Synthesis of digital holograms by direct binary search

Michael A. Seldowitz, Jan P. Allebach, and Donald W. Sweeney

A new approach to the design of computer-generated holograms makes optimal use of the available device resolution. An iterative search algorithm minimizes an error criterion by directly manipulating the binary hologram and observing the effect on the desired reconstruction. Several measures of error and efficiency useful in assessing the optimality of digital holograms are defined. Methods for designing digital holograms that are based on projections and error diffusion are presented as established techniques for comparison to direct binary search.

I. Introduction

Computer-generated holograms are widely employed in wavefront manipulation, synthesis, and retrieval. Present and potential applications include optical information processing, interferometry, synthesis of novel optical elements, laser scanning, and laser machining. A critical factor hindering the continued growth of digital holography is the limited space–bandwidth products presently achievable. This is due to both the limited resolution of the device used to write the holographic component and the large amount of computation required to determine the desired component. The use of e-beam lithography for hologram fabrication has led to large improvements in resolution; however, writing time may be extremely long. Alternatively, holograms can be written on spatial light modulators that are dynamically reconfigurable in real time but that have the disadvantage of low resolution.

We have developed a new method for hologram synthesis that efficiently utilizes the available hologram resolution. By employing an iterative search algorithm, we generate a binary hologram that satisfies a criterion of optimality. The algorithm works directly with the actual binary hologram and the desired reconstruction, avoiding problems related to traditional approaches of encoding the complex-valued object spectrum. Since the new method generates optimal or near optimal hologram configurations of any given number of addressable points, it is suitable for output devices having a wide range of resolutions.

In Sec. II, we present criteria used to assess the optimality of computer-generated holograms. Several measures of error and efficiency are defined. In Sec. III, we introduce our new method, referred to as direct binary search (DBS). In addition, we discuss the application of simulated annealing to hologram synthesis and compare it to the DBS algorithm. In Sec. IV, we consider several variations to our algorithm including alternate search ordering, forced efficiency, and an error measure based on only the magnitude of the reconstruction. To judge the merit of the DBS method relative to established methods for generating digital holograms, we describe a cell-oriented projection method and a point-oriented method employing error diffusion in Sec. V. Section VI contains experimental results for the procedures of the previous sections. The three methods are evaluated based on the optimality criteria of Sec. II; and both digital and optical reconstructions are presented as examples.

II. Problem Definition

The basic problem in the synthesis of binary digital holograms is to find a binary-valued transmittance function for the hologram so that a portion of the propagating wavefront at a given distance from the hologram plane has some desired form. In this paper, we consider an optical configuration in which propagation between the hologram and the plane of observation is described by a 2-D continuous inverse Fourier transform. The hologram $H(u,v)$ is considered to be in the frequency domain and the observed diffraction pattern $h(x,y)$ in the spatial domain. They are related by the continuous Fourier transform:

$$h(x,y) = \int\int H(u,v) \exp[-i2\pi(ux + vy)]dxdy. \quad (1)$$

The hologram structure is an $M \times N$ array of rectangu-

Donald Sweeney is with Sandia National Laboratories, Livermore, California 94550; the other authors are with Purdue University, School of Electrical Engineering, West Lafayette, Indiana 47907.

Received 1 April 1987.
0003-6935/87/142788-11$02.00/0.

lar cells, each of size $R \times S$ with binary transmittance H_{kl}, $-M/2 \le k < M/2$ and $-N/2 \le l < N/2$. Here for convenience we assume that M and N are even. Therefore,

$$H(u,v) = \sum_{k=-M/2}^{M/2-1} \sum_{l=-N/2}^{N/2-1} H_{kl} \operatorname{rect}\left(\frac{u-kR}{R}, \frac{v-lS}{S}\right), \quad (2)$$

where

$$\operatorname{rect}(a,b) = \begin{cases} 1, & \text{if } |a|,|b| < 1/2, \\ 0, & \text{else.} \end{cases}$$

The diffraction pattern is then given by

$$h(x,y) = RS \operatorname{sinc}(Rx,Sy) \sum_{k=-M/2}^{M/2-1} \sum_{l=-N/2}^{N/2-1} H_{kl} \exp[i2\pi(Rxk + Syl)], \quad (3)$$

where $\operatorname{sinc}(a,b) = \sin(\pi a) \sin(\pi b)/\pi^2 ab$. Sampling $h(x,y)$ at interval X in the x direction and at interval Y in the y direction, where $RX = 1/M$ and $SY = 1/N$, we obtain

$$h(mX,nY) = \left(\frac{RS}{XY}\right)^{1/2} \operatorname{sinc}\left(\frac{m}{M}, \frac{n}{N}\right) h_{mn}, \quad (4)$$

where h_{mn} is the inverse discrete Fourier transform of H_{kl}:

$$h_{mn} = \frac{1}{\sqrt{MN}} \sum_{k=-M/2}^{M/2-1} \sum_{l=-N/2}^{N/2-1} H_{kl} \exp\left\{i2\pi\left(\frac{mk}{M} + \frac{nl}{N}\right)\right\}. \quad (5)$$

Finally, define the digitally reconstructed image as

$$g_{mn} = \begin{cases} h_{m-m_0,n-n_0}, & \text{if } (m-m_0,n-n_0) \in \mathbf{R}, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where $\mathbf{R}$ is the support region of the desired reconstruction, and (m_0,n_0) is the desired reconstruction origin.

Throughout this paper, we work only with the sampled hologram H_{kl} and the digitally reconstructed image g_{mn}, which is simply related to the samples $h(mX,nY)$ from the diffraction pattern by Eqs. (4) and (6). Thus we ignore the sinc rolloff in Eq. (4) due to the nonzero size of the $R \times S$ hologram elements, which may easily be eliminated by precompensating the desired object.

More significantly, we also do not take into account interaction of the samples h_{mn} from the digitally reconstructed diffraction pattern that results in the optically reconstructed diffraction pattern[1]

$$h(x,y) = \left(\frac{RS}{XY}\right)^{1/2} \operatorname{sinc}(Rx,Sy) \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} h_{mn} \times s_{MN}\left(\frac{x-mX}{X}, \frac{y-nY}{Y}\right), \quad (7)$$

where

$$s_{MN}(x,y) = \frac{\operatorname{sinc}(x,y)}{\operatorname{sinc}(x/M,y/N)} \exp[-i\pi(x/M + y/N)]. \quad (8)$$

This interaction is responsible for the speckled appearance of images reconstructed from traditional computer-generated holograms in which a random phase code has been assigned to the object.

Ideally, the complex amplitude of the reconstructed image is proportional to some object of interest. However, because of the finite resolution of the output device and the binary transmittance of the resulting hologram, the reconstruction will be in error. This error along with the diffraction efficiency are essential criteria for evaluating the optimality of the synthesized hologram. In the remainder of this section, we discuss various measures of both error and efficiency useful in assessing reconstruction quality.

An error measure should reflect the similarity between the object and reconstructed image. Let f_{mn} be the object and $\tilde{f}_{mn}$ be the object scaled to have a peak spectral amplitude of unity. Assume that both the object and desired reconstructed image have $A \times B$ addressable points, where again A and B are taken to be even. We define the mean squared error e between the object and reconstructed image as

$$e = \frac{1}{AB} \sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} |\tilde{f}_{mn} - \lambda g_{mn}|^2, \quad (9)$$

where the parameter λ is a complex factor that scales the reconstruction for a minimum mean squared error fit to the original object. This scale factor effectively normalizes the reconstruction to remove the effects of differing diffraction efficiencies so that the error measures do not reflect these differences. It is derived by setting the partial derivatives of e with respect to the real and imaginary parts of λ equal to 0, resulting in

$$\lambda = \frac{\sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} \tilde{f}_{mn} g^*_{mn}}{\sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} |g_{mn}|^2}. \quad (10)$$

The * is the complex conjugate operation. We define the normalized rms error $\bar{e}_{\text{rms}}$ as

$$\bar{e}_{\text{rms}} = \frac{e^{1/2}}{\max_{mn} |\tilde{f}_{mn}|}, \quad (11)$$

where the overbar denotes normalization. The denominator of the above expression maintains consistency among the reconstructions for different objects; it is equivalent to the peak spatial magnitude of the object divided by its peak spectral magnitude.

We will define three efficiency measures for the assessment of reconstruction quality: object efficiency; total efficiency; and binarization efficiency. These measures are related as shown in Fig. 1. A hologram can only modulate an incident wavefront by selectively absorbing or delaying it. The ideal passive hologram would have complex-valued transmittance $H_{kl} = \tilde{F}_{kl}$. When illuminated by a unit amplitude plane wave, the fraction of the incident energy remaining in the wavefront emanating from behind this hologram is the object efficiency η_{obj}. All this energy would be diffracted into the desired reconstruction. Thus

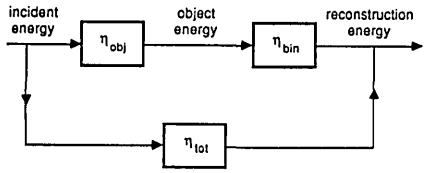

Fig. 1 Relationship between efficiency measures.

$$\eta_{obj} = \frac{1}{MN} \sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} |\tilde{f}_{mn}|^2. \quad (12)$$

A binary amplitude-only hologram modulates the incident wavefront by selectively blocking it. Of the energy passed by the hologram, only part will be diffracted into the desired reconstruction. The total efficiency η_{tot} is the fraction of energy incident on the hologram directed into this reconstruction:

$$\eta_{tot} = \frac{1}{MN} \sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} |g_{mn}|^2. \quad (13)$$

A measure of the efficiency of the binary amplitude-only hologram relative to the ideal hologram is the binarization efficiency

$$\eta_{bin} = \frac{\eta_{tot}}{\eta_{obj}}. \quad (14)$$

With a binary phase-only medium, this efficiency would be boosted by a factor of 4.

There exists a relationship between the mean squared error and the efficiency measures which becomes apparent when Eq. (10) is substituted into Eq. (9), resulting in

$$e = \frac{1}{AB} \left\{ \sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} |\tilde{f}_{mn}|^2 - |\lambda|^2 \sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} |g_{mn}|^2 \right\}. \quad (15)$$

Thus the mean squared error can be interpreted as the difference between the normalized object energy and the optimally scaled reconstruction energy. In terms of the efficiency measures,

$$e = \frac{MN}{AB} \eta_{obj}(1 - |\lambda|^2 \eta_{bin}). \quad (16)$$

Note that when $e = 0$,

$$\eta_{bin} = \frac{1}{|\lambda|^2}. \quad (17)$$

III. Direct Binary Search Hologram

Traditional approaches to the problem of generating binary Fourier transform holograms convert the complex-valued object spectrum into a binary real-valued transmittance function. The direct binary search method shifts the emphasis from spectrum encoding to manipulating the hologram transmittance directly to produce the best reconstruction. The new approach employs an iterative search algorithm to find a binary transmittance function that minimizes the mean squared error between the reconstructed image and original object.

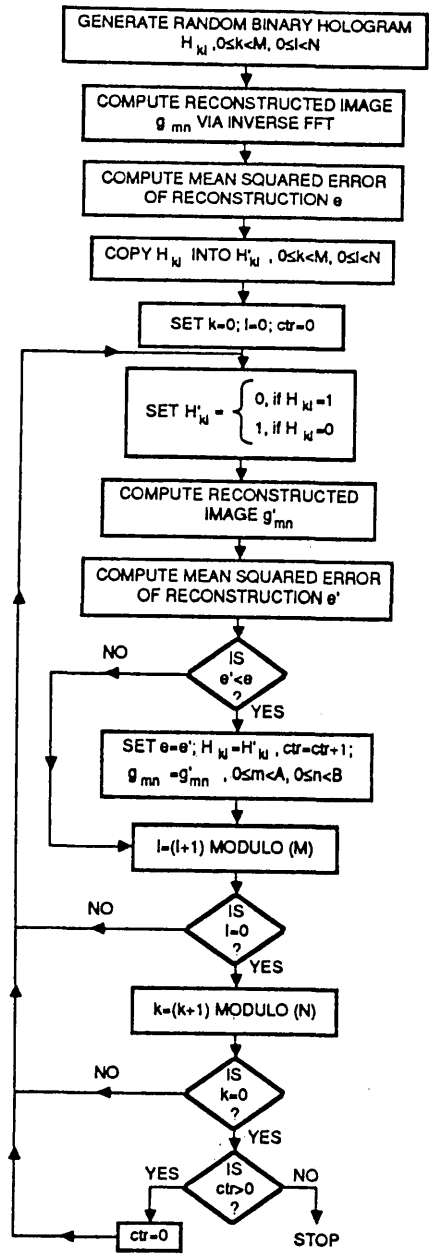

Fig. 2. Flow chart of the direct binary search algorithm.

The DBS algorithm is flowcharted in Fig. 2. We begin by generating a random binary hologram H_{kl} with a uniform distribution of 1s and 0s. The reconstructed image g_{mn} is calculated via the inverse FFT, and the mean squared error between it and the original object is computed using Eqs. (10) and (15). The hologram is then scanned in lexicographic order. For each hologram point, we invert the binary value of its transmittance and compute the resultant reconstructed image. The mean squared error between the new reconstruction and the object is then calculated and

compared to the previous error. If the error decreases, the altered hologram configuration and the new error value are retained. Otherwise, the hologram point is restored to its original value. Once every addressable point in the hologram has been considered, an iteration is said to be completed. The algorithm terminates when no inversions are kept during an entire iteration.

Because of the computational intensity of the FFT algorithm, its use in reconstructing the image after each hologram point inversion is impractical. A more efficient means of computing the reconstruction exploits the specific nature of the DBS algorithm. Since we are altering only a single point in the binary hologram per reconstruction and only interested in the reconstruction in a limited part of the diffraction pattern, the inverse discrete Fourier transform given by Eq. (5) simplifies to

$$g_{mn} = \begin{cases} \hat{g}_{mn} + \dfrac{1}{\sqrt{MN}} \exp\left[i2\pi\left(\hat{k}\dfrac{m}{M} + \hat{l}\dfrac{n}{N}\right)\right], & \text{if } H_{\hat{k}\hat{l}} = 1, \\ \hat{g}_{mn} - \dfrac{1}{\sqrt{MN}} \exp\left[i2\pi\left(\hat{k}\dfrac{m}{M} + \hat{l}\dfrac{n}{N}\right)\right], & \text{if } H_{\hat{k}\hat{l}} = 0, \end{cases} \tag{18}$$

where $(\hat{k},\hat{l})$ is the altered hologram point, and $\hat{g}_{mn}$ is the reconstructed image computed prior to the inversion of $H_{\hat{k}\hat{l}}$. This expression need only be evaluated for $(m - m_0, n - n_0) \in \mathbf{R}$. Repeated evaluation of trigonometric functions can be avoided by precomputing and storing the terms $\cos(2\pi p/M)$ and $\sin(2\pi p/M)$, $p = 0, \ldots, M-1$ and $\cos(2\pi q/N)$ and $\sin(2\pi q/N)$, $q = 0, \ldots, N-1$.

The hologram generated by the DBS procedure is near optimum in the sense that it minimizes the mean squared error between the reconstruction and object for a given random start. Consider an MN-dimensional hologram space, where MN is the number of addressable hologram points. Each possible hologram configuration corresponds to a corner of a unit hypercube with coordinates that are some permutation of MN 0s and 1s. The DBS algorithm starts at a random corner of the hypercube and attempts to move along an edge to a neighboring corner. If the move results in a decrease in reconstruction error, a further edge traversal is attempted from the new corner. Otherwise, the move is reversed, and a traversal along a different edge from the previous corner is attempted. The process continues until a corner is reached that has no neighbors with lower associated reconstruction errors. Thus the final hologram configuration has the minimum mean squared reconstruction error of all hologram configurations evaluated in the procedure. However, hologram configurations with lower reconstruction errors may correspond to corners of the hypercube never reached during the course of the search. The DBS algorithm, therefore, finds only local optima.

By applying the technique of simulated annealing,[2–4] we introduce a means of transitioning out of local optima to reach a global or more nearly global optimum. These transitions are accomplished by probabilistically accepting hologram point inversions that increase the mean squared error of the reconstruction. The probability of accepting these inversions is $\exp(-\Delta e/T)$, where Δe is the change in error due to the inversion and T is a control parameter. As before, hologram alterations which decrease the error are unconditionally accepted. We start with a high value of T so that essentially all proposed inversions are accepted and then slowly lower T at the beginning of each iteration until the number of accepted changes is small. This process is analogous to melting a physical system at a high temperature and then lowering the temperature slowly until the system freezes at a state of minimum energy. The inversion of a hologram point corresponds to a small perturbation of the physical system, and the resulting change in the mean squared error of the reconstruction corresponds to the resulting change in the energy of the system. Thus this process finds a hologram configuration which has a reconstructed image of minimum mean squared error.

Viewing simulated annealing in the context of searching the vertices of an MN-dimensional unit hypercube, edge transversals corresponding to increases in the reconstruction error are permitted on a probabilistic basis. Theoretically, this new degree of freedom allows the search to encompass near globally optimum corners of the hypercube that might otherwise not be reached. The significance of this advantage over the DBS procedure depends entirely on how the error varies with each possible hologram configuration. It may be conjectured that if a small number of configurations have reconstruction errors much less than those of all other configurations, i.e., only a few minima are close in value to the absolute minimum, simulated annealing should perform better than the DBS procedure. On the other hand, if many hologram configurations have reconstruction errors close to the absolute minimum error, the difference in performance between the two methods may not be significant. Two sets of experimental results for holograms of size 64×64 support the latter hypothesis.

In the first experiment, we used ten different random seeds to generate initial hologram configurations for both a binary and gray level object and observed the effect on the reconstructions from the holograms generated by the DBS algorithm. The difference between the best and worst rms reconstruction errors was only 6.5% of the average rms error value, and the standard deviation was only 2.0% of the average value. Hence, while different random starting configurations led to different final hologram configurations, there was little effect on the quality of the reconstructed image.

The second set of results compares rms reconstruction errors obtained by simulated annealing to those obtained by the DBS method. The performance of the simulated annealing procedure depends on the selection of an adequate annealing schedule and stopping rule. The annealing schedule specifies the initial temperature T_0 and the rate T_1/T_0 by which the system is cooled. These two parameters determine the temperature during the pth iteration according to the equation

$$T_p = (T_1/T_0)^p T_0. \tag{19}$$

For the object used in this experiment, we empirically found the optimum initial temperature to be 0.001 and the optimum rate to be 0.9. The stopping rule determines when the annealing is terminated. The rule we implemented stopped annealing when five (~2% of the total number of addressable hologram points) or fewer hologram point inversions were accepted per iteration for three consecutive iterations. Using these parameters, the simulated annealing algorithm converged in an average of forty-seven iterations using four different random seeds, while the DBS algorithm converged in an average of only fifteen iterations for the same object and random seeds. The quality of the reconstructions was similar, with the holograms generated by simulated annealing having rms reconstruction errors only an average of 3.7% lower than the DBS holograms.

IV. Variations on the Direct Binary Search Method

The direct binary search procedure introduced in the previous section scans the hologram in lexicographic order, scales the reconstruction to minimize the mean squared error, and computes error measures based on the complex amplitude of the reconstruction. Alternatives investigated include congruential scanning of the hologram, fixing the scale factor in the mean squared error calculation to force the efficiency to a prescribed value, and basing error measures on only the magnitude of the reconstruction.

The order in which hologram point inversions are attempted affects the final hologram configuration, since the inversion of a single hologram point changes the set of future inversions that will be accepted. In the DBS algorithm described in Sec. III, the hologram is scanned lexicographically; thus consecutive inversion attempts occur at contiguous points in the hologram. To determine if a better hologram configuration could be achieved by dispersing consecutive inversion attempts over the entire hologram, we implemented a congruential scanning pattern.[5] In this case, the qth trial inversion during an iteration is made at coordinates

$$(k_q,l_q) = (\Delta kq \bmod M, \Delta lq \bmod N), \tag{20}$$

where the hologram dimensions M and N and the displacement $(\Delta k,\Delta l)$ between successive points must satisfy $\gcd(M,N) = 1$, $\gcd(\Delta k,M) = 1$, and $\gcd(\Delta l,N) = 1$. Here gcd denotes the greatest common divisor. These conditions assure that each hologram point will be visited once and only once during an iteration. While congruential scanning yields different hologram configurations than lexicographic scanning, experimental results show that rms reconstruction errors do not differ significantly.

The second variation to the original DBS algorithm involves holding the scale factor λ in the mean squared error calculation at a fixed value. From Eq. (17), $\eta_{\text{bin}} \to 1/|\lambda|^2$ as the mean squared error $\to 0$; thus fixing λ will tend to drive η_{bin} to a prechosen value. One might expect that by setting λ low, η_{bin} could be increased above that resulting in the unconstrained case at the cost of a higher rms reconstruction error or that setting λ high might reduce the rms reconstruction error below that resulting from the unconstrained case. The latter situation would be consistent with the observation that it is possible to design an arbitrarily bandlimited 1-D signal that will have any specified set of zero crossings in a finite interval, if we let the fraction of the signal energy in that interval be sufficiently small.[6] To explore these possibilities, we first generated a hologram using the original DBS algorithm. Using the same object and identical parameters, we generated a number of holograms using the modified DBS algorithm with fixed values of λ. For the final calculation of the mean squared error, the optimal λ given by Eq. (10) was used. None of the holograms generated by the modified method had rms reconstruction errors as low as the rms reconstruction error of the hologram generated with optimal scaling. The fixed λ which yielded the lowest reconstruction error was closest to the optimum λ used to scale the final reconstructed image produced by the original procedure. As λ deviated further from its optimum value, the resultant mean squared reconstruction error increased significantly. Thus reconstruction efficiency may be increased at the expense of higher reconstruction error, but a lower reconstruction error cannot be achieved by decreasing the reconstruction efficiency.

The final and most significant variation to the original DBS algorithm bases the error measure on only the magnitude of the reconstruction instead of its complex amplitude. In many applications of holography, only the intensity of the reconstructed image is of interest. In these cases, an ideal hologram would generate a desired reconstruction with a magnitude at every point equal to that of the original object. The mean squared error measure e must, therefore, be redefined as

$$e = \frac{1}{AB} \sum_{m=-A/2}^{A/2-1} \sum_{n=-B/2}^{B/2-1} (|\tilde{f}_{mn}| - \lambda|g_{mn}|)^2. \tag{21}$$

The scaling factor λ is now real and defined by Eq. (10) if $|\tilde{f}_{mn}|$ and $|g_{mn}|$ are used in place of $\tilde{f}_{mn}$ and g_{mn}. While traditional methods for generating digital holograms choose explicit phase codes to reduce the spectral dynamic range when only the magnitude of the reconstruction is of interest,[7] the DBS algorithm simply allows the object phase to float to improve the reconstruction. The improvement that results will be discussed in Sec. VI.

V. Established Methods Used for Comparison

To assess the performance of the direct binary search hologram, it is necessary to compare the quality of images reconstructed from it to images reconstructed from holograms synthesized by more established methods. Traditionally, binary Fourier transform holograms are generated by a binary encoding of the complex-valued object spectrum. Approaches to this procedure can be classified as being either cell-oriented or point-oriented.[8] With a cell-oriented approach, the hologram is divided into $M/A \times N/B$ cells of resolvable points, where it is assumed that M/A and N/B are

integers. Each cell is used to encode a single sample of the object spectrum, which is calculated via the 2-D DFT of the $A \times B$ object. With point-oriented approaches, the object spectrum is calculated at $M \times N$ points via the 2-D DFT of the $A \times B$ object zero-padded to $M \times N$ points. The $M \times N$ spectrum is then encoded as an $M \times N$ binary array.

In another taxonomy,[9] cell- and point-oriented approaches as described above are referred to as Category 1 and 3 methods, respectively. An additional class of method which utilizes more than one spectral sample in the encoding of each hologram cell is described as a Category 2 method. Both theoretical[9] and experimental[10] studies have shown that in terms of reconstructed image quality, these methods generally rank in order of increasing category with Category 3 methods performing best. For comparison with the DBS method, we implemented both Category 2 and Category 3 methods.

The Category 2 method was the projection-type hologram[11] which has been found experimentally to perform best among several Category 2 methods and nearly as well or better than some Category 3 methods.[10] With this method, the object spectrum is calculated at $M \times N/A$ points via the 2-D DFT of the zero-padded object. Next, we clip the spectrum to reduce the spectral dynamic range and scale it to obtain a peak spectral amplitude of unity. We then project each sample onto one of the M/A basis vectors uniformly spaced in angle around the unit circle. The index of the basis vector used is the column index of the sample modulo M/A. If the projection is non-negative, that column of the cell is filled with 0s and 1s, so that the average transmittance is proportional to the magnitude of the projection; otherwise, the column is filled with 0s.

The point-oriented or Category 3 approach that we considered was based on error diffusion.[12] Here we zero pad the object and calculate its spectrum F_{kl} at $M \times N$ points. We then clip the spectrum and modulate to make it real-valued. The modulated spectrum $\hat{F}_{kl}$ is given by

$$\hat{F}_{kl} = |F_{kl}| \cos[2\pi(km_0/M + ln_0/N) - \Theta_{kl})], \tag{22}$$

where (m_0,n_0) is the reconstruction origin, and Θ_{kl} is the phase of F_{kl}. This corresponds to forming a composite version of the object in the spatial domain which possesses Hermitian symmetry. Next we add a real-valued bias function B_{kl}, so that

$$\tilde{F}_{kl} = \hat{F}_{kl} + B_{kl} \tag{23}$$

will be real-valued and non-negative. There are four bias functions we will consider: (1) constant bias; (2) intensity bias; (3) magnitude bias; and (4) halfwave rectification. In the first case, we add the maximum magnitude of the object spectrum, so that

$$B_{kl} = \max_{kl} |F_{kl}|. \tag{24}$$

In the second case, this bias is based on the way in which reference and object beams interact in conventional holography to form an intensity pattern recorded by the medium. We start with the intensity pattern

$$I_{kl} = \left| \frac{F_{kl}}{\sqrt{c}} + \frac{\sqrt{c}}{2} \exp[i2\pi(km_0/M + ln_0/N)] \right|^2, \tag{25}$$

where c is a positive constant. It can easily be shown that this quantity is equivalent to

$$I_{kl} = \hat{F}_{kl} + \frac{|F_{kl}|^2}{c} + \frac{c}{4}. \tag{26}$$

Since I_{kl} is real-valued and non-negative,

$$B_{kl} = \frac{|F_{kl}|^2}{c} + \frac{c}{4} \tag{27}$$

is a valid bias. To reduce the amount by which $\tilde{F}_{kl}$ must be attenuated to put it in the 0 to 1 range, we obtain a value for c which minimizes the maximum value of the bias. This is accomplished by setting the derivative of $\max_{kl} B_{kl}$ with respect to c equal to zero, resulting in

$$c = 2 \max_{kl} |F_{kl}|. \tag{28}$$

In the third case, we add the magnitude of the object spectrum, so that

$$B_{kl} = |F_{kl}|. \tag{29}$$

The last bias we will examine halfwave rectifies the signal. The resulting bias is

$$B_{kl} = \begin{cases} 0, & \text{if } \hat{F}_{kl} \geq 0, \\ -\hat{F}_{kl}, & \text{else.} \end{cases} \tag{30}$$

After the bias is added, we scale the resulting spectrum to obtain a real-valued function which lies between 0 and 1. For the first three biases, $\tilde{F}_{kl}$ is scaled by $2 \max_{kl} |F_{kl}|$; for the last bias, $\tilde{F}_{kl}$ is scaled by $\max_{kl} |F_{kl}|$. The final step is pointwise binarization using the error-diffusion procedure of Floyd and Steinberg.[13] The algorithm scans the spectrum row by row and compares each sample with a fixed threshold of 0.5. If the threshold is exceeded, the sample is set to 1.0; otherwise, it is set to 0.0. The resulting quantization error is then distributed to neighboring samples that have not yet been binarized. The values of these sample points are modified, so that the error just made tends to be corrected when these points are binarized. For example, if the sample value is 0.7, the binary value will be set to 1.0. The error will then be +0.3; and the neighboring samples will be reduced by a total of 0.3. The error is distributed according to the coefficients in a weighting matrix which sum to 1.0.

As previously noted, when only the magnitude of the reconstructed image is of interest, traditional methods assign specific phase codes to the zero-phase object to reduce its spectral dynamic range. We implemented a random phase code uniformly distributed over the interval $[0,2\pi)$ for both the projection and error diffusion methods. To maintain consistency with the case where both the magnitude and phase of the reconstructed image are of interest, the error measure $\bar{e}_{\text{rms}}$ is still normalized by the maximum amplitude of the

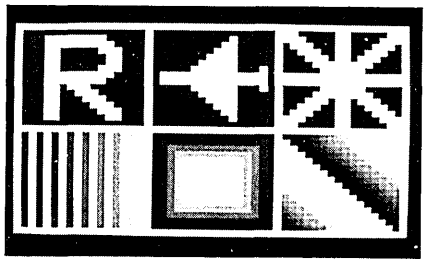

Fig. 3. Set of 16 × 16 objects used to generate experimental results.

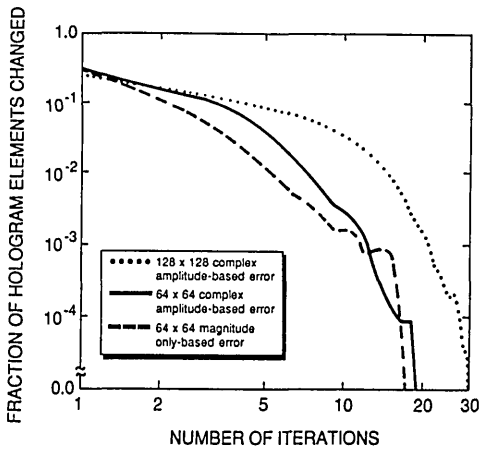

Fig. 4. Fraction of addressable hologram points changed during each iteration of the DBS algorithm.

zero-phased object scaled to have a peak spectral amplitude of unity.

VI. Experimental Results

The projection, error diffusion, and DBS methods were used to synthesize computer-generated holograms for the six 16 × 16 objects shown in Fig. 3. For each method, we examine three cases: (1) hologram size 64 × 64, complex amplitude-based error; (2) hologram size 128 × 128, complex amplitude-based error; and (3) hologram size 64 × 64, magnitude only-based error. Quantitative evaluation of the methods was based on the calculation of the error and efficiency measures described in Sec. II for digital reconstructions. Unless stated otherwise, results are averaged over all six objects. A more subjective assessment of reconstruction quality is made by visual examination of the digitally reconstructed images.

We begin by investigating the performance of the DBS algorithm as a function of the number of iterations executed. Figure 4 shows the fraction of addressable hologram points changed during each iteration. Initially, this fraction is quite large; but it decreases nearly monotonically with each iteration. The rate of decrease is significantly slower with the larger 128 × 128-point hologram and somewhat greater with the magnitude only-based error measure. The irregular behavior during the last few iterations is due to the small number of points being changed. The rms reconstruction error $\bar{e}_{\text{rms}}$ and the binarization efficiency η_{bin} vs the number of iterations completed are shown

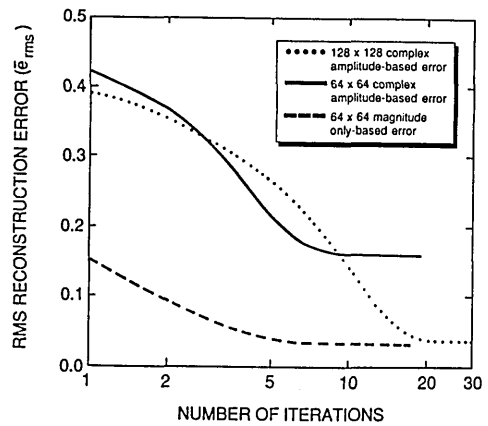

Fig. 5. Normalized rms error after each iteration of the DBS algorithm.

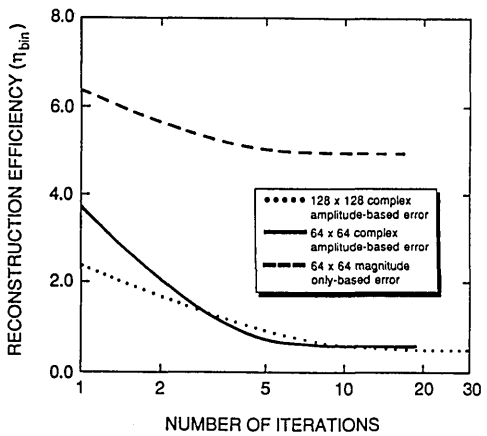

Fig. 6. Binarization efficiency after each iteration of the DBS algorithm.

in Figs. 5 and 6. The steady decrease in $\bar{e}_{\text{rms}}$ with each iteration is accompanied by a corresponding decrease in η_{bin}, reflecting the trade-off between reconstruction error and reconstruction efficiency. In the magnitude only-based error case, the overall decrease in reconstruction error and efficiency is significantly less than that for the other two cases.

The performance of the DBS algorithm described above is relatively independent of the initial hologram configuration. As discussed in Sec. III, using different random seeds to generate the initial hologram configuration has little effect on the resultant rms reconstruction errors. This conclusion is based on data presented in Table I, which shows the effect of ten random seeds on error and efficiency measures for both the complex amplitude-based error case and the magnitude only-based error case. Values are averaged over two objects, and the hologram size is 64 × 64. The reconstruction efficiency also proves to be relatively insensitive to changes in the random seeds.

Based on results obtained by Hauck and Bryngdahl[12] in optimizing the coefficients of the error diffusion weighting matrix, we applied a 1-D version of the error diffusion algorithm in the v direction. Figures 7 and 8 show rms reconstruction errors $\bar{e}_{\text{rms}}$ and recon-

Table I. Effect of Random Seeds on DBS

Error based on:	Complex Amplitude	Magnitude-only
Error		
$\langle \bar{e}_{rms} \rangle$ †	0.139	0.030
$\sigma_{\bar{e}_{rms}} / \langle \bar{e}_{rms} \rangle$	0.020	0.032
$\min(\bar{e}_{rms})$	0.134	0.028
$\max(\bar{e}_{rms})$	0.143	0.031
Efficiency		
$\langle \eta_{bin} \rangle$ ‡	0.465	4.237
$\sigma_{\eta_{bin}} / \langle \eta_{bin} \rangle$	0.022	0.037
$\min(\eta_{bin})$	0.445	4.023
$\max(\eta_{bin})$	0.478	4.496

† Normalized rms error averaged over two objects with 10 random seeds each.
‡ Binarization efficiency averaged over two objects with 10 random seeds each.

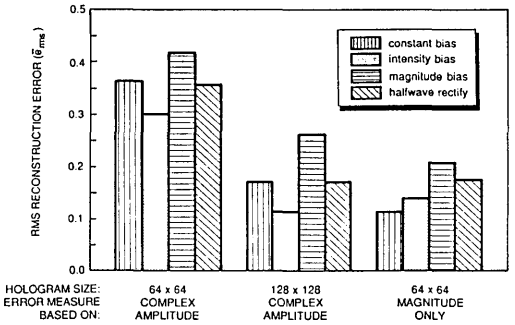

Fig. 7. Normalized rms error for each of the four biases implemented in the error diffusion algorithm.

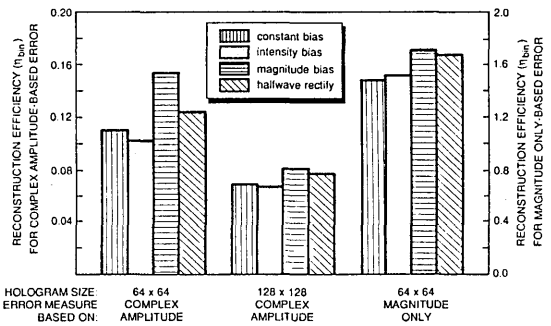

Fig. 8. Binarization efficiency for each of the four biases implemented in the error diffusion algorithm.

struction efficiencies η_{bin} for each of the four biases discussed in the last section. The clip level is set to unity in all cases. The intensity bias results in the lowest complex amplitude-based reconstruction errors, while the constant bias performs slightly better for magnitude only-based reconstruction error. The magnitude bias yields the highest reconstruction efficiency for each case, but the other biases do nearly as well.

We compare the performance of the three methods in Figs. 9 and 10. The parameters used for the projection and error diffusion methods are listed in Table II. For each case, DBS produces significantly better results than either of the other two methods. Averaged over the three cases, DBS gives a 67.5% reduction in rms error compared to the projection method and a

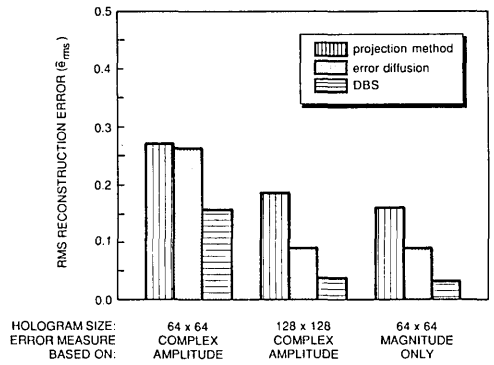

Fig. 9. Comparison of the normalized rms errors of the three methods investigated.

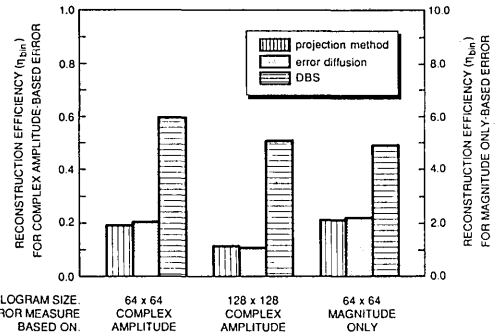

Fig. 10. Comparison of the binarization efficiencies of the three methods investigated.

55.2% reduction in rms reconstruction error compared to error diffusion. Letting the phase float decreases the error by approximately the same amount as increasing the hologram size by a factor of 2. The projection and error diffusion methods yield almost identical reconstruction efficiencies η_{bin}, which DBS increases by a factor of 3.35 averaged over the three cases. It should be noted that for the magnitude only-based error measure, the projection and error diffusion methods would perform better if the spectral phase were determined by an iterative procedure[7,14] rather than by simply assigning a random phase code.

Figures 11–15 are photographs of hologram transmittance functions, corresponding diffraction patterns, and digitally reconstructed images. The letter R is the object for the top row; the diagonal test pattern is the object for the bottom row. These objects are

Table II. Parameters Used for Generating Results of Figs. 7 and 8

Hologram size:	64 × 64	128 × 128	64 × 64
Error based on:	Complex Amplitude	Complex Amplitude	Magnitude-only
Projection Method			
Hologram Cell Size	4	8	4
Clip Level	0.5	0.75	0.75
Error Diffusion			
Bias*	Intensity	Intensity	Constant
Clip Level	0.5	0.75	0.75

* Bias that minimized $\bar{e}_{rms}$ for each case.

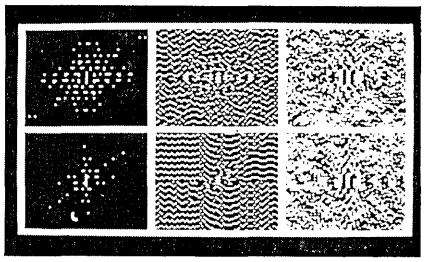

Fig. 11. Binary transmittance functions of 64 × 64 holograms for the complex amplitude-based error criterion. The left column was synthesized by the projection method, the middle column by error diffusion, and the right column by DBS. The letter R is the object for the top row; the diagonal test pattern is the object for the bottom row.

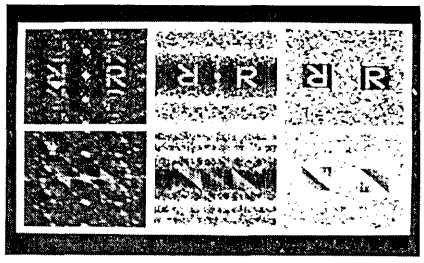

Fig. 12. Diffraction patterns of 64 × 64 holograms for the complex amplitude-based error criterion. The left column was synthesized by the projection method, the middle column by error diffusion, and the right column by DBS. The letter R is the object for the top row; the diagonal test pattern is the object for the bottom row.

Fig. 13. Digitally reconstructed images of 64 × 64 holograms for the complex amplitude-based error criterion. The left column was synthesized by the projection method, the middle column by error diffusion, and the right column by DBS. The letter R is the object for the top row; the diagonal test pattern is the object for the bottom row.

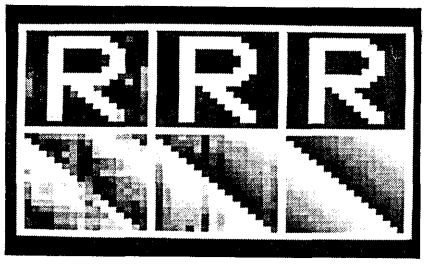

Fig. 14. Digitally reconstructed images of 128 × 128 holograms for the complex amplitude-based error criterion. The left column was synthesized by the projection method, the middle column by error diffusion, and the right column by DBS. The letter R is the object for the top row; the diagonal test pattern is the object for the bottom row.

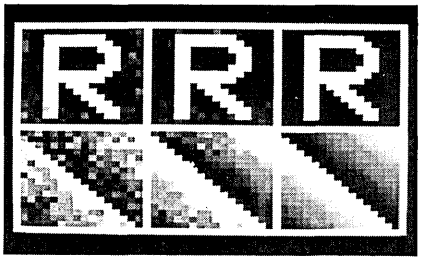

Fig. 15. Digitally reconstructed images of 64 × 64 holograms for the magnitude only-based error criterion. The left column was synthesized by the projection method, the middle column by error diffusion, and the right column by DBS. The letter R is the object for the top row; the diagonal test pattern is the object for the bottom row.

shown in Fig. 3. The left column is produced by the projection method, the middle column is produced by error diffusion, and the right column is produced by DBS. Figure 11 shows the binary hologram transmittance functions for the 64 × 64 complex amplitude-based error case. The projection holograms have the smallest number of bright elements, which are concentrated around the origin. The error diffusion holograms have a wormy appearance typical of the technique. The DBS holograms have the highest average transmittance and are the most unstructured. Figure 12 shows the diffraction patterns generated from the holograms in Fig. 11. Before being displayed, the diffraction patterns are clipped at the level $\max_{mn}|\tilde{f}_{mn}|$ corresponding to the normalization of the rms error in Eq. (11), so that the reconstructed images will be more visible. The diffraction patterns produced by the projection method have multiple reconstruction orders, while the other methods produce a single reconstruction order. Since error diffusion encodes the spectrum of a zero padded object, the diffraction pattern outside the desired reconstruction region is predominately dark except for the presence of noise clouds. Since DBS places no restrictions on the diffraction pattern outside the desired reconstruction, the entire pattern is noisy except for the reconstructed image and its Hermitian twin. Note that the reconstructed images are brightest for DBS, since they have the highest reconstruction efficiencies. Figures 13–15 show digitally reconstructed images for each of the three cases. The images are normalized by their maximum values so that differences in efficiency are not shown. Reconstruction quality improves as we move from left to right. By far the highest quality reconstructions are generated by the DBS algorithm using a 128 × 128 hologram size or magnitude only-based error measure.

Figures 16–20 are pictures of actual optical reconstructions of the letter R from holograms generated by DBS. The holograms were written on a 128 × 128

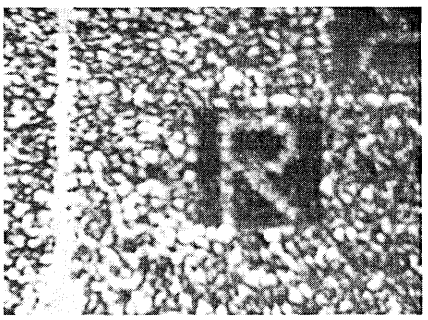

Fig. 16. Optical reconstruction of a 64 × 64 hologram synthesized by DBS under the complex amplitude-based error criterion.

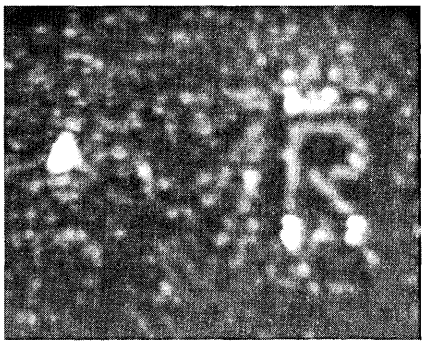

Fig. 17. Optical reconstruction of 128 × 128 hologram synthesized by DBS under the complex amplitude-based error criterion.

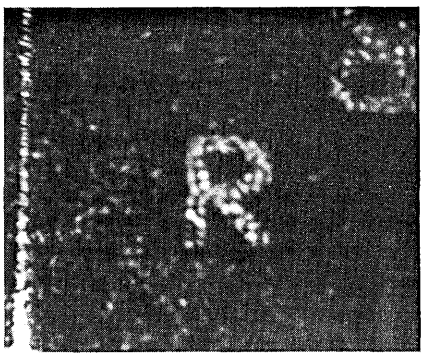

Fig. 18. Optical reconstruction of 64 × 64 hologram synthesized by DBS under the magnitude only-based error criterion.

element Litton magnetooptic[15] device used as a binary spatial light modulator under the control of a host DEC 11/73 processor. The Litton device was positioned at the front focal plane of a Fourier transform lens and illuminated with a collimated beam from a green Ar-ion laser. A video camera target was positioned at the back focal plane of the lens. The reconstructed images were acquired with a Datacube Video Acquisition and Display board.

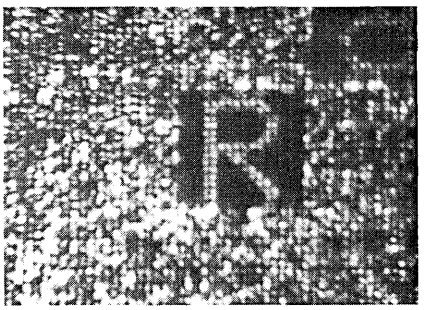

Fig. 19. Optical reconstruction of 64 × 64 hologram synthesized by DBS under the complex amplitude-based error criterion replicated 2 × 2 times on the Litton device.

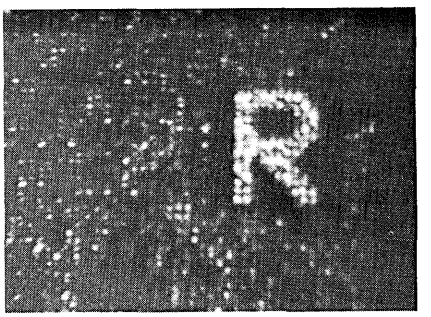

Fig. 20. Optical reconstruction of 64 × 64 hologram synthesized by DBS under the magnitude only-based error criterion replicated 2 × 2 times on the Litton device.

Compared to the corresponding digitally reconstructed images in Figs. 13–15, the optical reconstructions are significantly degraded by speckle. For the image corresponding to the magnitude-only error criterion shown in Fig. 15, this is to be expected. However, for the other two images, the phase is nearly zero at the sample points within the region of the desired reconstruction. Thus the speckle is entirely due to leakage from the sample points outside the desired reconstruction region, which are not controlled by the DBS algorithm.

That the speckle is indeed due to the interaction between hologram points may be verified for the 64 × 64 holograms by replicating them 2 × 2 times on the spatial light modulator. This effectively samples the optical reconstruction and does suppress the speckle as shown in Figs. 19 and 20.

VII. Summary and Conclusions

We have introduced a new technique for synthesizing computer-generated holograms called direct binary search. Unlike traditional approaches that encode

the complex-valued spectrum into a binary real-valued transmittance function, DBS works directly with the hologram and desired reconstruction. The DBS algorithm begins by generating a random binary hologram configuration which is subsequently scanned in lexicographic order. As each point is addressed, its value is inverted, and the resulting mean square reconstruction error is computed. If a reduction in reconstruction error is noted, the inverted hologram value at that point is retained; otherwise the original value is restored. This process is repeated until the entire hologram is scanned without retaining an inversion. The resultant hologram is optimal in the sense that it has the minimum mean squared reconstruction error of all the intermediate hologram configurations evaluated in the iterative search, and no single-point inversion will further decrease the reconstruction error. Experimental results indicate that the algorithm is relatively insensitive to the initial random configuration and that simulated annealing, which converges much more slowly than DBS, does not produce a reconstructed image of significantly higher quality.

The main drawback of the DBS algorithm is its computational intensity, which is primarily due to the mean squared error calculation done MN times in each iteration, where MN is the number of addressable hologram points. As a result, DBS, as presently implemented, may not be a practical method for generating extremely large holograms or holograms for which the desired reconstruction region is large. However, the mean squared error calculation is well suited for an array processor. This could significantly reduce run time.

For hologram sizes appropriate for low resolution devices such as spatial light modulators, experimental results have been very promising for DBS. Compared to holograms generated by the more established techniques we investigated, which were methods based on projections and error diffusion, the DBS holograms resulted in substantially lower reconstruction errors and higher reconstruction efficiencies. Since the DBS method does not manipulate the object spectrum, no time need be spent optimizing such parameters as the clip level, hologram cell size, bias, or diffusion weights. An additional advantage of DBS is its generality. Although in this paper we considered only Fourier transform holograms, the DBS method can easily be extended to the synthesis of any portion of a wavefront at any distance from the arbitrarily illuminated hologram plane according to any analytic criterion of optimality. The DBS algorithm should also be capable of compensating for faulty elements that might be present in the spatial light modulator.

To reduce leakage that can cause speckle in the reconstructed image, it may be desirable to base the error criterion for DBS on image samples calculated in the reconstruction region at a spacing less than the Nyquist interval. This will, of course, entail additional computation.

This work was supported by the National Science Foundation under grant ESC-8419997 and by the U. S. Department of Energy.

References

1. J. P. Allebach, N. C. Gallagher, and B. Liu, "Aliasing Error in Digital Holography," Appl. Opt. **15,** 2183 (1976).
2. S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by Simulated Annealing," Science **220,** 671 (1983).
3. P. Carnevali, L. Coletti, and S. Patarnello, "Image Processing By Simulated Annealing," IBM J. Res. Dev. **29,** 569 (1985).
4. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," J. Chem. Phys. **21,** 1087 (1953).
5. J. P. Allebach, "Design of Antialising Patterns for Time-Sequential Sampling of Spatiotemporal Signals," IEEE Trans. Acoust. Speech Signal Process. **ASSP-32,** 137 (1984).
6. J. A. Bucklew and B. E. A. Saleh, "Theorem for High-Resolution High Contrast Image Synthesis," J. Opt. Soc. Am. A **2,** 1233 (1985).
7. J. P. Allebach and B. Liu, "Minimax Spectrum Shaping with a Bandwidth Constraint," Appl. Opt. **14,** 3062 (1975).
8. W. J. Dallas, "Computer-Generated Holograms," in *The Computer in Optical Research*, B. R. Frieden, Ed. (Springer-Verlag, New York, 1980), pp. 291–366.
9. J. P. Allebach, "Representation-Related Errors in Binary Digital Holograms: A Unified Analysis," Appl. Opt. **20,** 290 (1981).
10. R. H. Squires and J. P. Allebach, "Digital Holograms: A Guide to Reducing Quantization and Phase Encoding Errors," Proc. Soc. Photo-Opt. Instrum. Eng. **437,** 12 (1983).
11. N. C. Gallagher and J. A. Bucklew, "Nondetour Phase Digital Holograms: An Analysis," Appl. Opt. **19,** 4266 (1980).
12. R. Hauck and O. Bryngdahl, "Computer-Generated Holograms with Pulse-Density Modulation," J. Opt. Soc. Am. A **1,** 5 (1984).
13. R. W. Floyd and L. Steinberg, "An Adaptive Algorithm for Spatial Greyscale," Proc. Soc. Inf. Disp. **17,** (1976).
14. N. C. Gallagher and B. Liu, "Method for Computing Kinoforms that Reduces Image Reconstruction Error," Appl. Opt. **12,** 2328 (1973).
15. W. E. Ross, K. M. Snapp, and R. H. Anderson, "Fundamental Characteristics of the Litton Iron Garnet Magneto-Optic Spatial Light Modulator," Proc. Soc. Photo-opt. Instrum. Eng. **388,** 55 (1983).