Resolution limits in practical digital holographic systems

Damien P. Kelly
Bryan M. Hennelly, MEMBER SPIE
Nitesh Pandey, MEMBER SPIE
National University of Ireland, Maynooth
Department of Computer Science
Maynooth, Kildare County
Ireland
E-mail: damienpk@cs.nuim.ie

Thomas J. Naughton, MEMBER SPIE
National University of Ireland, Maynooth
Department of Computer Science
Maynooth, Kildare County
Ireland
and
University of Oulu
RFMedia Laboratory
Oulu Southern Institute
Vierimaantie 5
84100 Ylivieska
Finland

William T. Rhodes, FELLOW SPIE
Florida Atlantic University
Imaging Technology Center
777 Glades Road
Building 43, Room 486
Boca Raton, Florida 33431

Abstract. We examine some fundamental theoretical limits on the ability of practical digital holography (DH) systems to resolve detail in an image. Unlike conventional diffraction-limited imaging systems, where a projected image of the limiting aperture is used to define the system performance, there are at least three major effects that determine the performance of a DH system: (*i*) The spacing between adjacent pixels on the CCD, (*ii*) an averaging effect introduced by the finite size of these pixels, and (*iii*) the finite extent of the camera face itself. Using a theoretical model, we define a single expression that accounts for all these physical effects. With this model, we explore several different DH recording techniques: off-axis and inline, considering both the dc terms, as well as the real and twin images that are features of the holographic recording process. Our analysis shows that the imaging operation is shift variant and we demonstrate this using a simple example. We examine how our theoretical model can be used to optimize CCD design for lensless DH capture. We present a series of experimental results to confirm the validity of our theoretical model, demonstrating recovery of super-Nyquist frequencies for the first time. [DOI: 10.1117/1.3212678]

Subject terms: Holography; digital imaging; interference.

Paper 080958RR received Dec. 10, 2008; revised manuscript received Jun. 27, 2009; accepted for publication Jul. 7, 2009; published online Sep. 4, 2009.

1 Introduction

Holography is a technique that enables the magnitude and phase of an optical field to be recorded.[1,2] Because recording media, in general, are sensitive only to the intensity of the light field incident upon them, a more complex optical setup is required to record the phase of the incident field. To record the phase information of the object field, it must be interfered with a so-called reference beam at the recording plane, thereby encoding the phase information as intensity variations.[1,3–6] In traditional holography, these intensity variations are recorded on a photosensitive material. Digital holography (DH) is an extension of this technique where the photosensitive material is replaced with a digital camera.[7–9] Recording the hologram electronically means that this information can be stored and manipulated in real time. The flexibility this allows the user in the recording, processing, and remote replaying of holograms is one of the primary advantages of a digital holographic approach. DH has growing applications in 3D recording and display devices,[10–12] as well as industrial metrology systems[13] with potentially far-reaching implications for cellular and microelectromechanical systems imaging.[14,15] One significant drawback of this technique however is the limited space-bandwidth product[16] of the recorded digital holograms, due in large part to the limited number of pixels in a typical CCD camera. These cameras are discrete devices, and therefore, sampling theory plays an important role in modeling and optimizing DH setups. Other factors, such as the finite size of the camera and the finite pixel size of the individual detectors, both act to significantly limit the performance of DH systems. This is not unexpected and has been discussed by other authors (see, for example, Refs. 17–21). What perhaps is not as well appreciated in the DH community is that it is not necessary for the field in the recording plane to be sampled at the Nyquist limit.[17–21] In the following analysis, we examine lensless DH exclusively.

In Fig. 1, a typical (in-line) optical setup for performing DH is illustrated. At the camera face, a reference wave and a field that has been scattered from the object interfere and the resulting intensity pattern is recorded by the camera. This pattern contains the dc terms, the real and virtual image. Typically, the user is interested in one of the image terms and the other elements only act to degrade the quality of the reconstructed hologram. Removal of the dc terms and the virtual (or real) image is thus an important practical consideration in all holographic systems, and accordingly, many different techniques have been proposed to get rid of these unwanted terms.[9,22–24] Let us assume for the moment

0091-3286/2009/$25.00

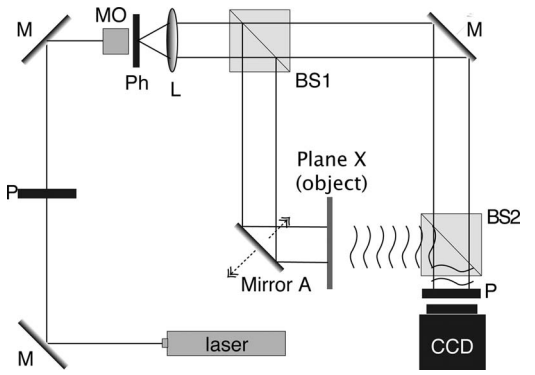

Fig. 1 Schematic depicting a typical inline DH setup: M, mirror; P, polarizer; BS, beamsplitter; Ph, pinhole; L lens; and MO, microscope objective.

that the complex field (the real image term), $u_z(x)$, is available to us (i.e., that we have somehow removed the dc and virtual terms). We note then that once the complex field, $u_z(x)$, at the camera plane has been obtained, the field at the object plane, $u(X)$ (the reconstructed image), can be calculated digitally using numerical techniques by a computer. Suppose that this field $u(X)$ has a maximum spatial frequency given by $f_{\max}$. $u(X)$ now propagates (described under paraxial conditions with the Fresnel transform) to the camera face, where its complex amplitude, $u_z(x)$, is recorded. It is a property of the Fresnel transform that the magnitude of the field's spectral distribution remains invariant under propagation and so $u_z(x)$ must also have a maximum spatial frequency $f_{\max}$ (see Refs. 25 and 26). A camera's spatial frequency bandwidth, $B_c=2f_c$, and thus, its ability to resolve a spatial frequency at the camera plane is determined by the distance between the centers of adjacent pixels on the camera face. Therefore, on initial consideration, one might conclude that the camera must be able to resolve $f_{\max}$ (i.e., $f_c \geqslant f_{\max}$) if the object field is to be reconstructed properly. However, we note that several authors[27–32] have shown that this sampling criterion may be too strict and that it is possible to recover the object signal when its Fresnel transform is sampled in the camera plane (over an infinite extent) at a rate lower than the Nyquist limit. These results are clearly of interest in DH and may have practical implications: the camera could be placed closer to the object with a resultant increase in numerical aperture and, thus, an improvement in 3D perspective. In Ref. 30, the authors describe recovery of these super-Nyquist frequencies in terms of a generalized sampling theory (GST). In Ref. 29, the authors apply the GST to DH systems however do not examine the effect of the finite pixel size. In a later paper,[33] these authors do consider the effect of pixel size, concluding that the maximum recoverable frequency, $f_{\max}=1/2\zeta$, where 2ζ is the width of the pixel [see Eq. (7) in Ref. 33]. As we shall see, however, by suitably designing a camera, this limitation can be overcome. In Refs. 20 and 21 resolution limits in DH systems are also examined, however only for signals that are considered well sampled in the Nyquist sense. In this paper, we develop a theoretical model that describes the limitations on resolution that are imposed by (*i*) the finite extent of the camera, (*ii*) the sampling rate, and (*iii*) the finite extent of the pixels. We note that if the camera pixels can be considered point detectors (described by a comb of Dirac δ functions), then the analysis we present for the real image term reduces to that discussed in Ref. 34. Also, in Refs. 20 and 21 the authors derive a similar expression to that discussed in Ref. 34. In our model, we treat the averaging effect introduced by finite-size pixels in a different manner to that discussed in Refs. 20, 21, and 34, providing an alternative means of understanding this effect.

The paper is organized as follows; In Sec. 2, we develop a theoretical model describing the imaging process that incorporates (*i*) the finite extent of the CCD, (*ii*) the reduction of power in higher spatial frequencies due to averaging introduced by rectangular size pixels, and (*iii*) the effective sampling rate imposed by the spacing between adjacent pixels on the CCD face. We specifically examine the dc terms and the twin image. Using this theoretical framework in Sec. 3, we examine the predictions of our model for several different recording architectures, off-axis and inline, and show that recovery of super-Nyquist frequencies is not possible using an off-axis configuration. We then examine how to optimally design CCD sensors to resolve frequencies much higher than the Nyquist limit and that limit defined in Ref. 33, by balancing these three different effects [i.e., (*i*), (*ii*), and (*iii*)]. In Sec. 4, we demonstrate that the imaging operation performed by a DH system is a shift variant using a simple example. In Sec. 5, we provide a series of experimental results that clearly demonstrate the limitations on resolution in a DH system imposed by the three different factors identified in our theoretical model. We provide what we believe is the first experimental evidence demonstrating that we can recover frequencies above the Nyquist limit of the CCD camera. A fast reconstruction algorithm based on the fast Fourier transform (FFT) is discussed. Finally, we close the paper with a brief conclusion.

2 Theoretical Analysis

In Fig. 1, an object is illuminated with a temporally and spatially coherent monochromatic plane wave. We describe the resulting scattered field at plane X (see Fig. 1) by the function $u(X)$. This field then propagates to the camera plane (located in the plane $z=z_c$), where it interferes with a reference wavefield $u_R(x)$, and the resulting intensity is recorded by the CCD. Through a numerical reconstruction process, where we simulate free-space propagation back to the object plane (plane X, see Fig. 1), we can approximately recover $u(X)$. There are several features of the recording process, however, that limit the accuracy of our recovered signal: (*i*) The finite extent of the camera, $2D$, (*ii*) the spacing between the centers of adjacent pixels, T, and (*iii*) the finite extent, 2ζ of the pixels themselves. In this section, we investigate each of these effects. We begin by writing the continuous and instantaneous field intensity, $I_c(x;t)$, at the camera face as[17,18,35]

$$I_c(x;t) = |u_z(x) + u_R(x)|^2 = I_z(x) + I_R(x) + u_z^*(x)u_R(x) + u_z(x)u_R^*(x), \quad (1)$$

where $I_z(x)$ and $I_R(x)$ are the intensities of the object and reference fields, respectively, and are referred to as the dc terms. The two latter terms in Eq. (1) contain the virtual

and real images, respectively, and the superscripted asterisk denotes the complex conjugate operation. Under paraxial conditions, we may relate the field $u_z(x)$ to $u(X)$ using a Fresnel transform,[3] which we define as

$$u_z(x) = \chi_z\{u(X)\}(x),$$

$$u_z(x) = \frac{1}{\sqrt{j\lambda z}} \int u(X)\exp\left[\frac{j\pi}{\lambda z}(x-X)^2\right]dX, \tag{2}$$

$$u(X) = \frac{1}{\sqrt{-j\lambda z}} \int u_z(x)\exp\left[\frac{-j\pi}{\lambda z}(X-x)^2\right]dx, \tag{3}$$

where χ is the Fresnel transform operator. In Eq. (3), the inverse operation is defined. In what follows, we drop the constant term $1/\sqrt{\pm j\lambda z}$ from Eqs. (2) and (3). The CCD has a finite number of pixels, with an assumed uniform structure, that average the light energy incident upon them (see Chap. 9 of Ref. 36). The measured quantity is referred to as the integrated intensity and has units of energy. A camera consists of an array of $N \times N$ pixels separated from each other by a distance T. We can represent (in one dimension) the integrated intensity array using the vector

$$W^n = [W^1, W^2, \ldots \ldots W^N] = W(x';t)\delta_T(x')p_D(x'), \tag{4}$$

where $p_D(x')=1$, when $|x'|<D$ and is 0 otherwise. Thus, $p_D(x')$ is an aperture function that defines the extent of the camera face. The function $\delta_T(x)$ in Eq. (4) is a train of Dirac δ functions[37] and is defined as $\delta_T(x)=\Sigma_{n=-\infty}^{\infty}\delta(x-nT)$. Because the function $\delta_T(x)$ is periodic, it may also be expressed mathematically using a Fourier series representation,[38,39]

$$\delta_T(x) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \exp\left(\frac{j2\pi nx}{T}\right). \tag{5}$$

The function $W(x';t)$ is continuous and is related to the temporally and spatially varying intensity $I_c(x;t)$ at the camera face by

$$W(x';t) = \frac{1}{2\zeta}\int_t^{t+\Delta t}\int_{x'-\zeta}^{x'+\zeta} I_c(x;t)dxdt, \tag{6}$$

where Δt is the integration time of the camera and 2ζ is the width (area) of the pixel that is sensitive to light, and is related to the fill factor of the camera; $\zeta \leq T/2$.[17,18,35,40] If we assume a stationary object and note that the illumination source is coherent and monochromatic, then the intensity of the light field will not vary over the integration time of the camera. Thus, we need only consider the spatial variation of intensity over each pixel area, and thus, we rewrite Eq. (6) as

$$W(x') = \frac{C}{2\zeta}\int_{x'-\zeta}^{x'+\zeta} I_c(x)dx, \tag{7}$$

where C is an unimportant constant. We now reinterpret Eq. (7) as a convolution relation;

$$W(x') = \frac{C}{2\zeta}\int_{-\infty}^{\infty} p_\zeta(x-x')I_c(x)dx$$

$$= \frac{1}{2\zeta}I_c(x') * p_\zeta(x'), \tag{8}$$

where $p_\zeta(x')=1$, when $|x'|<\zeta$ and is 0 otherwise, and where $*$ indicates a convolution operation. Substituting Eqs. (1) and (8) into Eq. (4) and dropping the unimportant scaling constant C, we arrive at the following result:[20,21,34]

$$W^n = \frac{1}{2\zeta}p_D(x)\delta_T(x)\{p_\zeta(x) * [I_z(x) + I_R(x) + u_z^*(x)u_R(x) + u_z(x)u_R^*(x)]\}, \tag{9}$$

where for notational simplicity, we set $x' \rightarrow x$. Thus, we see that the discrete vector of values returned by the camera arise due to contributions from four separate sources: the two dc terms and both the real and twin image. If we simply apply an inverse Fresnel transform to Eq. (9), then we find that we will indeed arrive at an approximation to $u(X)$, which arises due to the contribution of the real image [i.e., the fourth term in Eq. (9)]; however, this result will, in general, be effected by the contributions of both the dc and twin image terms. We note that the Fresnel operation is linear, and thus, in the following subsections, we consider the terms in Eq. (9), separately.

2.1 *DC Terms*

The dc terms in Eq. (9) can be removed by recording the intensities of the reference and object fields separately and then subtracting them digitally from the captured hologram. We note that other numerical approaches can also be used to suppress these terms.[8,41] Phase-shifting interferometric techniques can also be used to remove both the twin image and the dc terms. Although these dc terms do effect the quality of the reconstructed hologram, they are relatively unimportant compared to the behavior of the twin and real-image terms.

2.2 *Real Image Term*

We now turn our attention to numerically reconstructing an approximation to the continuous field $u(X)$. First, however, we simplify the fourth term in Eq. (9) further by assuming an ideal unit amplitude in-line reference beam, $u_R(x) = \exp[j2\pi(z-z_c)/\lambda]$. Setting $z=z_c$ and applying an inverse Fresnel transform on the real-image term, we get the following result for the reconstructed image:

$$u^s(X) = \frac{1}{2\zeta}\int [u_z(x) * p_\zeta(x)]\delta_T(x)p_D(x) \times \exp\left[\frac{-j\pi}{\lambda z}(X-x)^2\right]dx. \tag{10}$$

Using the results from Appendix A, we can rewrite Eq. (10) as

$$u^s(X) = \exp\left(\frac{-j\pi}{\lambda z}X^2\right)\int u(X_1)\exp\left(\frac{j\pi}{\lambda z}X_1^2\right)\Psi(X,X_1)dX_1, \tag{11}$$

where

$$\Psi(X,X_1) = \int B(x,X_1)\,\delta_T(x)p_D(x)\exp\left[\frac{j2\pi}{\lambda z}x(X-X_1)\right]dx = \int_{-D}^{D} B(x,X_1)\,\delta_T(x)\exp\left[\frac{j2\pi}{\lambda z}x(X-X_1)\right]dx, \tag{12}$$

and where

$$B(x,X_1) = \frac{1}{2\zeta}\int_{-\zeta}^{\zeta}\exp\left(\frac{j\pi}{\lambda z}u^2\right)\exp\left[\frac{-j2\pi}{\lambda z}(x-X_1)u\right]du. \tag{13}$$

From Eqs. (10)–(13), we can see that there is a complex interaction between the finite camera extent $p_D(x)$, the sampling rate $\delta_T(x)$, and the averaging due to the finite pixel extent $p_\zeta(x)$. In order to gain some insight into how these different factors effect the numerically calculated reconstruction, $u^s(X)$, we apply a series of limiting operations to Eq. (10). Initially, we will allow the camera extent approach infinity (i.e., $D\to\infty$) and the finite extent of the pixels to approach Dirac δ functions (i.e., $\zeta\to 0$). Following this, we examine, separately, how the finite camera aperture and how the finite pixel size change this initial result. Finally, we will discuss the interaction of all three factors.

2.2.1 *Infinitely large camera face with infinitely narrow pixels: $D\to\infty$, $\zeta\to 0$*

Making use of Eq. (5), letting $D\to\infty$ and $\zeta\to 0$ reduces Eq. (10) to the following:

$$u^s(X) = \frac{1}{T\sqrt{-j\lambda z}}\int u_z(x)\,\delta_T(x)\exp\left[\frac{-j\pi}{\lambda z}(X-x)^2\right]dx = \frac{1}{T\sqrt{-j\lambda z}}\sum_{n=-\infty}^{\infty}\int u_z(x)\exp\left(\frac{j2\pi nx}{T}\right)\times\exp\left[\frac{-j\pi}{\lambda z}(X-x)^2\right]dx. \tag{14}$$

We also note the shifting property[27,42] of the Fresnel transform (for an arbitrary linear phase ξ), for some analytical signal $f(X)$,

$$\chi_z\{f(X)\exp(j2\pi\xi X)\}(x) = \exp\left(\frac{-j\pi\xi^2}{\lambda z}\right)\exp(j2\pi x\xi)\chi_z\{f(X)\}(x-\xi\lambda z). \tag{15}$$

Combining the results from Eqs. (14) and (15), we arrive at

$$u^s(X) = \frac{1}{T\sqrt{-j\lambda z}}\sum_{n=-\infty}^{\infty}\int u_z(x)\exp(j2\pi nx/T)\times\exp\left[\frac{-j\pi}{\lambda z}(X-x)^2\right]dx,$$

$$u^s(X) = \frac{1}{T}\sum_{n=-\infty}^{\infty}\chi_{-z}\left\{u_z(x)\exp\left[j2\pi\left(\frac{n}{T}\right)x\right]\right\}(X),$$

$$u^s(X) = \frac{1}{T}\sum_{n=-\infty}^{\infty}\exp\left[\frac{-j\pi(n/T)^2}{\lambda z}\right]\exp(j2\pi Xn/T)\times\chi_{-z}\{u_z(x)\}\left(X-\frac{n\lambda z}{T}\right). \tag{16}$$

Thus, from Eq. (16), we can relate $u^s(X)$ to the actual field $u(X)$. The sampling process however causes differences between the actual signal and our approximation to it. We note several points in relation to this: (*i*) the sampling process creates an infinite number of replicas in the object plane, (*ii*) the centers of adjacent replicas are separated by a distance $\lambda z/T$, (*iii*) each of the replicas is also multiplied by a different linear phase as well as some unimportant constant phase factor.

If we impose the constraint that our object (field) has a finite support Δ in the object plane, then this field can be imaged without overlapping replicas provided that $T \leqslant (\lambda z)/\Delta$. This important result is known[27,28,30,32–34,43] and means that, under certain conditions, it is possible to sample the diffracted field at rates below the Nyquist limit and to recover, through a generalized interpolation formula, super-Nyquist frequencies. In Ref. 26, some implications of this result are explored in more detail using a simple analytical example. As we will see, however, the effect of the finite pixel size and camera extent impose resolutions limits in addition to this constraint.

2.2.2 *Infinitely large camera face and averaging due to finite pixel extent: $D\to\infty$*

In this section, we look at the effect of averaging due to the finite extent of the pixels in the camera plane and examine how this impacts on our reconstructed approximation $u^s(X)$. In particular, we note the effect on the resolution obtainable in DH systems. If we now include the effect of pixel averaging in our analysis, then we replace $u_z(x)$ with $u_z(x)*p_\zeta(x)$ in Eq. (16), to give

$$u^s(X) = \frac{1}{T}\sum_{n=-\infty}^{\infty}\exp\left[\frac{-j\pi(n/T)^2}{\lambda z}\right]\exp(j2\pi Xn/T)\times\chi_{-z}\{u_z(x)*p_\zeta(x)\}\left(X-\frac{n\lambda z}{T}\right). \tag{17}$$

We now consider the inverse Fresnel operation

$$\chi_{-z}\{u_z(x) * p_\zeta(x)\}(X) = u_z(x) * p_\zeta(x) * \exp\left(\frac{-j\pi}{\lambda z}x^2\right)$$
$$= u_z(x) * \exp\left(\frac{-j\pi}{\lambda z}x^2\right) * p_\zeta(x)$$
$$= u(X) * p_\zeta(X). \quad (18)$$

The conclusion we draw from this result is that the averaging introduced by the finite size pixels acts to degrade the quality of the reconstructed hologram by convolving it with a narrow rectangular function that is the same size as the pixel. As $\zeta \to 0$, this rectangular function narrows, reducing the distorting effect that it has on the reconstruction. Provided that the distribution $u(X)$ is approximately constant over the width of the function, $p_\zeta(X)$, there will be little distortion of $u(X)$. However, if we are attempting to recover spatial frequencies higher than the Nyquist limit, then $u(X)$ will, by definition, vary significantly over the width of the pixel and thus will act to make these distortions increasingly pronounced. We also note that the effect of convolving $p_\zeta(X)$ with $u(X)$ is to increase the spatial extent of the resultant signal from Δ to $\Delta+2\zeta$. Thus, this new signal may be recovered provided that $T \leqslant (\lambda z)/(\Delta+2\zeta)$.

We may also examine the effect of the convolution of $p_\zeta(X)$ with $u(X)$ in the spatial frequency domain. From Ref. 37 (p. 128), we find that the Fourier transform of $p_\zeta(X)$ is given by

$$F\{p_\zeta(x)\}(f_x) = \int p_\zeta(x)\exp(-j2\pi x f_x)dx$$

$$F\{p_\zeta(x)\}(f_x) = \mathrm{sinc}[2\pi f_x(\zeta)], \quad (19)$$

where $\mathrm{sinc}(x)=\sin(x)/x$. Therefore, the effect of the convolution operation is to multiply the spatial frequency content of the signal, $F\{u(X)\}(f_x)$ by a sinc function. We note that for values of f_x such that $f_x=n/(2\zeta)$, then Eq. (19) is zero, and therefore, these spatial frequencies will be entirely removed from the signal. A more detailed interpretation of this result is discussed in Ref. 26.

2.2.3 *Finite camera extent, neglecting averaging due to pixels: $\zeta \to 0$*

We now examine the third factor that impacts on the quality of our reconstructed hologram $u^s(X)$, the finite extent of the camera. For simplicity, we assume that we are sampling with point detectors (i.e., we allow $\zeta \to 0$). In this instance, Eq. (12) can be expressed as[34]

$$\Psi(X,X_1) = \int_{-D}^{D} \delta_T(x)\exp\left[\frac{j2\pi}{\lambda z}x(X-X_1)\right]. \quad (20)$$

Using Eq. (5) in conjunction with the Fourier shift theorem (see Ref. 37, p. 104), we find that

$$\Psi(X,X_1) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \mathrm{sinc}\left[2\pi\frac{D}{\lambda z}\left(X-X_1-\frac{\lambda z n}{T}\right)\right]. \quad (21)$$

Subbing Eq. (21) into Eq. (11)

$$u^s(X) = \exp\left(\frac{-j\pi}{\lambda z}X^2\right)\frac{1}{T}\sum_{n=-\infty}^{\infty}\int u(X_1)\exp\left(\frac{j\pi}{\lambda z}X_1^2\right)$$
$$\times \mathrm{sinc}\left[\frac{2\pi D}{\lambda z}\left(X-X_1-\frac{\lambda z n}{T}\right)\right]dX_1$$
$$= \exp\left(\frac{-j\pi}{\lambda z}X^2\right)\frac{1}{T}\sum_{n=-\infty}^{\infty} R(X,n), \quad (22)$$

where

$$R(X,n) = u(X)\exp\left(\frac{j\pi}{\lambda z}X^2\right) * \mathrm{sinc}\left[\frac{2\pi D}{\lambda z}\left(X-\frac{\lambda z n}{T}\right)\right]. \quad (23)$$

Thus, we can see from Eqs. (22) and (23) that the effect of the finite camera extent is to reduce the resolving ability of the DH system by convolving the product of the initial input and a quadratic phase term, with a sinc function, whose width is determined by the wavelength of the light, the size of the aperture, and the distance the camera is placed from the object plane. However, as we shall shortly demonstrate in Sec. 4, it is important to note that this "convolution" relationship is not a shift-invariant operation due to the presence of the quadratic phase factor. Nevertheless, as a useful "rule-of-thumb" approximation that we use later when examining experimental results, it is convenient to reinterpret Eq. (23) in the spatial frequency domain as a low-pass filtering operation,

$$R(X,n) = F^{-1}\left\{F\left[u(X)\exp\left(\frac{j\pi}{\lambda z}X^2\right)\right]P_{D/\lambda z}(v)\right.$$
$$\left.\times\exp\left(j2\pi v\frac{nD}{T}\right)\right\}(X), \quad (24)$$

where $P_{D/\lambda z}(v)=1$, when $|v|<D/\lambda z$ and is 0 otherwise, and where F and F^{-1} indicate Fourier and inverse Fourier transform operations.

Again assuming that our input field $u(X)$ has a finite spatial extent Δ, we can see from Eq. (23) that this input extent will be increased due to the presence of the sinc function. Strictly speaking, a sinc function spans an infinite spatial extent, implying that our recovered signal will inevitably suffer from aliasing. Practically, however, we may assume that the sinc function can be effectively limited in space. We therefore define the effective extent of the sinc function, Δ_S, as being twice the distance from its maximum value to its first null in keeping with the analysis presented in Ref. 34 (i.e., $\Delta_S=\lambda z/D$). Therefore, in order to ensure the successive replicas do not overlap, we require that $T \leqslant (\lambda z)/(\Delta+\Delta_S)$.

2.2.4 *Finite camera extent and averaging due to pixels*

In this section, we look at how all three factors interact with each other to limit the resolution of a practical DH system. We begin by examining Eq. (13) and discuss how we may simplify the expression considerably. Substituting this simplified expression into Eqs. (11) and (12), we then investigate how finite pixel size and the finite extent of the camera limit the resolution of the imaging system, identifying the

regimes where one effect dominates over the other. For the readers convenience, we again present Eq. (13),

$$B(x,X_1)=\frac{1}{2\zeta}\int_{-\zeta}^{\zeta}\exp\left(\frac{j\pi}{\lambda z}u^2\right)\exp\left[\frac{-j2\pi}{\lambda z}(x-X_1)u\right]du,$$

and note the similarity of the integral equation to that used to describe Fresnel diffraction from a square aperture (1-D slit) [see Refs. 3 (Sec. 4.51 on p. 84) and 44]. We further note that, under certain conditions, the quadratic phase factor $\exp[(j\pi/\lambda z)u^2]$ may be neglected. This approximation is identical to the Fraunhofer approximation, which, from Eq. 4.24 in Ref. 3, is satisfied when

$$z \gg \frac{\pi u^2|_{\max}}{\lambda}$$

$$\frac{\zeta^2}{\lambda z} \ll 1/\pi. \tag{25}$$

To examine whether use of this approximation is justified for practical DH systems, we put some typical values into Eq. (25). Setting $\lambda=633$ nm, $\zeta=T/2=5$ μm, and $z=100$ mm, we find that Eq. (25) gives ≈ 0.0004, which is approximately three orders of magnitude lower than $1/\pi \approx 0.32$. Once we make the Fraunhofer approximation, Eq. (13) reduces to

$$B(x,X_1)=\mathrm{sinc}\left[\frac{2\pi\zeta}{\lambda z}(x-X_1)\right], \tag{26}$$

which we now substitute back into Eq. (12) to give

$$\Psi(X,X_1)=\int_{-D}^{D}\mathrm{sinc}\left[\frac{2\pi\zeta}{\lambda z}(x-X_1)\right]\delta_T(x)\times\exp\left[\frac{j2\pi}{\lambda z}x(X-X_1)\right]dx. \tag{27}$$

Unfortunately, we are not aware of an analytic solution for Eq. (27) and, thus, we now examine three different regimes where the above integral is effectively limited by (*i*) the finite extent D, (*ii*) a combination of both D and ζ, and (*iii*) where the pixel-related sinc function effectively limits the integration range.

In order for Eq. (27) to be limited by the finite extent of the camera, then the term $\mathrm{sinc}[(2\pi\zeta/\lambda z)(x-X_1)]$ should remain approximately constant over the range of integration (i.e., $\pm D$). We arbitrarily define this to be the case when $\mathrm{sinc}[(2\pi\zeta/\lambda z)(x-X_1)] \geq 0.85$ over this integration range (see Fig. 2). Assuming fixed values for λ, z, and ζ, we now determine the restrictions on the variables x and X_1 [i.e., what is the maximum value of $(x-X_1)$ such that $\mathrm{sinc}[(2\pi\zeta/\lambda z)(x-X_1)] \geq 0.85$?]. To determine this, we note from Eq. (27) that x ranges from $-D \leq x \leq D$, thus $x_{\max}=D$. To determine the values that X_1 can take, we must once more consider Eq. (11) and note that we have assumed our input field $u(X)$ has a finite extent given by Δ. Examining Eq. (11), we can see that the values X_1 can take are limited by the assumed finite extent of input field. Using the maximum (worst-case) value for $(x-X_1)$ gives the following inequality:

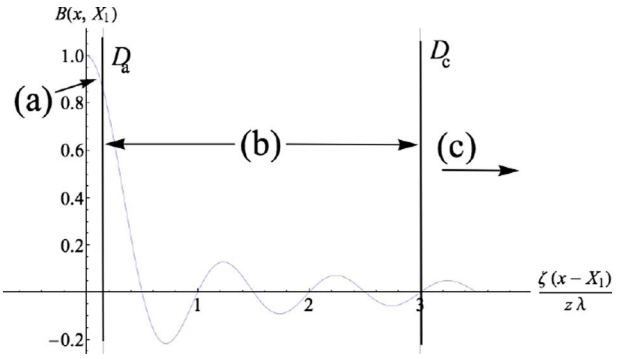

Fig. 2 Plot of $B(x,X_1)$ for different values of camera extent D corresponding to the three different regions: (a) where D dominates the resolution of the system, (b) intermediate region where both ζ and D limit the resolution, and (c) where the ζ limits the resolution.

$$\left(D+\frac{\Delta}{2}\right) \leq \frac{3}{20}\left(\frac{\lambda z}{\zeta}\right) \tag{28}$$

or equivalently

$$\zeta \leq \frac{3}{20}\left[\frac{\lambda z}{(D+\Delta/2)}\right]. \tag{29}$$

We note that the 3/20 factor was chosen so that $B(D, -\Delta/2) \approx 0.858$. Once the inequality [Eq. (29)] is satisfied, we ignore the effect of pixel averaging on the spatial resolution of the system. Thus, the limiting factor on the resolution of the DH system is determined by the finite camera extent D and the sampling rate, as was discussed in Sec. 2.2.3.

If the range of integration in Eq. (27) is limited somewhere in region (b) (see Fig. 2), then both the averaging effect of the pixels and the finite camera both contribute to limiting the resolution in DH imaging systems. Because of the complex interaction of all three factors, it is difficult to provide a strict guideline over this region as to what exactly makes the greatest contribution to limiting the resolution.

Finally, we identify region (c) as where the averaging effect of the pixels is shown to dominate the maximum recoverable frequency. Region (c) is again arbitrarily defined as occurring once the third null of $B(x,X_1)$ occurs before the finite extent of the camera limits the integration in Eq. (27). In this instance, we assume that the finite extent of the camera can be effectively ignored once

$$\zeta \geq 3\left[\frac{\lambda z}{(D+\Delta/2)}\right]. \tag{30}$$

2.3 *Twin Image Term*

In this section, we wish to consider the twin image term in Eq. (9). We have already noted that the Fresnel transform is linear, and therefore, we may use the results derived in Sec. 2.2.4 to examine the effect of the twin-image term. Again, we wish to consider how each of the factors—sampling

rate, finite pixel size, and finite camera extent—modify the twin image term. First, we will assume an infinitely large camera face with infinitely narrow pixels: $D\to\infty$, $\zeta\to 0$. These assumptions are identical to those made in Sec. 2.2.1 and allow us to effectively determine the spatial extent, $\tilde{\Delta}$ of the twin image. Once we have defined $\tilde{\Delta}$, we examine how it imposes additional constraints on the sampling rate T to ensure that the twin-image replicas do not overlap each other. Substituting $u_z^*(x)$ for $u_z(x)$ in Eq. (16), we get

$$\tilde{u}^s(X) = \chi_{-z}[u_z^*(x)\delta_T(x)](X),$$

$$\tilde{u}^s(X) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \exp\left[\frac{-j\pi(n/T)^2}{\lambda z}\right]\exp(j2\pi Xn/T) \times \chi_{-z}[u_z^*(x)]\left(X-\frac{n\lambda z}{T}\right). \quad (31)$$

Note that we use the expressions $\tilde{u}(X)$ and $\tilde{u}^s(X)$ to refer to the twin and the numerically reconstructed twin terms, respectively. We recall from the derivation of Theorem 3 in Ref. 27 that $u_z^*=\chi_{-z}(x)[u^*(X)](x)$, and thus rewrite Eq. (31) as

$$\tilde{u}^s(X) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \exp\left[\frac{-j\pi(n/T)^2}{\lambda z}\right]\exp(j2\pi Xn/T) \times \chi_{-2z}[u^*(X)]\left(X-\frac{n\lambda z}{T}\right). \quad (32)$$

The observations made about Eq. (16) also apply to Eq. (32); in particular, we recall Observation (i): the sampling process produces an infinite number of replicas, separated from each other by a distance $\lambda z/T$. Unlike in the previous case, however, where object field $u(X)$ spanned a finite extent Δ, the field $\chi_{-2z}[u^*(X)](X)$ cannot have finite support. This follows from Theorem 2 in Ref. 27, which states that a function can have a finite support in only one Fresnel plane. It must follow, therefore, that signal power from adjacent $\tilde{u}^s(X)$ replicas leak into each other, distorting the true signal, $\tilde{u}(X)$. Quantifying this distortion is not necessarily straightforward and is signal dependent. For example, in Ref 45, it is shown that the rate of change of a signal's distribution (under Fresnel propagation) is dependent on the magnitude of spatial frequency content of the signal [see Eq. (14) therein]. Similarly, it is expected that the increase in spatial extent of a real wave field under Fresnel propagation will also depend on the physical spatial frequency content of the signal, or if simulating wave propagation numerically, will further depend on spatial frequency extent imposed by the chosen sampling rate. Nevertheless, if we accept that a signal can be approximately bounded in both space and spatial frequency,[16,46] then under certain conditions[16,46,47] the power contained in the signal $u(X)$ after Fresnel propagation is approximately confined within a spatial extent $\tilde{\Delta}$, where

$$\tilde{\Delta} = \Delta + \lambda ZB, \quad (33)$$

and where Z and B are the propagation distance and the spatial frequency extent of the signal, respectively (see Sec. 3 in Ref. 46 and Eq. (4) in Ref. 48). From Eq. (32) and (33), we can see that the spatial extent spanned by the virtual image is approximately given by $\tilde{\Delta}=\Delta+2\lambda zB$, because, after numerical reconstruction, the twin image has propagated a distance of $Z=2z$. Thus, recovery of the twin image imposes the following constraint on the sampling rate: $T\leqslant(\lambda z)/\tilde{\Delta}$.

Now we turn our attention to the averaging effect introduced by the finite size of the pixels. Using results from Sec. 2.2.2, in particular substituting $u_z^*(x)$ for $u_z(x)$ in Eqs. (16) and (17), we find that the effect of sampling with finite pixels is to increase the extent of the reconstructed twin image such that the sampling rate must be further increased so that $T\leqslant(\lambda z)/(\tilde{\Delta}+2\zeta)$, to ensure that successive twin-image replicas do not overlap.

We note from Sec. 2.2.3 that, when the finite size of the camera is taken into consideration $(\zeta\to 0)$, that an appropriate sampling rate is given by $T\leqslant(\lambda z)/(\tilde{\Delta}+\Delta_S)$. Finally, when considering the interaction of these three factors, D, T, and ζ, one must substitute $\tilde{\Delta}$ for Δ in Eqs. (29) and (30).

3 DH Architectures and Optimal Camera Design for Lensless DH Microscopy

In this section, we examine how the theoretical results from Sec. 2 apply to different DH systems. We then discuss some guidelines, implied by our theoretical model, for the optimal design of microscopic cameras for use in DH systems.

3.1 *DH Architectures*

The purpose of this section is to investigate two different but well-known DH architectures: inline and off-axis. In off-axis DH, a tilted plane wave is used as the reference field. The use of an off-axis reference wave spatially separates the virtual and real images in the reconstruction (object) plane. For simplicity, let us once more assume that $D\to\infty$ and $\zeta\to 0$. To ensure that there is sufficient space between replicas in the reconstruction plane so that the twin and the real image may be spatially separated, we require that the replicas be separated by a distance greater than twice $\tilde{\Delta}$, which imposes a minimum sampling rate of $T\leqslant(\lambda z)/(2\tilde{\Delta})$. With some manipulation of Eq. (33) and noting that $B=2f_{\max}$, we may express the latter result as

$$f_{\max} \leqslant \frac{1}{2T} - \frac{\Delta}{2\lambda z},$$

$$f_{\max} \leqslant f_{NQ} - \frac{\Delta}{2\lambda z}, \quad (34)$$

where f_{NQ} is the Nyquist frequency. Thus, we conclude from Eq. (34) that the maximum recoverable spatial frequency, $f_{\max}$, is lower than the Nyquist frequency for off-axis configurations.

There is a phase-shifting interferometric (PSI) technique for removing the twin and dc terms[9,49] for use in inline DH. This technique typically requires several captures of some scene whereby a series of phase shifts are introduced between the object and the reference arms. This can be achieved using a high-precision piezomotor[9,50] or alternatively using wave plates.[51] Using the four-step algorithm described in Ref. 49 [see Eq. (3) therein] requires that four separate holograms are captured where the phase between captures is stepped by precisely $\pi/2$ radians. Assuming a unit amplitude flat reference wave $u_R=\exp(j\Theta)$, we can calculate u_z from[49]

$$u_z(x)=[W^n_{\Theta=0}-W^n_{\Theta=\pi}+j(W^n_{\Theta=\pi/2}-W^n_{\Theta=3\pi/2})]/4,$$

$$u_z(x)=p_D(x)\,\delta_T(x)\{p_\zeta(x)*[I^0_c-I^\pi_c+j(I^{\pi/2}_c-I^{3\pi/2}_c)]\}/(8\zeta), \tag{35}$$

where

$$I^\Theta_c(x;t)=I_z(x)+1+u^*_z(x)\exp(j\Theta)+u_z(x)\exp(-j\Theta). \tag{36}$$

The disadvantage to the PSI approach is that, generally, several captures are required and, thus, real-time PSI imaging is difficult. Nevertheless, once the twin and dc terms have been removed, it paves the way to recovering spatial frequencies higher than the Nyquist limit.

In Ref. 52, the authors investigate the effect of finite pixel size on reconstructions for a Fourier DH setup. In Fourier DH, the reference beam is a displaced point source adjacent to the object. This architecture allows the hologram reconstruction to be calculated using a single FFT algorithm. The reconstruction plane is therefore made up of a dc term occupying the central region and the two twin images on each side, which are both in-focus reconstructions of the object. By an analytical approach similar to that outlined in this paper, the authors examine the effect of finite pixel size and camera extent on the reconstructed object field and refer the interested reader to this source for a more detailed examination. This approach may also allow the recovery of super-Nyquist frequencies.

3.2 *Optimal Camera Design for Lensless Microscopic DH Systems*

Here, we examine how to optimize camera design for lensless DH systems, with the assumption that some technique has first been used to recover the real-image term. Let us consider imaging microscopic objects where the extent of the object field is limited to a very small region. Ideally, we should strive to maximize the resolvable detail in our object [i.e., we should ensure that the DH system is operating in region (c) as defined by the inequality in Eq. (30)]. In this region, it is possible in theory to recover frequencies far higher than the Nyquist limit and, thus, D should be increased as much as is feasible. It may also be possible to artificially increase D using synthetic aperture techniques. Once λ and z are fixed, we note that the limitation on the sampling rate is then determined solely by the region of space (object extent) that we wish to look at. As Δ is made smaller, T can be reduced accordingly. The finite pixel extent reduces, and in some cases eliminates, the power in higher spatial frequencies [see Eqs. (18) and (19), and for a more thorough discussion, see Ref. 26]. To ensure that all spatial frequencies can be recovered, we suggest designing a camera that has two different size pixels, ζ_1 and ζ_2, associated with it. Two different pixel sizes will modulate the spatial frequency content of the object field in different manners in accordance with Eq. (19). By combining the information from both sets of pixels and performing some numerical processing, it may be possible to satisfactorily undo the filtering operation performed by the pixel or to deconvolve the signal. In closing, we note that while choosing a small pixel size for ζ increases the spatial frequency response of the system, it also results in less power being incident on the light-sensitive region of the camera. To balance these counteracting effects, we would suggest that a small value for ζ_1 and a relatively larger value for ζ_2 be chosen for optimal performance.

4 Shift Variant Properties of DH Imaging

We now wish to demonstrate the shift-variant nature of the DH system described by Eqs. (22) and (23). We begin by examining how two point sources are imaged by the DH system.[53] We thus set

$$u(X)=\delta(X+d_1)+\delta(X+d_2). \tag{37}$$

For simplicity, we initially consider the case where $n=0$ and substitute Eq. (37) back into Eq. (22) to give

$$u^s(X)=\exp\left[\frac{-j\pi}{\lambda z}(X^2-d_1^2)\right]\left\{\operatorname{sinc}\left[\frac{2\pi D}{\lambda z}(X-d_1)\right]+\exp(j\phi)\operatorname{sinc}\left[\frac{2\pi D}{\lambda z}(X-d_2)\right]\right\}, \tag{38}$$

where

$$\phi=\left(\frac{\pi}{\lambda z}\right)(d_2^2-d_1^2). \tag{39}$$

From Eqs. (38) and (39), we can see that two displaced point sources are mapped to two displaced sinc functions. We note however that there is a space-dependent phase difference, ϕ, between these two terms. In the following example, we keep the distance between the two point sources fixed at $2d=\lambda z/(2D)$. However, we now allow them to shift in the input plane and examine the resultant distribution in the output plane. We consider three special input plane locations for our point sources: (*i*) $d_1=d$, $d_2=-d$, (*ii*) $d_1=d+D/2$, $d_2=-d+D/2$, and (*iii*) $d_1=d+D$, $d_2=-d+D$ and plot their distributions in Fig. 3. We have carefully chosen our displacements so that ϕ in Eq. (40) has the value of (*i*) $\phi=0$, (*ii*) $\phi=\pi/2$, and (*iii*) $\phi=\pi$ in keeping with a similar example presented on page 157 of Ref. 3 (p. 157) and specifically addressed in Ref. 53. In case (*i*), the point sources are located symmetrically about the optical axis; however, they are not considered to be resolved in the Raleigh sense.[3] In case (*ii*), even though the distance between the point sources has not changed, we have shifted their location in the input plane a distance $D/2$, which has the effect of introducing a phase shift $\phi=\pi/2$ between the two contributions (In Fig. 3, we overlay the plots $|u^s(X$

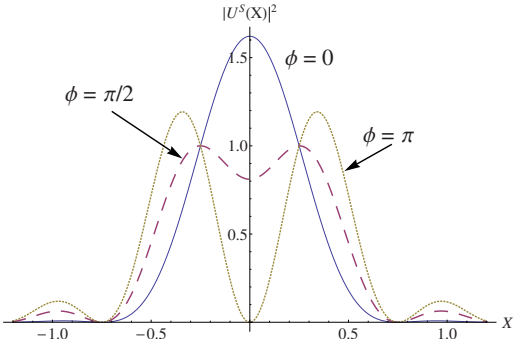

Fig. 3 Plot of $|u^s(X)|^2$ for (*i*) $\phi=0$ (solid line), (*ii*) $\phi=\pi/2$ (dashed line), and (*iii*) $\phi=\pi$ (dotted line). Because (*ii*) and (*iii*) are imaged at different locations they have been displaced so that the three plots can be over-layed on top of each other for comparative purposes.

$-D/2)|^2$ and $|u^s(X-D)|^2$ for presentation purposes). In this instance, we can see that the point sources are deemed to be resolved in the Raleigh sense. Finally, in case (*iii*), where the point sources are shifted a distance D, they appear to be even more clearly resolved than in the previous two cases. To generate these results, we have used the following values: $z=200$ mm, $\lambda=785$ nm, $D=3.354$ mm, and $T=6.45$ μm, to give $d=11.7$ μm. We also note that this choice of T ensures that the replicas generated in the output plane (i.e., the contributions arising from $n=\pm 1,2,3,\ldots$) are well separated from each other.

5 Experimental Results

In this section, we present some experimental results that verify the theoretical description of DH imaging process that we presented in Sec. 2. The wave field emerging from the laser in Fig. 1 is, after passage through several optical elements, focused through a spatial filter and collimated to form an approximately flat plane wave. It is then incident upon BS1 (see Fig. 1) and separated into the object and reference wave fields. The object wave field is reflected from mirror A in Fig. 1 and illuminates our object, a standard transmissive USAF chart. The transmitted object field then passes through BS2 and is combined at the camera face with the reference field and the resulting intensity recorded. The distance from the object to the camera face is given by $z=108$ mm. We use a PSI technique to recover the real-image term, $(2\zeta)^{-1}\delta_T(x)p_D(x)[u_s(x)*p_\zeta(x)]$, by capturing four separate images where the object beam has been phase shifted in steps of $\lambda/4$ between successive captures. The phase shifting is introduced by moving mirror A in the reference beam arm (see Fig. 1) with a calibrated piezoelectric motor supplied by PiezoSystemJena PZ38CAP and driven using Controller NV40/1CLE. These four captures were then processed using a standard phase-shifting algorithm,[9] and the real-image term was recovered. Our laser source is a CrystalLaser with $\lambda=785$ nm, and the camera used in the experiment was an AVT Dolphin F-145B camera with 1392×1040 pixels of pitch, $T=6.45$ μm. Unfortunately, for the camera we are using, the fill factor is not specified. We can however still demonstrate the effect of varying the pixel size. This results in a light-sensitive area of width and height: $2D_x=(1392)T\approx 8.978$ mm and $2D_y=(1040)T\approx 6.708$ mm, respectively.

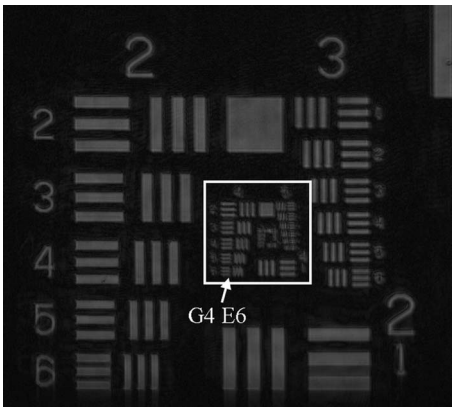

Fig. 4 Reconstruction of captured hologram. The reconstructed image has a size 2048×2048 pixels. For ease of viewing, we show a section of size 626×701 pixels. The bars of element 6 of group 4 (G4 E6) can just be resolved in this reconstruction.

We note that a sampling rate of $T=6.45$ μm corresponds to a Nyquist frequency of $f_{NQ}=77.5$ lines/mm.

To numerically reconstruct the hologram, we first zero pad it so that our matrix size is now 2048×2048. We now propagate this matrix a distance $z=-108$ mm using the direct method implementation of the Fresnel transform that makes use of the FFT algorithm.[46,48] A section (size: 626×701 pixels) of the reconstructed hologram is presented in Fig. 4. It is just about possible to discern both the horizontal and vertical lines (although not the text) of element 6 in group 4 on our USAF resolution chart, indicating that we can resolve a spatial frequency, $f_R=28.5$ lines/mm. We now wish to examine the effect of reducing the aperture size D_x from its maximum value of $D_x=4.489$ mm to a lower value of $D_x\approx 1.123$ mm, which is 25% of the actual camera extent. We affect this change by setting to zero all the values in our hologram matrix that lie outside the rectangular window determined by $2D_x$. We recall from Eq. (21) that as a rule of thumb the spatial frequency that the DH system can resolve is approximately given by $D_x/(\lambda z)$, which for the cases indicated in the caption of Fig. 5, cor-

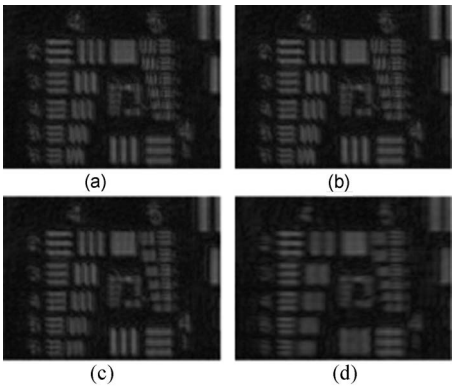

Fig. 5 The effect varying D_x has on the perceived resolution of the DH system. A specific region from Fig. 4 was examined (see the white box) for different values of D_x: (a) $D_x\approx4.489$ mm, (b) $D_x\approx3.367$ mm, 75% of (a); (c) $D_x\approx2.245$ mm, 50% of (a); and (d) $D_x\approx1.123$ mm, 25% of (a).

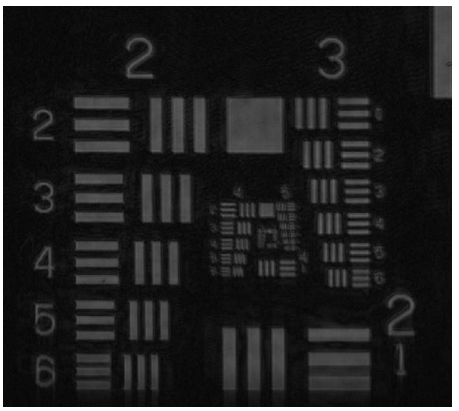

Fig. 6 Effect varying T on the resolvable detail in a reconstructed hologram. The region is the same as that in Fig. 4; however, the sampling rate along the x direction has been reduced so that $\tilde{T}=2T$.

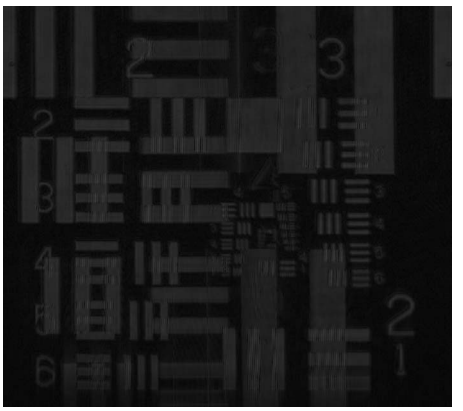

Fig. 7 The effect varying T on the resolvable detail in a reconstructed hologram. The region is the same as that in Fig. 5; however, the sampling rate along the x; direction has been reduced so that $\tilde{T}=3T$.

respond to spatial frequencies of (a) 53.9 lines/mm, (b) 39.7 lines/mm, (c) 26.5 lines/mm, and (d) 13.2 lines/mm. On examination of Fig. 5, we can see that, broadly speaking, the resolvable spatial frequencies conform to this rule of thumb. Note, however, that we have left D_y unchanged, and accordingly, we do not see a corresponding reduction in detail in the vertical direction. When $D_x=4.489$ and assuming a 100% fill factor, we are now operating in region (b) as discussed in Sec. 2.2.4; however, when $D_x \approx 1.123$ mm, we move to region (a) and, thus for the latter example Fig. 4(d), the finite extent of the camera exclusively determines the resolution of the DH imaging system.

Setting $D_x=4.489$ mm, we now examine the effect of varying the sampling rate T. We are now operating in region (b); however, from the inequalities defined in Eqs. (29) and (30), we note that the sampling rate will not effect the resolution. The sole effect of the sampling rate is to determine the distance between neighboring replicas. We also note that reducing the sampling rate is equivalent to reconstructing the hologram using a fewer number of samples from the hologram matrix. We can do this by setting to zero the values in hologram matrix that we do not want to contribute to the reconstruction, allowing us to maintain the same number of samples in the output domain. This is discussed in more detail in Appendix B. The main result of reducing the sampling rate is that the higher-order replicas move into the region of space that we wish to view, corrupting the data therein. This can seen clearly by examining Figs. 6–8. In Fig. 6, we present the same region in space as depicted in Fig. 5; however in this instance, the sampling rate along the x direction has been set to $\tilde{T}=2T$. In the section shown in Fig. 6, no replicas are visible and we note that the resolution of the system remains unchanged in this window. In Fig. 7, we set $\tilde{T}=3T$ and we can now see the presence of higher-order replicas. In any areas where the replicas do not overlap, the resolution remains unchanged. The encroachment of higher-order replicas is perhaps most dramatically illustrated in Fig. 8, $\tilde{T}=4T$, where nearly the whole image is distorted. However, in the highlighted region, group 4 (see white box in Fig. 8) is visible and it is clear that the resolvable resolution has not been affected by the reduction in the sampling rate. We note that decreasing the sampling rate by a factor of four reduces f_{NQ} from 77.5 to 19.37 lines/mm. Nevertheless, we can still resolve element 4 of group 4 (see square box in Fig. 8), which is rated as having a spatial frequency of 22.62 lines/mm. This demonstrates that we are able to resolve frequencies higher than the Nyquist limit. According to Eqs. (29) and (30), our experimental system is currently operating in region (b), however, with a differently designed camera, it may be possible to recover spatial frequencies far higher than we are able to experimentally demonstrate. In closing this discussion on sampling, we would like to draw the readers attention to the interference pattern that occurs when replicas overlap with each other (see Fig. 7). We remind the reader that we have only reduced the sampling along the x-axis direction, and here we see that the interference modulation also occurs along the x-axis. This is perhaps related to linear phase terms associated with higher-order replicas in accordance with Eq. (16).

Finally, we wish to examine the effect of increasing the effective pixel size. We have already noted that we are

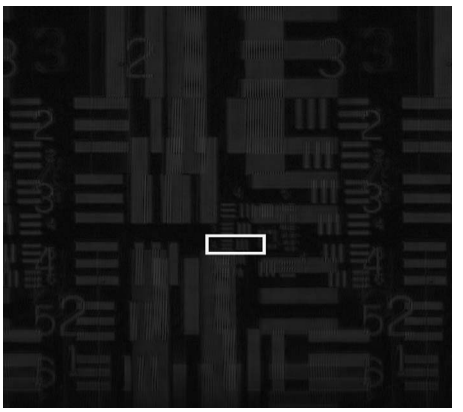

Fig. 8 Effect varying T on the resolvable detail in a reconstructed hologram. The region is the same as that in Fig. 5; however, the sampling rate along the x direction has been reduced so that $\tilde{T}=4T$.

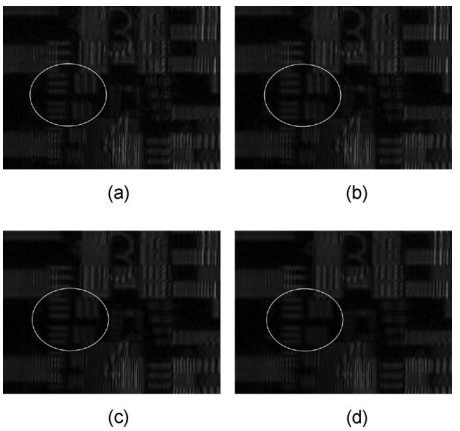

Fig. 9 We are looking at group 4 and examining how varying the pixel size 2ζ effects the resolvable detail in a reconstructed hologram: (a) $2\zeta \approx T$, (b) $2\zeta \approx 2T$, (c) $2\zeta \approx 4T$, and (d) $2\zeta \approx 6T$. Note the increase in blurring over the horizontal bars for the elements 3 and 4 as one increases the finite pixel size.

operating in region (b), and thus, we expect that for a given object extent both the finite camera extent and finite pixel size will act together to limit the resolution of our system. Here we will demonstrate the deleterious effect of the finite pixel size by reconstructing several holograms that have been recorded by a camera with different pixel sizes. Suppose our camera returns the following 1-D array:

$$W^n = [1,2,3,4,5,6,7,8,9,\ldots]. \tag{40}$$

To downsample our signal sixfold, we only consider every sixth sample in the matrix, setting all other values to zero

$$W^n = [1,0,0,0,0,0,7,\ldots]. \tag{41}$$

In Eq. (41), the pixel size is given by $2\zeta = T$. We can however change the size of our pixel by averaging the values of neighboring pixels. For example, to set $2\zeta = 2T$, Eq. (41) becomes

$$W^n = [1.5,0,0,0,0,0,7.5,\ldots]. \tag{42}$$

Thus, the first two elements of the array given by Eq. (40) are averaged to give the first element of Eq. (41). Using this approach to vary the effective size of our pixels, we generate four different reconstructions (see Fig. 9). For Fig. 9(a), we set $2\zeta = T$, Fig. 9(b) $2\zeta = 2T$, Fig. 9(c) $2\zeta = 4T$, and Fig. 9(d) $2\zeta = 6T$. We wish to direct the readers attention to group 4, elements 3 and 4 (see circled region in figures) and note the increased blur on the vertical bars as one moves from Fig. 9(a)–9(d). The effect is more pronounced by comparing the blur over the vertical bars to those of the horizontal bars, which are unchanged over the figures.

6 Conclusion

DH is a technique for capturing and storing the complex amplitude of a light field. This technique has many applications in modern science, including metrology, microscopy, as well as the capture of the 3-D scenes, which can be subsequently replayed in a different physical location. A particularly attractive feature of DH is that the hologram can be processed numerically, allowing for the extraction of 3-D information, speckle removal, identification of structures in a 3-D volume, as well as many other image-processing techniques. A major limitation in DH, however, is the relatively low spatial frequency response of modern digital cameras, which in turn severely limits the resolution of reconstructed holograms. Although in many applications high spatial resolution may not be necessary, it becomes a critical factor when trying to accurately record 3-D scenes. As such, much effort is being expended both practically in the improvement of camera technology and theoretically in researching optimal methods of representing sampled data (see, for example, Ref. 54). In this paper, we have examined three different and independent factors that limit the ability of DH systems to resolve detail in a reconstructed image. We identified these factors as the finite extent of the camera, the sampling rate, and the finite size of the pixels. By developing a theoretical model to describe each of these effects, we examined fundamental resolution limits in DH systems. We extended the analysis to examine the effect of both the dc terms and the twin (virtual) image. Several different recording architectures were studied, and it was shown that, in order to recover frequencies higher than the Nyquist limit, it is necessary to use a PSI setup. We demonstrated that DH imaging is shift variant. We used our theoretical model to provide guidelines to optimally design cameras for microscopic DH systems. Finally, a series of experiments were conducted that verify the usefulness of the theoretical model. These experiments provide the first evidence to our knowledge that frequencies above the Nyquist limit may recovered in DH systems, verifying a well-known theoretical result.

Much work remains to be done to further improve DH imaging techniques and to fully take advantage of some of the characteristics of this imaging modality. For example, we note that because DH is a coherent imaging process, it is subject to the deleterious effects of speckle noise. Therefore, robust techniques for reducing speckle noise while maintaining detail in the reconstructed image constitute an important numerical processing consideration. Several different techniques have shown some promise in this area.[55] Another promising approach is the use of wavelets,[54,56] which allows more freedom in processing the digital hologram once it has been captured. Wavelets show promise in several other areas, including speckle reduction techniques,[54] the development of propagation algorithms,[57] as well as autofocusing techniques.[58]

Appendix A

The purpose of this appendix is to simplify the expression in Eq. (10),

$$u^s(X) = \exp\left(\frac{-j\pi}{\lambda z}X^2\right)\int [u_z(x) * p_\zeta(x)]\delta_T(x)p_D(x) \times \exp\left(\frac{-j\pi}{\lambda z}x^2\right)\exp\left(\frac{-2j\pi}{\lambda z}xX\right)dx. \tag{43}$$

We begin by expanding the convolution expression in Eq. (43),

$$u_z(x) * p_\zeta(x) = \int u_z(x-u)p_\zeta(u)du. \tag{44}$$

Using Eq. (2) and making the variable change $x \rightarrow x-u$, we rewrite Eq. (44) as

$$u_z(x) * p_\zeta(x) = \int \exp\left[\frac{j\pi}{\lambda z}(x^2+u^2)\right]\exp\left(\frac{-j2\pi}{\lambda z}xu\right)\int u(X_1) \times \exp\left\{\frac{j\pi}{\lambda z}[X_1^2 - 2X_1(x-u)]\right\}dX_1 p_\zeta(u)du. \tag{45}$$

After some manipulation, this can be rearranged to give

$$u_z(x) * p_\zeta(x) = \exp\left(\frac{j\pi}{\lambda z}x^2\right)\int B(x,X_1)\times u(X_1) \times \exp\left[\frac{j\pi}{\lambda z}(X_1^2 - 2X_1x)\right]dX_1, \tag{46}$$

where

$$B(x,X_1) = \int_{-\zeta}^{\zeta} \exp\left(\frac{j\pi}{\lambda z}u^2\right)\exp\left[\frac{-j2\pi}{\lambda z}u(x-X_1)\right]du. \tag{47}$$

Substituting Eq. (46) into Eq. (43), changing the order of integration, and simplifying, we get

$$u^s(X) = \exp\left(\frac{-j\pi}{\lambda z}X^2\right)\int \Psi(X,X_1)u(X_1)\exp\left(\frac{j\pi}{\lambda z}X_1^2\right)dX_1, \tag{48}$$

$$\Psi(X,X_1) = \int B(x,X_1)\delta_T(x)p_D \exp\left[\frac{j2\pi}{\lambda z}x(X-X_1)\right]dx. \tag{49}$$

Appendix B

In this appendix, we examine the numerical algorithm we use to generate the results presented in Sec. 4. Let there be two fields $f(X)$ and f_z that are related to each other by a Fresnel transform,

$$f_z(x) = \frac{1}{\sqrt{j\lambda z}}\exp\left(\frac{j\pi x^2}{\lambda z}\right)\int f(X)\exp\left(\frac{j\pi X^2}{\lambda z}\right) \times \exp\left(\frac{-j2\pi xX}{\lambda z}\right)dX. \tag{50}$$

We drop the leading quadratic phase factor and scaling term for the remainder of this section. The field $f_z(x)$ may be approximately calculated from a discrete set of N samples in which the case of Eq. (50) changes to

$$f_z(x) \approx \sum_{n=-N/2}^{n=N/2-1} f(n\delta_X)\exp\left(\frac{j\pi n^2\delta_X^2}{\lambda z}\right)\exp\left(\frac{-j2\pi xn\delta_X}{\lambda z}\right), \tag{51}$$

where $X \rightarrow n\delta_X$. To make use of the efficient FFT algorithm, we must map N samples in one domain to an equal number of M samples in the output domain. Therefore, we define the vector $\bar{f}_z$ as

$$\bar{f}_z^{\mathrm{T}} = \left[f_z\left(\frac{-M\delta_x}{2}\right), f_z\left(\frac{-(M-1)\delta_x}{2}\right), \dots\right]$$

$$\bar{f}_z^{\mathrm{T}} = [f_{z1}, f_{z2}, \dots], \tag{52}$$

where T indicates a transpose operation. $\bar{f}_z$ may be related to the input function with the following relation:

$$\bar{f}_z = \sum_{m=-M/2}^{n=M/2-1}\sum_{n=-N/2}^{n=N/2-1} f(n\delta_X)\exp\left(\frac{j\pi n^2\delta_X^2}{\lambda z}\right)\exp\left(\frac{-j2\pi mn\delta_x\delta_X}{\lambda z}\right), \tag{53}$$

where $x \rightarrow m\delta_x$. Making use of the FFT algorithm imposes the following constraint:

$$M\delta_x = \lambda z/\delta_X, \tag{54}$$

and now we can rewrite Eq. (53)

$$\bar{f}_z = \bar{A}\bar{f}, \tag{55}$$

where

$$\bar{A} = \begin{pmatrix} A_{11} & A_{12} & \dots \\ A_{21} & A_{22} & \dots \\ \vdots & \vdots & \end{pmatrix} \tag{56}$$

with

$$A_{mn} = \exp\left(\frac{-j2\pi mn\delta_x\delta_X}{\lambda z}\right), \tag{57}$$

and

$$\bar{f}^{\mathrm{T}} = [f_{-N/2}, f_{-(N-1)/2}, \dots, f_{-n}, \dots] \tag{58}$$

Expanding Eq. (55) for f_{zm}, we get

$$f_{zm} = A_{m1}f_1 + A_{m2}f_2 + A_{m3}f_3 + A_{14}f_4 + A_{15}f_5 + \cdots. \tag{59}$$

The FFT algorithm limits the user's control over the output domain space; for example, one must map N samples from one domain to an equal number of samples in the output domain. In Sec. 4, we wish to maintain the same number of samples in the output domain (for comparative purposes) while reducing the number of contributing samples in the input domain. From Eq. (59), we can see that, by setting individual values in the matrix f to zero, they will not contribute to the calculated output value f_{zm}. We can therefore reduce the number of contributing samples in the input domain without affecting the number of samples in the output domain. We note that it is possible to calculate the continuous distribution $f_z(x)$; however, the process is relatively slow compared to the efficiency of the FFT algorithm. The FFT-based Fresnel calculation described here is based on what is referred to as the DM.[46,48] This technique is also discussed in Ref. 29.

Acknowledgments

This research is funded from the European Community's Seventh Framework Programme FP7/2007-2013 under Grant No. 216105 "Real 3D."

References

1. D. Gabor, "A new microscopic principle," *Nature (London)* **161**, 777–778 (1948).
2. A. W. Lohmann, *Optical Information Processing*, Universitätsverlag Ilmeanau (2006).
3. J. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1966).
4. P. Hariharan, *Basics of Holography*, Cambridge University Press, New York (2002).
5. E. N. Leith and J. Upatnieks, "Wavefront reconstruction with continuous-tone objects," *J. Opt. Soc. Am.* **53**, 1377–1381 (1963).
6. M. Testorf and A. W. Lohmann, "Holography in phase space," *Appl. Opt.* **47**, A70–A77 (2008).
7. J. W. Goodman and R. Lawrence, "Digital image formation from electronically detected holograms," *Appl. Phys. Lett.* **11**, 77–79 (1967).
8. U. Schnars and W. Juptner, "Digital recording and numerical reconstruction of holograms," *Meas. Sci. Technol.* **13**, R85–R101 (2002).
9. I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," *Opt. Lett.* **22**, 1268–1270 (1997).
10. O. Matoba, T. J. Naughton, Y. Frauel, N. Bertaux, and B. Javidi, "Real-time three-dimensional object reconstruction by use of a phase-encoded digital hologram," *Appl. Opt.* **41**, 6187–6192 (2002).
11. U. Gopinathan, D. S. Monaghan, B. M. Hennelly, C. P. M. Elhinney, D. P. Kelly, J. McDonald, T. J. Naughton, and J. T. Sheridan, "A projection system for real world three-dimensional objects using spatial light modulators *J. Disp. Technol.* **4**, 254–261 (2008).
12. A. Michalkiewicz, M. Kujawinska, J. Krezel, L. Salbut, X. Wang, and P. J. Bos, "Phase manipulation and optoelectronic reconstruction of digital holograms by means of LCOS spatial light modulator," *Proc. SPIE* **5776**, 144–152 (2005).
13. G. Pedrini, I. Alexeenko, W. Osten, and U. Schnars, "On-line surveillance of a dynamic process by a moving system based on pulsed digital holographic interferometry," *Appl. Opt.* **45**, 935–943 (2006).
14. F. Charrière, T. Colomb, F. Montfort, E. Cuche, P. Marquet, and C. Depeursinge, "Shot-noise influence on the reconstructed phase image signal-to-noise ratio in digital holographic microscopy," *Appl. Opt.* **45**, 7667–7673 (2006).
15. P. Ferraro, S. Grilli, D. Alfieri, S. D. Nicola, A. Finizio, G. Pierattini, B. Javidi, G. Coppola, and V. Striano, "Extended focused image in microscopy by digital holography," *Opt. Express* **13**, 6738–6749 (2005).
16. A. W. Lohmann, R. G. Dorsch, D. Mendolovic, Z. Zalevsky, and C. Ferreira, "Space-bandwidth product of optical signals and systems," *J. Opt. Soc. Am. A* **13**, 470–473 (1996).
17. T. M. Kreis, "Frequency analysis of digital holography *Opt. Eng.* **41**, 771–778 (2002).
18. T. M. Kreis, "Frequency analysis of digital holography with reconstruction by convolution," *Opt. Eng.* **41**, 1829–1839 (2002).
19. J. Garcia-Sucerquia, W. Xu, S. K. Jericho, P. Klages, M. H. Jericho, and H. J. Kreuzer, "Digital in-line holographic microscopy," *Appl. Opt.* **45**, 836–850 (2006).
20. P. Picart, J. Leval, D. Mounier, and S. Gougeon, "Some opportunities for vibration analysis with time averaging in digital Fresnel holography," *Appl. Opt.* **44**, 337–343 (2005).
21. P. Picart and J. Leval, "General theoretical formulation of image formation in digital Fresnel holography *J. Opt. Soc. Am. A* **25**, 1744–1761 (2008).
22. T. Latychevskaia and H.-W. Fink, ArXiv Physics e-prints (2006), arXiv:physics/0610048.
23. E. N. Leith and J. Upatnieks, "Reconstructed wavefronts and communication theory *J. Opt. Soc. Am.* **52**, 1123–1128 (1962).
24. C. Liu, Y. Li, X. Cheng, Z. Liu, F. Bo, and J. Zhu, "Elimination of zero-order diffraction in digital holography, " *Opt. Eng.* **41**, 2434–2437 (2002).
25. F. Dubois, C. Schockaert, N. Callens, and C. Yourassowsky, "Focus plane detection criteria in digital holography microscopy by amplitude analysis," *Opt. Express* **14**, 5895–5908 (2006).
26. D. P. Kelly, B. M. Hennelly, C. McElhinney, and T. J. Naughton, "A practical guide to digital holography and generalized sampling," *Proc. SPIE* **7072**, 707215 (2008).
27. F. Gori, "Fresnel transform and sampling theorem," *Opt. Commun.* **39**, 293–297 (1981).
28. L. Onural, "Exact analysis of the effects of sampling of the scalar diffraction field," *J. Opt. Soc. Am. A* **24**(2), 359–367 (2007).
29. A. Stern and B. Javidi, "Improved-resolution digital holography using the generalized sampling theorem for locally band-limited fields," *J. Opt. Soc. Am. A* **23**, 1227–1235 (2006).
30. A. Stern and B. Javidi, "Sampling in the light of Wigner distribution," *J. Opt. Soc. Am. A* **21**, 360–366 (2004).
31. A. Stern and B. Javidi, "Sampling in the light of Wigner distribution: errata," *J. Opt. Soc. Am. A* **21**, 2038 (2004).
32. L. Onural, "Sampling of the diffraction field," *Appl. Opt.* **39**, 5929–5935 (2000).
33. A. Stern and B. Javidi, "Space-bandwith conditions for efficient phase-shifting digital holographic microscopy," *J. Opt. Soc. Am. A* **25**, 736–741 (2008).
34. A. Stern and B. Javidi, "Analysis of practical sampling and reconstruction from Fresnel fields," *Opt. Eng.* **43**, 239–250 (2004).
35. C.-S. Guo, L. Zhang, Z.-Y. Rong, and H.-T. Wang, "Effect of the fill factor of CCD pixels on digital holograms: comment on the papers "Frequency analysis of digital holography" and "Frequency analysis of digital holography with reconstruction by convolution"," *Opt. Eng.* **42**, 2768–2771 (2003).
36. J. W. Goodman, *Statistical Optics*, Wiley, Hoboken, NJ, (1985).
37. R. Bracewell, *The Fourier Transform and its Applications*, McGraw-Hill, New York (1965).
38. A. Stern, "Sampling of linear canonical transformed signals," *Sig. Proc.* **86**, 1421–1425 (2006).
39. S. Haykin, *Communications Systems*, 4th ed., Wiley, Hoboken, NJ (2000).
40. T. M. Kreis, "Response to "Effect of the fill factor of CCD pixels on digital holograms: comment on the papers 'Frequency analysis of digital holography' and 'Frequency analysis of digital holography with reconstruction by convolution'"," *Opt. Eng.* **42**, 2772 (2003).
41. T. Kreis and W. Juptner, "Suppression of the dc term in digital holography," *Opt. Eng.* **36**, 2357–2361 (1997).
42. B. M. Hennelly and J. T. Sheridan, "Fast numerical algorithm for the linear canonical transform," *J. Opt. Soc. Am. A* **22**, 928–937 (2005).
43. L. Onural, "Some mathematical properties of the uniformly sampled quadratic phase function and associated issues in digital Fresnel diffraction simulations," *Opt. Eng.* **43**, 2557–2563 (2004).
44. L. Leushacke and M. Kirchner, "Three-dimensional correlation coefficient of speckle intensity for rectangular and circular apertures," *J. Opt. Soc. Am. A* **7**, 827–832 (1990).
45. A. W. Lohmann and J. Ojeda-Castañeda, "Fresnel similarity," *Opt. Commun.* **249**, 397–405 (2005).
46. D. P. Kelly, B. M. Hennelly, W. T. Rhodes, and J. T. Sheridan, "Analytical and numerical analysis of linear optical systems," *Opt. Eng.* **45**, 008201-1-008201-12 (2006).
47. B. M. Hennelly and J. T. Sheridan, "Generalizing, optimizing, and inventing numerical algorithms for the fractional Fourier, Fresnel, and linear canonical transforms," *J. Opt. Soc. Am. A* **22**, 917–927 (2005).
48. D. Mas, J. Garcia, C. Ferreira, L. M. Bernardo, and F. Marinho, "Fast numerical calculation of Fresnel patterns in convergent systems," *Opt. Commun.* **227**, 245–258 (2003).
49. I. Yamaguchi, J. ichi Kato, S. Ohta, and J. Mizuno, "Image formation in phase-shifting digital holography and applications to microscopy," *Appl. Opt.* **40**, 6177–6186 (2001).
50. U. Schnars and W. P. Jueptner, *Digital Holography: Digital Recording, Numerical Reconstruction and Related Techniques*, Springer-Verlag, Berlin (2005).
51. T. Nomura, S. Murata, E. Nitanai, and T. Numata, "Phase-shifting digital holography with a phase difference between orthogonal polarizations," *Appl. Opt.* **45**, 4873–4877 (2006).
52. H. Jin, H. Wan, Y. Zhang, Y. Li, and P. Qiu, "The influence of structural parameters of CCD on the reconstruction image of digital holograms," *J. Mod. Opt.* **55**, 2989–3000 (2008).
53. W. T. Rhodes, *Lecture Series: Fourier Optics and Holography*, Imaging Technology Center, Florida Atlantic University (2006).
54. M. Liebling, "A Fresnelet approach to digital holography," in *New Directions in Holography and Speckles*, pp. 69–86, American Scientific Publishers, Fisk University, Nashville, USA (2008).
55. J. Maycock, B. M. Hennelly, J. B. McDonald, Y. Frauel, A. Castro, B. Javidi, and T. J. Naughton, "Reduction of speckle in digital holography by discrete Fourier filtering," *J. Opt. Soc. Am. A* **24**, 1617–1622 (2007).
56. M. Liebling, PhD thesis, Swiss Federal Institute of Technology Lausanne (2004).
57. M. Liebling, T. Blu, and M. Unser, "Fresnelets: new multiresolution wavelet bases for digital holography," *IEEE Trans. Image Process.* **12**, 29–43 (2003).
58. M. Liebling and M. Unser, "Autofocus for digital Fresnel holograms by use of a Fresnelet-sparsity criterion," *J. Opt. Soc. Am. A* **21**, 2424–2430 (2004).

Biographies and photographs of the authors not available.