcome by using an inclined mirror to form double
images with perpendicular views (Fig. 26), thus providing isotropic 3-D resolutions. Analysis
of the holographic movies revealed the velocity field and trajectories of particles entrained by
the copepod having a recirculating pattern in the copepod’s frame of reference. The pattern is
caused by the copepod sinking at a rate that is lower than its terminal sinking speed, due to
the propulsive force generated by the feeding current. Propulsive force generated by its feeding
appendages was measured to be $1.8 \times 10^{-8}$ N.

Deformation of tissues under mechanical impact has been imaged using double-pulse digital
holography [183]. Recognition and identification of biological cells by digital holography was
studied [184].

### 7.2 Particle Field Holography

Particle imaging velocimetry (PIV) has been an important application area of conventional
holography [185–187], but several factors have limited its wider practical applications, such
as difficulty with real-time imaging and instrument complexity. Digital Gabor holography
can remove much of these limitations by allowing real-time monitoring of the 3-D positions
and distribution of the particle field with significantly reduced instrument complexity and
maintenance [114,188]. Fundamental issues of holographic PIV (HPIV) are reviewed in Ref. 115, and it is seen that digital HPIV can revitalize holographic particle imaging [88,89,109].

Xu et al. studied the 3-D position measurements of microspheres by DGH and obtained 50-nm accuracy in all three directions [116]. Instead of using the intensity profile of the reconstructed image, the minimization of the variance of the imaginary part of the complex amplitude was found to improve the measurement of the location of the focal plane [110,170].

The single-beam two-view setup, as in Fig. 26, has been used in general particle imaging applications [111], and a low-coherence source is used for particle imaging in Ref. 106 to reduce the coherent noise. Dubois and Grosfils introduced dark-field digital holography for studying nanoparticles smaller than the optical resolution [189]. DGH is also used for measuring the 3-D position and orientation of a microfiber segment [112]. Biomedical particulate imaging applications of DGH have been noted earlier.

7.3 Microscopy and Metrology of Microstructures

DHM is particularly useful for MEMS characterization because of the relatively smooth and well-defined surface profile [60,190]. Microcantilever beams, bridges, and membranes are imaged by quantitative phase holography [22]. The multiwavelength optical phase unwrapping of phase images yields shape and deformation measurements with submicron precision over a many-micron range [151,169,190]. Various optical techniques, including digital holography, for characterization of MEMS devices are reviewed in Ref. 191.

Quantitative phase microscopy by DH also provides a unique tool for monitoring nonlinear optics processes, such as ferroelectric crystal domains [61,192]. Digital holography was used to measure the refractive index changes during femtosecond laser writing of waveguides in Ti$^{3+}$-doped sapphire [193], as well as waveguides written in glass substrates using a Ti:sapphire laser [194]. DH is used in high precision 3-D angle measurements by analyzing the angular spectrum of the holographic interference [195]. Precision of 0.005 arc sec was obtained. Polished coal particle surfaces were analyzed [59]. Tunable liquid microlens arrays are characterized by digital holography [57]. In Ref. 196, the second-harmonic generation at the glass-air surface was studied by setting up an interference between the second harmonic of the reference wave and the SHG signal from the interface using focused femtosecond laser illumination. Imaging through a nonlinear medium was demonstrated in Ref. 197.

Birefringence of a material can be imaged by using two references with orthogonal polarizations [198]. The two references are angle multiplexed, as in two-wavelength holography, to yield two pairs of spectra in the Fourier domain.

7.4 Holographic Tomography

Although holography reproduces the 3-D optical field of the object, a holographic image at a given distance contains not only the in-focus image but also contributions from out-of-focus planes. Quantitative phase microscopy by DHM also profiles the accumulated phase variation along the direction of light propagation. On the other hand, the 3-D scattering amplitude or refractive index variation over the object volume is of great interest in many applications in biomedical imaging, metrology, particle field analysis, etc. There are two main approaches for achieving such tomographic images. One is by acquiring the transmitted optical field from numerous different directions and combining the projections by a back-propagation algorithm. The other is by using the low-coherence interference effects to isolate the interaction of light with a cross section of the object volume.

Analogous to computer-aided x-ray tomography (CAT), holographic tomography of a 3-D phase structure has been achieved by rotating the cellular specimen in the object volume [180,181,199]. The technique is also demonstrated for tomographic profiles of microfiber [200,201]. Instead of full rotation, the illumination angle is scanned over a finite range in [202,203]. In contrast to shadow projection in other tomography methods, digital holography
allows back-propagation that includes diffraction effects for more accurate and higher resolution reconstruction—also referred to as optical diffraction tomography [204,205].

7.5 Low-Coherence Holography

Use of low-coherence light in general holography has a long history, and it has been applied in digital holography mainly for two purposes: reducing spurious coherent noise and generating tomographic or topographic images. The short coherence length can be used in tomographic or topographic imaging by extracting only the part of an object that is within the coherence length with respect to the reference beam. The first such demonstrations were light-in-flight digital holography in Refs. 206 and 207, where parts of the reference beam were delayed by different amounts, and the reconstructed image displayed different parts of the object with correspondingly different distances.

For tomographic and topographic imaging, depth resolution is determined by the laser coherence length. For example, a laser spectral width of 30 nm corresponds to about 10-μm coherence length. Topographic imaging with 20-μm depth resolution was demonstrated using a diode laser [208]. Unlike low-coherence interference imaging, the hologram or the camera is focused an arbitrary distance from the object plane, and the image of the object is reconstructed by numerical propagation. Therefore, interference imaging can be considered a special case of holography.

Another useful effect of low-coherence light is the reduction of spurious interference noise from parts of the optical system that are not within the coherence length [83]. A light source with controllable spatial coherence is possible by focusing a laser light onto a rotating diffuser glass plate [105]. The method is applied to particle field analysis [106] and in fluid concentration measurements [58]. In Ref. 209, several holograms using different wavelengths are superposed to observe the reduction in speckle noise.

7.6 Total Internal Reflection Holographic Microscopy

Microscopy of cellular adhesion is important for a deeper understanding of cellular motion and morphogenesis. The primary tools for imaging and studying these surface processes have been total internal reflection fluorescence microscopy (TIRFM) and interference reflection microscopy. In TIRFM, the evanescent field is used to selectively illuminate the contact layer of the cell. In contrast, in TIR holographic microscopy (TIRHM), the interaction of the evanescent field with the contact layer results in the modulation of the phase profile in the TIR-reflected light [160,164]. Phase modulation is then detected and imaged by the quantitative phase microscopy of DHM. The method is noninvasive, does not require fluorophores, and makes full use of the input irradiance. The geometry of the optical system necessitates numerical imaging over a plane at a large inclination angle, but again digital holography provides a unique capability for reconstruction over an inclined plane as described earlier. We note that the TIR has also been used for illumination of the object in digital holography for dark-field microscopy [210]. A TIRHM apparatus based on a Mach-Zehnder interferometer is depicted in Fig. 27, and an example set of the amoeba proteus images is shown in Fig. 28.

7.7 Optical Scanning Holography

Instead of setting up interference of object and reference fields at the hologram plane, optical scanning holography (OSH) [211–213] illuminates the object with an interference field created by the overlap of two coherent beams, which is then scanned across the object (Fig. 29). The light transmitted or reflected by the object is then collected and detected by a point detector. A detailed theoretical description is given in Refs. 214 and 215. The detected light may in fact be a secondary emission, such as fluorescence, and does not need to be coherent. For this reason, OSH has a unique potential for 3-D fluorescence microscopy [216]. In fact, the detection
Fig. 27  Apparatus for TIRHM: (a) BS: beamsplitters; M: mirrors; A: object plane; H: hologram plane. (b) Detail of TIR prism.

Fig. 28  TIRHM images of live *Amoeba proteus*: (a) hologram, (b) amplitude image, and (c) phase image. FOV = 250 μm (tilt-compressed) × 125 μm [164].

Fig. 29  Fresnel zone patterns at two different distances. Scanning of the patterns across the points $P_1$ or $P_2$ records the 3-D positions of these points.
mode can be varied so that the coherence property of the imaging process ranges from linear in amplitude to linear in intensity [217].

Another technique for holography with incoherent light, called space-time holography, is based on time-modulated interference between the object wave and a delta-function reference pupil [218]. A time series of the interference images is analyzed to extract the modulation frequency component from each pixel, which constitutes the single sideband complex hologram [219]. A low-coherence gating method is demonstrated to produce tomographic images of an object (an insect wing) behind a ground glass [220].

7.8 Digital Interference Holography

Low-coherence tomography can be synthesized by using a diversity of wavelengths [221,222]. In digital interference holography (DIH), the tomography of a 3-D volume is constructed by acquiring a series of holograms while the wavelength is scanned over a range [223,224] (Fig. 30). Suppose an object is illuminated by a laser beam of wavelength $\lambda$. A point $r_0$ on the object scatters the light into a Huygens wavelet,

$$A(r_0) \exp(ik |r - r_0|),$$

where the object function $A(r_0)$ is proportional to the amplitude and phase of the wavelet scattered or emitted by object points. For an extended object, the field at $r$ is

$$E(r) \sim \int A(r_0) \exp(ik |r - r_0|)d^3r_0,$$

where the integral is over the object volume. The amplitude and phase of this field at the hologram plane $z = 0$ is recorded by the hologram as $H(x_h, y_h; \lambda)$. The holographic process is repeated using $N$ different wavelengths, generating the holograms $H(x_h, y_h; \lambda_1), H(x_h, y_h; \lambda_2), \ldots, H(x_h, y_h; \lambda_N)$. From each of the holograms, the field $E(x, y, z; \lambda)$ is calculated as a complex 3-D array over the volume in the vicinity of the object (Fig. 30). Superposition of these $N$ 3-D arrays results in

$$\sum_k \int A(r_0) \exp(ik |r - r_0|)d^3r_0 \sim \int A(r_0)\delta(r - r_0)d^3r_0 \sim A(r).$$

That is, for a large enough number of wavelengths, the resultant field is proportional to the field at the object and is nonzero only at the object points. In practice, if one uses a finite number...
Fig. 31 Examples of DIH topography and tomography images. (a), (b), and (c) show perspective rendering of a 3-D volume image data of a finger imprint by DIH from few different viewing angles. The image volume is 4.86 mm × 4.86 mm × 0.210 mm [225]. (d) and (e) show the reconstructed volume of the human optic nerve sample: (d) x-y cross section, FOV = 1100 × 1100 μm²; (e) y-z cross sections, 1100 × 280.35 μm², at x1, x2, and x3; (f) x-z cross sections, 280.35 × 1100 μm², at y1, y2, and y3 [226].

N of wavelengths with uniform increment Δ(1/λ) of the inverse wavelengths, then the object image $A(r)$ repeats itself (other than the diffraction/defocusing effect of propagation) at a beat wavelength $\Lambda = [\Delta(1/\lambda)]^{-1}$, with axial resolution $\delta = \Lambda / N$. By use of appropriate values of $\Delta(1/\lambda)$ and $N$, the beat wavelength $\Lambda$ can be matched to the axial range of the object, and $\delta$ to the desired level of axial resolution.

DIH has been applied to tomographic imaging of retinal tissues [226] and for generating 3-D profiles of fingerprint [225] (Fig. 31). DIH reconstruction over an arbitrary inclined plane has been demonstrated [166,167]. A superluminescent diode (SLD) and an acousto-optic tunable filter was used as the variable wavelength source in Ref. 85. Submicron tomographic depth resolution has been demonstrated [227], and it was used for 3-D profiling of erythrocyte membranes [228].

7.9 Heterodyne Holography

Heterodyne digital holography is an extension of phase-shifting digital holography [229]. Using a pair of AOMs, the reference is frequency shifted relative to the object beam by one quarter of the CCD frame rate (Fig. 32), so that four consecutive frames acquire one set of the quadrature phase-shifted images. Heterodyne holography in off-axis configuration was used for filtering of residual zero-order and spurious components to achieve ultimate sensitivity [179]. Heterodyne
digital holography, together with TIR illumination, is used to image subwavelength-size gold particles [210].

In time-averaged digital holography, a vibrating surface displays dark fringes at the zeros of the Bessel function $J_0(z)$, where $z$ is the amplitude of vibration [92,230]. For large-amplitude vibrations, the fringes become too numerous and difficult to resolve. Sideband digital holography uses heterodyne digital holography with the frequency offset of the reference tuned to any $n$'th harmonic of the vibration frequency, allowing selective observation of the large-amplitude areas [231].

8 Conclusions
We present a review of a subset of research and development in digital holography, focusing on microscopy techniques and applications. Over a relatively short history, interests and activities have grown exponentially, as a cursory look at publication statistics shows. An increasing number of researchers—not only in optical physics and optical engineering, but also in diverse applications areas such as microbiology, medicine, marine science, particle analysis, microelectromechanics, and metrology—are realizing the new capabilities of digital holography. First of all, with the replacement of cumbersome photochemical processes with optoelectronic imaging, DH allows orders-of-magnitude improvement in acquisition speed and sensitivity. Second, the numerical representation of optical fields yields a host of interference and other manipulation techniques that are difficult or infeasible in real-space holography. Lastly, the availability of the complex phase of optical fields leads to quantitative phase microscopy in a direct, highly sensitive, and versatile manner. Indeed, digital holography is seen to be a breakthrough that can revitalize the general area of holography, whose development has been somewhat stagnant and applications constrained. With the ever-accelerating development of electronic imaging and computing technologies, it is easy to envision a near future when, for example, pixel resolution will be comparable to photochemical media and computational load will become not much of an issue. With such technological developments, new holographic imaging capabilities yet to be conceptually imagined will emerge. Acceleration of the advances will also require re-examination of many preconceptions inherited from conventional holography, such as general approaches to apparatus setup, behavior of noise, and range of application areas. This is indeed an exciting time for digital holography in general, and it is hoped that this review will help whet the curiosities of more capable students and researchers for the field.

Acknowledgment
The author gratefully acknowledges the financial support of the National Science Foundation.
References


[211] T.C. Poon, “Scanning holography and two-dimensional image-processing by acoustoopt-


Dimensional microscopy by optical-scanning holography,” Opt. Eng. 34, 1338–1344


“Twin-image elimination experiments for three-dimensional images in optical scanning


graphy as a technique for high-resolution three-dimensional biological microscopy,” J.


three-dimensional display using an electron-beam-addressed spatial light modulator,”


1508 (1997).


microscopic imaging scheme with an arbitrary degree of spatial coherence,” Appl. Phys.


[220] G. Indebetouw and P. Klysubun, “Imaging through scattering media with depth resolution


214 (2000).

[221] J.C. Marron and T.J. Schulz, “3-dimensional, fine-resolution imaging using laser fre-


[223] M.K. Kim, “Tomographic three-dimensional imaging of a biological specimen using


[227] F. Montfort, T. Colomb, F. Charriere, J. Kuhn, P. Marquet, E. Cuche, S. Herminjard,

and C. Depeursinge, “Submicrometer optical tomography by multiple-wavelength digital


[228] J. Kuhn, F. Montfort, T. Colomb, B. Rappaz, C. Moratal, N. Pavillon, P. Marquet, and

C. Depeursinge, “Submicrometer tomography of cells by multiple-wavelength digital


[229] F. Le Clerc, L. Collot, and M. Gross, “Numerical heterodyne holography with two-


[231] F. Joud, F. Laloe, M. Atlan, J. Hare, and M. Gross, “Imaging a vibrating object by

Myung K. Kim is a professor of physics at the University of South Florida (USF), Tampa, and directs the Digital Holography and Microscopy Laboratory. His PhD degree is from the University of California, Berkeley in 1986. After postdoctoral research at the SRI International in Menlo Park, California, he went on to an assistant professor position at Wayne State University, Detroit, Michigan, in 1988, where he stayed until his move to USF in 1995. His current research interests are in the development of novel techniques and applications in digital holography, microscopy, interference imaging, optical tomography, and biomedical imaging, as well as manipulation of particles and microbes by optical force. His publication list includes more than 120 publications and eight patents granted or pending.