

Iterative Fourier transform algorithm with regularization for the optimal design of diffractive optical elements

Hwi Kim, Byungchoon Yang, and Byoungho Lee

National Research Laboratory of Holography Technologies, School of Electrical Engineering, Seoul National University, Kwanak-Gu Shinlim-Dong, Seoul 151-744, Korea

Received February 4, 2004; revised manuscript received June 11, 2004; accepted June 15, 2004

There is a trade-off between uniformity and diffraction efficiency in the design of diffractive optical elements. It is caused by the inherent ill-posedness of the design problem itself. For the optimal design, the optimum trade-off needs to be obtained. The trade-off between uniformity and diffraction efficiency in the design of diffractive optical elements is theoretically investigated based on the Tikhonov regularization theory. A novel scheme of an iterative Fourier transform algorithm with regularization to obtain the optimum trade-off is proposed. © 2004 Optical Society of America

OCIS codes: 050.1970, 090.1970, 140.3300.

1. INTRODUCTION

A fundamental objective of diffractive optics is the optimal design of a diffractive optical element (DOE).^{1–3} A DOE is a device that forms a target diffraction image in a specified image plane (output plane) through diffracting an incident optical wave in a pre-given DOE plane (input plane). A general quadratic phase system as well as free space can be placed between the DOE plane and the image plane.⁴ In most cases, a DOE is used as a phase-only element, which modulates only the phase of the incident optical wave without disturbing the amplitude. A phase-only element is in practice attractive for its high transmission efficiency and simple fabrication.

The most important task in the design of a phase-only DOE is to find an optimal phase profile to be given to the optical wave by the DOE. However, the design problem of the DOE is difficult to solve, since it is basically a non-linear ill-posed inverse problem.⁵ Though mathematical fundamentals of the DOE design have been investigated continuously during recent decades, a complete understanding has not been reached yet. The origin of the difficulty is the inherent ill-posedness of the DOE design problem. Ill-posedness means that (1) a solution does not exist, (2) the solution may not be unique, or (3) the solution may be unstable.^{5,6} In the DOE design problem, it is seen that a band-limited phase profile does not exist, which indicates nonexistence of the solution,^{7,8} and many quite different candidates for the DOE phase profile may generate the same diffraction image.

In the DOE design, several degrees of freedom must be adopted to well pose the design problem.^{7,9} The amplitude and phase freedoms are essential in the DOE design irrespective of practical design methods.⁷ The design problem can be described as a phase retrieval problem of the DOE to form a target diffraction image in the image plane with the amplitude and phase freedoms. In the scalar domain, the phase retrieval problems can be solved

approximately by various numerical methods. Many efficient methods to find the phase profile of a DOE have been studied, such as the iterative Fourier transform algorithm (IFTA),^{9–14} the genetic algorithm, etc.^{15,16} The most preferred algorithm among them is the IFTA. The amplitude and phase freedoms were adopted in Wyrowski's IFTA scheme to overcome the stagnation effect.⁹

In addition, the relaxation parameter is an important freedom for improving the performance of the IFTA, since the relaxation parameter can control the convergence rate of the IFTA. An example of the parametric version of the IFTA is the input-output method.^{10,11} Recently the IFTA with a variable relaxation parameter for steepest-descent convergence was reported.¹²

Kotlyar *et al.*, in particular, formulated a version of the IFTA based on the Tikhonov regularization scheme.¹³ Through one of the variational methods, the Landweber iteration method,¹⁷ the IFTA is derived from the least-squares object functional with a regularization term. In that scheme, the variation of the iterate is selected to maximize the negative variation of the object functional. Their IFTA version can be viewed as a refined generalization of almost all historic variants of the IFTA, and it includes all key concepts of the IFTA, such as (1) parameterization by a relaxation parameter to control the convergence rate, (2) extra degrees of freedom for exploiting amplitude freedom outside the signal area in the image plane, i.e., the noise area, and phase freedom in the whole area of the image plane to eliminate the stagnation effect, and (3) regularization. In particular, the concept of regularization will be discussed in a precise manner in this paper.

On the other hand, the quality of a DOE can be estimated with several evaluation features of the generated diffraction image, such as mean square error (MSE), diffraction efficiency, and uniformity. They are defined, respectively, as

$$\text{Uniformity} = \frac{|F|_{\max} - |F|_{\min}}{|F|_{\max} + |F|_{\min}}, \quad (1a)$$

$$\text{Diffraction efficiency} = \frac{\int_S |F|^2 dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F|^2 dx dy} \times 100 (\%), \quad (1b)$$

$$\text{MSE} = \int_S ||F| - F_0|^2 dx dy, \quad (1c)$$

where F and F_0 denote the complex amplitude of the optical signal and the target image, respectively, and S indicates the signal area in the image plane. In particular, the uniformity designates flatness or smoothness of the intensity distribution of the diffraction image. In this paper, uniform target images are considered. It is seen that uniformity is better as its value gets smaller. The diffraction efficiency is the ratio of the power focused on the signal area to the total power of the incident optical wave. The MSE as defined in Eq. (1c) is the integral of the MSE of the resulting diffraction image with respect to the target image. As will be shown, the conventional IFTA is made to minimize only the MSE, since the form of the generating functional of the conventional IFTA is just the MSE.

Generally the minimization of the MSE through the IFTA results in a reliable solution with fair uniformity and diffraction efficiency. However, it is well-known that there is a direct trade-off between uniformity and diffraction efficiency. A common understanding about the trade-off is that lower efficiency allows more noise outside the signal area, ensures more amplitude freedom, leads to reduced errors, and overcomes the stagnation and improves uniformity within the signal region. However, through such an understanding, we can only get an intuitive explanation about the trade-off between uniformity and diffraction efficiency and cannot obtain any clue on the concept of the optimum trade-off. The optimum trade-off means the minimum trade-off, which designates the capability of obtaining the best uniformity for a specific value of diffraction efficiency. For the optimal design, the optimum trade-off needs to be obtained. The conventional IFTA tries to minimize the MSE value, but the minimization of only the MSE does not guarantee the best uniformity of the diffraction image, since the mathematical properties of the MSE and the uniformity are so different. Hence an elaborate innovation of the conventional IFTA is necessary to minimize the trade-off, namely, to improve uniformity for a specific goal of diffraction efficiency. The ultimate objective of the study on optimization techniques is surely to find a global optimum solution, but the ill-posedness of our problem encourages us to devise effective ways to reduce the trade-off and, furthermore, reach the optimum trade-off.

In this paper, a novel scheme of the IFTA to minimize the trade-off between diffraction efficiency and uniformity is proposed. We do not guarantee that the trade-off obtained by the proposed technique is optimum but will

prove that the proposed scheme is effective for improving uniformity of the diffraction image.

This paper is organized as follows. In Section 2, as preliminaries, the setup of an optical system with a DOE is described, and the problem to be treated is formulated. In Section 3, a theoretical analysis of the trade-off between uniformity and diffraction efficiency of the IFTA scheme is considered, based on the Tikhonov regularization theory. In Section 4, a novel IFTA scheme for minimizing the trade-off is devised with use of the first-order Tikhonov regularization theory. In Section 5, some comments on practical implementation of the proposed algorithm are given. In Section 6, comparisons between the conventional and proposed IFTA schemes are made with the aid of numerical simulations. Finally, concluding remarks are given in Section 7.

2. OPTICAL SYSTEM WITH A DIFFRACTIVE OPTICAL ELEMENT

In general, since a phase-only DOE comprises a very thin surface relief profile on a substrate, the thin-element approximation (TEA) is used for connecting the surface relief profile with the transmittance of the DOE.¹⁸ In the thin-element approximation, the phase modulation at a spatial point is proportional to the local thickness of the optical element at that point. Although a more real situation such as multiple internal reflections inside the DOE can be considered with the IFTA,¹⁹ in this paper, for simplicity, the thin-element approximation is assumed.

In this paper, we are concerned with the design of a phase-only DOE placed in the system indicated in Fig. 1. The schematic of the optical system with a DOE assumed in this paper is presented in Fig. 1. A thin lens of focal length f and a phase-only DOE constitute the optical system. The DOE is placed in the DOE plane. Let the distance from the DOE plane to the lens and that from the lens to the image plane be d_1 and d_2 , respectively, and let the wavelength of the optical wave be λ . The incident optical wave impinges on the back side of the DOE and passes through the DOE with its phase modified. The modulated optical wave signal is transformed by the optical system to form a diffraction image in the image plane.

The IFTA is mainly considered in this paper. As shown in Fig. 1, a general quadratic phase system can be placed between the DOE plane and the image plane in the IFTA-based design. The general quadratic phase system

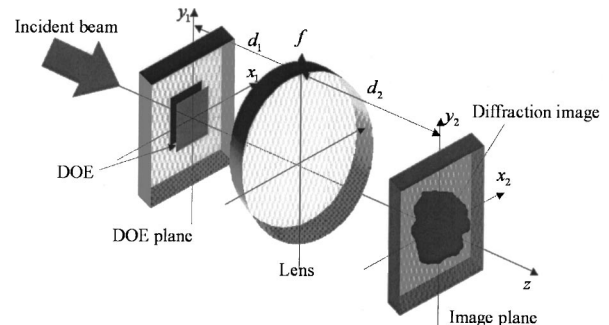


Fig. 1. Schematic of a paraxial optical system with a DOE.

is mathematically represented by the generalized Fresnel transform or the fractional Fourier transform.^{20,21} The main transform of the IFTA is extended to the fractional Fourier transform or the generalized Fresnel transform. Since the generalized Fresnel and fractional Fourier transforms can be practically implemented based on the fast Fourier transform (FFT) algorithm, the IFTA can be easily established on a digital computer. The transform of the optical system is described as the generalized Fresnel transform with (x_1, y_1) and (x_2, y_2) denoting the coordinates of a point in the DOE plane and the image plane, respectively. Then, by the linear canonical transform method,²¹ the forward Fresnel transform, denoted by $\text{Fr}(\cdot)$, can be obtained:

$$\begin{aligned} F(x_2, y_2) &= \text{Fr}[G(x_1, y_1)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x_2, y_2, x_1, y_1) G(x_1, y_1) dx_1 dy_1, \end{aligned} \quad (2)$$

where $h(x_2, y_2, x_1, y_1)$ is the propagator of the forward Fresnel transform and takes the form

$$\begin{aligned} h(x_2, y_2, x_1, y_1) &= \frac{-j}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \exp \left\{ \frac{j\pi}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \right. \\ &\quad \times \left[\left(1 - \frac{d_1}{f} \right) (x_2^2 + y_2^2) - 2(x_2 x_1 + y_2 y_1) \right. \\ &\quad \left. \left. + \left(1 - \frac{d_2}{f} \right) (x_1^2 + y_1^2) \right] \right\}. \end{aligned} \quad (3)$$

The inverse Fresnel transform, denoted by $\text{Fr}^{-1}[\cdot]$, can be expressed as

$$\begin{aligned} G(x_1, y_1) &= \text{Fr}^{-1}[F(x_2, y_2)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h^{-1}(x_1, y_1, x_2, y_2) \\ &\quad \times F(x_2, y_2) dx_2 dy_2, \end{aligned} \quad (4)$$

where $h^{-1}(x_1, y_1, x_2, y_2)$ is the propagator of the inverse Fresnel transform and takes the form

$$\begin{aligned} h^{-1}(x_1, y_1, x_2, y_2) &= \frac{j}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \exp \left\{ \frac{-j\pi}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \right. \\ &\quad \times \left[\left(1 - \frac{d_2}{f} \right) (x_1^2 + y_1^2) - 2(x_2 x_1 + y_2 y_1) \right. \\ &\quad \left. \left. + \left(1 - \frac{d_1}{f} \right) (x_2^2 + y_2^2) \right] \right\}. \end{aligned} \quad (5)$$

It is noted that when the focal length f of the lens is infinite, the propagator of the generalized Fresnel transform

leads to the free-space propagator and that when d_1 and d_2 are equal to f , the generalized Fresnel transform becomes a Fourier transform realized by a thin lens.

With the unitary condition satisfied,⁴ the forward and inverse Fresnel transforms can be implemented based on the FFT. The size of the computation grid is set to $N \times N$, where N is commonly 2^m for the FFT. The forward Fresnel transform is rewritten in the form of a Fourier transform of the signal $G(x_1, y_1)$ multiplied by the chirped phase function:

$$\begin{aligned} F(x_2, y_2) &= \frac{-j}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \exp \left[\frac{j\pi(1 - d_1/f)(x_2^2 + y_2^2)}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \right] \\ &\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \exp \left[\frac{j\pi(1 - d_2/f)(x_1^2 + y_1^2)}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \right] \right. \\ &\quad \left. \times G(x_1, y_1) \right\} \exp \left[\frac{-j2\pi(x_2 x_1 + y_2 y_1)}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \right] dx_1 dy_1. \end{aligned} \quad (6)$$

The instantaneous spatial angular frequency of the quadratic chirped phase function near the aperture edge in one direction parallel to the x_1 axis is

$$\begin{aligned} \omega_M(x_1) &= \frac{\partial}{\partial x_1} \left[\frac{\pi(1 - d_2/f)(x_1^2 + y_1^2)}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \right] \\ &= \frac{2\pi(1 - d_2/f)x_1}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]}. \end{aligned} \quad (7)$$

By the Nyquist sampling theory, the sampling spatial angular frequency ω_S is determined:

$$\omega_S > 2|\omega_M(x_1)| = \left| \frac{4\pi(1 - d_2/f)x_1}{\left[\lambda(d_1 + d_2) - \frac{\lambda d_1 d_2}{f} \right]} \right|, \quad (8)$$

where x_1 is assumed to be positive and viewed as the distance between the origin and the point $(x_1, 0)$. With Δx_1 and L_{x_1} denoting, respectively, the spatial sampling inter-

val in the DOE plane and the length of the computation grid along the x_1 direction, the following inequality should hold for Δx_1 :

$$\Delta x_1 = \frac{2\pi}{\omega_S} < \left| \frac{\lambda[(d_1 + d_2) - d_1 d_2 / f]}{2(1 - d_2 / f)x_1} \right|. \quad (9)$$

With the aid of the relation $L_{x_1} = N\Delta x_1$, and by setting $x_1 = R$ as the aperture radius of the DOE, we can obtain the relation between the upper bound of the dimension of the DOE plane, the aperture radius of the DOE, and the size N of the computation grid, that is, the number of sampling points along the horizontal or longitudinal direction:

$$L_{x_1} R < \left| \frac{\lambda(d_1 + d_2 - d_1 d_2 / f)}{2(1 - d_2 / f)} \right| N, \quad R \leq \frac{L_{x_1}}{2}. \quad (10)$$

The discretization of the inverse Fresnel transform need not be considered seriously, since the chirped phase term will be canceled out by the conjugate chirped phase term outside the integrals of Eq. (6).

In the following sections, the main topic of reducing the trade-off between uniformity and diffraction efficiency is investigated in a precise manner. We approach the problem from a theoretical point of view rather than a practical point of view. We concentrate on the IFTA itself rather than practical design issues²² of DOEs. Therefore the issue of phase quantization is not touched. The construction of the continuous phase profile is considered in this paper.

3. ANALYSIS OF TRADE-OFF BETWEEN UNIFORMITY AND DIFFRACTION EFFICIENCY

In practice, the balance of uniformity and diffraction efficiency can be controlled by a simple scaling up or down of the signal distribution in the area outside the signal area, viz., the noise area, at the step of updating the amplitude distribution in the image plane during the IFTA process. If the scaling factor is larger than 1, the amplitude freedom increases, and uniformity is improved but diffraction efficiency decreases; whereas, if the constant scaling factor is smaller than 1, the amplitude freedom decreases, so uniformity is degraded but diffraction efficiency is improved.

In this section, a theoretical analysis of the trade-off between uniformity and diffraction efficiency is contemplated based on the Tikhonov regularization theory,^{6,13,17} which is a basis for the discussion in Section 4. To analyze the relationship between uniformity and diffraction efficiency, we build an object functional to be minimized with two distinct regularization parameters α_S and α_N :

$$J(F) = \int \int_{-\infty}^{\infty} |D_S F| - F_0|^2 dx_2 dy_2 + \alpha_S \int \int_S |F|^2 dx_2 dy_2 + \alpha_N \int \int_N |F|^2 dx_2 dy_2, \quad (11)$$

where D_S is the area-limiting operator defined as

$$D_S F = \begin{cases} F & \text{for } (x_2, y_2) \in S \\ 0 & \text{for } (x_2, y_2) \notin S \end{cases} \quad (12)$$

F is the calculated complex diffraction image, F_0 is the target image, which is real, and S and N denote the signal area and the noise area, respectively. Equation (11) can be viewed as a Tikhonov functional parameterized by two regularization parameters. It should be noted that energy conservation is assumed in the design of DOEs:

$$\begin{aligned} \int \int_{-\infty}^{\infty} |F|^2 dx_2 dy_2 &= \int \int_S |F|^2 dx_2 dy_2 + \int \int_N |F|^2 dx_2 dy_2 \\ &= \int \int_{-\infty}^{\infty} |F_0|^2 dx_2 dy_2. \end{aligned} \quad (13)$$

We can derive an iterative algorithm to minimize the functional (11) by using a variational method, the Landweber iteration method.¹⁷ The variation of the functional (11) is given by

$$\begin{aligned} \delta J(F) &= 2 \int \int_{-\infty}^{\infty} \text{Re}\{[-|F_0| \exp(i\psi) + D_S F + D_S \alpha_S F \\ &\quad + (1 - D_S) \alpha_N F] \delta F^*\} dx_2 dy_2, \end{aligned} \quad (14)$$

where $\text{Re}(\cdot)$ is the real part of the complex number. According to the Landweber iteration method, we should select the variation of F to generate the maximum negative variation of δJ ; then δF takes the form

$$\begin{aligned} \delta F &= \bar{F} - F = -\tau[D_S(1 + \alpha_S)F + (1 - D_S)\alpha_N F \\ &\quad - F_0 \exp(j\psi)], \end{aligned} \quad (15)$$

where τ is the relaxation parameter and ψ is the phase distribution of F . Hence the n th iterate modified in the image plane is obtained:

$$\begin{aligned} \bar{F}_n &= F_n - \tau[D_S(1 + \alpha_S)F_n + (1 - D_S)\alpha_N F_n \\ &\quad - F_0 \exp(j\psi_n)]. \end{aligned} \quad (16)$$

To satisfy the constraints in the DOE plane, we obtain the $(n + 1)$ th diffracted field F_{n+1} by applying the error-reduction operator to Eq. (16):

$$F_{n+1} = \text{Fr } D_{\text{DOE}} \text{Fr}^{-1}(\bar{F}_n), \quad (17)$$

where the operator D_{DOE} , which expresses the surface boundary condition in the DOE plane, is given by

$$D_{\text{DOE}} G = \begin{cases} A_0 \exp[j \arg(G)] & \text{for } (x_1, y_1) \in \Omega \\ 0 & \text{for } (x_1, y_1) \notin \Omega \end{cases} \quad (18)$$

in which Ω denotes the encoding area in the DOE plane and $\arg(G)$ is the phase function of the complex function G . Equations (16) and (17) describe a generalized IFTA for the DOE design. Equation (16) can be rewritten as

$$\bar{F}_n = \begin{cases} \tau F_0 \exp(j\psi_n) + (1 - \tau - \tau\alpha_S)F_n & \text{for } (x_2, y_2) \in S \\ (1 - \tau\alpha_N)F_n & \text{for } (x_2, y_2) \notin S \end{cases} \quad (19)$$

Even if Eq. (19) is divided by a constant $(1 - \tau\alpha_S)$, the behavior of the IFTA does not change, since the IFTA is invariant for constant scaling of the signal distribution in the whole image plane. Therefore we obtain the modified form of Eq. (19):

$$\bar{F}_n = \begin{cases} \tau' F_0 \exp(j\psi_n) + (1 - \tau')F_n & \text{for } (x_2, y_2) \in S \\ \beta F_n & \text{for } (x_2, y_2) \notin S \end{cases} \quad (20)$$

where

$$\tau' = 1 - \tau\alpha_S, \quad (21)$$

$$\beta = \frac{1 - \tau\alpha_N}{1 - \tau\alpha_S}. \quad (22)$$

If the two regularization parameters are equal ($\alpha_S = \alpha_N$), we can see that $\beta = 1$ and that Eq. (20) leads to the versions of the IFTA of Kotlyar *et al.*¹³ and Wyrowski.⁹ As stated above, the practical heuristic to improve uniformity at the cost of lower diffraction efficiency is to make β slightly larger than 1. Conversely, if β is smaller than 1, diffraction efficiency is improved but uniformity is degraded.

We can see the difference between the two conditions $\beta > 1$ and $\beta < 1$ in the structure of the object functional (11). Let us consider the object functional (11) again. We can see that the functional is the sum of three integrals, which can be viewed, respectively, as the MSE estimated over the signal area, the total energy allocated to the signal area weighted by α_S , and the total energy allocated to the noise area weighted by α_N . It is confirmed that the conditions $\beta > 1$ and $\beta < 1$ correspond to $\alpha_S > \alpha_N$ and $\alpha_N > \alpha_S$, respectively. In the former case, the IFTA will try to further allocate signal energy to the noise area to minimize the functional (11), which means decreasing diffraction efficiency. On the contrary, in the latter case, the IFTA will try to allocate signal energy to the signal area to minimize the functional (11), which means increasing diffraction efficiency. Next, the object functional (11) can be rearranged with a little manipulation as

$$J(F) = (\alpha_S + 1) \int \int_S \left(|F| - \frac{F_0}{\alpha_S + 1} \right)^2 dx_2 dy_2 + \alpha_N \int \int_N |F|^2 dx_2 dy_2 + \frac{\alpha_S}{\alpha_S + 1} \int \int_S F_0^2 dx_2 dy_2. \quad (23)$$

The newly arranged functional (23) is the sum of three integrals. The first term is considered the MSE of the signal distribution F with respect to the target signal down-scaled by $(\alpha_S + 1)^{-1}$. The second term is the total energy allocated to the noise area. And the third integral is the total energy of the incident optical wave multiplied by a constant $\alpha_S (\alpha_S + 1)^{-1}$. We observe that the weight factor of the first integral is $\alpha_S + 1$. Therefore as α_S increases, the weight factor of the MSE between the signal distribution and the scaled target signal function increases and the IFTA makes an exertion to minimize the first integral further. It is comprehensible that the minimization of this MSE will lead to the improvement of uniformity of the signal distribution in the signal area, since an even distribution of the signal will minimize the MSE mathematically. The increase of the amount of the energy allocated to the noise area induces the decrease in diffraction efficiency.

However, it is known that the minimization of MSE alone does not guarantee the best uniformity of the diffraction signal, since the functional structures of MSE and uniformity are correlated weakly. Observing the functional structures of MSE and uniformity defined above, we can think that the MSE in the case in which all the samples of F except one exactly match the target values and the excluded sample has a relatively big difference from the target value can be smaller than that of the case in which all the samples of F have a relatively small difference from the corresponding target values. Then the uniformity of the former case is worse than that of the latter case. Conversely, it is easily conceived that the optimization of only the uniformity will not guarantee a decrease in the MSE. That is, uniformity and MSE are only weakly correlated. Recently, based on a similar analysis, a regularization technique, the so-called adaptive regularization parameter distribution (ARPD), was devised to strongly correlate uniformity with MSE to alleviate the trade-off between uniformity and diffraction efficiency.¹⁴ The ARPD $\alpha_S(x_2, y_2)$ is defined as

$$\alpha_S(x_2, y_2) = \frac{2\gamma}{\pi} \tan^{-1} \left[\frac{|F(x_2, y_2)| - F_0(x_2, y_2)}{F_0(x_2, y_2)} \right] + \gamma - 1, \quad (24)$$

where γ is a tuning parameter. The ARPD of Eq. (24) is substituted into the IFTA scheme (19) and results in a modified form of the IFTA:

$$\bar{F}_n = \begin{cases} \tau F_0 \exp(j\psi_n) + \left(1 - \tau - \tau \left\{ \frac{2\gamma}{\pi} \tan^{-1} \left[\frac{|F(x, y)| - F_0(x, y)}{F_0(x, y)} \right] + \gamma - 1 \right\} \right) F_n & \text{for } (x, y) \in S \\ F_n & \text{for } (x, y) \notin S \end{cases} \quad (25)$$

With use of the IFTA scheme with the ARPD, uniformity can be improved more than with the conventional IFTA for the same diffraction efficiency.¹⁴ However, we make an effort to refine the regularized scheme for the IFTA based on the first-order Tikhonov regularization theory so as to reduce the trade-off more.

4. ITERATIVE FOURIER TRANSFORM ALGORITHM WITH THE ADAPTIVE REGULARIZATION PARAMETER DISTRIBUTION AND THE FIRST-ORDER TIKHONOV REGULARIZATION

Our objective in this paper is to maximally mitigate the disadvantage of the trade-off between uniformity and diffraction efficiency in the IFTA-based design of DOEs. In this section, based on the first-order Tikhonov regularization theory and the analysis of the trade-off of Section 3, a novel scheme of the IFTA to reduce the trade-off more is proposed. The validity of the proposed algorithm will be proved with numerical simulations in Section 5.

First, we contrive a new object functional to generate the IFTA with the aid of the first-order Tikhonov regularization theory. It should be recalled that the object functional (11) generates the IFTA scheme through the variation method. In the same manner, we will derive an iterative algorithm from the proposed object functional. As stated in Section 3, the minimization of only the MSE does not guarantee the best uniformity of the diffraction signal in the conventional IFTA scheme. However, we think that if an appropriate functional form for the measure of uniformity exists, it is possible to devise a new IFTA to directly optimize the uniformity during the iteration process. Hence we need a direct measure of uniformity having an analytic form that can be manipulated easily in the discrete domain and combined with the object functional. Furthermore, it is desirable that the measure be independent of the diffraction efficiency. Fortunately, we can get such a measure of uniformity from the generalization of the objective functional (11) with the help of the first-order Tikhonov regularization theory. At this stage, we move consideration of the ARPD technique back to the end of this section. Hence let us assume that the regularization parameter α_S in Eq. (11) is a constant and not a distribution.

The sum of the second and third integrals of Eq. (11) can be viewed as a W_2^0 Sobolev space norm (whose subscript and superscript stand for space dimension and derivative order, respectively), seen in the conventional Tikhonov regularization scheme except that here it has two distinct regularization parameters. Let this form be referred to as the zeroth Tikhonov regularization functional. As a course of mathematical generalization to higher-order regularization, we can build an objective functional with a W_2^1 Sobolev space norm with three distinct regularization parameters:

$$\begin{aligned} J(F) = & \int_{-\infty}^{\infty} |D_S|F| - F_0|^2 dx_2 dy_2 + \alpha_S \int_S |F|^2 dx_2 dy_2 \\ & + \alpha_N \int_N |F|^2 dx_2 dy_2 + \alpha_D \int_S [(\partial_{x_2}|F|)^2 \\ & + (\partial_{y_2}|F|)^2] dx_2 dy_2. \end{aligned} \quad (26)$$

The proposed functional (26) is made by adding the integral of the square of the first-order partial derivative terms of the signal amplitude $|F|$ through the third regularization parameter α_D to the integrand of the second integral of the functional (11). The integrand of the additional integral is independent of the absolute magnitude distribution of the signal $|F|$ but is determined by the derivative of the magnitude distribution of the signal. Its integrand, $D_S[(\partial_{x_2}|F|)^2 + (\partial_{y_2}|F|)^2]$, is considered a direct and sensitive measure of the roughness of the two-dimensional signal amplitude distribution $|F|$ in the signal area of the image plane. We can see that a critical point is the dependency on diffraction efficiency of the first integral and the newly added fourth integral. The integrand of the additional fourth integral is almost independent of the ratio of the assigned energy between the signal area and the noise area, i.e., diffraction efficiency. Hence the fourth integral of Eq. (26) is also independent of the diffraction efficiency. On the contrary, the first integral of Eq. (26), MSE, is directly dependent on the ratio of the assigned energy between the signal area and the noise area and can be regarded as only an indirect measure of the roughness of the signal amplitude distribution. Therefore we can expect that the fourth term of Eq. (26) will improve uniformity without considerable loss of diffraction efficiency, since there is no dependence between the proposed measure of uniformity [the fourth term of Eq. (26)] and the measure of diffraction efficiency [the second term of Eq. (26)]. The regularization parameters α_S , α_N , and α_D of Eq. (26) can be used to balance uniformity and diffraction efficiency. As a result, we strongly correlate the measure of uniformity with the object functional to be minimized.

The inherent ill-posedness of the phase retrieval problem raises the possibility that there exist several candidate solutions of the phase profile of the DOE showing the same diffraction efficiency but different uniformity. It is desirable to obtain one having the best uniformity among them, viz., the optimum trade-off. The IFTA derived from the proposed objective functional (26) is expected to produce a solution with the desired properties. Furthermore, after deriving a stable iterative algorithm generated from the functional (26), we will combine the ARPD technique with the established algorithm. To construct a stable iterative process to minimize Eq. (26), we apply the Landweber iteration method to Eq. (26). Let us rewrite Eq. (26) as

$$J(F) = T(F) + \alpha_D \int_S [(\partial_{x_2}|F|)^2 + (\partial_{y_2}|F|)^2] dx_2 dy_2, \quad (27)$$

where $T(F)$ denotes the sum of the first, second, and third integrals of Eq. (26). First, the variation of Eq. (27) should be found. This takes the form

$$\begin{aligned} \delta J(F) = & J(F + \delta F) - J(F) \\ = & \delta T(F) + \alpha_D \int_S [(\partial_x|F + \delta F|)^2 \\ & + (\partial_y|F + \delta F|)^2] - [(\partial_x|F|)^2 + (\partial_y|F|)^2] dx dy, \end{aligned} \quad (28)$$

where $\delta T(F)$ has already been derived in Section 3 in the form

$$\delta T(F) = 2 \int \int_{-\infty}^{\infty} \text{Re}\{[-F_0 \exp(i\psi) + D_S F + D_S \alpha_S F + (1 - D_S) \alpha_N F] \delta F^*\} dx dy. \quad (29)$$

In Eqs. (28) and (29) and also in the following Eqs. (30)–(32), we use the notation of x and y instead of x_2 and y_2 , respectively, for simplicity. With the assumption of phase conservation of the signal to the variation operation, the second term of the integrand of Eq. (28) is manipulated as follows:

$$\begin{aligned} & (\partial_x |F + \delta F|^2 + \partial_y |F + \delta F|^2 - (\partial_x |F|^2 - (\partial_y |F|^2) \\ & = (\partial_x |F| + \partial_x |\delta F|)^2 + (\partial_y |F| + \partial_y |\delta F|)^2 - (\partial_x |F|)^2 \\ & \quad - (\partial_y |F|)^2 \\ & = 2(\partial_x |F| \partial_x |\delta F| + \partial_y |F| \partial_y |\delta F|). \end{aligned} \quad (30)$$

We proceed to calculate the second integral of Eq. (28) with the aid of Eq. (30) by the following manipulation:

$$\begin{aligned} & \int \int_S [(\partial_x |F + \delta F|^2 + \partial_y |F + \delta F|^2) - (\partial_x |F|^2 \\ & \quad + (\partial_y |F|^2)] dx dy \\ & = \int \int_S 2(\partial_x |F| \partial_x |\delta F| + \partial_y |F| \partial_y |\delta F|) dx dy \\ & = 2 \int \int_S \partial_x |F| \partial_x |\delta F| dx dy + 2 \int \int_S \partial_y |F| \partial_y |\delta F| dx dy \\ & = \int_{S_y} \left\{ [2(\partial_x |F|) |\delta F|]_{-\partial S_x}^{+\partial S_x} - 2 \int_{S_x} (\partial_{xx} |F|) |\delta F| dx \right\} dy \\ & \quad + \int_{S_x} \left\{ [2(\partial_y |F|) |\delta F|]_{-\partial S_y}^{+\partial S_y} \right. \\ & \quad \left. - 2 \int_{S_y} (\partial_{yy} |F|) |\delta F| dy \right\} dx \\ & = -2 \int \int_S (\partial_{xx} |F| + \partial_{yy} |F|) |\delta F| dx dy \\ & = -2 \int \int_S \nabla^2 |F| \exp(i\psi) \exp(-i\psi) |\delta F| dx dy \\ & = -2 \int \int_S \nabla^2 |F| \exp(i\psi) \delta F^* dx dy \\ & = -2 \int \int_S \text{Re}[\nabla^2 |F| \exp(i\psi) \delta F^*] dx dy, \end{aligned} \quad (31)$$

where it is assumed that both $\partial_x |F|$ and $\partial_y |F|$ are equal to 0 at the boundary of the signal area, ∂S . In Section 5, it will be further explained that this boundary condition is used to preserve the edge of the signal distribution and will be implemented numerically. Combining Eqs. (29) and (31) leads to the following variation of Eq. (27):

$$\begin{aligned} \delta J(F) & = 2 \int \int_{-\infty}^{\infty} \text{Re}\{[-F_0 \exp(i\psi) + D_S F + D_S \alpha_S F \\ & \quad + (1 - D_S) \alpha_N F \\ & \quad - D_S \alpha_D \nabla^2 |F| \exp(i\psi)] \delta F^*\} dx dy. \end{aligned} \quad (32)$$

According to the Landweber iteration method, the variation of F to generate the maximum negative variation of δJ is selected; then δF takes the form

$$\begin{aligned} \delta F & = \bar{F} - F \\ & = -\tau [D_S (1 + \alpha_S) F + (1 - D_S) \alpha_N F - F_0 \\ & \quad \times \exp(j\psi) - D_S \alpha_D \nabla^2 |F| \exp(i\psi)], \end{aligned} \quad (33)$$

where τ is the relaxation parameter and ψ is the phase distribution of F . Hence the n th iterate modified in the image plane is obtained:

$$\begin{aligned} \bar{F}_n & = F_n - \tau [D_S (1 + \alpha_S) F_n + (1 - D_S) \alpha_N F_n - F_0 \\ & \quad \times \exp(j\psi_n) - D_S \alpha_D \nabla^2 |F_n| \exp(i\psi_n)]. \end{aligned} \quad (34)$$

To satisfy the constraints in the DOE plane, we obtain the $(n + 1)$ th diffracted field F_{n+1} by applying the error-reduction operation to Eq. (34):

$$F_{n+1} = \text{Fr } D_{\text{DOE}} \text{Fr}^{-1}(\bar{F}_n). \quad (35)$$

Then Eq. (34) can be rewritten as

$$\bar{F}_n = \begin{cases} \tau F_0 \exp(j\psi_n) + (1 - \tau - \tau \alpha_S) F_n + \tau \alpha_D \nabla^2 |F_n| \exp(i\psi_n) & \text{for } (x_2, y_2) \in S, \\ (1 - \tau \alpha_N) F_n & \text{for } (x_2, y_2) \notin S, \end{cases} \quad (36)$$

where it is noted that the phase value of the Laplacian is equal to that of the signal, according to the phase conservation assumption. Equations (35) and (36) describe a novel IFTA for the design of DOEs.

We can refine the obtained IFTA scheme (36) further by substituting the ARPD $\alpha_S(x_2, y_2)$ into Eq. (36) along the same line as that described in Section 3. The obtained algorithm takes the form

$$\bar{F}_n = \begin{cases} \tau F_0 \exp(j\psi_n) + \left(1 - \tau - \tau \left[\frac{2\gamma}{\pi} \tan^{-1} \left[\frac{|F(x_2, y_2)| - F_0(x_2, y_2)}{F_0(x_2, y_2)} \right] + \gamma - 1 \right] \right) F_n \\ \quad + \tau \alpha_D \nabla^2 |F_n| \exp(i\psi_n) & \text{for } (x_2, y_2) \in S, \\ (1 - \tau \alpha_N) F_n & \text{for } (x_2, y_2) \notin S \end{cases} \quad (37)$$

5. NUMERICAL IMPLEMENTATION

For computation, a finite-difference representation of Eq. (36) is obtained. First, the finite-difference representation of the Laplacian of the signal amplitude $|F|$ takes the form

$$\begin{aligned} (\nabla^2 |F_n|)(k, l) = & \frac{1}{\Delta^2} [|F_n(k+1, l)| + |F_n(k, l+1)| \\ & + |F_n(k-1, l)| \\ & + |F_n(k, l-1)| - 4|F_n(k, l)|], \end{aligned} \quad k, l = 0, 1, 2, \dots, N, \quad (38)$$

where Δ is the spatial sampling interval and (k, l) is the index pair of the computation grid. In the derivation shown in Eq. (31), it is assumed that both $\partial_x |F|$ and $\partial_y |F|$ are equal to 0 at the boundary of the signal area, ∂S . This constraint is used to conserve the sharpness of the boundary of the signal area in a practical implementation. In particular, for practical calculation of the Laplacian at a point (k, l) on the boundary of the signal area, $F_n(k, l)$ is in the signal area, but one more sample among the signal samples $F_n(s, t)$ for four points, $(s, t) = (k+1, l), (k-1, l), (k, l+1)$, and $(k, l-1)$, may

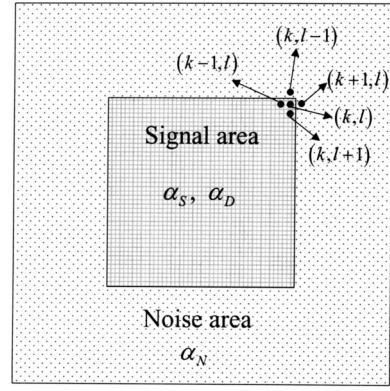


Fig. 2. Computation grid used for the DOE design. The signal area is equal to the target image of nonzero magnitude. Three regularization parameters are indicated in the corresponding regions.

dition, the replacement operation prevents local averaging over the boundary between the signal area and the noise area and conserves the edge discontinuity of the diffraction image in the image plane. Substituting Eq. (38) into Eq. (36), we have the finite-difference representation of the iterative algorithm:

$$\bar{F}_n(k, l) = \begin{cases} \tau F_0(k, l) \exp[j\psi_n(k, l)] + (1 - \tau - \tau\alpha_S) F_n(k, l) & \text{for } (x_2(k), y_2(l)) \in S \\ + \tau\alpha_D \frac{1}{\Delta^2} [|F_n(k, l)| + |F_n(k, l+1)| + |F_n(k, l)| + |F_n(k, l-1)| - 4|F_n(k, l)|] \exp[j\psi_n(k, l)], & \\ (1 - \tau\alpha_N) F_n(k, l) & \text{for } (x_2(k), y_2(l)) \notin S \end{cases} \quad (40)$$

$$F_{n+1} = \text{Fr } D_{\text{DOE}} \text{Fr}^{-1}(\bar{F}_n).$$

be outside the signal area. In this case, the value of the signal $F_n(k, l)$ replaces $F_n(s, t)$ in Eq. (38). For example, the Laplacian of the signal at the point (k, l) in Fig. 2 is calculated by the equation

$$\begin{aligned} (\nabla^2 |F_n|)(k, l) = & \frac{1}{\Delta^2} [|F_n(k, l)| + |F_n(k, l+1)| \\ & + |F_n(k, l)| \\ & + |F_n(k, l-1)| - 4|F_n(k, l)|], \end{aligned} \quad (39)$$

where the value of the signal $F_n(k, l)$ replaces $F_n(k, l-1)$ and $F_n(k+1, l)$ at the same time. The replacement operation permits the boundary condition used in the manipulation (31) to be satisfied. We can interpret the practical role of the Laplacian itself as a local averaging operation as perceived in the structure of the finite-difference representation of the Laplacian in Eq. (38). The calculation of the Laplacian of the signal is not defined in the noise area, as seen in Eq. (37). Hence the Laplacian acts to improve smoothness, i.e., uniformity, of the signal distribution in the signal area only and does not concern itself with any feature in the noise area. In ad-

The generalized Fresnel transform $\text{Fr}(\cdot)$ can be efficiently implemented by use of the FFT algorithm over the discrete domain. If the ARPD $\alpha_S(x_2(k), y_2(l))$ is substituted into the iterative algorithm (40), we have a discrete representation of the iterative algorithm (37). In our scheme, since appropriate regularization parameters cannot be determined *a priori*, it is necessary to search for the optimal regularization parameters.

6. NUMERICAL SIMULATIONS

In this section, the validity of the proposed IFTA scheme is proved with several numerical simulations. We compare the improvement of the trade-off between uniformity and diffraction efficiency of the resulting diffraction image of the proposed IFTA with that of the conventional IFTA. The obtained values of the quality factors such as uniformity and diffraction efficiency are significantly influenced by several factors, such as the initial phase profile of the DOE, the size of the computation grid, the relative size of the DOE aperture to the whole computation grid, the structure of the intermediate optical system, the

value of the relaxation parameter, and the target image itself. That is to say, the results of the IFTA are strongly dependent on the problem itself. The effects of those factors on the results of the IFTA are well understood.

Concentrating on the role of the regularization technique developed in this paper, we set all the simulation conditions and circumstances to be the same except those related to the regularization. In particular, the relative size of the DOE aperture to the whole computation grid is considered in this section together with the effect of the regularization. Usually, for beam-shaping problems, the aperture of the DOE should be imbedded on at least a double-sized computation grid,²² and the area padded by zeros in the DOE domain should be set up large enough to realize the physical situation properly. Freedom in the DOE plane cannot be fully used in that case. On the other hand, for the beam-splitting problem, the aperture of the DOE fills almost the whole computation grid. Then almost full freedom is available in the DOE plane. Here we examine two DOE design problems. The one does use the almost full freedom of the DOE plane, and the other does not. For convenience, let the former problem be referred to as design B and the latter referred to as design A. It is more difficult to obtain fair uniformity of the diffraction image at a certain level of diffraction efficiency in the case of design B than in the case of design A, since the optimization circumstance of design B is coarser than that of design A.

For simulations, the computation grid size is taken as 128×128 . Let the wavelength λ , the distances d_1 and d_2 , and the focal length f for the generalized Fresnel transform be 633 nm, 1 m, 2 m, and 1 m, respectively. The radius of the DOE is taken as 1.5 mm. The sizes of the DOE planes are set to $3 \text{ mm} \times 3 \text{ mm}$ and $3.6 \text{ mm} \times 3.6 \text{ mm}$, respectively, for each case. The target image is selected as a rectangular pattern. A normal incident optical wave is taken as the incident plane wave. The initial phase profile of the DOE is taken as quadratic. For simplicity, it is assumed that in the expression for α_D/Δ^2 of Eqs. (40), the term Δ is set to 1 and the regularization parameter α_N and the relaxation parameter τ are set to 0 and 1, respectively.

In the simulation, the trade-off between uniformity and diffraction efficiency of the resulting diffraction images obtained by the four types of the IFTA are compared. The first IFTA is the conventional IFTA of the form of Eq. (36) with α_D equal to 0. By adjusting the value of the regularization parameter α_S , we can control uniformity and diffraction efficiency. As the value of $1/(1 - \alpha_S)$ [see Eq. (22)] increases, the uniformity improves and the diffraction efficiency worsens. The second type is the IFTA with the ARPD, as represented by Eq. (25). In this scheme, uniformity and diffraction efficiency can be controlled by the tuning parameter γ . As the value of $1/(2 - \lambda)$ increases, the uniformity improves and the diffraction efficiency worsens. The third type is the IFTA with only the first Tikhonov regularization, as given by Eq. (36). The only difference between the first and the third type of the IFTA is the Laplacian of the signal amplitude multiplied by the regularization α_D . The first-order regularization term is not related to the diffraction efficiency. In the same manner as that for the first type, the

balance between uniformity and diffraction efficiency can be adjusted by the change of the regularization parameter α_S . The fourth type is the IFTA with both the ARPD and the first-order Tikhonov regularization. In the case of the fourth type, the balance between uniformity and diffraction efficiency is adjusted by the tuning parameter γ in the same way as that for the second type.

The trade-off curves obtained through 200 iterations of the four types of IFTA with the regularization parameter α_D equal to 0.2 are presented in Figs. 3(a) and 3(b) for designs A and B, respectively. Since the regularization parameter α_D cannot be determined *a priori*, the value 0.2 of α_D was searched to improve uniformity and confirm stability of the IFTA for all cases concerned. The trade-off curve indicates the change in the obtained uniformity with the change in the obtained diffraction efficiency through the adjusting process explained above. Seeing the trade-off curves, we can compare the obtainable uniformity for a goal of diffraction efficiency of the four types

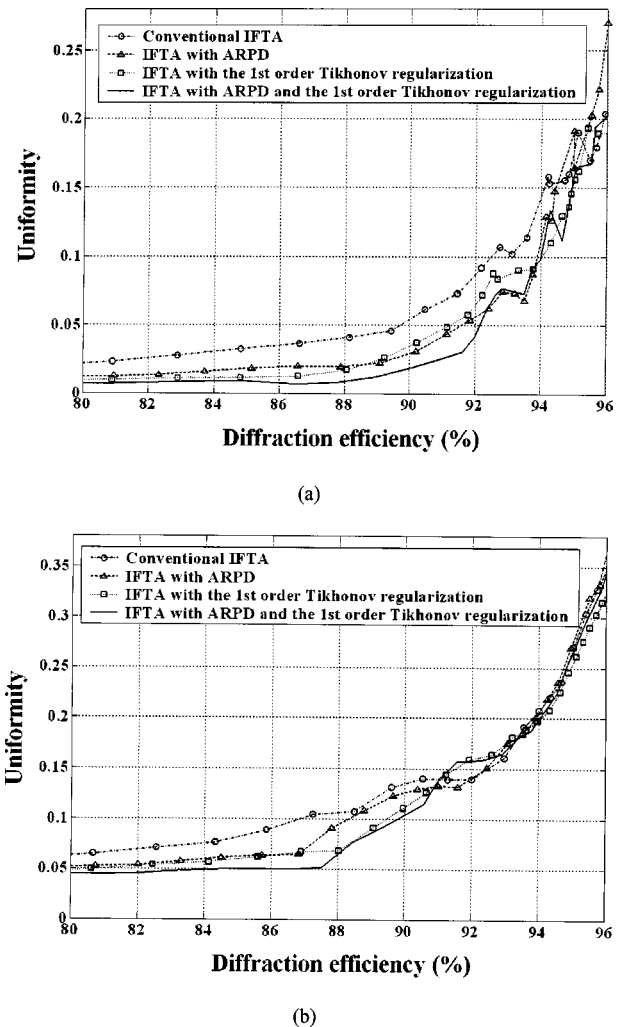


Fig. 3. Comparison of the trade-off curves of the four types of IFTA: conventional IFTA, IFTA with the ARPD, IFTA with the first-order Tikhonov regularization, and IFTA with the ARPD and the first-order Tikhonov regularization for (a) design A with full freedom in the DOE plane and (b) design B with decreased freedom. For a specific value of diffraction efficiency, the obtained uniformity is different for each case. The lowest value of uniformity indicates the best solution.

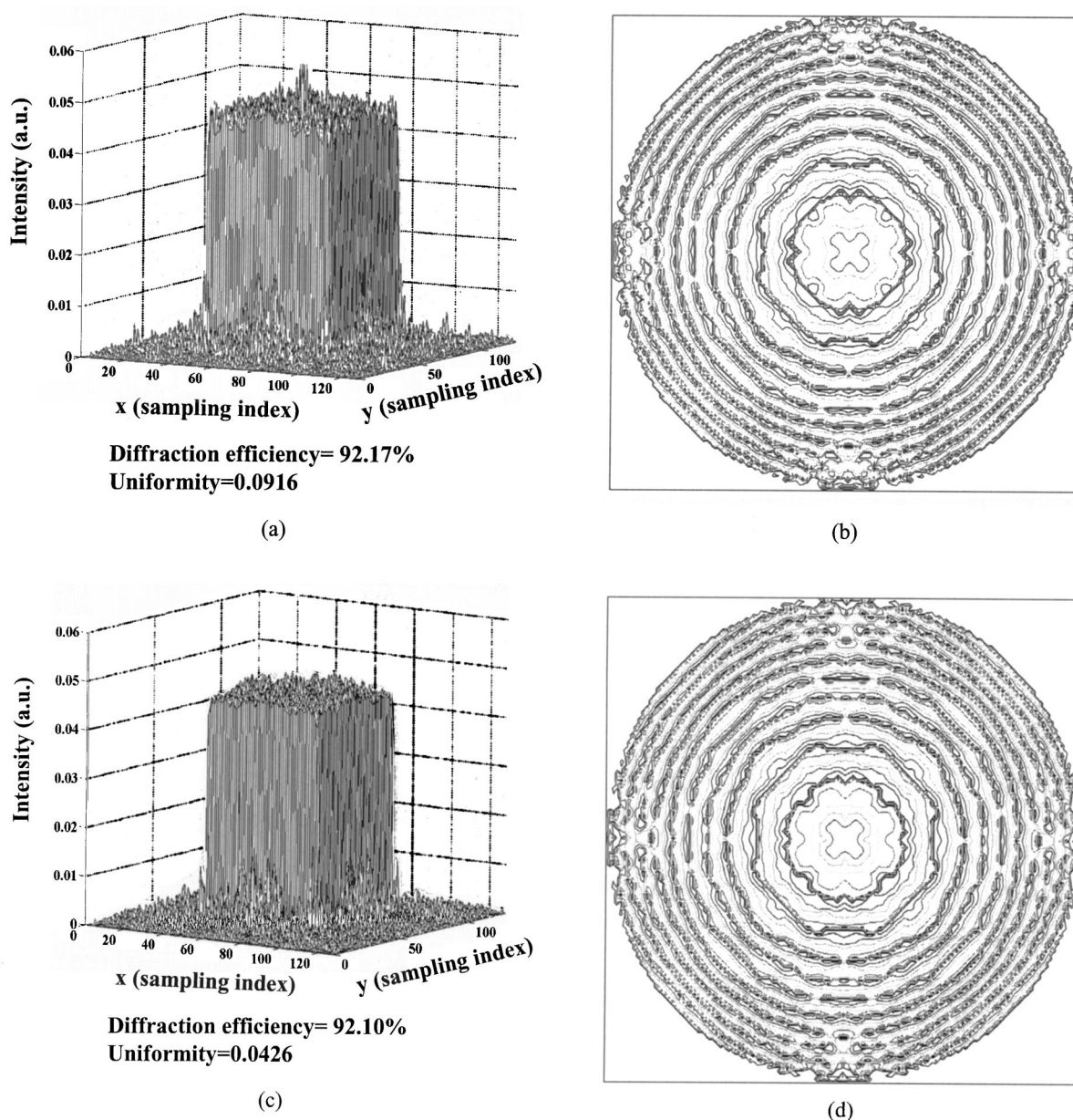


Fig. 4. Comparative illustration of the design results with use of the conventional IFTA and the proposed IFTA with the ARPD and the first-order Tikhonov regularization in the case of design A with full freedom in the DOE plane. (a) Intensity distribution of the resulting diffraction image and (b) phase profile of the DOE obtained by the conventional IFTA. (c) Intensity distribution of the resulting diffraction image and (d) phase profile of the DOE obtained by the proposed IFTA with the ARPD and the first-order Tikhonov regularization. For both IFTAs, the iteration number is 200.

of IFTA. In this paper, the IFTA giving the lowest uniformity for a specific value of diffraction efficiency is regarded as the best one. Comparing the four trade-off curves in Fig. 3, we can see that the performance of the fourth type of IFTA, i.e., IFTA with both the ARPD and the first Tikhonov regularization, is surely preeminent in the range below 92% diffraction efficiency and 91% diffraction efficiency for the cases of designs A and B, respectively. In these regions, the second and third types of IFTA compete with each other, while the four IFTAs compete for the best performance in the range above 92% and 91% of diffraction efficiency for the cases of designs A and B, respectively. Toward the upper regions, the difference among the trade-off curves becomes smaller. In the case

of design B, the difference among the values of uniformity of the four types of IFTA is indistinguishable, as seen in Fig. 3(b). From this observation, we can conjecture that the maintenance of the amplitude freedom outside the signal area in the image plane is a necessary condition for the devised regularization technique to demonstrate its effectiveness.

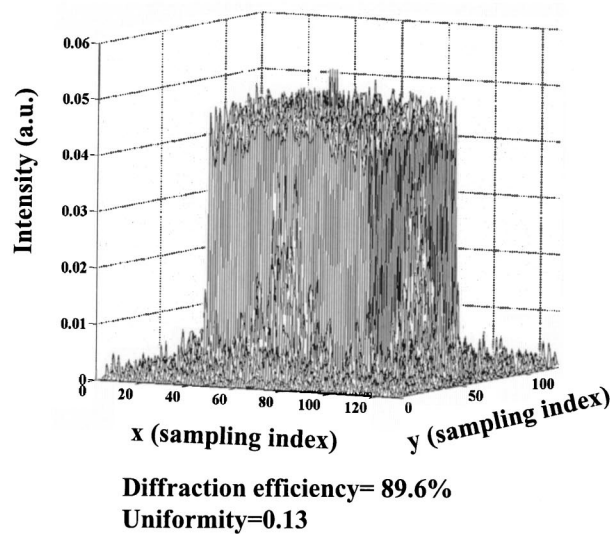
Comparisons between the results of the conventional and proposed IFTA schemes are illustrated in Figs. 4 and 5. In case of design A, the intensity distribution of the diffraction image and the phase profile of the DOE obtained by using the first type of conventional IFTA are shown in Figs. 4(a) and 4(b), respectively. The corresponding results obtained by the proposed IFTA with the

ARPD and the first-order Tikhonov regularization are presented in Figs. 4(c) and 4(d). To illustrate the detailed differences between the obtained phase profiles precisely, contour maps of the phase profiles are displayed. Comparing the diffraction images in Figs. 4(a) and 4(c), we can see that the uniformity of the diffraction image designed by the proposed algorithm (0.0426) is superior to that designed by the conventional algorithm (0.0916) with the same diffraction efficiency (92%).

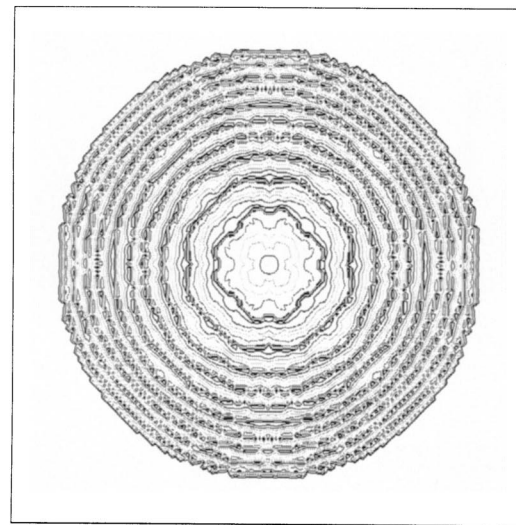
On the other hand, in the case of design B, because of lack of freedom in the DOE plane, it is more difficult to obtain fair uniformity of the diffraction image than in the case of the design for the same diffraction efficiency. However, as seen in Fig. 3(b), the proposed IFTA with regularization is effective in the improvement of uniformity in such a coarse circumstance as that of design B.

The intensity distribution of the diffraction image and the phase profile of the DOE obtained by using the first type of conventional IFTA are shown in Figs. 5(a) and 5(b), respectively. The corresponding results obtained by the proposed IFTA are shown in Figs. 5(c) and 5(d). The uniformity of the diffraction image obtained by the proposed algorithm (0.09) is better than that obtained by the conventional algorithm (0.13) with almost the same diffraction efficiency (89%).

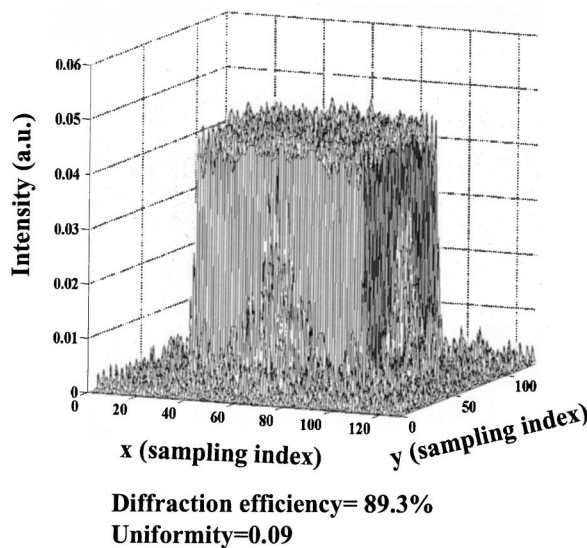
Next, we show the validity of the claim mentioned in Section 4 that the first-order Tikhonov regularization term is almost independent of the diffraction efficiency, so that, as the regularization parameter α_D of the regularization term increases, the uniformity improves without considerable loss of diffraction efficiency. We illustrate only the case of design A, since similar results are ob-



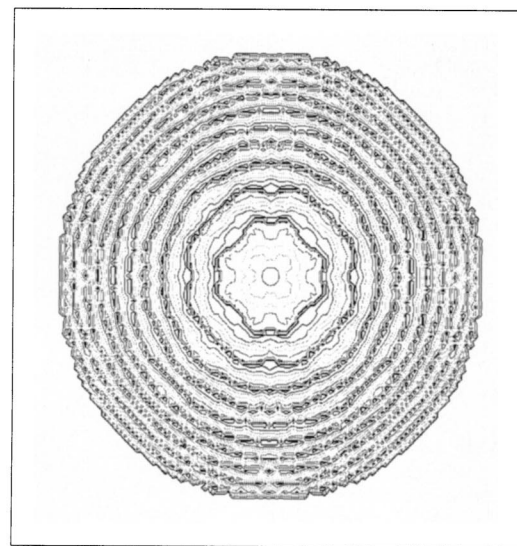
(a)



(b)



(c)



(d)

Fig. 5. Same as Fig. 4 but for the case of design B with decreased freedom in the DOE plane.

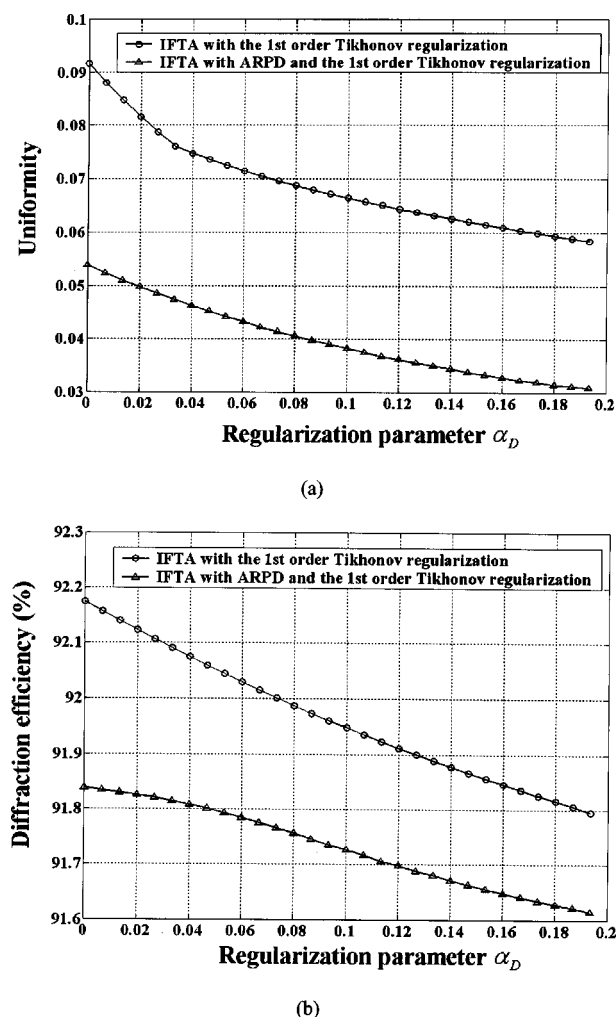


Fig. 6. Comparison of the changes in (a) uniformity and (b) diffraction efficiency for several values of the regularization parameter α_D (in the range 0–0.2) with use of the IFTA with the first-order Tikhonov regularization and of the IFTA with the ARPD and the first-order Tikhonov regularization in the case of design A with full freedom in the DOE plane.

served in the case of design B. Figures 6(a) and 6(b) show, respectively, the variations of the value of uniformity and of the diffraction efficiency of the diffraction image obtained after 200 iterations of the third and fourth types of IFTA as the regularization parameter α_D changes from 0 to 2, where the values of α_S and τ are tuned to result in a diffraction efficiency of approximately 92%. As indicated in Fig. 6(a), the first type of IFTA (conventional IFTA) and the second type of IFTA (with the ARPD) ($\alpha_D = 0$) yield uniformity values of 0.092 and 0.054, respectively. When $\alpha_D = 0.2$, the uniformity of both IFTAs reaches the values of 0.058 and 0.031, respectively, but the decrease in the diffraction efficiency is less than 0.3% for both cases. The IFTA with the ARPD (the second type) must reduce a considerable amount of diffraction efficiency to improve uniformity, as seen in Fig. 3(a). On the contrary, it is noteworthy that the first-order Tikhonov regularization can improve uniformity without considerable loss of diffraction efficiency.

7. CONCLUSIONS

In conclusion, based on the Tikhonov regularization theory, we investigated the trade-off between uniformity and diffraction efficiency in the IFTA-based design of DOEs. We proposed a novel IFTA scheme with the combination of two effective regularization techniques: the ARPD and the first-order Tikhonov regularization to minimize the trade-off. The main idea of the proposed method is to correlate the object functional with the uniformity strongly. For this, we devised the combination of the ARPD and the first-order Tikhonov regularization. We substantiated the validity of the proposed algorithm theoretically and numerically. It was shown that the amplitude freedom outside the signal area is necessary for the devised regularization technique to be effective. It was confirmed that the combination of the ARPD and the first-order Tikhonov regularization is a reliable strategy to approach the optimum trade-off.

ACKNOWLEDGMENT

The authors acknowledge the support by the Ministry of Science and Technology of Korea through the National Research Laboratory Program.

Address correspondence to Byoungho Lee: e-mail, byoungho@snu.ac.kr; fax, 82-2-873-9953.

REFERENCES

1. J. Turunen and F. Wyrowski, *Diffraction Optics for Industrial and Commercial Applications* (Wiley, New York, 1997).
2. B. Kress and P. Meyrueis, *Digital Diffractive Optics: An Introduction to Planar Diffractive Optics and Related Technology* (Wiley, New York, 2000).
3. J. Bengtsson, P. Modh, J. Backlund, H. Lindberg, and A. Larsson, "Progress in diffractive integrated optics," in *Diffractive Optics and Micro-Optics*, R. Magnusson, ed., Vol. 75 of OSA Trends in Optics and Photonics Series (Optical Society of America, Washington, D.C., 2002), pp. 17–19.
4. B. Dong, Y. Zhang, B. Gu, and G. Yang, "Numerical investigation of phase retrieval in a fractional Fourier transform," *J. Opt. Soc. Am. A* **14**, 2709–2714 (1997).
5. V. A. Soifer, V. V. Kotlyar, and L. Doskolovich, *Iterative Methods for Diffractive Optical Elements Computation* (Taylor & Francis, London, 1997).
6. A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems* (Kluwer Academic, Boston, Mass., 1995).
7. F. Wyrowski, "Upper bound of the diffraction efficiency of diffractive phase elements," *Opt. Lett.* **16**, 1915–1917 (1991).
8. F. Wyrowski, "Considerations on convolutions and phase factors," *Opt. Commun.* **81**, 353–358 (1991).
9. F. Wyrowski, "Diffractive optical elements: iterative calculation of quantized blazed phase structure," *J. Opt. Soc. Am. A* **7**, 961–969 (1990).
10. J. R. Fienup, "Phase-retrieval algorithms: a comparison," *Appl. Opt.* **21**, 2758–2769 (1982).
11. J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," *Appl. Opt.* **32**, 1737–1746 (1993).
12. S. Buhling and F. Wyrowski, "Improved transmission design algorithms by utilizing variable-strength projections," *J. Mod. Opt.* **49**, 1871–1892 (2002).
13. V. V. Kotlyar, P. G. Seraphimovich, and V. A. Soifer, "An iterative algorithm for designing diffractive optical elements with regularization," *Opt. Lasers Eng.* **29**, 261–268 (1998).
14. H. Kim and B. Lee, "Iterative Fourier transform algorithm

- with adaptive regularization parameter distribution for the optimal design of diffractive optical elements," *Jpn. J. Appl. Phys., Part 1* **43**, L702–L705 (2004).
15. S. Rudnaya, "Analysis and optimal design of diffractive optical elements," Ph.D. thesis (University of Minnesota, Minneapolis, Minn., 1999).
 16. E. G. Johnson and M. A. G. Abushagur, "Microgenetic-algorithm optimization methods applied to dielectric gratings," *J. Opt. Soc. Am. A* **12**, 1152–1160 (1995).
 17. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Kluwer Academic, Boston, Mass., 1996).
 18. J. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996).
 19. H. Kim and B. Lee, "Calculation of the transmittance function of a multilevel diffractive optical element considering multiple internal reflections," *Opt. Eng.* (to be published).
 20. A. Sahin, H. M. Ozaktas, and D. Mendlovic, "Optical implementations of two-dimensional fractional Fourier transforms and linear canonical transforms with arbitrary parameters," *Appl. Opt.* **37**, 2130–2141 (1998).
 21. H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing* (Wiley, New York, 2001).
 22. P. Senthilkumaran and F. Wyrowski, "Phase synthesis in wave-optical engineering: mapping- and diffuser-type approaches," *J. Mod. Opt.* **49**, 1831–1850 (2002).