Spatial bandwidth analysis of fast backward Fresnel diffraction for precise computer-generated hologram design

Jinyang Liang[1,2] and Michael F. Becker[2,*]

[1]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA

[2]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712, USA

*Corresponding author: becker@uts.cc.utexas.edu

Received 2 April 2014; revised 25 June 2014; accepted 27 June 2014;
posted 27 June 2014 (Doc. ID 209345); published 31 July 2014

Designing near-field computer-generated holograms (CGHs) for a spatial light modulator (SLM) requires backward diffraction calculations. However, direct implementation of the discrete computational model of the Fresnel diffraction integral often produces inaccurate reconstruction. Finite sizes of the SLM and the target image, as well as aliasing, are major sources of error. Here we present a new design prescription for precise near-field CGHs based on comprehensive analysis of the spatial bandwidth. We demonstrate that, by controlling two free variables related to the target image, the designed hologram is free from aliasing and can have minimum error. To achieve this, we analyze the geometry of the target image, hologram, and Fourier transform plane of the target image to derive conditions for minimizing reconstruction error due to truncation of spatial frequencies lying outside of the hologram. The design prescription is verified by examples showing reconstruction error versus controlled parameters. Finally, it is applied to precise three-dimensional image reconstruction. © 2014 Optical Society of America

OCIS codes: (090.1760) Computer holography; (050.1940) Diffraction.
http://dx.doi.org/10.1364/AO.53.000G84

1. Introduction and Background

Holography has been used widely in many areas [1], including optical trapping and tweezing [2], laser beam shaping [3], and biomedical imaging [4]. To produce a high-quality holographic image at a certain distance, spatial amplitude and phase of the light must be precisely controlled. Modern spatial light modulators (SLMs) are becoming the standard device to create arbitrary complex wavefronts by modulating the wave magnitude and/or phase. Methods exist that utilize two modulators to achieve fully complex modulation [5–7], and some schemes can achieve complex modulation utilizing a single SLM with perhaps some trade-offs such as aperture sharing [8–10]. Having complex wave modulation readily available, the design of a desired near-field computer-generated hologram (CGH) becomes largely a matter of calculating the inverse diffraction from the target image back to the SLM. Backward propagation calculations can be performed using exact point-by-point algorithms [11,12] or various implementations of the Rayleigh–Sommerfeld (R-S) diffraction formula [13–18]. In addition, given the pixel size of current SLMs (typically 7–20 μm), diffraction within the paraxial approximation is justified for referenceless holography. Thus, diffraction calculations related to SLMs also can be performed with the Fresnel diffraction formula using a single Fourier transform (FT) [19–22]. Taking advantage of the discrete Fourier transformation (DFT), a single FT implementation of the Fresnel diffraction integral (FDI) possesses a much faster computation

1559-128X/14/270G84-11$15.00/0
© 2014 Optical Society of America

speed, as compared with direct implementations of the R-S formula, and can have accuracy comparable to a R-S computation within the Fresnel region [23].

However, direct implementation of these computational methods can yield inaccurate hologram reconstruction results. The first problem relates to sizing the two regions: SLM and target image. For a finite-sized target image, its spatial bandwidth determines the size of the hologram generated from backward propagation. Quite often, this area is larger than the finite size of the SLM, resulting in truncation of spatial frequency content. Those spatial frequency components that miss the SLM cannot be reproduced by the CGH encoded onto the SLM and are therefore lost in the reconstruction process. Consequently, the reconstruction cannot reproduce the original target image. Second, when the backward FDI is digitized so that the DFT can be used for numerical computations, the possibility of errors resulting from aliasing also arises. Previous work by Liu [21] discussed aliasing when using the forward FDI with a fixed bandwidth source image. However, this work cannot be readily applied to backward propagation to design a precise CGH for two main reasons. First, boundary conditions for the SLM (primarily its finite dimension; see Section 2 for a complete specification of the boundary conditions) were not considered in a backward diffraction calculation. Second, the target image bandwidth is now a free variable, and it will affect error and the setting of other parameters. For backward diffraction calculation, either of these reasons could cause significant error in hologram design and in the reconstructed image.

Our objective in this paper is to analyze and minimize these limitations and error sources, so as to synthesize a desired wavefront for a referenceless Fresnel hologram using backward propagation from a target image. This design process needs to be computationally fast and efficient. The designed hologram can then be encoded onto an appropriate SLM, and the reconstructed holographic image should have minimum error. We solve the problem of sizing the SLM and target image by requiring the target image to be accurate in irradiance but with its phase as a free variable. A quadratic phase is added to the target image, so that, when a hologram is designed, its diffracted light fills the target image region with minimum error. We present a comprehensive spatial bandwidth analysis of hologram design using backward FDI, which allows us to quantify the resulting hologram reconstruction errors versus both target image bandwidth and quantity of added quadratic phase. The result is a design prescription for precise hologram generation with minimum reconstruction error. Example simulations are shown that demonstrate practical hologram design within specific limits.

This paper is organized as follows. In Section 2, we take a detailed look at the boundary conditions of the backward diffraction problem. First, the relationship between the target image bandwidth, sizes of the SLM and target image, and their separation distance is examined. Second, we derive the conditions for adding quadratic phase to the target image as a free variable in order to properly scale between the two planes and to reduce error in the reconstructed image. Third, the geometry of the light path from the target image, past the SLM, and onto the FT plane of the target image allows the path of particular spatial frequency components to be traced. We perform this geometrical analysis, which identifies the regions of the target image that experience truncation of some spatial frequency components, and derive conditions that minimize error in the reconstructed images. Last, the conditions to avoid aliasing are derived. We will show that the conditions for minimum error due to truncation of spatial frequency components at the SLM plane are more stringent than the conditions to avoid aliasing. Section 3 is devoted to numerical simulations where the backward diffraction method to design the hologram is verified and the errors are quantified. First, the backward diffraction routine derived in Section 2 is validated against an exact point-by-point diffraction computation. The major content of this section examines hologram designs for different cases of optical layout. The root-mean-square (RMS) error calculated for each reconstructed image and the point of abruptly increasing error are compared with the theoretical predictions of Section 2. A discussion that summarizes our findings concludes the paper in Section 4.

2. Principles of Hologram Generation by Backward Propagation

We first consider the problem in the continuous, scalar-wave domain. As illustrated in Fig. 1, plane-wave light, of wavelength λ, illuminates a pixelated SLM in the $z = 0$ plane. The SLM is $L_o \times L_o$ in size with a square pixel pitch of Δ_o and a pixel number of $M_o \times M_o$. The closest possible hologram reconstruction plane is equal in size to the SLM. At this minimum distance, every image point can be reconstructed by a properly sampled Fresnel phase encoded lens at the SLM hologram plane

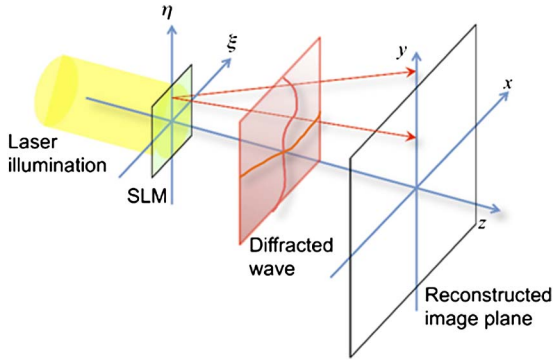

Fig. 1. General forward diffraction problem of generating a desired target wavefront (amplitude and phase) by propagation from an SLM to the target plane.

(i.e., R-S scalar diffraction). In other words, there is no vignetting of the reconstructed image. This minimum distance is defined as twice the critical length, z_c, defined by

$$z_c = \frac{M_o \Delta_o^2}{\lambda}. \tag{1}$$

Previous authors have introduced the critical length and used $2z_c$ to define the closest distance to employ the FDI method [14,21]. Subsequently, Davis and Cottrell also identified z_c as the minimum Nyquist sampled focal length of a phase-encoded lens centered on axis [24].

However, the problem illustrated in Fig. 1 is somewhat ill posed. The difficulty arises from the readily apparent boundary conditions at the target image and SLM planes; i.e., each is bounded in lateral extent to a region of interest that will later be sampled for digital computations. Specifically, for forward diffraction calculations from the bounded SLM plane, the extent of the diffraction solution, which stays within the right half plane from $-\pi/2$ to $+\pi/2$ in diffraction angle, extends outside the bounded target image, as shown in Fig. 1. When the SLM is pixelated, the referenceless hologram reconstructed image is located at the zeroth diffraction order. Due to a relatively large pixel size, the majority of the zeroth-order diffracted light stays in a relatively small area. Strictly speaking, however, because of the sharp pixel edge, the diffracted light contains high spatial frequencies. This high-spatial-frequency light, although weak, diffracts at large angles outside the bounded target image. This is most easily visualized using the R-S formulation, where the diffracted field is given by a convolution of the bounded wave, leaving the SLM with the unbounded spherical phase kernel. For the forward diffraction calculation, the area outside the region of interest is ignored as a "don't care" region. Furthermore, contributions to the convolution integral far from the center of the kernel are insignificant, and a practical implementation with minimum energy outside the region of interest can be formulated [23]. For a paraxial diffraction problem that employs the Fresnel approximation, bounded support at the SLM and target image planes creates a similar problem. But again, energy outside the region of interest can be minimized by correctly sizing the observation region and/or source function bandwidth to fit the problem.

Fresnel diffraction may be expressed using the FT integral as

$$U(x,y) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k}{2z}(x^2+y^2)} \iint_{-\infty}^{+\infty} \left\{ U(\xi,\eta) e^{j\frac{k}{2z}(\xi^2+\eta^2)} \right\} e^{-j\frac{2\pi}{\lambda z}(x\xi+y\eta)} \mathrm{d}\xi \mathrm{d}\eta, \tag{2}$$

where $k = 2\pi/\lambda$. The source amplitude distribution is given by $U(\xi,\eta)$, and the diffracted field at a distance z is given by $U(x,y)$. This forward Fresnel transform can be expressed compactly using Fourier operator notation as

$$U(x,y) = \frac{e^{\mathrm{jkz}}}{j\lambda z} e^{j\frac{\pi}{\lambda z}(x^2+y^2)} F\left\{ U(\xi,\eta) e^{j\frac{\pi}{\lambda z}(\xi^2+\eta^2)} \right\}. \tag{3}$$

In Eq. (3), $F\{\}$ denotes the forward FT, and the spatial frequency variables are given in terms of target plane coordinates (x,y) as $f_x = x/\lambda z$; $f_y = y/\lambda z$. We make the following definitions for the quadratic phase or chirp functions: the term, $e^{j\pi/\lambda z(\xi^2+\eta^2)}$, within the FT is denoted as $\phi_1(\xi,\eta)$, and the premultiplier term, $e^{j\pi/\lambda z(x^2+y^2)}$, is denoted as $\phi_2(x,y)$.

When using Eq. (3) for forward diffraction, the source function $U(\xi,\eta)$ has bounded support and so does the product of $U(\xi,\eta)$ and $\phi_1(\xi,\eta)$. This product function cannot have a bounded spectrum (i.e., bounded FT) and thus cannot be rigorously expressed in a bounded region at the target image plane. In practice, limiting the bandwidth of the source function narrows the maximum diffraction angle, so that the source and diffraction regions of interest can be bounded and produce minimal error. The error arises from the fact that the source cannot be band limited without making its spatial extent infinite. To solve this problem in practice, the usual approach is to apply an infinite impulse response filter to the source function that leaves support for the spectrum unbounded but still confines the major spectral energy to a usefully limited spectral region. Truncation at the edge of the diffraction region of interest and disregarding the lost energy can now result in a tolerably insignificant error.

To design the hologram that will be coded onto the SLM to reconstruct the target image, the problem is reversed, and backward diffraction is used to determine the desired light field at the SLM plane. For backward propagation, Eq. (3) is inverted [25] as

$$U(\xi,\eta) = (-j\lambda z) e^{jkz} e^{j\frac{\pi}{\lambda z}(\xi^2+\eta^2)} F^{-1}\left\{ U(x,y) e^{j\frac{\pi}{\lambda z}(x^2+y^2)} \right\} (z<0), \tag{4}$$

where $F^{-1}\{\}$ denotes the inverse FT, and the spatial frequency variables are given in terms of the SLM plane coordinates (ξ,η) as $f_\xi = \xi/\lambda z$; $f_\eta = \eta/\lambda z$.

As the target function and SLM are necessarily finite in size, the backward diffraction solution will extend beyond the SLM edges due to the same bounded support issue as faced by forward diffraction. Thus, in practical design problems, the bandwidth of the target function must be subjected to a bandwidth-limiting process. In order to have a tractable target function, the hard edges of the image also must be low-pass filtered and zero padded. An important difference for hologram design by backward diffraction, as opposed to forward diffraction, is that some important information may extend beyond the SLM boundary, and this information is necessary for rigorously accurate target function

reconstruction when the SLM is illuminated. Handling and bounding this hologram reconstruction error is an important result to be presented in this paper.

A. Sizing the Hologram

Based on the discussion above, the first aspect of hologram design, which can be dealt with in the continuous domain, is the sizing of the SLM and the target image with respect to each other and their separation. For backward diffraction, a phase-flat target image illuminated by a plane wave will not diffract back to a smaller source area. A quadratic phase, equivalent to a positive lens in the Fresnel approximation, is required in the target image. This requirement is illustrated in Fig. 2, and it is equivalent to illuminating a phase-flat target image with a converging spherical wave. An FT plane will exist at the center of curvature of this converging wave. In addition, if this phase function is in the form of a positive lens, it is opposite in sign to $\phi_2(x,y)$. By adjusting the quadratic phase in $U(x,y)$, the phase of the product $\{U(x,y)\phi_2(x,y)\}$ can be reduced to zero or to a small residual quadratic phase of either a positive or negative sign.

Figure 2 also illustrates the geometrical optical path of the highest spatial frequency in the target image. The SLMs shown in Fig. 2 are sufficiently wide, as to not truncate any of this high spatial frequency light. When the SLM is encoded based on this computed backward-propagated light field, all the light at this spatial frequency can be diffracted by the SLM to the reconstructed image plane. The heavy solid lines in Fig. 2 illustrate the minimum size of the SLM to avoid high-frequency truncation. Two general cases are illustrated: the SLM to the right of the FT plane (requiring quadratic phase radius of curvature larger than the separation between the SLM and target planes), and the SLM to the left of the FT plane (requiring a radius of curvature smaller than that separation).

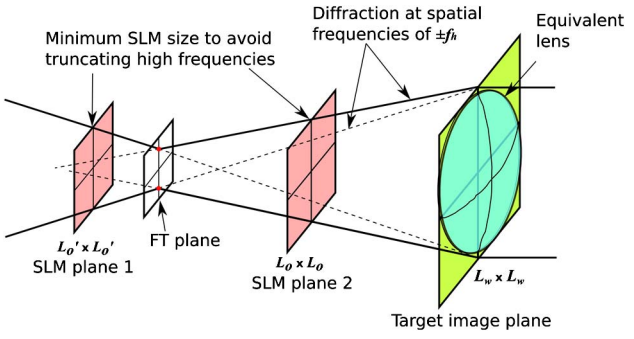

Fig. 2. Effect of a quadratic phase in the target image expressed as an equivalent lens. Backward propagation results in an FT plane located at the phase center of curvature. Two of the many possible SLM planes are shown that accommodate the highest spatial frequency light. These possible SLM sizes are bounded by the heavy solid lines in the figure that connect to the highest spatial frequency ($\pm f_h$) locations in the FT plane (red dots).

To fit the backward diffraction field onto the finite SLM area, we introduced a quadratic phase in the target image. The term in Eq. (4) to be inverse Fourier transformed is given by

$$U(x,y)\phi_2(x,y) = U_R(x,y)\exp\left[+j\frac{\pi}{\lambda(z_f z_c + a z_c)}(x^2+y^2)\right] \times \exp\left[-j\frac{\pi}{\lambda z_f z_c}(x^2+y^2)\right], \quad (5)$$

where a dimensionless propagation distance z_f is defined as $z_f = |z|/z_c$. The target image amplitude is now the explicitly real function $U_R(x,y)$. The parameter a controls the curvature of the quadratic phase added in the target image. The sign of a can be either $+$ or $-$, and, for $a = 0$, the quadratic phase added to the target image exactly cancels the phase of $\phi_2(x,y)$. In this case, backward Fresnel propagation reduces to the inverse FT of the real target image.

Although placing the SLM at the FT plane is attractive because it has the minimum required size for the SLM (Fig. 2), it is difficult to implement in practice. The strong DC component [26] of the FT requires a very large dynamic range for the light at this plane and for the amplitude modulation capability of the SLM. For practical reasons, the SLM is best placed a short distance away from the FT plane to alleviate this dynamic range problem. Thus, the parameter a should be small, to avoid high-frequency truncation, but not be exactly zero.

The truncation effect in Fresnel holograms was discussed previously by Stern and Javidi [27]. From the spatial-frequency perspective, the spectrum of the target image was low-pass filtered by the hard-edged aperture function, resulting in error in the reconstructed image. Without phase added to the object, the aperture effect was simple and space invariant. In contrast, we introduced an additional quadratic phase to the object that interacted with $\phi_2(x,y)$ to alter the overall phase profile. For this case, the effect of truncation on the reconstructed irradiance image is spatially dependent, since $U_R(x,y)$ with the added phase being low-pass filtered and converted to irradiance. Phase was added to the target image in order to scale the diffracted field to the SLM aperture. It also allows us to minimize the truncation problem and therefore ensure more precise image reconstruction.

We now proceed to analyze the requirements to avoid high-frequency truncation using geometrical analysis. The lateral size of the target image and the hologram are denoted by L_w and L_o. The bounds on a to avoid high-frequency truncation can be derived using similar triangles shown in Fig. 3 for positive a and negative a. For both cases, similar triangles have the common solid (green) and dashed (red) sides. The heights of the similar triangles are h and $L_w/2 \pm L_o/2$ for the smaller and larger triangles, respectively. At the FT plane, the highest spatial frequency in the target image diffracts to height x_0

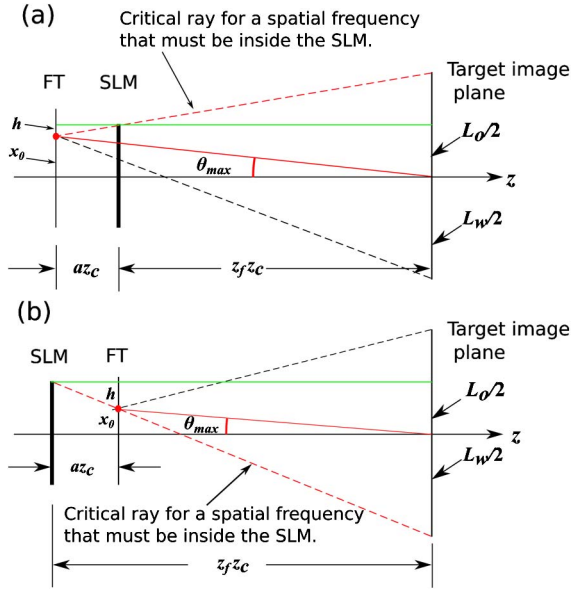

Fig. 3. Light paths for backward propagation between the target image and SLM planes when there exists an FT plane located at a finite distance from the target image. $\theta_{\max}$ is the maximum diffraction angle from the target. Critical rays for high-frequency truncation are shown. In (a) a positive value a is shown and in (b) a negative value.

from the z axis. The value of a must allow the dashed (red) critical ray for the highest spatial frequency to reside within the SLM area.

From geometry, the condition on a to avoid high spatial frequency truncation is given by

$$|az_c| \le \frac{z_f z_c}{\frac{L_w/2}{L_o/2 - z_f z_c \theta_{\max}} \mp 1}, \tag{6}$$

where the minus sign in $\mp$ should be used for a positive value of a and vice versa. The meaning of this inequality will be explored further when the problem is discretized.

B. Discretization

To implement our method by digital computation, the problem was digitized. The geometry of the discretized diffraction problem is shown in Fig. 4. The computation window is M pixels wide at both planes. The width of the active windows for the hologram and target image are M_o and M_w pixels, respectively. The computation window, an $M \times M$ matrix, is introduced along with the zero padding factors $\mu_o = M/M_o$ for the SLM plane and $\mu_w = M/M_w$ for the target image plane. Both padding factors are greater than or equal to 1. The sample interval at the target image in relation to the SLM pixel size is determined from the FT spatial frequency definition as

$$\Delta = \frac{\lambda z}{M\Delta_o} = \Delta_o \frac{z_f}{\mu_o}. \tag{7}$$

The normalized bandwidth of the target image, b, determines the maximum diffraction angle from the

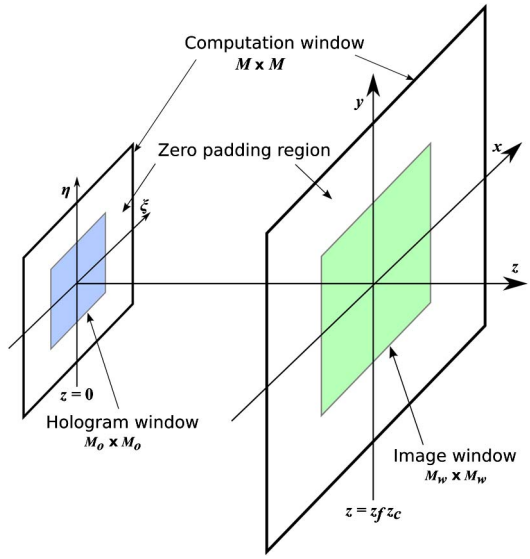

Fig. 4. Relationship of the hologram (SLM) plane (ξ, η) at $z = 0$ with pixel pitch Δ_o to the reconstructed image plane (x, y) at $z = z_f z_c$ with pixel pitch Δ. The computation window is M pixels wide at both planes. The active hologram window is M_o pixels wide, and the active window of the target image is M_w pixels wide.

target, which is $\theta_{\max} = b\theta_{Nyq}$, where $\theta_{Nyq} = \lambda/2\Delta$ denotes the paraxial approximation to the diffraction angle at the Nyquist sampling frequency.

For backward propagation, correct sampling of the $\{U(x,y)\phi_2(x,y)\}$ product in Eq. (5) is determined by its bandwidth. This in turn depends on the bandwidths of $U_R(x,y)$ and the product of the two exponentials in Eq. (5). The product of the two exponentials can be reduced to a single quadratic exponential with an equivalent radius of curvature, z_{eq}, given by

$$\frac{1}{z_{\text{eq}}} = \frac{1}{z}\left(\frac{a}{z_f + a}\right) = \frac{1}{z} f(a). \tag{8}$$

The total available sampling bandwidth at the target image plane, $1/2\Delta$, must be greater than or equal to the sum of the bandwidths of the two functions in the product $\{U(x,y)\phi_2(x,y)\}$. This relation is expressed as

$$\frac{M_w \Delta}{2\lambda z} f(a) + \frac{b}{2\Delta} \le \frac{1}{2\Delta}, \tag{9}$$

where the first term represents the bandwidth of the combined product of the quadratic exponentials, and the second is the bandwidth allocated to the target image function. Equation (9) is the expression of the bandwidth constraint to prevent aliasing of the $\{U(x,y)\phi_2(x,y)\}$ product with the boundary conditions noted above. This may be solved for the conditions on a for a given target image bandwidth, b, as

$$|a| \le \frac{z_f}{\frac{z_f}{\mu_w \mu_o (1-b)} \mp 1}, \tag{10}$$

where the minus sign in $\mp$ should be used for a positive value of a and vice versa, both here and in Eq. (11).

The condition to avoid high-frequency truncation [Eq. (6)] can also be converted to the variables of the discrete problem. Substituting for L_o, L_w, and $\theta_{\max}$ in Eq. (6), one obtains

$$|a| \le \frac{z_f}{\frac{z_f}{\mu_w(1-b\mu_o)} \mp 1}. \tag{11}$$

For $\mu_o = 1$, Eq. (11) is identical to Eq. (10). In the DFT domain, the function calculated at the SLM plane by the inverse DFT is one period in a plane of sampled, replicated units [28]. Clearly, this makes no sense when the objective is to study the effects of high spatial frequency components that would miss the SLM edge. Instead, these high-frequency components would be aliased back toward the center. For this reason, the case of $\mu_o = 1$ is not of interest. For $\mu_o > 1$, Eqs. (10) and (11) are no longer identical, and Eq. (11) becomes a more stringent condition on a. This implies that proper sampling of $\{U(x,y)\phi_2(x,y)\}$ is guaranteed when the condition for no high-frequency truncation at the SLM edges is met.

At the SLM plane, the full complex field is needed to program the SLM values. This requires that the product of the pre-exponential $\phi_1(\xi,\eta)$ and the inverse FT, $F^{-1}\{U(x,y)\phi_2(x,y)\}$, in Eq. (4) also be properly sampled. This in turn requires that the SLM plane sampling bandwidth be sufficient as given by

$$\frac{M_o\Delta_o}{2\lambda z} + \frac{1}{2\mu_w\Delta_o} \le \frac{1}{2\Delta_o}, \tag{12}$$

where the first term represents the bandwidth required for $\phi_1(\xi,\eta)$, and the second term is the bandwidth of the inverse FT of the $\{U(x,y)\phi_2(x,y)\}$ product. The second term is independent of a but depends on the spatial width of the target image region. Choosing the target plane padding factor to be $\mu_w = 2$ apportions the available bandwidth equally between the quadratic phase, $\phi_1(\xi,\eta)$, and $F^{-1}\{U(x,y)\phi_2(x,y)\}$. After noting that $z = z_f z_c$, and substituting for z_c in Eq. (12), this simply states that $z_f \ge 2$. This is always true since $z \ge 2z_c$ is required to avoid vignetting of the reconstructed image. It is important to notice that the right hand side indicates that the full bandwidth of the SLM is utilized. To correctly sample the hologram without aliasing for the subsequent forward propagation calculation, however, it is required that the sampled data have bandwidth of half the Nyquist frequency of sampling [21]. This may be achieved by resampling the matrix for the forward propagation calculation by a factor of two. At this point, we choose $\mu_o = 2$ for equal size and symmetry of the regions of interest. As a result of Eq. (11), the bandwidth of the target image must be strictly less than half the Nyquist frequency of the sampling because, if $(1 - b\mu_o)$ goes to zero, the allowed range of a also goes to zero.

In summary, the conditions for applying a single DFT implementation of the inverse FDI have been derived to obtain the correctly sampled magnitude and phase of the diffracted field. Besides the requirement of padding factors μ_o and μ_w, proper choice of a for a given target image bandwidth b is sufficient for using the FDI to solve the backward propagation problem. A quadratic phase must multiply the flat-phase target image, and its curvature must satisfy Eq. (11) to guarantee that no high-frequency truncation occurs at the SLM and for proper sampling at the target plane. Also, $z \ge 2z_c$ guarantees proper sampling at the SLM plane. Finally, note that the right side of Eq. (12) requires the use of the whole available bandwidth of the SLM, and that this sampling period must be considered for any subsequent forward FDI diffraction computations.

3. Simulations

To design a hologram and validate its design or to study the effects of encoding algorithms at the SLM plane, at least two approaches are possible. First, we validate the SLM design by doing a point-by-point backward R-S computation from the target function. This is computationally slow, so only a few cases were studied. It is also important to propagate the designed hologram back to the reconstruction plane and compare it with the original target image. In the next section (Section 3.A), our backward design FDI algorithm with added quadratic phase is validated against an exact point-by-point diffraction calculation from points in the target image back to the SLM plane. Following that, example holograms are designed and reconstructed using the FDI method, and the RMS error dependence on the various parameters is determined.

The following parameters were used throughout the simulations: $\lambda = 633$ nm, $\Delta_o = 13.68$ μm, and for backward propagation, $\mu_o = \mu_w = 2$ and $M = 1024$. For forward propagation, the matrix $(M \times M)$ was resampled by a factor of two, as described in Section 2.B, becoming 2048×2048. The quality of reconstructed images was quantified by using the RMS error, defined by

$$\sigma_{\text{RMS}} = \frac{1}{\bar{I}_t}\sqrt{\iint [I_o(x,y) - I_t(x,y)]^2 \mathrm{d}x\mathrm{d}y}, \tag{13}$$

where $I_o(x,y)$ and $I_t(x,y)$ are the intensity of the reconstructed image and the target image, respectively, with equalized irradiance. $\bar{I}_t$ is the averaged irradiance of the target image.

A. Validating the Backward Propagation Algorithm

The algorithm for fast backward propagation using a single DFT, as described in Eq. (4), and the sampling scheme of the previous section was validated for accuracy against an exact R-S diffraction calculation.

Fig. 5. Double Lena target image (512×512) with a $b = 0.35$ low-pass Butterworth filter applied.

The exact computation used a point-by-point summation of the backward diffracted fields from each pixel of the target image. Although computationally slow, this technique is rigorously correct. Complex wave fields at the SLM plane computed by both methods were compared using a double Lena target image, shown in Fig. 5. A fourth-order, digital Butterworth low-pass filter was applied to the target image to confine its bandwidth [29]. The irradiance transmittance of the filter was 0.1 at the cut-off frequency given by b, and 95% of the energy was contained within the circle of spatial frequencies less than b. When normalized to the same total energy, the comparison diffracted wave amplitudes for all pixels differed by less than one part in 10^{-12}, and the phase differed by less than 10^{-10} radians. When the backward and forward FDI algorithms were used together without any truncation of the fields at the SLM plane, the exact target image was recovered ($<10^{-10}$% RMS error).

B. Hologram Design Examples

To verify and illustrate the assertions of the previous sections, we conducted a series of hologram design simulations using backward and forward diffraction. A complete cycle of backward and forward diffraction without SLM encoding recovers the exact target image. Truncating the forward wave to occupy the SLM window will result in small or large errors. Because the inequality governing high-frequency truncation is more stringent than the one governing aliasing, the reconstructed image will be examined with various values of added quadratic phase, a, and target image bandwidth, b, to see the effects of violating the inequality of Eq. (11) on the reconstructed image quality.

Equation (11) specifies that, to avoid high-frequency truncation, the phase chirp factor a must be limited to a region that depends on target image bandwidth, b. We plot this region as a function of the normalized propagation distance z_f in Fig. 6. Three different target image bandwidth parameters ($b = 0.25$, 0.35, and 0.45) were chosen for simulation; all were less than the absolute limit of 0.5 indicated by Eq. (11). These curves illustrate that a larger

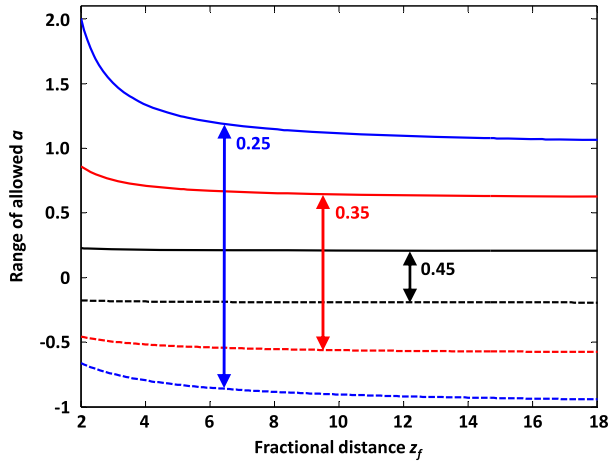

Fig. 6. Range of allowed phase chirp factor a to avoid high-frequency truncation for three bandwidths, $b = 0.25$ (blue), 0.35 (red) and 0.45 (black). The minimum and maximum boundaries of a are plotted as dashed and solid lines, respectively.

bandwidth tolerates a smaller range of a. For example, at the fractional distance $z_f = 8$, the range of a is $\Delta a_{0.25} = [-0.88, 1.15]$ for $b = 0.25$, $\Delta a_{0.35} = [-0.55, 0.65]$ for $b = 0.35$, and $\Delta a_{0.45} = [-0.19, 0.21]$ for $b = 0.45$, respectively.

The effect of high-frequency truncation on the quality of the reconstructed image was first examined by using one-dimensional (1D) spatial cosine images at $z_f = 8$. The amplitude of these cosine images is defined by $U(x, y) = \cos(\pi(xb/\Delta_o))$. Cosine images serve as good candidates to verify Eq. (11) because the majority of their energy is concentrated at a single spatial frequency determined by b. Therefore, the conformity between the reconstructed and target images will be broken when a specific value of a causes high-frequency truncation. Figure 7 shows the RMS error of reconstructed cosine images versus chirp factor a. The high-frequency truncation effect is pronounced for all three cosine frequencies, as all three curves for RMS error show a sharp increase when the corresponding spatial frequency moves off

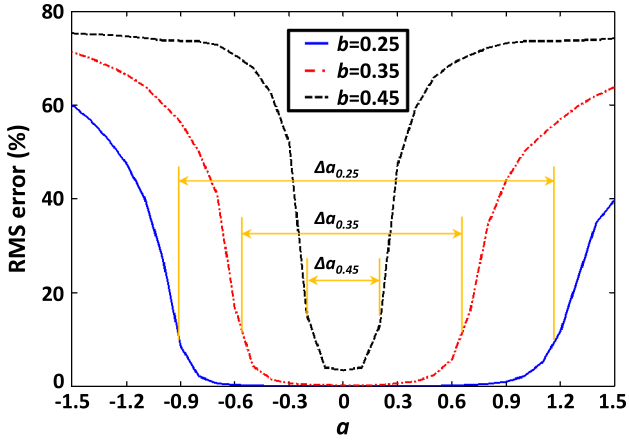

Fig. 7. RMS error of reconstructed 1D spatial cosine images versus phase chirp factor a. The range allowed for a to avoid high-frequency truncation for each bandwidth is identified as $\Delta a_{0.25}$ for $b = 0.25$, $\Delta a_{0.35}$ for $b = 0.35$, and $\Delta a_{0.45}$ for $b = 0.45$.

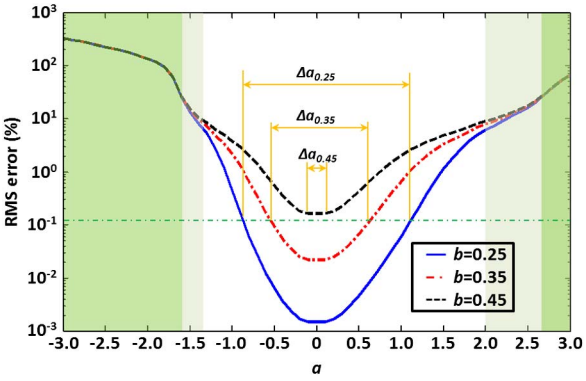

Fig. 8. RMS error of the reconstructed image versus phase chirp factor a. The allowed range of a without high-frequency truncation for each bandwidth is identified as $\Delta a_{0.25}$ for $b = 0.25$, $\Delta a_{0.35}$ for $b = 0.35$, and $\Delta a_{0.45}$ for $b = 0.45$. The truncation region for the lowest 10% of spatial frequencies is marked in gray. The truncation region of the DC component is marked in light green.

of the SLM region and becomes truncated. The simulation results are in good agreement with Eq. (11).

To implement our proposed method with a grayscale image, we conducted simulations using the double Lena target image at a distance of $z_f = 8$. The target image was band limited by the Butterworth low-pass filter. The RMS error of the reconstructed images versus a is plotted in Fig. 8. Different from the RMS error curves for cosine target images, no abrupt increase in error was observed at the high-frequency truncation boundary. At the boundary of the high-frequency truncation region, the reconstructed images for all three bandwidths have around 0.1% RMS error. This is because the main spatial frequency content of the target image is concentrated at low frequencies, and the small residual error arises from the spectral energy beyond the cut-off frequency of the Butterworth low-pass filter. High-frequency truncation does not cause a significant degradation of the reconstructed image near the onset value of a. The RMS error curves increase steeply when major spectral energy in the low-frequency range (from the DC component [26] to $b = 0.1$) is truncated. This is illustrated in Fig. 8, where the RMS error increases to 10%–30% when the truncated frequency reaches $b = 0.1$. Finally, the three RMS error curves converge when truncation of the DC component begins at the image edges resulting in an even larger RMS error.

To examine the reconstruction quality for different cases, a mosaic of representative reconstructed images is displayed in Fig. 9(a). At the onset of the high-frequency truncation region, reconstructed images resemble the target image. Truncation causes minor ripples to appear at the outer edges and blurs some fine spatial structures, resulting in a small RMS error. For example, Figs. 9(b)–9(d) show reconstructed images for $a = 1$ and bandwidths of

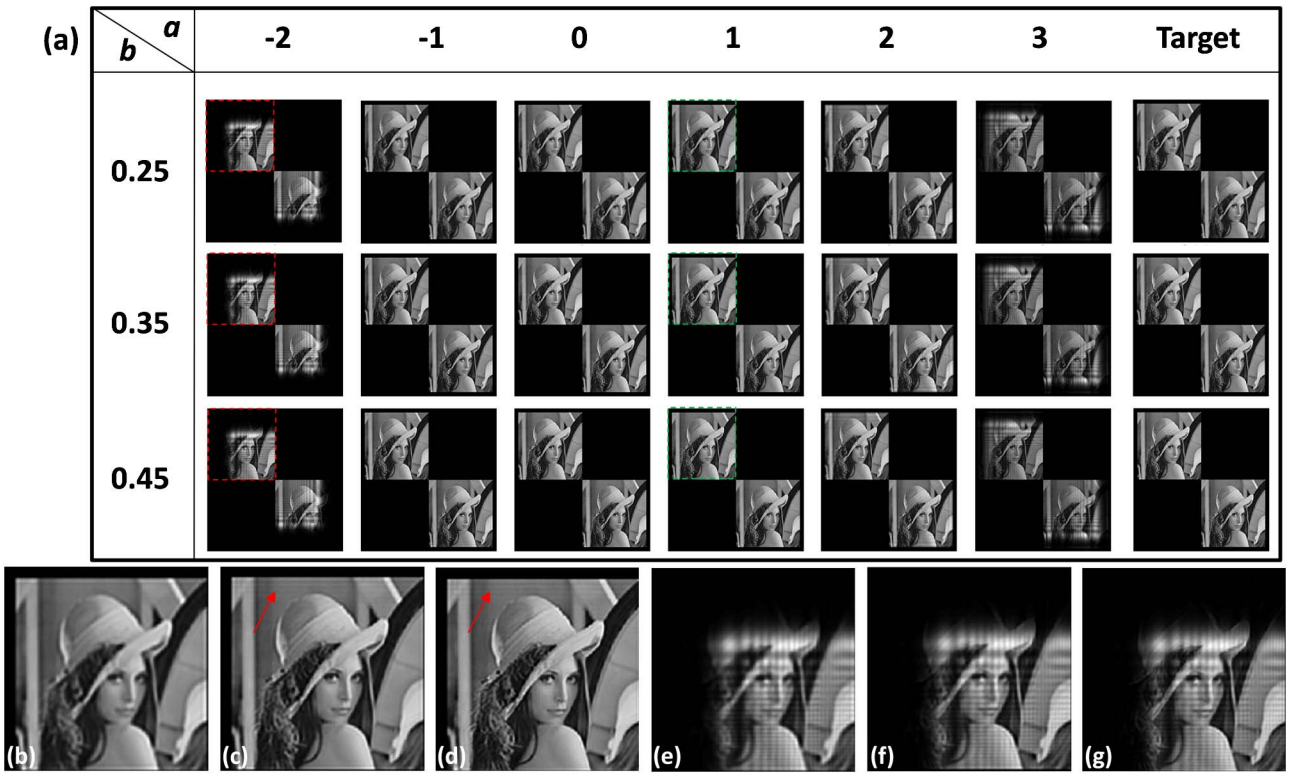

Fig. 9. (a) A mosaic of representative reconstructed images for various values of the phase chirp factor a (columns) and fractional bandwidth b (rows). Below this, in the bottom row, left to right are images (b)–(d) that show zoomed views of reconstructed images [green dashed squares in mosaic (a)] for $a = 1$ and bandwidths of $b = 0.25$ in (b), $b = 0.35$ in (c), and $b = 0.45$ in (d), respectively. Red arrows indicate the ripples caused by high-frequency truncation. In the bottom row, further to the right, images (e)–(g) are zoomed views of reconstructed images [red dashed squares in mosaic (a)] for $a = -2$ and bandwidths of $b = 0.25$ in (e), $b = 0.35$ in (f), and $b = 0.45$ in (g), respectively.

$b = 0.25$, 0.35, and 0.45. For $a = 1$, the target image with a bandwidth of $b = 0.25$ remains within the truncation-free region, while target images with the other two bandwidths experience high-frequency truncation during the reconstruction process. As a result, small ripples are visible in Figs. 9(c) and 9(d) near the outer edges. The small, single ripple visible in Fig. 9(b) is also present in the target image and results from the Butterworth low-pass filter, not high-frequency truncation.

At larger values of a, low frequency and then DC content are lost. In Fig. 9(a), this truncation results in severe degradation at the edges of the reconstructed images. More specifically, Figs. 9(e)–9(g) show reconstructed images for $a = -2$ and bandwidths of $b = 0.25$, 0.35, and 0.45. All three images suffered a severe aperture effect. The upper edge and left edge are completely lost, and the entire image is affected by significant ripples.

By examining the reconstructed images for different values of a, we notice that the truncation effect is spatially dependent. At the onset of high-frequency truncation, the minor ripples are only observed at the outermost image boundaries as seen in Figs. 9(b)–9(d). Similarly, truncation of the DC component also starts at the outer edges and proceeds toward the center. In addition, the ripple fringes in truncation of the DC component have the largest period at the corner, and the period decreases as they approach the center as seen in Figs. 9(e)–9(g). This phenomenon is predicted by Fresnel diffraction and also by the geometric analysis, shown in Figs. 2 and 3, which approximates Fresnel diffraction.

Finally, we demonstrated three-dimensional (3D) holographic image reconstruction. As illustrated in Fig. 10(a), two Lenas in the target image are separated in distance along the z axis. The upper-left Lena image is placed at $z_f = 4$; the lower-right image is placed at $z_f = 16$. Both images have a normalized bandwidth of $b = 0.35$. In the simulation, the chirp factor a was set to be $a = 0.6$, so that the backward FDI calculation for both images stayed within the truncation-free region defined by Eq. (11). The two hologram fields generated by these images were added and encoded on the SLM. The holographic images were reconstructed at the two preset propagation distances of $z_f = 4$ and $z_f = 16$, as shown in Figs. 10(b) and 10(c). At each distance, the corresponding Lena image was successfully reconstructed with minimum error while the other image was blurred.

In summary, these simulations validate the fast backward diffraction routine that we implemented. When the computed hologram field was not truncated at the SLM edge, the RMS error of the reconstructed image was less than $10^{-10}\%$. The size of the field distribution at the SLM was controlled by the phase chirp added to the target image using the parameter a. The chief error mechanisms were related to truncating information at the SLM edges as a became larger. High frequencies were lost first at the edges of the reconstructed image. As lower frequencies became truncated, larger errors appeared, mostly in the form of ripples and resolution loss localized at the image edges. Severe errors resulted when low spatial frequencies and the DC component were truncated for large values of a. This error also began at the image edges and progressed inward, blocking more of the reconstructed image. The limits on a derived in Section 2 were demonstrated in the simulated examples. The onset of aliasing was not reached in any of the test cases, since it was a less severe limit, as shown by Eqs. (10) and (11). Most importantly, limiting the bandwidth of the target image was shown to be important in reducing error in the reconstructed image. Target image bandwidth is strictly limited to be less than half the Nyquist frequency of the image sampling, and, for practical cases, it should be less than this.

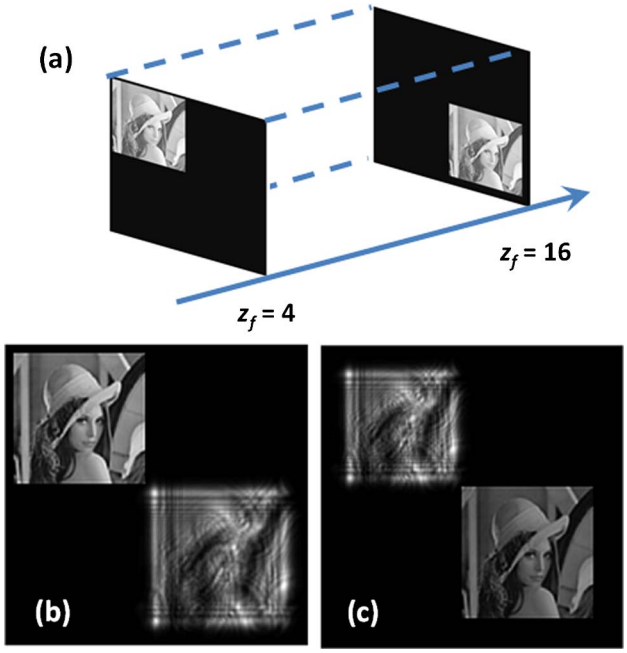

Fig. 10. Illustration of 3D holographic image reconstruction. (a) The target image with two Lena images displaced along the z axis at $z_f = 4$ and $z_f = 16$. (b) The reconstructed image at the plane $z_f = 4$. (c) The reconstructed image at the plane $z_f = 16$. Intensity images are shown with their dynamic range expanded to fill the entire gray scale, 0–255.

4. Conclusion

We have presented a design prescription and simulated the performance of fast backward Fresnel diffraction for precise CGH design. Our method used a single DFT implementation of the FDI. Careful consideration was given to the spatial bandwidths of the target image and the hologram encoded onto the SLM and their boundary conditions. The target image bandwidth and added quadratic phase were treated as free variables that could be adjusted to minimize reconstruction error.

Truncation of the backward diffracted field at the edges of the SLM was found to play a major role in

introducing error in the reconstructed image. This error resulted from the spatially limited support (finite width) of the target image and the SLM. The conditions to avoid this error were derived and shown to be more stringent than those used to avoid aliasing of the sampled functions. The quadratic phase added to the target image can reduce this error by sizing the backward diffracted light to better fit the SLM area. The bandwidth of the target image also played a role in the spatial distribution of the backward propagated light at the SLM. This bandwidth was strictly limited to less than half the Nyquist sampling frequency of the target image. An inequality was derived that computed the onset conditions for error due to truncation of spectral frequency components at the SLM edge. The truncation onset depended on the parameters governing the added phase and the target image bandwidth as well as the zero-padding factors used at the target and SLM planes. Zero padding was required at both planes. Finally, proper sampling in the backward FDI algorithm required that the distance between the target image and SLM must be greater than twice the critical distance, $z \geq 2z_c$. We showed that the designed hologram pattern used the entire spatial bandwidth of the SLM. To correctly sample this SLM pattern for forward diffraction computation with the FDI, it must be resampled at the SLM by a factor of two to avoid aliasing when the reconstructed image was computed.

The accuracy of the fast FDI method for hologram design was verified by precise agreement between the simulations of backward diffraction by the fast FDI method and by an exact point-by-point R-S calculation. Simulations of typical optical hologram design arrangements illustrated reconstructed image quality in the low error region and for mild and severe spatial frequency truncation error. RMS error for the images was used to quantify the error versus the control parameters of target image bandwidth and added phase chirp. For the three bandwidths that were investigated, the low reconstruction error region for the phase chirp parameter was narrow for a target image normalized bandwidth of 45% of the Nyquist frequency, but the low error region increased substantially as the image bandwidth was decreased. Regions of RMS error $\leq 0.1\%$ could be found for all target image normalized bandwidths less than 45%. Careful choice of the bandwidth reduction method for the target image is required, so that this process does not introduce additional unwanted image artifacts. Finally, a 3D target image was encoded using our prescription, and accurate 3D reconstruction at the different in-focus image planes was demonstrated.

This work was supported by one grant jointly from the Army Research Office and the Defense Advanced Research Projects Agency (DARPA) as part of the Optical Lattice Emulator Initiative (OLE) program.

References

1. S. A. Benton and V. M. Bove, Jr., *Holographic Imaging* (Wiley-Interscience, 2008).
2. D. G. Grier, "A revolution in optical manipulation," Nature **424**, 810–816 (2003).
3. S. Ngcobo, I. Litvin, L. Burger, and A. Forbes, "A digital laser for on-demand laser modes," Nat. Commun. **4**, 2289 (2013).
4. Y. M. Wang, B. Judkewitz, C. A. DiMarzio, and C. H. Yang, "Deep-tissue focal fluorescence imaging with digitally time-reversed ultrasound-encoded light," Nat. Commun. **3**, 928 (2012).
5. R. D. Juday and J. M. Florence, "Full complex modulation with 2 one-parameter SLMs," Proc. SPIE **1558**, 499–504 (1991).
6. L. G. Neto, D. Roberge, and Y. L. Sheng, "Full-range, continuous, complex modulation by the use of two coupled-mode liquid-crystal televisions," Appl. Opt. **35**, 4567–4576 (1996).
7. J. M. Herrera-Fernandez and L. M. Sanchez-Brea, "Double diffractive optical element system for near-field shaping," Appl. Opt. **50**, 4587–4593 (2011).
8. A. Jesacher, C. Maurer, A. Schwaighofer, S. Bernet, and M. Ritsch-Marte, "Near-perfect hologram reconstruction with a spatial light modulator," Opt. Express **16**, 2597–2603 (2008).
9. J. P. Liu, W. Y. Hsieh, T. C. Poon, and P. Tsang, "Complex Fresnel hologram display using a single SLM," Appl. Opt. **50**, H128–H135 (2011).
10. S. Reichelt, R. Haussler, G. Futterer, N. Leister, H. Kato, N. Usukura, and Y. Kanbayashi, "Full-range, complex spatial light modulator for real-time holography," Opt. Lett. **37**, 1955–1957 (2012).
11. I. Mertz and N. O. Young, "Fresnel transformation of images," in *International Conference on Optical Instruments and Techniques*, K. J. Habell, ed. (Chapman & Hall, 1961), pp. 305–310.
12. T. Nishitsuji, T. Shimobaba, T. Kakue, N. Masuda, and T. Ito, "Fast calculation of computer-generated hologram using the circular symmetry of zone plates," Opt. Express **20**, 27496–27502 (2012).
13. A. Vanderlugt, "Optimum sampling of Fresnel transforms," Appl. Opt. **29**, 3352–3361 (1990).
14. D. Mendlovic, Z. Zalevsky, and N. Konforti, "Computation considerations and fast algorithms for calculating the diffraction integral," J. Mod. Opt. **44**, 407–414 (1997).
15. R. Silvennoinen and M. Mozerov, "Controlled effects of aliasing synthetic Fresnel holograms with pixel phase error function," Opt. Eng. **36**, 558–565 (1997).
16. L. Onural, "Sampling of the diffraction field," Appl. Opt. **39**, 5929–5935 (2000).
17. T. M. Kreis, "Frequency analysis of digital holography with reconstruction by convolution," Opt. Eng. **41**, 1829–1839 (2002).
18. 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).
19. D. Mas, J. Garcia, C. Ferreira, L. M. Bernardo, and F. Marinho, "Fast algorithms for free-space diffraction patterns calculation," Opt. Commun. **164**, 233–245 (1999).
20. D. G. Voelz and M. C. Roggemann, "Digital simulation of scalar optical diffraction: revisiting chirp function sampling criteria and consequences," Appl. Opt. **48**, 6132–6142 (2009).
21. J. P. Liu, "Controlling the aliasing by zero-padding in the digital calculation of the scalar diffraction," J. Opt. Soc. Am. A **29**, 1956–1964 (2012).
22. N. Okada, T. Shimobaba, Y. Ichihashi, R. Oi, K. Yamamoto, M. Oikawa, T. Kakue, N. Masuda, and T. Ito, "Band-limited double-step Fresnel diffraction and its application to computer-generated holograms," Opt. Express **21**, 9192–9197 (2013).
23. J. W. Goodman, *Introduction to Fourier Optics* (Roberts & Company, 2005).
24. J. A. Davis and D. M. Cottrell, "Ray matrix analysis of the fast Fresnel transform with applications towards liquid crystal displays," Appl. Opt. **51**, 644–650 (2012).

25. G. C. Sherman, "Integral-transform formulation of diffraction theory," J. Opt. Soc. Am. **57**, 1490 (1967).
26. A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, 2nd ed. (Prentice Hall, 1997), p. 207.
27. A. Stern and B. Javidi, "Analysis of practical sampling and reconstruction from Fresnel fields," Opt. Eng. **43**, 239–250 (2004).
28. Y. Xiao, X. H. Tang, Y. X. Qin, H. Peng, and W. Wang, "Band-limited angular spectrum numerical propagation method with selective scaling of observation window size and sample number," J. Opt. Soc. Am. A **29**, 2415–2420 (2012).
29. A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, 2nd ed. (Prentice Hall, 1997), pp. 703–706.