Synthesis of three-dimensional light fields with binary spatial light modulators

Erdem Ulusoy,* Levent Onural, and Haldun M. Ozaktas

Department of Electrical and Electronics Engineering, Bilkent University, TR-06800 Bilkent, Ankara, Turkey
*Corresponding author: eulusoy@bilkent.edu.tr

Computation of a binary spatial light modulator (SLM) pattern that generates a desired light field is a challenging quantization problem for which several algorithms have been proposed, mainly for far-field or Fourier plane reconstructions. We study this problem assuming that the desired light field is synthesized within a volumetric region in the non-far-field range after free space propagation from the SLM plane. We use Fresnel and Rayleigh–Sommerfeld scalar diffraction theories for propagation of light. We show that, when the desired field is confined to a sufficiently narrow region of space, the ideal gray-level complex-valued SLM pattern generating it becomes sufficiently low pass (oversampled) so it can be successfully halftoned into a binary SLM pattern by solving two decoupled real-valued constrained halftoning problems. Our simulation results indicate that, when the synthesis region is considered, the binary SLM is indistinguishable from a lower resolution full complex gray-level SLM. In our approach, free space propagation related computations are done only once at the beginning, and the rest of the computation time is spent on carrying out standard image halftoning.   © 2011 Optical Society of America

OCIS codes: 070.0070, 090.0090, 070.6120, 090.1760, 090.2870, 100.2810.

1. INTRODUCTION

Computer-generated holograms (CGHs) have been studied since 1960s for applications such as beam shaping, optical data storage, optical information processing, optical metrology, nondestructive testing, optical interconnections, and three-dimensional (3D) holographic display [1–4]. In the early days, CGHs were physically implemented as one-time fabricated optical masks named diffractive optical elements (DOEs) [5,6]. Since the 1990s, with the advancement in spatial light modulator (SLM) technologies, CGHs are usually written on SLMs, which are basically dynamically programmable optical masks [7–11]. Although they are convenient to use, most SLMs do not provide full complex modulation, thus it is not possible to directly write CGHs with arbitrary complex values on them. For instance, phase-only SLMs can only provide phase modulation on the incoming light, so, on such SLMs, we can only write a phase-only CGH. In addition, there are quantization constraints: pixels of most SLMs can be set only to a finite number of different values. Such constraints impose a limit on the range of the light fields that can be synthesized with these devices. Given a desired field, determination of the best hologram pattern subject to the SLM constraints is a widely studied problem [12–17]. As the constraints get harsher, the problem becomes more interesting.

The most constrained SLMs are the binary ones. Pixels of binary SLMs can be set to only two possible distinct values, such as $(0, 1)$ or $(-1, 1)$. In this case, the quantization constraint on the SLM is quite harsh. Not surprisingly, when other parameters, such as number of pixels or pixel periods, are kept the same, the range of the light fields that can be synthesized is the most limited when a binary SLM is used. In addition, determination of a binary hologram that generates a desired light field is more difficult than determining a multilevel hologram. However, binary SLMs have some advantages over others that make them attractive to use. For instance, amplitude-only binary SLMs (such as the digital micromirror devices produced by Texas Instruments [18,19]) provide the same $(0, 1)$ modulation independent of the wavelength of the illumination wave. On the other hand, the pixel values of most multilevel SLMs change with the wavelength. This makes their use difficult in multicolor applications, such as 3D displays. Second, most multilevel phase-only (or amplitude-only) SLMs are imperfect in the sense that, in addition to the phase (amplitude) modulation that they provide, they perform an uncontrollable amplitude (phase) modulation. In this sense, binary SLMs are much more robust. As a third factor, miniaturization of binary SLMs, that is, manufacturing binary SLMs with small pixel pitches and high pixel counts, seems to have a higher potential compared to other types of SLMs.

The first significant progress in the research on binary CGH computation and implementation was achieved with the detour phase method [20–28]. This method was developed to obtain binary DOEs which generate desired complex-valued monochromatic light fields within a small region (centered around the optical axis) of the far field or on the Fourier plane of a $2f$ setup. These binary DOEs are actually opaque masks in which holes are cut. In particular, the DOE is broken down into a number of cells and, in each cell, a rectangular hole is placed. The position and the dimensions of this hole are adjusted such that when illuminated with an oblique wave, the cell behaves no different than a complex gray-valued pixel when the synthesis region is considered. Hence, the entire DOE behaves like a gray-level DOE. Later, the basic method was improved and modified to operate in the non-far-field range where Fresnel diffraction model is applicable [29,30]. Such methods are called cell-oriented methods.

With the advancement in the pixelated SLM technologies, the research on binary CGHs shifted toward pixel-oriented

methods since direct application of cell-oriented methods became difficult. In these methods, the SLM is taken as a collection of binary pixels and the goal is to determine the discrete binary CGH pattern to be written on the SLM. Similar to the cell-oriented case, the research initially focused on reconstructions at the far field or on the Fourier plane of a $2f$ setup. The reason is that, in the far-field or the $2f$ setup case, the relation between the SLM pixels and output field samples are simply given by a discrete Fourier transform (FT), which is easy to understand and manipulate. Many iterative and noniterative algorithms have been designed or adapted and applied to this problem [31–38]. Exploiting the intrinsic connection to the classical halftoning problem of image processing, researchers also adapted and applied halftoning algorithms such as error diffusion [39–43] and direct binary search [44–48]. In addition, projection onto convex sets (POCS) or Gercshberg–Saxton-like algorithms, such as the iterative FT algorithm, have been proposed [49–51]. Such algorithms have been extensively analyzed in terms of reconstruction error, diffraction efficiency, computational performance, etc. [52–56]. However, minor work has been done to develop algorithms for the non-far-field range where Rayleigh–Sommerfeld (RS) or Fresnel diffraction models are valid, perhaps due to the difficulty in the involved analytical relations [57,58].

In this paper, we develop the theory of 3D light field synthesis with a finite-size binary SLM. We assume that the binary SLM is illuminated by a plane wave and the desired field is synthesized within a volumetric region in the non-far-field range after free space propagation from the SLM plane. We use RS and Fresnel diffraction theories for free space propagation. In Section 2, we review the basics of diffraction. In Section 3, we analyze the light field generated by a finite-size SLM and discuss the constraints that the pixellated SLM structure impose on the output field. In Section 4, we show that an SLM pattern and its low-pass filtered version essentially produce the same light field within a certain region of space. We use this observation in Section 5 to find binary SLM patterns that generate desired light fields specified within an appropriately defined volumetric region. Using computer simulations, we show that binary SLM patterns computed with our approach successfully generate planar as well as volumetric (3D) light fields.

2. BASICS OF DIFFRACTION

As explained in [59], according to scalar wave theory of light, free space propagation of monochromatic light between two parallel planes having a distance z in between is a linear shift invariant system. If $u_0(x,y)$ and $u_z(x,y)$, respectively, denote the light fields at the input and output planes, we have

$$u_z(x,y) = u_0(x,y) **h_z(x,y)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_0(x',y')h_z(x-x',y-y')dx'dy', \quad (1)$$

where $**$ denotes two-dimensional (2D) analog convolution and $h_z(x,y)$ denotes the impulse response of free space propagation. According to the RS theory, considering only the propagating waves, $h_z(x,y)$ is given as

$$h_z(x,y) = \frac{z}{j\lambda}\frac{e^{jkR}}{R^2}, \quad (2)$$

where $R = \sqrt{x^2+y^2+z^2}$, $k = \frac{2\pi}{\lambda}$. The FT of Eq. (2), i.e., the frequency response of free space propagation is given as [60,61]

$$H_z(f_x,f_y)$$
$$= \mathcal{F}\{h_z(x,y)\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_z(x,y)\exp\{-j2\pi(xf_x+yf_y)\}dxdy$$
$$= \exp\left\{jkz\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right\}\mathrm{rect}\left(\frac{\sqrt{(\lambda f_x)^2+(\lambda f_y)^2}}{2}\right), \quad (3)$$

where $\mathrm{rect}(x) = 1$ for $|x| < 0.5$, $\mathrm{rect}(x) = 0.5$ for $|x| = 0.5$, and $\mathrm{rect}(x) = 0$ for $|x| > 0.5$. The $\mathrm{rect}(\cdot)$ function in Eq. (3) appears since we only consider propagating waves and ignore the evanescent waves.

Under the commonly used Fresnel diffraction theory, which is accurate for paraxial cases, the impulse response is approximated with a chirp (quadratic phase exponential) signal:

$$h_z(x,y) = \frac{e^{jkz}}{j\lambda z}e^{\frac{j\pi}{\lambda z}(x^2+y^2)}. \quad (4)$$

The corresponding frequency response becomes

$$H_z(f_x,f_y) = e^{jkz}\exp\{-j\pi\lambda z(f_x^2+f_y^2)\}, \quad (5)$$

which is also a chirp.

In Section 3, we will analyze the light field generated by an SLM. We will carry out that analysis assuming that the complex transmittance of the SLM is obtained by sampling and reinterpolating a light field that is bandlimited to a rectangular region in the frequency plane. Since the space of bandlimited signals with rectangular frequency support is spanned by $\mathrm{sinc}(\cdot)$ functions, it is important for us to understand the nature of the diffraction field produced by the input $u_0(x,y) = B_x B_y \mathrm{sinc}(xB_x)\mathrm{sinc}(yB_y)$, whose FT is given by $U_0(f_x,f_y) = \mathrm{rect}\left(\frac{f_x}{B_x}\right)\mathrm{rect}\left(\frac{f_y}{B_y}\right)$, where $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$. As in Eq. (1), the output is given by $u_z(x,y) = u_0(x,y) **h_z(x,y)$. Note that $u_z(x,y)$ can also be interpreted as the low-pass filtered version of $h_z(x,y)$, where $u_0(x,y)$ denotes the impulse response and B_x and B_y denote the bandwidths of the low-pass filter. In [59], exact expressions for $u_z(x,y)$ under Fresnel approximation in terms of Fresnel sine and cosine integrals are developed. However, here we will focus on a more useful approximate formula. Consider the $h_z(x,y)$ given in Eq. (4). The instantaneous spatial frequencies of $h_z(x,y)$ along the x and y directions are given, respectively, as $f_X(x,y) = \frac{x}{\lambda z}$ and $f_Y(x,y) = \frac{y}{\lambda z}$. When $|x| \leq \frac{\lambda z B_x}{2}$ and $|y| \leq \frac{\lambda z B_y}{2}$, we have $|f_X(x,y)| \leq \frac{B_x}{2}$ and $|f_Y(x,y)| \leq \frac{B_y}{2}$. On the other hand, when $|x| > \frac{\lambda z B_x}{2}$ or $|y| > \frac{\lambda z B_y}{2}$, we have $|f_X(x,y)| > \frac{B_x}{2}$ or $|f_Y(x,y)| > \frac{B_y}{2}$. Therefore, after convolution with $u_0(x,y)$, it is natural to expect the portion of $h_z(x,y)$ lying in the $|x| \leq \frac{\lambda z B_x}{2}$ and $|y| \leq \frac{\lambda z B_y}{2}$ region to be preserved, and the portion lying in the $|x| >$
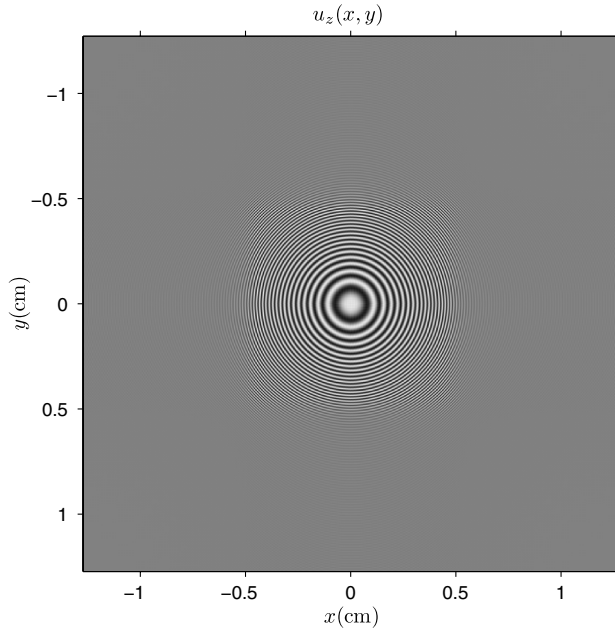
Fig. 1.   Real part of $u_z(x,y)$.

$\frac{\lambda z B_x}{2}$ or $|y| > \frac{\lambda z B_y}{2}$ region to be eliminated. This is indeed the case, as can be seen from Fig. 1, where $u_z(x,y)$ is displayed for $z = 1\,\mathrm{m}$, $\lambda = 632.9\,\mathrm{nm}$, and $B_x = B_y = \frac{0.01}{\lambda}$. As we can see, $u_z(x,y)$ is approximately obtained by windowing $h_z(x,y)$ with a rect$(\cdot)$ function whose widths are equal to $\lambda z B_x = \lambda z B_y = 1\,\mathrm{cm}$. Therefore, we can write

$$h_z(x,y) * * B_x B_y \mathrm{sinc}(xB_x)\mathrm{sinc}(yB_y)$$
$$\approx h_z(x,y)\mathrm{rect}\left(\frac{x}{\lambda z B_x}\right)\mathrm{rect}\left(\frac{y}{\lambda z B_y}\right). \qquad (6)$$

Actually, careful examination of Fig. 1 reveals that the window is not a perfect rect$(\cdot)$ function. However, when B_x, B_y, and z are kept within suitable ranges, the approximation in Eq. (6) works fine. For instance, the approximation holds with a mean squared error that is less than 5% when $\frac{1}{B_x}$ and $\frac{1}{B_y}$ are between 10λ and 100λ, and z is greater than about $7.5 \times 10^5 \lambda$. These ranges for B_x, B_y, and z are of interest to us in this paper, and the indicated approximation error is acceptable for our purposes. Hence, we assume that the approximation is successful and we will use it frequently from now on. We carried the discussion for the impulse response of Fresnel approximation, but a similar windowing effect (with a slight change in the window dimensions) is observed for the exact $h_z(x,y)$ given in Eq. (2) as long as B_x, B_y, and z stay within the ranges mentioned above.

3. ANALYSIS OF LIGHT FIELD GENERATED BY A SPATIAL LIGHT MODULATOR

SLMs can be viewed as programmable 2D optical masks. Most of the SLMs today have a pixelated structure and in this paper we will be interested only in such SLMs. Let Δ_x and Δ_y denote the pixel periods of a pixelated SLM. Typical values for Δ_x and Δ_y are $8\,\mu\mathrm{m}$, $10\,\mu\mathrm{m}$, etc. Let $a(x,y)$ denote the pixel aperture function. For practical cases, $a(x,y) = 0$ for $|x| > \frac{\Delta_x}{2}$ or $|y| > \frac{\Delta_y}{2}$. Mostly, $a(x,y) = \mathrm{rect}(\frac{x}{W_x})\mathrm{rect}(\frac{y}{W_y})$, where $W_x \leq \Delta_x$,

$W_y \leq \Delta_y$. Let $\bar{s}[m,n]$ (m, $n \in \mathcal{Z}$) denote the complex value of the (m,n)th SLM pixel. When viewed as a discrete function of m and n, $\bar{s}[m,n]$ denotes the 2D complex-valued pattern that we write on the SLM. We will call $\bar{s}[m,n]$ the SLM pattern from now on. We will assume that the SLM has $(2M+1) \times (2N+1)$ pixels such that $\bar{s}[m,n] = 0$ for $|m| > M$ and $|n| > N$. For practical SLMs, M and N are around 500–1000, so the physical dimensions of the SLM are around 1–2 cm by 1–2 cm. If we denote the complex transmittance of the SLM with $s^a(x,y)$, we have

$$s^a(x,y) = \sum_{m=-M}^{M}\sum_{n=-N}^{N} \bar{s}[m,n]a(x - m\Delta_x, y - n\Delta_y). \qquad (7)$$

Assuming that the SLM is placed at the $z = 0$ plane and illuminated by a normally incident plane wave, we can write the light field produced by it at a distance z as $u_z^a(x,y) = s^a(x,y) * * h_z(x,y)$. We wish to understand the nature of $u_z^a(x,y)$.

Pixelated SLMs are inherently associated with sampling and reinterpolation of light fields. Consider a light field $s(x,y)$, which is defined as

$$s(x,y) = \sum_{m=-M}^{M}\sum_{n=-N}^{N} \bar{s}[m,n]\mathrm{sinc}\left(\frac{x - m\Delta_x}{\Delta_x}\right)\mathrm{sinc}\left(\frac{y - n\Delta_y}{\Delta_y}\right). \qquad (8)$$

Note that $s(x,y)$ is bandlimited to the $|f_x| \leq \frac{1}{2\Delta_x}$ and $|f_y| \leq \frac{1}{2\Delta_y}$ band, so $s(x,y)$ can be sampled with periods Δ_x, Δ_y without any aliasing. Actually, when we sample $s(x,y)$ with Δ_x, Δ_y, we obtain the discrete function $\bar{s}[m,n]$, i.e., $s(m\Delta_x, n\Delta_y) = \bar{s}[m,n]$. Hence, we can consider the SLM pattern $\bar{s}[m,n]$ as being obtained by sampling $s(x,y)$ with Δ_x, Δ_y without any aliasing. Then, we can obtain $s^a(x,y)$ from $\bar{s}[m,n]$ by applying a discrete-to-analog converter whose interpolating function is $a(x,y)$. Therefore, $s^a(x,y)$ is obtained by sampling and reinterpolating $s(x,y)$. In mathematical terms,

$$s^a(x,y) = a(x,y) * * \left\{ s(x,y) \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \delta(x - m\Delta_x, y - n\Delta_y) \right\}$$
$$= \frac{1}{\Delta_x \Delta_y} a(x,y) * * s(x,y)$$
$$+ \frac{1}{\Delta_x \Delta_y} \sum_{\substack{p=-\infty \\ (p,q) \neq (0,0)}}^{\infty}\sum_{q=-\infty}^{\infty} a(x,y)$$
$$* * \left[s(x,y)\exp\left\{ j2\pi\left(\frac{px}{\Delta_x} + \frac{qy}{\Delta_y}\right) \right\} \right], \qquad (9)$$

where we wrote the second expression using the well-known identity $\sum_m \delta(x - m\Delta) = \frac{1}{\Delta}\sum_p \exp\{j2\pi(\frac{px}{\Delta})\}$. Note that, in the second expression, the terms at the bottom line would disappear if we had $a(x,y) = \mathrm{sinc}(\frac{x}{\Delta_x})\mathrm{sinc}(\frac{y}{\Delta_y})$, so we would have $s^a(x,y) = s(x,y)$. However, this is not the case for the interpolating function $a(x,y)$ of a practical SLM. Thus, the terms at the bottom line remain and, at the output, they give rise to the well-known diffraction orders.

To understand the nature of the SLM output $u_z^a(x,y)$, let us first assume that $a(x,y) = \delta(x,y)$. Let $u_z(x,y)$ denote the light

field produced by $s(x, y)$ so that $u_z(x, y) = s(x, y) * * h_z(x, y)$. Examining the second expression in Eq. (9), we see that, at the output plane, the term at the top line produces $\frac{1}{\Delta_x \Delta_y} u_z(x, y)$, which is commonly called the central diffraction order. The terms at the bottom line produce the so-called higher diffraction orders. In [62] it is shown that higher diffraction orders are essentially translated, modulated, and dispersed versions of the central diffraction order when the RS diffraction model is used. In [63], it is shown that, under the Fresnel approximation, the (p, q)th diffraction order is given by

$$\frac{1}{\Delta_x \Delta_y} e^{-j\pi\lambda z \left(\frac{p^2}{\Delta_x^2} + \frac{q^2}{\Delta_y^2} \right)} u_z \left(x - p\frac{\lambda z}{\Delta_x}, y - q\frac{\lambda z}{\Delta_y} \right)$$
$$\times \exp\left\{ j2\pi \left(\frac{px}{\Delta_x} + \frac{qy}{\Delta_y} \right) \right\}, \quad (10)$$

which is essentially a shifted and modulated version of the central diffraction order. Hence, the SLM output $u_z^a(x, y)$ consists of diffraction orders that are all related to $u_z(x, y)$. We can get more insight about $u_z^a(x, y)$ if we examine $u_z(x, y)$. Applying the approximation in Eq. (6) to Eq. (8), we see that $u_z(x, y)$ is approximately confined in space to the region given as $|x| < \frac{\lambda z}{2\Delta_x} + M\Delta_x$ and $|y| < \frac{\lambda z}{2\Delta_y} + N\Delta_y$. When practical values are considered for λ, Δ_x, Δ_y, M, and N, for distances greater than about $80\,\mathrm{cm}$, this region can be approximated as

$$|x| < \frac{\lambda z}{2\Delta_x}, \qquad |y| < \frac{\lambda z}{2\Delta_y}. \quad (11)$$

We will call this region the central diffraction order region. When z is viewed as a varying parameter, Eq. (11) defines a 3D pyramid (whose tip is at the origin and whose base expands in the $+z$ direction), which we will name the central diffraction order pyramid. By Eqs. (6), (8), and (11), we can write

$$u_z(x, y) \approx \Delta_x \Delta_y \mathrm{rect}\left(\frac{x\Delta_x}{\lambda z} \right) \mathrm{rect}\left(\frac{y\Delta_y}{\lambda z} \right)$$
$$\times \sum_{m=-M}^{M} \sum_{n=-N}^{N} \bar{s}[m, n] h_z(x - m\Delta_x, y - n\Delta_y). \quad (12)$$

Next, by Eq. (10), we see that the (p, q)th diffraction order of the SLM is approximately centered around $(\frac{p\lambda z}{\Delta_x}, \frac{q\lambda z}{\Delta_y})$, and has dimensions of $\frac{\lambda z}{\Delta_x}$ and $\frac{\lambda z}{\Delta_y}$. Therefore, for sufficiently large distances, diffraction orders of the SLM do not overlap in space (approximately), meaning that higher diffraction orders make no contribution to the central diffraction order region. Hence, we have

$$u_z^a(x, y) \approx \frac{1}{\Delta_x \Delta_y} u_z(x, y) \quad \text{for } |x| < \frac{\lambda z}{2\Delta_x}, \quad |y| < \frac{\lambda z}{2\Delta_y}. \quad (13)$$

Since higher diffraction orders do not contain any new information, it suffices to examine the SLM output only in the central diffraction order region.

Assuming that the Fresnel approximation is valid within the central diffraction order region, using Eqs. (4), (12), and (13),

we can write

$$u_z^a(x, y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2 + y^2)} \sum_{m=-M}^{M} \sum_{n=-N}^{N} \bar{s}[m, n] e^{\frac{j\pi}{\lambda z}(m^2\Delta_x^2 + n^2\Delta_y^2)}$$
$$\times e^{-j\frac{2\pi}{\lambda z}(xm\Delta_x + yn\Delta_y)} \quad (14)$$

for $|x| < \frac{\lambda z}{2\Delta_x}$ and $|y| < \frac{\lambda z}{2\Delta_y}$. Since $\bar{s}[m, n]$ has $(2M + 1) \times (2N + 1)$ degrees of freedom, $u_z^a(x, y)$ also has $(2M + 1) \times (2N + 1)$ degrees of freedom. In fact, it can be shown that $u_z^a(x, y)$ of Eq. (14) can fully be represented by its $(2M + 1) \times (2N + 1)$ samples taken uniformly within the central diffraction order region. Letting $M_s = 2M + 1$ and $N_s = 2N + 1$ for convenience, these samples can be computed as

$$\bar{u}_z^a[m, n] = u_z^a\left(\frac{m\lambda z}{M_s\Delta_x}, \frac{n\lambda z}{N_s\Delta_y} \right) = \frac{e^{jkz}}{j\lambda z} e^{j\pi\lambda z \left(\frac{m^2}{M_s^2\Delta_x^2} + \frac{n^2}{N_s^2\Delta_y^2} \right)}$$
$$\times \sum_{m'=-M}^{M} \sum_{n'=-N}^{N} \bar{s}[m', n'] e^{\frac{j\pi}{\lambda z}(m'^2\Delta_x^2 + n'^2\Delta_y^2)} e^{-j2\pi \left(\frac{mm'}{M_s} + \frac{nn'}{N_s} \right)} \quad (15)$$

for $m, n \in \mathcal{Z}$ and $|m| \le M$, $|n| \le N$. As seen, computation of the samples of SLM output within the central diffraction order region involves multiplying the SLM pattern $\bar{s}[m, n]$ with a discrete chirp, taking a centered 2D discrete FT, and then multiplying with another discrete chirp [64]. Interpolation of $u_z^a(x, y)$ from $\bar{u}_z^a[m, n]$ is discussed in [65] and is slightly different than the classical sinc interpolation. In light field synthesis problems, desired fields are usually specified through $\bar{u}_z^a[m, n]$. For a given $\bar{u}_z^a[m, n]$, the required SLM pattern $\bar{s}[m, n]$ can be computed as

$$\bar{s}[m, n] = \frac{j\lambda z e^{-jkz}}{M_s N_s} e^{-\frac{j\pi}{\lambda z}(m^2\Delta_x^2 + n^2\Delta_y^2)}$$
$$\times \sum_{m'=-M}^{M} \sum_{n'=-N}^{N} \bar{u}_z^a[m', n'] e^{-j\pi\lambda z \left(\frac{m'^2}{M_s^2\Delta_x^2} + \frac{n'^2}{N_s^2\Delta_y^2} \right)} e^{j2\pi \left(\frac{mm'}{M_s} + \frac{nn'}{N_s} \right)} \quad (16)$$

for $|m| \le M$ and $|n| \le N$. Indeed, Eq. (16) is just the inverse of Eq. (15).

Finally, up to now, we assumed that $a(x, y) = \delta(x, y)$. In practice, $a(x, y)$ extends over a nonzero area but is confined to the $|x| \le \frac{\Delta_x}{2}$ and $|y| \le \frac{\Delta_y}{2}$ region. Mostly, $a(x, y) = \mathrm{rect}(\frac{x}{W_x})\mathrm{rect}(\frac{y}{W_y})$ with $W_x \le \Delta_x$ and $W_y \le \Delta_y$. In such a case, convolution with $a(x, y)$ must be incorporated into Eqs. (10) and (12)–(14). The result will be a blurring in the SLM output relative to the $a(x, y) = \delta(x, y)$ case. Higher diffraction orders suffer more from this blurring, so they are attenuated relative to the central diffraction order. But the locations and spatial supports of diffraction orders will not change, since $a(x, y)$ is narrow. There will also be a slow amplitude variation within the central diffraction order region, but it is negligible since $a(x, y)$ is narrow. Therefore, we can assume that Eqs. (11)–(16) are valid in the case of a practical $a(x, y)$ within constant multiplicative factors.

4. EFFECTS OF APPLYING A DISCRETE REAL-VALUED LOW-PASS FILTER TO THE SLM PATTERN

In this section, we apply a discrete real-valued low-pass filter with an impulse response $\bar{g}[m,n]$ to the SLM pattern $\bar{s}[m,n]$ and write the resulting discrete signal on the SLM instead of $\bar{s}[m,n]$. Let $\bar{s}_L[m,n]$ denote the discrete output after filtering such that

$$\bar{s}_L[m,n] = \bar{g}[m,n] \star\star \bar{s}[m,n]$$
$$= \sum_{m'=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} \bar{g}[m',n'] \bar{s}[m-m', n-n']. \quad (17)$$

Here, $\star\star$ denotes 2D discrete convolution. Let $u_{z_L}^a(x,y)$ denote the new SLM output. We wish to understand the relation between $u_z^a(x,y)$ and $u_{z_L}^a(x,y)$, where $u_z^a(x,y)$ denotes the output produced when we write $\bar{s}[m,n]$ on the SLM, as in Section 3.

As we did for $\bar{s}[m,n]$ in Section 3, we can think of $\bar{g}[m,n]$ as being obtained by sampling a continuous bandlimited signal $g(x,y)$ with the pixel periods Δ_x and Δ_y of the SLM. Let $g(x,y)$ and $G(f_x,f_y)$, respectively, denote the impulse and frequency responses of the ideal analog low-pass filter with bandwidths B_x and B_y such that

$$g(x,y) = B_x B_y \mathrm{sinc}(xB_x)\mathrm{sinc}(yB_y), \quad (18)$$

$$G(f_x,f_y) = \mathrm{rect}\left(\frac{f_x}{B_x}\right)\mathrm{rect}\left(\frac{f_y}{B_y}\right). \quad (19)$$

Then, we can assume that $\bar{g}[m,n]$ is obtained as $\bar{g}[m,n] = g(m\Delta_x, n\Delta_y)$. Note that the bandwidths B_x and B_y should satisfy

$$B_x < \frac{1}{\Delta_x}, \qquad B_y < \frac{1}{\Delta_y}, \quad (20)$$

so that aliasing is avoided. Suppose this is the case.

Let $s_L(x,y) = s(x,y) \ast\ast g(x,y)$, where $s(x,y)$ is as given in Eq. (8). Using the theory of discrete processing of band-limited continuous signals [66], it is easy to see that $\bar{s}_L[m,n] = s_L(m\Delta_x, n\Delta_y)$. Hence, the new SLM pattern $\bar{s}_L[m,n]$ is obtained by sampling $s_L(x,y)$. Then, by the analysis in Section 3, we know that, within the central diffraction order region given by Eq. (11), we have $u_{z_L}^a(x,y) \approx \frac{1}{\Delta_x \Delta_y} u_{z_L}(x,y)$, where $u_{z_L}(x,y)$ denotes the diffraction field produced by $s_L(x,y)$ such that $u_{z_L}(x,y) = s_L(x,y) \ast\ast h_z(x,y)$. We can understand the relation between $u_z^a(x,y)$ and $u_{z_L}^a(x,y)$ if we examine $u_{z_L}(x,y)$. Using Eq. (8), and noting that $g(x,y) \ast\ast \mathrm{sinc}(\frac{x}{\Delta_x})\mathrm{sinc}(\frac{y}{\Delta_y}) = \Delta_x \Delta_y g(x,y)$ because of Eq. (20), we can write

$$s_L(x,y) = \Delta_x \Delta_y \sum_{m=-M}^{M} \sum_{n=-N}^{N} \bar{s}[m,n] g(x - m\Delta_x, y - n\Delta_y). \quad (21)$$

Next, applying Eq. (6) to Eq. (21), we can see that $u_{z_L}(x,y)$ is approximately confined in space to the region given as $|x| < \frac{\lambda z B_x}{2} + M\Delta_x$ and $|y| < \frac{\lambda z B_y}{2} + N\Delta_y$. When practical values are considered for λ, Δ_x, Δ_y, M, and N, for distances greater than

about 80 cm, this region can be approximated as

$$|x| < \frac{\lambda z B_x}{2}, \qquad |y| < \frac{\lambda z B_y}{2}. \quad (22)$$

Note that, because of Eq. (20), the region given above is a sub-region of the central diffraction order region that is given in Eq. (11). Within this region, by Eqs. (6) and (21), we can write

$$u_{z_L}(x,y) \approx \Delta_x \Delta_y \sum_{m=-M}^{M} \sum_{n=-N}^{N} \bar{s}[m,n] h_z(x - m\Delta_x, y - n\Delta_y)$$
$$\approx u_z(x,y), \quad (23)$$

where the second line follows from Eq. (12). Therefore, by Eq. (13), within central diffraction order region we get

$$u_{z_L}^a(x,y) \approx u_z^a(x,y)\mathrm{rect}\left(\frac{x}{\lambda z B_x}\right)\mathrm{rect}\left(\frac{y}{\lambda z B_y}\right). \quad (24)$$

Hence, within the region specified by Eq. (22), $\bar{s}[m,n]$ and $\bar{s}_L[m,n]$ approximately produce the same field. Moreover, $\bar{s}_L[m,n]$ approximately produces nothing in the rest of the central order diffraction region. As an illustration of this effect, consider the SLM pattern $\bar{s}[m,n]$ shown in Fig. 2, which is computed according to Eq. (16) in order to synthesize the light field displayed in Fig. 3 within the central diffraction order region. Here, the SLM size is 1024×1024, $\Delta_x = \Delta_y = 8\,\mu\mathrm{m}$, $\lambda = 632.9\,\mathrm{nm}$, and $z = 1\,\mathrm{m}$, so the physical size of the SLM is $8.2\,\mathrm{mm} \times 8.2\,\mathrm{mm}$ and the physical size of the central diffraction order is $7.91\,\mathrm{cm} \times 7.91\,\mathrm{cm}$. As seen, the light field in Fig. 3 consists of an image in the middle surrounded by text. If only the image were present, $\bar{s}[m,n]$ would be a low-pass pattern, because only low-angle rays from the SLM would be sufficient to produce the image. However, the presence of the text, which requires high-angle rays from the SLM, causes $\bar{s}[m,n]$ to be a full-band discrete signal. Next, consider Fig. 4,
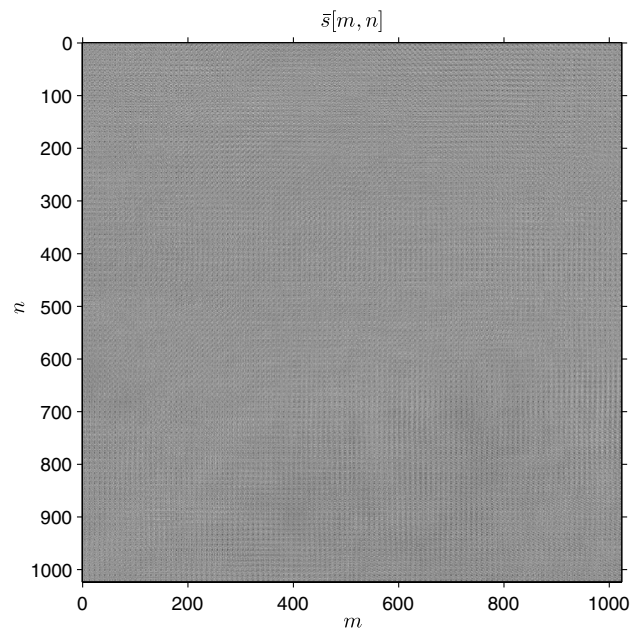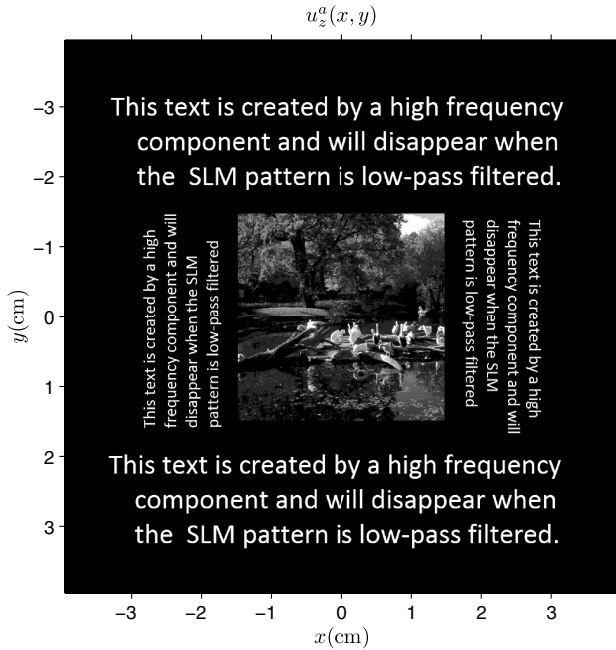

Fig. 2.   Actual SLM pattern $\bar{s}[m,n]$.

Fig. 3.   Output produced when $\bar{s}[m,n]$ is written on the SLM (only the central diffraction order is displayed).

Fig. 5.   Output produced when $\bar{s}_L[m,n]$ is written on the SLM. Also shown are the borders of the synthesis region given in Eq. (22).

which shows the SLM pattern $\bar{s}_L[m,n]$ obtained with filtering the SLM pattern in Fig. 2 with $\bar{g}[m,n]$ (B_x and B_y are taken such that $B_x \Delta_x = B_y \Delta_y = 0.375$). The resulting output is shown in Fig. 5. Also shown in Fig. 5 are the borders of the region specified in Eq. (22). As seen, the image is preserved, while the text is eliminated. Hence, as predicted, output approximately remains unchanged within the region specified in Eq. (22) and approximately vanishes outside this region.

In Section 5, we will exploit the result stated above to find binary SLM patterns that generate desired light fields confined to the region specified in Eq. (22) for appropriate selections of the parameters B_x and B_y. For a single z, Eq. (22) defines a rectangle, which we will call the synthesis region. Viewed as a
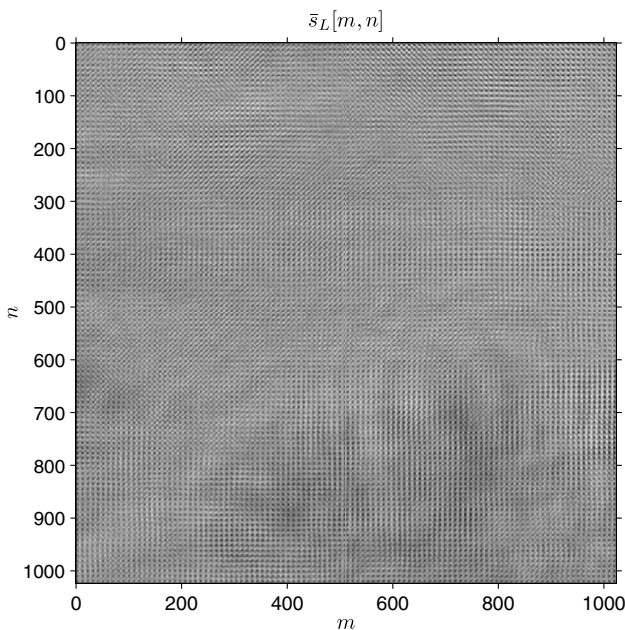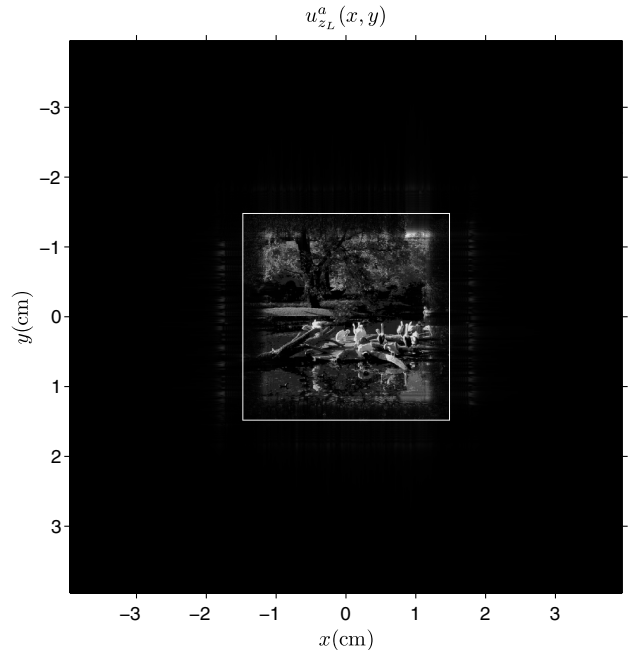
volumetric region, it defines a pyramid lying in the central diffraction order pyramid and we will call it the synthesis pyramid. The dimensions of the synthesis region increase as B_x and B_y increase. Therefore, as the low-pass effect of the filter gets stronger, the region that the stated equality holds gets narrower. On the other hand, as bandwidths approach the upper limits allowed by Eq. (20) ($B_x \rightarrow \frac{1}{\Delta_x}$ and $B_y \rightarrow \frac{1}{\Delta_y}$), the borders approach the borders of the central diffraction order given in Eq. (11). When light fields are to be synthesized with binary SLMs, a rational choice is to take $B_x \approx \frac{1}{4\Delta_x}$ and $B_y \approx \frac{1}{4\Delta_y}$ so that the area of the synthesis region is about $1/16$ of the area of the central diffraction order region.

5. ENCODING COMPLEX-VALUED OVERSAMPLED HOLOGRAMS ON BINARY SLMs

In this section, we turn to the main problem of this paper: how can we find binary SLM patterns that reconstruct desired 3D light fields? Findings of Section 4 will guide us.

First, consider a gray-level SLM pattern $\bar{s}[m,n]$ of size $(2M+1) \times (2N+1)$, that is, $\bar{s}[m,n]$ can be equal to any complex value for $|m| \leq M$ and $|n| \leq N$, but $\bar{s}[m,n] = 0$ for other (m,n). Suppose we write this pattern on an SLM that we illuminate with a normally incident plane wave. Let $\bar{u}_z^a[m,n]$ for $|m| \leq M$ and $|n| \leq N$ denote the samples of the output field taken uniformly within the central diffraction order region, as in Eq. (15). Suppose we wish to determine $\bar{s}[m,n]$ such that the output samples within the synthesis region specified by Eq. (22) are equal to some desired discrete signal $\bar{d}[m,n]$. In this section, it is important for the synthesis region to be sufficiently small, and the reason will become evident as we continue. For now, let us simply assume that the B_x and B_y parameters in Eq. (22) are chosen as $B_x = \frac{1}{4\Delta_x}$ and $B_y = \frac{1}{4\Delta_y}$ so that the area of the synthesis region is about $1/16$ of the area of the central diffraction order region. Hence,

Fig. 4.   Low-pass filtered SLM pattern $\bar{s}_L[m,n]$.

we wish to control about 1/16 of the output field samples. Suppose we are not concerned about the values of output samples that fall outside the synthesis region, i.e., they are "do not care" samples. Let us call the part of central diffraction order outside the synthesis region the "do not care" region. Since the number of samples that we wish to control is less than the degrees of freedom that we have in $\bar{s}[m,n]$, this problem does not have a unique solution but has many solutions. An easy solution can be found by setting

$$\bar{u}_z^a[m,n] = \begin{cases} \bar{d}[m,n] & \text{for samples within the synthesis region} \\ 0 & \text{for samples within the do not care region} \end{cases},$$
(25)

and then finding $\bar{s}[m,n]$ according to Eq. (16). Let us denote this particular solution with $\bar{s}^i[m,n]$ and call it the ideal SLM pattern since it produces $\bar{d}[m,n]$ within the synthesis region with maximum efficiency, i.e., output samples within the do not care region are zero, so essentially no power is spent on the do not care region.

The ideal SLM pattern $\bar{s}^i[m,n]$ found as above is a low-pass (oversampled) discrete signal. To see this, suppose we apply a real-valued discrete low-pass filter $\bar{g}[m,n]$ with bandwidth parameters $B_x = \frac{1}{4\Delta_x}$ and $B_y = \frac{1}{4\Delta_y}$ to $\bar{s}^i[m,n]$. As shown in Section 4, at the output, nothing should change in the synthesis region, and the field should vanish in the do not care region. But the field produced by $\bar{s}^i[m,n]$ already vanishes in the do not care region. Therefore, we should have $\bar{s}^i[m,n] \star\star \bar{g}[m,n] \approx \bar{s}^i[m,n]$, indicating that $\bar{s}^i[m,n]$ is the output of a low-pass filter, hence it is a low-pass (oversampled) SLM pattern.

Now let us try to reconstruct $\bar{d}[m,n]$ within the same synthesis region with a binary SLM pattern $\bar{s}_b[m,n]$, which again has a size of $(2M+1) \times (2N+1)$. Let us assume that $\bar{s}_b[m,n] = \pm 1$ for $|m| \le M$ and $|n| \le N$, but $\bar{s}_b[m,n] = 0$ for other (m,n). We now have a harsh constraint on the SLM pattern. Solving this problem is not as straightforward as for the gray-level $\bar{s}[m,n]$. For instance, it is common experience that direct quantization of $\bar{s}^i[m,n]$ does not produce satisfactory results. Smarter strategies are necessary.

From Section 4, we know that $\bar{s}_b[m,n]$ and its low-pass filtered version $\bar{s}_b^L[m,n] = \bar{s}_b[m,n] \star\star \bar{g}[m,n]$ should produce approximately the same field within the synthesis region. This means we should determine $\bar{s}_b[m,n]$ such that $\bar{s}_b^L[m,n]$ produces $\bar{d}[m,n]$ within the synthesis region. In particular, suppose we find $\bar{s}_b[m,n]$ such that $\bar{s}_b^L[m,n]$ is equal to $\bar{s}^i[m,n]$. (This is possible since $\bar{s}^i[m,n]$ is already low-pass as we discussed above.) Since $\bar{s}^i[m,n]$ already generates the desired field within the synthesis region, we see that $\bar{s}_b[m,n]$ performs the desired synthesis. Therefore, if we can find $\bar{s}_b[m,n]$ such that

$$\bar{s}_b[m,n] \star\star \bar{g}[m,n] \approx \bar{s}^i[m,n],$$
(26)

we can achieve the desired synthesis.

Above, we used $\approx$ instead of $=$ for two reasons. The first reason is to stress that, in general, the problem may not have an exact solution, i.e., there may be no $\bar{s}_b[m,n]$ that exactly gives $\bar{s}^i[m,n]$ when low-pass filtered, so we may need to seek for the best solution instead of an exact solution. The second reason is, according to our definitions, $\bar{s}^i[m,n]$ and $\bar{s}_b[m,n]$ are finite-sized patterns of size $(2M+1) \times (2N+1)$, but

strictly speaking $\bar{s}_b[m,n] \star\star \bar{g}[m,n]$ is not because of the infinite tails of low-pass filter $\bar{g}[m,n]$. However, we will not bother ourselves with this technical detail. We will simply assume that it is sufficient for Eq. (26) to hold only over the support of $\bar{s}^i[m,n]$ and $\bar{s}_b[m,n]$, i.e., for $|m| \le M$ and $|n| \le N$.

From the above discussion, we can realize that actually the low-pass component of $\bar{s}_b[m,n]$ (that is, $\bar{s}_b^L[m,n]$) is responsible for generating $\bar{d}[m,n]$ within the synthesis region. The high-pass component $(\bar{s}_b[m,n] - \bar{s}_b^L[m,n])$ only effects the output samples in the do not care region. Note that, when $\bar{g}[m,n]$ is applied to $\bar{s}_b[m,n]$, the high-pass component is eliminated, so nothing is generated within the do not care region. Actually, we can think that high-pass component is added to the low-pass component just to satisfy the binary SLM pattern constraint.

We recognize the problem stated in Eq. (26) as the well-known halftoning problem of image processing in which one tries to compute a binary image that produces a desired low-pass (oversampled) gray-level image when low-pass filtered. Notice that, in Eq. (26), both $\bar{s}_b[m,n]$ and $\bar{g}[m,n]$ are real valued, hence their convolution is also real valued. If $\bar{s}^i[m,n]$ is also real valued, we can easily find $\bar{s}_b[m,n]$ using any of the well-established halftoning algorithms, such as ordered dither, error diffusion, or direct binary search [67]. The problem is that, in general, $\bar{s}^i[m,n]$ is complex valued. Hence, we need to consider a modification to the halftoning procedure.

As a solution, we can partition the pixels of the SLM among two groups, such that the first group of pixels is responsible for halftoning the real part and the second group of pixels is responsible for halftoning the imaginary part.

Suppose we place an optical mask with complex transmittance $T(x,y)$ in front of the SLM such that

$$T(x,y) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} \bar{T}[m,n] a(x - m\Delta_x, y - n\Delta_y),$$

so that placing the mask is equivalent to writing the SLM pattern $\bar{s}_T[m,n] = \bar{s}_b[m,n]\bar{T}[m,n]$, which we call the effective SLM pattern. Suppose $\bar{T}[m,n]$ is given as

$$\bar{T}[m,n] = \begin{cases} 1 & \text{when } m+n \text{ is even} \\ j & \text{when } m+n \text{ is odd} \end{cases}.$$
(27)

The reader can understand that we used this particular pattern to obtain imaginary binary values as well as real binary values. (Later we will discuss other physical alternatives to using the mask.) We now have the following constraint on $\bar{s}_T[m,n]$:

$$\bar{s}_T[m,n] = \begin{cases} \pm 1 & \text{when } m+n \text{ is even} \\ \pm j & \text{when } m+n \text{ is odd} \end{cases}.$$
(28)

Note that we can write $\bar{s}_T[m,n] = \bar{s}_T^R[m,n] + j\bar{s}_T^I[m,n]$ where

$$\bar{s}_T^R[m,n] = \begin{cases} \pm 1 & \text{when } m+n \text{ is even} \\ 0 & \text{when } m+n \text{ is odd} \end{cases},$$

$$\bar{s}_T^I[m,n] = \begin{cases} 0 & \text{when } m+n \text{ is even} \\ \pm 1 & \text{when } m+n \text{ is odd} \end{cases}.$$
(29)

$\bar{s}_T^R[m,n]$ and $\bar{s}_T^I[m,n]$ can be considered as binary SLM patterns, some of whose pixels are eliminated. In particular, odd pixels (pixels for which $m+n$ is odd) of $\bar{s}_T^R[m,n]$ and even pixels (pixels for which $m+n$ is even) of $\bar{s}_T^I[m,n]$ are eliminated. Now, we can encode the real part of $\bar{s}^i[m,n]$ on $\bar{s}_T^R[m,n]$ and the imaginary part on $\bar{s}_T^I[m,n]$. In other words, we can try to find $\bar{s}_T^R[m,n]$ and $\bar{s}_T^I[m,n]$ such that

$$
\begin{aligned}
\bar{s}_T^R[m,n] \star\star \bar{g}[m,n] &\approx \mathcal{R}\{\bar{s}^i[m,n]\}, \\
\bar{s}_T^I[m,n] \star\star \bar{g}[m,n] &\approx \mathcal{I}\{\bar{s}^i[m,n]\},
\end{aligned}
\tag{30}
$$

where $\bar{s}_T^R[m,n]$ and $\bar{s}_T^I[m,n]$ are subject to the constraints of Eq. (29). Once we find $\bar{s}_T^R[m,n]$ and $\bar{s}_T^I[m,n]$, we can add them up to obtain a complete binary SLM pattern such that $\bar{s}_b[m,n] = \bar{s}_T^R[m,n] + \bar{s}_T^I[m,n]$. This binary pattern performs the desired synthesis when written on an SLM in front of which the mask $T(x,y)$ is placed. Note that, in this manner, we have converted the complex-valued halftoning problem of Eq. (26) (which was problematic in that form) into two decoupled real-valued constrained halftoning problems, as in Eq. (30) (which can now easily be solved with standard halftoning algorithms).

The constraints on $\bar{s}_T^R[m,n]$ and $\bar{s}_T^I[m,n]$ given in Eq. (29) will not cause significant halftoning error if $\bar{s}^i[m,n]$ (hence its real and imaginary components) is sufficiently low pass. This is the case if the synthesis region is selected sufficiently small.

Let us illustrate these ideas. Suppose Fig. 6 shows the desired field within the central diffraction order region. Only the samples within the synthesis region are nonzero. Figure 7 shows the real part of the ideal gray-level SLM pattern $\bar{s}^i[m,n]$ that exactly reconstructs the desired field of Fig. 6. The imaginary part is not displayed, but it has similar characteristics with the real part. As discussed above, the ideal gray-level SLM pattern is a low-pass (oversampled) discrete signal. In this example, the SLM size is 1024×1024, Δ_x, $\Delta_y = 8\,\mu\text{m}$, $\lambda = 632.9\,\text{nm}$, and $z = 1\,\text{m}$, so the physical size of the SLM is

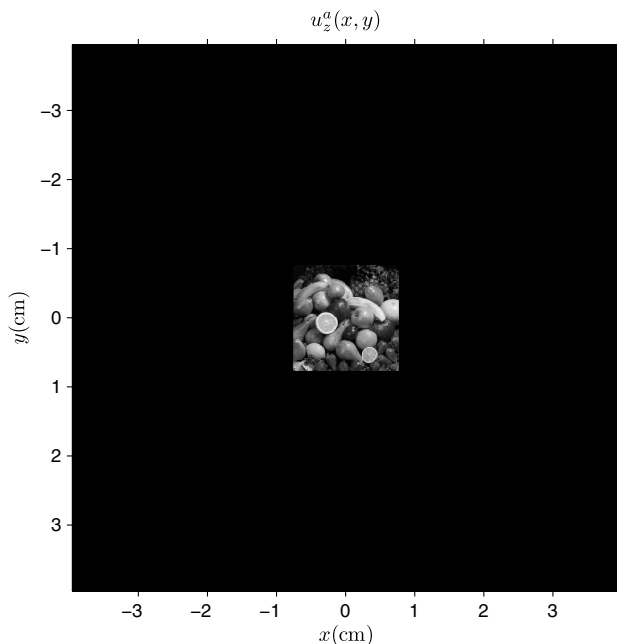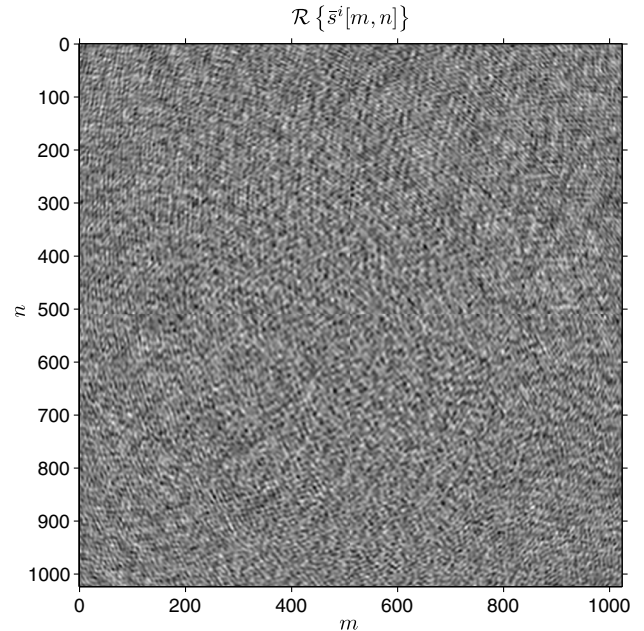Fig. 7.   Real part of ideal gray-level SLM pattern.

$8.2\,\text{mm} \times 8.2\,\text{mm}$ and the physical size of the central diffraction order region is $7.91\,\text{cm} \times 7.91\,\text{cm}$. The synthesis region consists of 200×200 samples, which corresponds to a physical size of $1.55\,\text{cm} \times 1.55\,\text{cm}$.

Next, we considered the computation of a binary pattern that generates the desired field. As the halftoning algorithm, we used the standard error diffusion algorithm [67]. As in Eqs. (29) and (30), we separately halftoned the real and imaginary parts of $\bar{s}^i[m,n]$ and computed $\bar{s}_T^R[m,n]$ and $\bar{s}_T^I[m,n]$. We display $\bar{s}_T^R[m,n]$ in Fig. 8 (black pixels have value –1, gray pixels have value 0, and white pixels have value 1). Note that, as imposed by Eq. (29), in this pattern, odd pixels are 0 while even pixels are either –1 or 1. $\bar{s}_T^I[m,n]$ is not displayed, but it
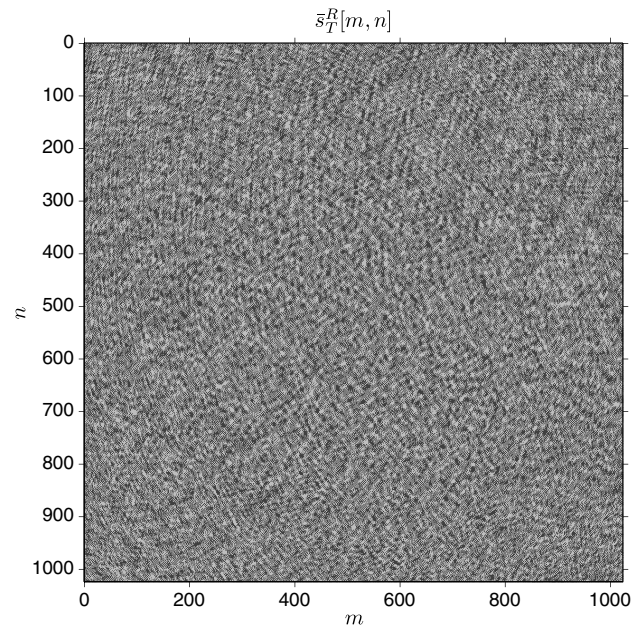

Fig. 6.   Desired field. Entire central diffraction order is displayed.

Fig. 8.   Three-level SLM pattern for real part obtained by solving the first constrained halftoning problem in Eq. (30). Even pixels are ± 1, odd pixels are 0.
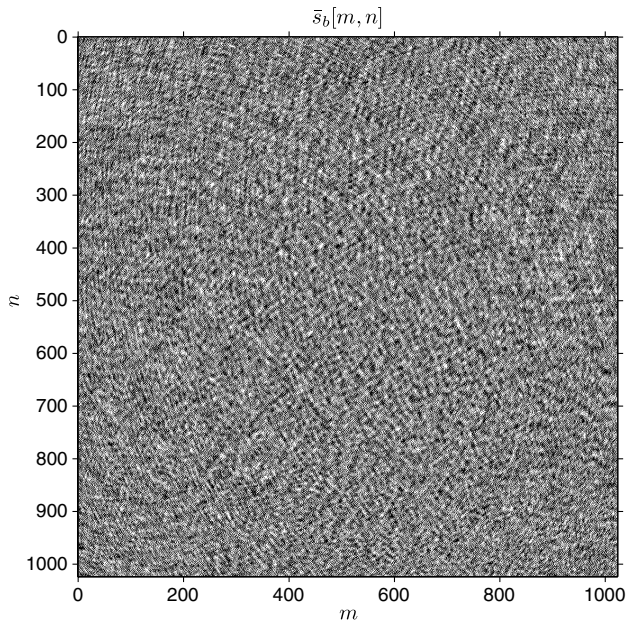
Fig. 9.   Binary SLM pattern obtained by adding the three-level SLM patterns for real and imaginary parts.

has similar characteristics with $\bar{s}_T^R[m,n]$, except for the fact that even pixels are 0 and odd pixels are either –1 or 1. Figure 9 shows the binary pattern obtained as $\bar{s}_b[m,n] = \bar{s}_T^R[m,n] + \bar{s}_T^I[m,n]$. Note that this pattern only contains –1 or 1, as it should. Next we multiply this pattern with $\bar{T}[m,n]$ given in Eq. (27). The resulting output field is shown in Fig. 10. As seen, the desired field is generated successfully within the synthesis region, where we also see the noise components that appear in the do not care region due to the high-pass component of $\bar{s}_b[m,n]$.

Above, we obtained imaginary values using a mask placed after the SLM. This option theoretically works, but complicates the optical setup in the sense that the mask should
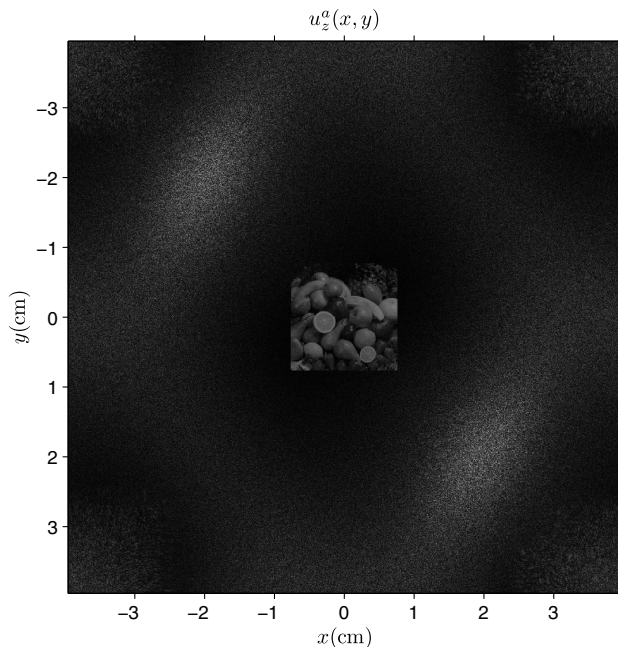
be physically produced and aligned properly with the SLM. The main purpose of using the mask is to obtain the effective SLM pattern of Eq. (28). This pattern can be obtained with alternative physical arrangements. As an easier option, suppose we illuminate the SLM with an oblique plane wave instead of a normally incident plane wave. In particular, let the illumination wave $I(x,y)$ be

$$I(x,y) = \exp\left\{j\frac{\pi}{2}\left(\frac{x}{\Delta_x}+\frac{y}{\Delta_y}\right)\right\}. \qquad (31)$$

When practical values are considered for Δ_x, Δ_y, and λ, the incidence angle of this wave is a few degrees less than 90 deg. On the SLM pixels (which are spaced by Δ_x, Δ_y), this wave creates the discrete pattern

$$\bar{I}[m,n] = \exp\left\{j\frac{\pi}{2}(m+n)\right\}, \qquad (32)$$

so that, when illuminated by $I(x,y)$, the binary pattern on the SLM is effectively multiplied with $\bar{I}[m,n]$. $\bar{I}[m,n]$ is a slightly modified version of $\bar{T}[m,n]$ given in Eq. (27), but handling this difference is trivial. In fact, it is easy to show that $\frac{\bar{T}[m,n]}{\bar{I}[m,n]}$ is always –1 or 1. If we consider a modified binary pattern $\bar{s}_{b_1}[m,n]$ such that

$$\bar{s}_{b_1}[m,n] = \begin{cases} \bar{s}_b[m,n] & \text{when } \frac{\bar{T}[m,n]}{\bar{I}[m,n]} = 1 \\ -\bar{s}_b[m,n] & \text{when } \frac{\bar{T}[m,n]}{\bar{I}[m,n]} = -1 \end{cases}, \qquad (33)$$

then we see that $\bar{s}_{b_1}[m,n]\bar{I}[m,n] = \bar{s}_b[m,n]\bar{T}[m,n]$. Therefore, in the second option, we can first compute $\bar{s}_b[m,n]$ as in Eqs. (29) and (30), then update it as in Eq. (33), then write the resulting $\bar{s}_{b_1}[m,n]$ on the SLM, and illuminate the SLM with $I(x,y)$ of Eq. (31). Figure 11 shows the updated SLM pattern obtained from the SLM pattern in Fig. 9. This SLM pattern also produces Fig. 10 when multiplied with $\bar{I}[m,n]$ of Eq. (32). This

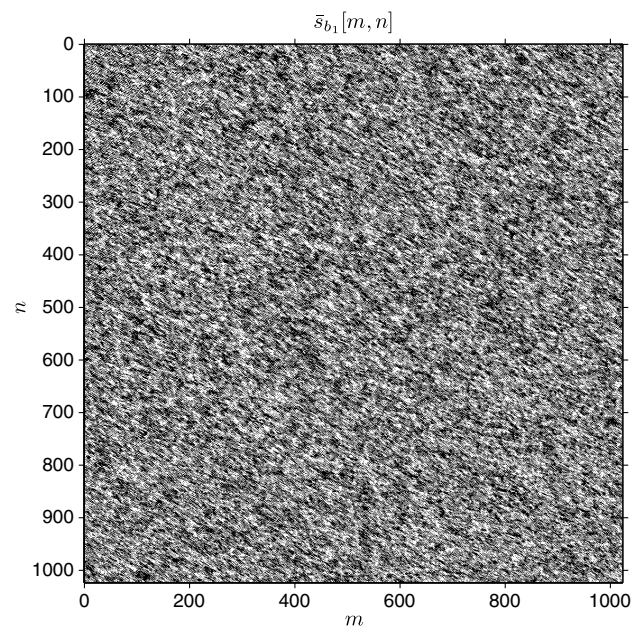Fig. 10.   Output produced by the binary SLM pattern in Fig. 9.

Fig. 11.   Updated version of the binary SLM pattern in Fig. 9 to be used when the mask is removed and oblique illumination is used.

option is much simpler than the first option since we do not need to produce any mask. However, we need to properly adjust the angle between the illumination wave and the SLM.

Finally, in practice, small deviations in the angle of the illumination wave are of no significance. The effect of such small deviations on the diffraction field generated by the SLM is merely a spatial shift. Hence, it is not critical to use exactly the illumination wave given in Eq. (31). For instance, instead of that wave, we can use a normally incident plane wave. This case is illustrated in Fig. 12. As seen, the desired field appears within a shifted version of the synthesis region. It can be shown that the center of the synthesis region is shifted to $(-\frac{\lambda z}{4\Delta_x}, -\frac{\lambda z}{4\Delta_y})$. For typical practical values of Δ_x, Δ_y, and λ, the amount of the shift is a few centimeters for z of around 1 m. Hence, as the third option, we can proceed as in the second option until we compute $\bar{s}_{b_1}[m, n]$, but then use a normally incident plane wave [or any other plane wave that makes a small angle with the wave in Eq. (31)] and accept obtaining the desired field within a shifted version of the synthesis region.

Up to now, we assumed that the binary SLM pixels can be set to ± 1. In a more general case, the pixels are set to two different complex numbers c_1 and c_2 rather than ± 1. Such a case is fundamentally no different than the ± 1 case. To see this, suppose on the physical binary SLM we can write c_1 and c_2, where $c_1 \neq c_2$. Suppose that, given a desired field, we first compute $\bar{s}_{b_1}[m, n]$ that consists of ± 1s as described above. Now, let $\bar{s}_{b_2}[m, n]$ denote the actual binary SLM pattern. Assume that we set $\bar{s}_{b_2}[m, n] = c_1$ when $\bar{s}_{b_1}[m, n] = -1$ and $\bar{s}_{b_2}[m, n] = c_2$ when $\bar{s}_{b_1}[m, n] = 1$. It is easy to show that $\bar{s}_{b_2}[m, n] = \frac{c_2 - c_1}{2}\bar{s}_{b_1}[m, n] + \frac{c_1 + c_2}{2}$. Assume the SLM is illuminated by a normally incident plane wave. The effect of the $\frac{c_2 - c_1}{2}$ term that multiplies $\bar{s}_{b_1}[m, n]$ is a trivial change in the output complex amplitude. The additive $\frac{c_1 + c_2}{2}$ term, which is nonzero when $c_1 \neq -c_2$, is more problematic and creates the so called undiffracted DC beam that propagates around the optical axis and has dimensions approximately equal to those
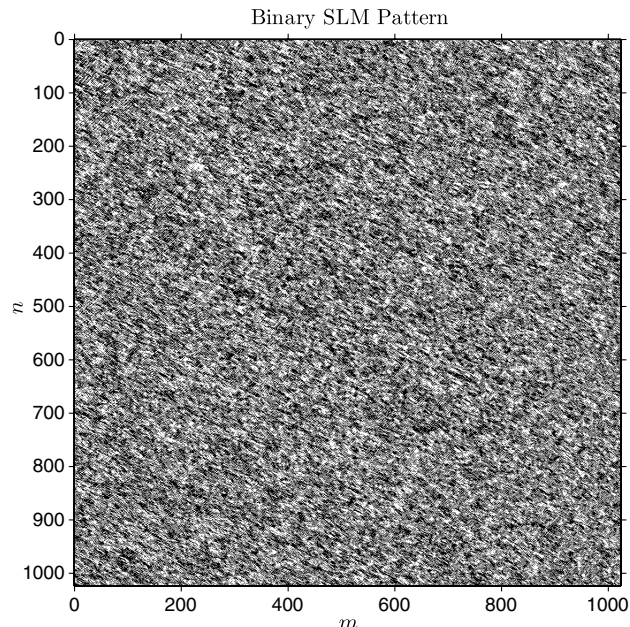

Fig. 13.   Binary SLM pattern.

of the SLM. However, when the reconstruction is performed sufficiently away from the SLM, this DC beam does not interfere with the reconstruction since the synthesis is performed in an off-axis window.

We conclude this section by noting that our approach can be used to compute binary SLM patterns for 3D applications, for instance, to synthesize 3D objects floating in air as in 3D display applications. Recall that, in Section 4, we stated that within the pyramid given in Eq. (22) (which we named the synthesis pyramid), an SLM behaves the same as its low-pass filtered version. When this pyramid is sufficiently narrow, a desired field specified within it can be synthesized with a sufficiently low-pass gray-level SLM pattern that can be halftoned
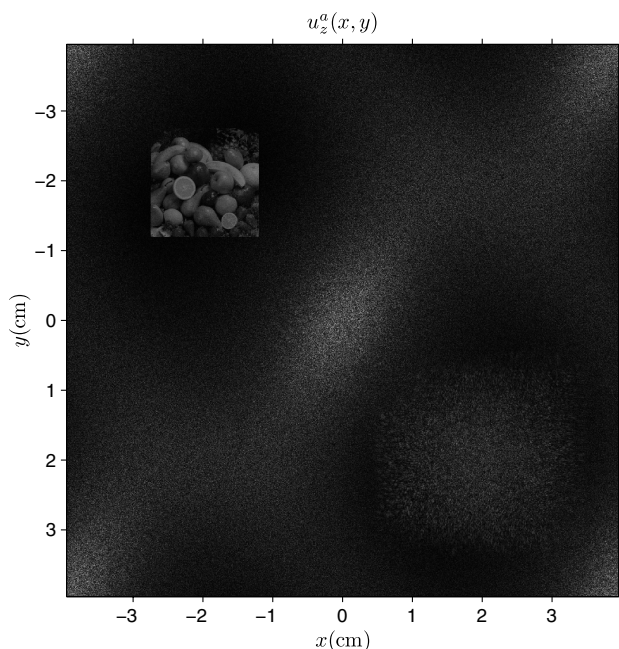

Fig. 12.   Output produced by the binary SLM pattern in Fig. 11 when normally incident illumination is used instead of oblique illumination.
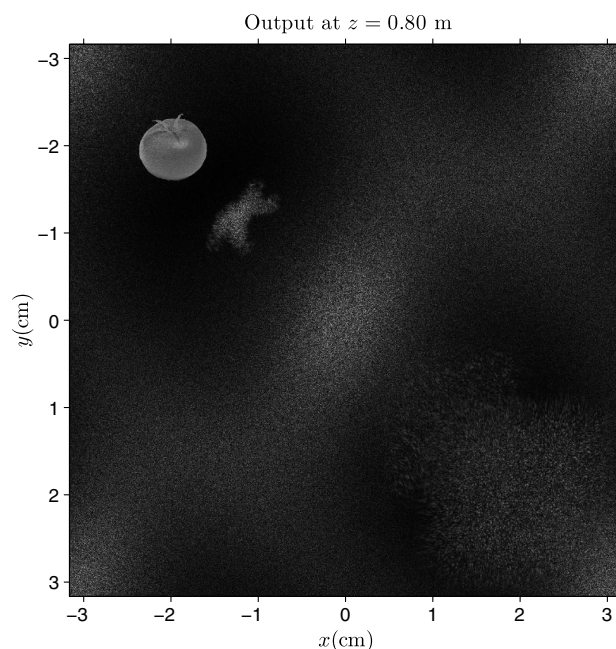

Fig. 14.   Output produced by the binary SLM pattern in Fig. 13 at $z = 0.8$ m.
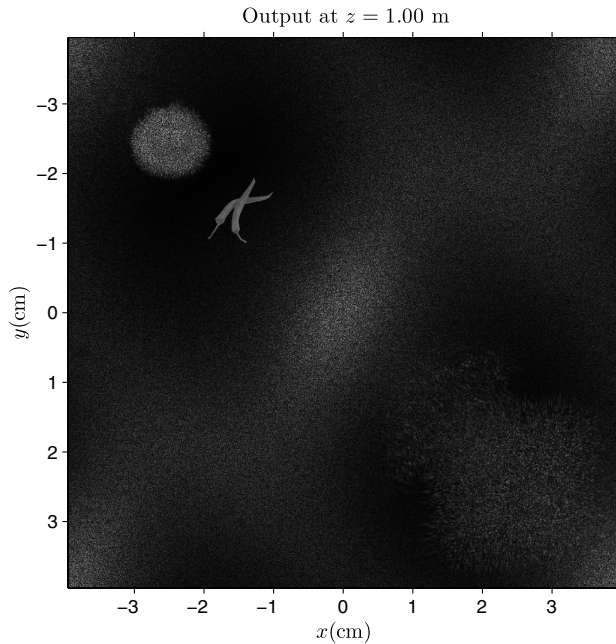
Output at $z = 1.00$ m

Fig. 15. Output produced by the binary SLM pattern in Fig. 13 at $z = 1.0$ m.

with low error into a binary SLM pattern. As an example, consider the 1024×1024 binary SLM pattern shown in Fig. 13. When written on an SLM with pixel periods Δ_x, $\Delta_y = 8\,\mu m$, and when $\lambda = 632.9$ nm, this pattern generates the fields displayed in Figs. 14 and 15 at $z = 0.8$ m and $z = 1$ m, respectively. (In this example, the physical size of the SLM is 8.2 mm $\times 8.2$ mm, the physical size of the central diffraction order is 6.33 cm $\times 6.33$ cm at $z = 0.8$ m and 7.91 cm $\times 7.91$ cm at $z = 1$ m. The physical sizes of the objects are around 1 cm.) As seen, a tomato is focused at $z = 0.8$ m, while two peppers are focused at $z = 1$ m. We also see that the quantization noise is successfully distributed over the do not care region. Note that, since the objects were chosen small enough, the ideal gray-level SLM pattern generating them was sufficiently low pass, so we managed to successfully halftone it into the binary SLM pattern in Fig. 13.

6. CONCLUSION

In this paper, we show that, when the desired light fields are suitably specified, a binary SLM can be used to synthesize them just after free space propagation without the need to use any complicated optical setup. By suitably specified, we mean that the desired fields should obey the constraints due to pixelated SLM structure discussed in Section 3 and they should be confined within the synthesis region or pyramid specified by Eq. (22) where B_x and B_y are selected sufficiently small. We showed that, if these constraints are satisfied, the ideal gray-level complex-valued SLM pattern becomes sufficiently oversampled, and it can be successfully halftoned with acceptable error. Although at a first glance the halftoning issue seemed problematic due to the fact that a binary SLM pattern is essentially real valued but the desired ideal SLM pattern is in general complex valued, we showed that this problem can be overcome with the simple technique proposed in Section 5. Our technique essentially decomposes the complex-valued halftoning problem to two decoupled real-valued constrained halftoning problems for the real and imaginary

parts of the ideal SLM pattern. Using our simulations, we show that the proposed method can be used to generate planar as well as volumetric light field distributions. Our results indicate that when ideal SLM patterns use about $1/16$ of the available bandwidth (that is, $B_x = \frac{1}{4\Delta_x}$, $B_y = \frac{1}{4\Delta_y}$), quite satisfactory results are obtained.

An important property of our approach is that, as long as the desired light field is specified properly, computation of a suitable binary SLM pattern is reduced to solving the complex-valued halftoning problem in Eq. (26). That is, free space propagation related computations can be handled separately from the halftoning related computations. This is an important advantage over many existing algorithms, especially ones that use iterative POCS-like methods similar to the Gerschberg–Saxton algorithm. Usually in such algorithms, during a typical iteration, the output field produced by some current binary SLM pattern is computed, and then that binary SLM pattern is updated according to the error between the output and desired fields. Such calculations greatly increase the computational complexity of those iterative algorithms. On the other hand, in our approach, given the desired field, it is sufficient to compute the ideal gray-level complex-valued SLM pattern only once. Then, all the computations can be carried out for solving the halftoning problem. Since any error that we have on the SLM surface after the halftoning process is directly reflected to the synthesis region, we do not need to separately incorporate free space propagation in the optimization procedure for halftoning. To solve the real-valued halftoning problems given in Eq. (30), one can use any of the many existing halftoning algorithms, depending on the expectations about the computational performance, reconstruction accuracy, binarization efficiency, etc. [67]. In our simulations, we used the standard error diffusion algorithm, which is a simple noniterative algorithm, and we found its performance quite satisfactory both in terms of reconstruction quality and computational speed.

REFERENCES

1. G. Tricoles, "Computer-generated holograms: an historical review," Appl. Opt. **26**, 4351–4360 (1987).
2. O. Bryngdahl and F. Wyrowski, "Digital holography—computer-generated holograms," in *Progress in Optics*, E. Wolf, ed. (Elsevier, 1990), Vol. 28, pp. 1–86.
3. C. Slinger, C. Cameron, and M. Stanley, "Computer-generated holography as a generic display technology," Computer **38**, 46–53 (2005).
4. W. J. Dallas, "Computer-generated holograms," in *Digital Holography and Three-Dimensional Display*, T. C. Poon, ed. (Springer, 2006), pp. 1–49.
5. F. Wyrowski, "Diffractive optical elements: iterative calculation of quantized, blazed phase structures," J. Opt. Soc. Am. A **7**, 961–969 (1990).
6. J. N. Mait, "Understanding diffractive optic design in the scalar domain," J. Opt. Soc. Am. A **12**, 2145–2158 (1995).
7. J. A. Neff, R. A. Athale, and S. H. Lee, "Two-dimensional spatial light modulators: a tutorial," Proc. IEEE **78**, 826–855 (1990).

8. M. Lucente, "Interactive three-dimensional holographic displays: seeing the future in depth," ACM SIGGRAPH Comput. Graph. **31**, 63–67 (1997).
9. , H. M. Ozaktas and L. Onural, eds., *Three-Dimensional Television: Capture, Transmission, Display*, Springer Series in Signals and Communication Technology (Springer, 2008).
10. F. Yaras, H. Kang, and L. Onural, "State of the art in holographic displays: A survey," J. Display Technol. **6**, 443–454 (2010).
11. L. Onural, F. Yaras, and H. Kang, "Digital holographic three-dimensional video displays," Proc. IEEE **99**, 576–589 (2011).
12. M. Lucente and T. A. Galyean, "Rendering interactive holographic images," in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (ACM, 1995), pp. 387–394.
13. D. Abookasis and J. Rosen, "Computer-generated holograms of three-dimensional objects synthesized from their multiple angular viewpoints," J. Opt. Soc. Am. A **20**, 1537–1545 (2003).
14. Y. Frauel, T. J. Naughton, O. Matoba, E. Tajahuerce, and B. Javidi, "Three-dimensional imaging and processing using computational holographic imaging," Proc. IEEE **94**, 636–653 (2006).
15. M. Janda, I. Hanak, and L. Onural, "Hologram synthesis for photorealistic reconstruction," J. Opt. Soc. Am. A **25**, 3083–3096 (2008).
16. L. Onural, A. Gotchev, H. M. Ozaktas, and E. Stoykova, "A survey of signal processing problems and tools in holographic three-dimensional television," IEEE Trans. Circuits Syst. Video Technol. **17**, 1631–1646 (2007).
17. L. Onural and H. M. Ozaktas, "Signal processing issues in diffraction and holographic 3DTV," Signal Process. Image Commun. **22**, 169–177 (2007).
18. T. Kreis, P. Aswendt, and R. Hofling, "Hologram reconstruction using a digital micromirror device," Opt. Eng. **40**, 926–933 (2001).
19. D. Dudley, W. Duncan, and J. Slaughter, "Emerging digital micromirror device (DMD) applications," White Paper (Texas Instruments, 2003).
20. B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," Appl. Opt. **5**, 967–969 (1966).
21. A. W. Lohmann and D. P. Paris, "Binary Fraunhofer holograms, generated by computer," Appl. Opt. **6**, 1739–1748 (1967).
22. J. P. Waters, "Three-dimensional Fourier-transform method for synthesizing binary holograms," J. Opt. Soc. Am. **58**, 1284–1288 (1968).
23. R. A. Gabel and B. Liu, "Minimization of reconstruction errors with computer-generated binary holograms," Appl. Opt. **9**, 1180–1191 (1970).
24. A. R. Sass, "Binary intensity holograms," J. Opt. Soc. Am. **61**, 910–915 (1971).
25. P. L. Ransom and R. F. Henton, "Analysis of a computer-generated binary-phase hologram," Appl. Opt. **13**, 2765–2767 (1974).
26. W. Lee, "Binary synthetic holograms," Appl. Opt. **13**, 1677–1682 (1974).
27. R. A. Gabel, "Reconstruction errors in computer-generated binary holograms: a comparative study," Appl. Opt. **14**, 2252–2255 (1975).
28. W. Lee, "Binary computer-generated holograms," Appl. Opt. **18**, 3661–3669 (1979).
29. Y. H. Wu and P. Chavel, "Cell-oriented on-axis computer-generated holograms for use in the Fresnel diffraction mode," Appl. Opt. **23**, 228–238 (1984).
30. M. Li, A. Larsson, N. Eriksson, and M. Hagberg, "Continuous-level phase-only computer-generated hologram realized by dislocated binary gratings," Opt. Lett. **21**, 1516–1518 (1996).
31. R. Hauck and O. Bryngdahl, "Computer-generated holograms with pulse-density modulation," J. Opt. Soc. Am. A **1**, 5–10 (1984).
32. O. K. Ersoy, J. Zhuang, and J. Brede, "Iterative interlacing approach for synthesis of computer-generated holograms," Appl. Opt. **31**, 6894–6901 (1992).
33. P. Thorston, F. Wyrowski, and O. Bryngdahl, "Importance of initial distribution for iterative calculation of quantized diffractive elements," J. Mod. Opt. **40**, 591–600 (1993).
34. C. Wu, C. Chen, and M. A. Fiddy, "Iterative procedure for improved computer-generated hologram reconstruction," Appl. Opt. **32**, 5135–5140 (1993).
35. N. Yoshikawa and T. Yatagai, "Phase optimization of a kinoform by simulated annealing," Appl. Opt. **33**, 863–868 (1994).
36. E. Zhang, S. Noehte, C. H. Dietrich, and R. Manner, "Gradual and random binarization of gray-scale holograms," Appl. Opt. **34**, 5987–5995 (1995).
37. L. Legeard, P. Refregier, and P. Ambs, "Multicriteria optimality for iterative encoding of computer-generated holograms," Appl. Opt. **36**, 7444–7449 (1997).
38. H. H. Suh, "Color-image generation by use of binary phase holograms," Opt. Lett. **24**, 661–663 (1999).
39. R. Eschbach, "Comparison of error diffusion methods for computer-generated holograms," Appl. Opt. **30**, 3702–3710 (1991).
40. A. Kirk, K. Powell, and T. Hall, "A generalization of the error diffusion method for binary computer-generated hologram design," Opt. Commun. **92**, 12–18 (1992).
41. R. Eschbach and Z. Fan, "Complex valued error diffusion for off-axis computer-generated holograms," Appl. Opt. **32**, 3130–3136 (1993).
42. F. Fetthauer, S. Weissbach, and O. Bryngdahl, "Equivalence of error diffusion and minimal average error algorithms," Opt. Commun. **113**, 365–370 (1995).
43. K. Heggarty and R. Chevallier, "Signal window minimum average error algorithm for computer-generated holograms," J. Opt. Soc. Am. A **15**, 625–635 (1998).
44. M. A. Seldowitz, J. P. Allebach, and D. W. Sweeney, "Synthesis of digital holograms by direct binary search," Appl. Opt. **26**, 2788–2798 (1987).
45. B. K. Jennison and J. P. Allebach, "Efficient design of direct-binary-search computer-generated holograms," J. Opt. Soc. Am. A **8**, 652–660 (1991).
46. J. Zhuang and O. K. Ersoy, "Fast decimation-in-frequency direct binary search algorithms for synthesis of computer-generated holograms," J. Opt. Soc. Am. A **11**, 135–143 (1994).
47. J. Zhuang and O. K. Ersoy, "Optimal decimation-in-frequency iterative interlacing technique for synthesis of computer-generated holograms," J. Opt. Soc. Am. A **12**, 1460–1468 (1995).
48. B. B. Chhetri, S. Yang, and T. Shimomura, "Stochastic approach in the efficient design of the direct-binary-search algorithm for hologram synthesis," Appl. Opt. **39**, 5956–5964 (2000).
49. F. Wyrowski, "Iterative quantization of digital amplitude holograms," Appl. Opt. **28**, 3864–3870 (1989).
50. L. Bigue and P. Ambs, "Optimal multicriteria approach to the iterative Fourier transform algorithm," Appl. Opt. **40**, 5886–5893 (2001).
51. S. H. Tao and X. Yuan, "Practical implementation of the phase-quantization technique in an iterative Fourier-transform algorithm," Appl. Opt. **43**, 2089–2092 (2004).
52. J. P. Allebach, "Representation-related errors in binary digital holograms: a unified analysis," Appl. Opt. **20**, 290–299 (1981).
53. B. K. Jennison and J. P. Allebach, "Analysis of the leakage from computer-generated holograms synthesized by direct binary search," J. Opt. Soc. Am. A **6**, 234–243 (1989).
54. F. Wyrowski, "Diffraction efficiency of analog and quantized digital amplitude holograms: analysis and manipulation," J. Opt. Soc. Am. A **7**, 383–393 (1990).
55. Y. Chang, P. Zhou, and J. H. Burge, "Analysis of phase sensitivity for binary computer-generated holograms," Appl. Opt. **45**, 4223–4234 (2006).
56. C. Maurer, A. Schwaighofer, A. Jesacher, S. Bernet, and M. Ritsch-Marte, "Suppression of undesired diffraction orders of binary phase holograms," Appl. Opt. **47**, 3994–3998 (2008).
57. R. Piestun, B. Spektor, and J. Shamir, "On-axis binary-amplitude computer-generated holograms," Opt. Commun. **136**, 85–92 (1997).
58. J. A. Davis, K. O. Valadez, and D. M. Cottrell, "Encoding amplitude and phase information onto a binary phase-only spatial light modulator," Appl. Opt. **42**, 2003–2008 (2003).
59. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, 1996).
60. G. C. Sherman, "Application of the convolution theorem to Rayleigh's integral formulas," J. Opt. Soc. Am. **57**, 546–547 (1967).
61. E. Lalor, "Conditions for the validity of the angular spectrum of plane waves," J. Opt. Soc. Am. **58**, 1235–1237 (1968).
62. L. Onural, "Exact analysis of the effects of sampling of the scalar diffraction field," J. Opt. Soc. Am. A **24**, 359–367 (2007).

63. L. Onural, "Sampling of the diffraction field," Appl. Opt. **39**, 5929–5935 (2000).

64. S. B. Tucker, J. O. Castaneda, and W. T. Cathey, "Matrix description of near-field diffraction and the fractional Fourier transform," J. Opt. Soc. Am. A **16**, 316–322 (1999).

65. F. Gori, "Fresnel transform and sampling theorem," Opt. Commun. **39**, 293–297 (1981).

66. D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing* (Prentice Hall, 1990).

67. R. Ulichney, *Digital Halftoning* (MIT Press, 1987).