14

Rigorous Diffraction Theory for 360° Computer-Generated Holograms

Toyohiko Yatagai, Yusuke Sando and Boaz Jessie Jackin
Center for Optical Research and Education, Utsunomiya University, Japan

14.1 Introduction

In computer-generated holography (CGH) [1] the fast Fourier transform (FFT) algorithm is commonly used to reduce the calculation time. Based on FFT, various types of algorithms for calculating diffraction have been developed [2]. Yoshikawa *et al.* have proposed a fast calculation method for large sized holograms by interpolation [3].

   Most algorithms for a *computer-generated hologram* (CGH) using FFT are effective only under the condition that both the input and observation surfaces are finite planes parallel to each other. Some authors proposed fast calculation methods using FFT that can be applied to the case where the input plane is not parallel to the observation plane [4, 5]. These methods are very useful for the calculation of reconstructed images observed from different points of view [6]. However, since the observation surfaces, in any of the methods, are assumed to be planes, very high-resolution display devices are necessary to enlarge the viewing angles. Although such devices are available, the reconstructed images cannot be observed from the opposite side of a hologram.

   A remarkable technique developed to achieve 360° field of view is 360° holography [7]. To synthesize a 360° hologram on a computer, a numerical simulation of diffraction on the non planar observation surfaces is required. Rosen synthesized a CGH on a spherical observation surface [8]. However, this method does not yield a 360° field of view because the origin of the object does not correspond to the center of the observation sphere, and also considerable computing time is required, since the FFT algorithm cannot be applied in this method.

To date, methods that enable the fast calculation of diffraction on spherical or cylindrical observation surfaces have been proposed based on the convolution theorem using the FFT algorithm. In this method, both the object and observation surfaces are cylindrical and concentric about the cylindrical axis [9, 10]. These methods are valid only in specific cases of geometrical configurations.

At first, we propose a simple but rigorous equation that describes the relationship between the diffracted wavefront of a 3D object and its 3D Fourier spectrum. In the present method, an exact solution of the diffraction integral is given by the Green function [11]. This principle gives us an intuitive understanding of calculation processes for various diffraction situation. To verify this method and confirm its effectiveness, CGHs with full view-angles are reconstructed using simulated experiments.

Alternatively, fast computation solutions for spherical computer generated hologram employing PSF (convolution method) was proposed by Tachiki *et al.* [12]. Here, the object and hologram were assumed to be concentric spherical surfaces in order to achieve shift invariance and hence, enable fast calculation. However, though the object was assumed to be a concentric spherical surface with the hologram surface, shift invariance does not exist due to unequal sampling points on a spherical grid (i.e., the grid points are more crowded at the poles). To facilitate fast calculation using FFT, an approximation to the convolution integral was proposed, which forced the PSF to be spatially invariant. Hence this calculation produced errors that were quantified in the same report. To solve the problem, we start the Helmholtz equation again, considering a boundary value problem in spherical coordinates. The solution will define the transfer function and the spectral decomposition of the wave field on the spherical surface. Using the transfer function and the wave spectrum we can develop a spectral propagation formula (for spherical surfaces in spherical coordinates) analogous to the angular spectrum formula (for plane surfaces in Cartesian coordinates).

This chapter explains how such theories for spherical CGH are devised.

14.2   Three-Dimensional Object and Its Diffracted Wavefront

Consider the geometry of a 3D object and an observation space as shown in Fig. 14.1. The coordinate system in the object is (x_o, y_o, z_o) and that of the observation space (x, y, z).
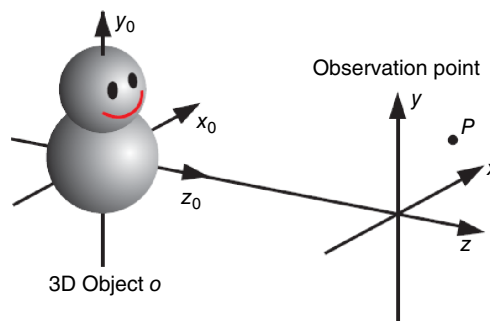

Figure 14.1   Geometry of 3-D object coordinates (x_o, y_o, z_o) and observation coordinates (x, y, z). x_o- and y_o-axes are parallel to the x- and y-axes, respectively, and the z_o-axis is the same as z-axis

The complex amplitude $u(\mathbf{r})$ of a wave in a homogeneous field at a position $\mathbf{r}(x, y, z)$ satisfies the following Helmholtz equation

$$(\nabla^2 + k_0^2)u(\mathbf{r}) = 0, \tag{14.1}$$

where $k_0 = 2\pi/\lambda$ is a constant wavenumber in a homogeneous field and λ denotes the wavelength.

The diffracted wavefront from the object $o(\mathbf{r})$ satisfies the following equation,

$$(\nabla^2 + k_0^2)u(\mathbf{r}) = -o(\mathbf{r}) \tag{14.2}$$

To solve Eq. (14.2), we employ the Green function. Green function is a solution of

$$(\nabla^2 + k_0^2)g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'), \tag{14.3}$$

and the Green function is given by

$$g(\mathbf{r}, \mathbf{r}') = \frac{\exp(ik_0|\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|}. \tag{14.4}$$

Therefore, the solution of Eq. (14.2) is given by

$$u(\mathbf{r}) = \int o(\mathbf{r}')g(\mathbf{r} - \mathbf{r}')d\mathbf{r}. \tag{14.5}$$

Equation (14.5) is a convolution integral and is rewritten by

$$u(x, y, z) = \mathcal{F}^{-1}[O(k_x, k_y, k_z) \cdot G(k_x, k_y, k_z)], \tag{14.6}$$

where

$$O(k_x, k_y, k_z) = \mathcal{F}[o(x, u, z)] \tag{14.7}$$

$$G(k_x, k_y, k_z) = \mathcal{F}[g(x, u, z)] \tag{14.8}$$

and $\mathcal{F}[\dots]$ denotes the 3D Fourier transform operator and $\mathcal{F}^{-1}[\dots]$ its inverse transform operator.

To obtain $G(k_x, k_y, k_z) = \mathcal{F}[g(x, y, z)]$, we Fourier-transform both sides of Eq. (14.3)

$$[-(k_x^2 + k_y^2 + k_z^2) + k_0^2]G(k_x, k_y, k_z) = -1. \tag{14.9}$$

So we have

$$G(k_x, k_y, k_z) = \frac{1}{k_x^2 + k_y^2 + k_z^2 - k_0^2}. \tag{14.10}$$

We use spatial frequencies (u,v,w) and since $k_x = 2\pi u, k_y = 2\pi v, k_z = 2\pi w$, we have

$$G(u, v, w) = \frac{1}{4\pi^2(u^2 + v^2 + w^2 - 1/\lambda^2)}. \tag{14.11}$$

Finally, from Eqs. (14.6) and (14.11), we have

$$u(x, y, z) = \frac{1}{4\pi^2} \iiint \frac{O(u, v, w)}{u^2 + v^2 + w^2 - 1/\lambda^2} \cdot$$
$$\times \exp[i2\pi(ux + vy + wz)]dudvdw \qquad (14.12)$$

Next, we will find the integral with respect to w. Integral (14.12) has a singularity for

$$w_\pm = \pm\sqrt{1/\lambda^2 - u^2 - v^2} \qquad (14.13)$$

By using contour integration, the integral with respect to w along the path shown in Fig. 14.2 is performed, where a complex value is defined by $\zeta = w + i\eta$. The residues at the singularities $w_\pm$ are given by

$$A_\pm = \lim_{\zeta \to w_\pm} (\zeta - w_\pm)\frac{O(u, v, \zeta)}{\zeta^2 + u^2 + v^2 - 1/\lambda^2} \exp(i2\pi\zeta z)$$
$$= \pm\frac{i2\pi O(u, v, \pm\sqrt{1/\lambda^2 - u^2 - v^2})}{2\sqrt{1/\lambda^2 - u^2 - v^2}}$$
$$\times \exp(\pm i2\pi\sqrt{1/\lambda^2 - u^2 - v^2}z). \qquad (14.14)$$

So we have

$$u(x, y) = \frac{i}{4\pi} \iint \frac{O(u, v, \sqrt{1/\lambda^2 - u^2 - v^2})}{\sqrt{1/\lambda^2 - u^2 - v^2}}$$
$$\times \exp[i2\pi(ux + vy + \sqrt{1/\lambda^2 - u^2 - v^2}z)]dudv$$
$$- \frac{i}{4\pi} \iint \frac{O(u, v, -\sqrt{1/\lambda^2 - u^2 - v^2})}{\sqrt{1/\lambda^2 - u^2 - v^2}}$$
$$\times \exp[i2\pi(ux + vy - \sqrt{1/\lambda^2 - u^2 - v^2}z)]dudv. \qquad (14.15)$$

Since only the forward propagating wave is considered, only the singularity w_+ is employed to calculate the contour integral. Suppose the observation plane is located at a plane $z = R$ and
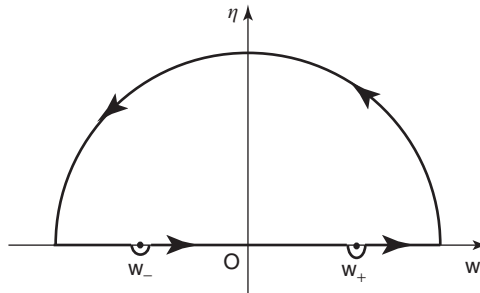

Figure 14.2   Singularities ($w_\pm$) and integration path in the complex plane $\zeta = w + i\eta$

we have

$$u(x, y)|_{z=R} = \frac{\mathrm{i}}{4\pi} \iint \frac{O(u, v, \sqrt{1/\lambda^2 - u^2 - v^2})}{\sqrt{1/\lambda^2 - u^2 - v^2}}$$

$$\times \exp[\mathrm{i}2\pi(ux + vy + R\sqrt{1/\lambda^2 - u^2 - v^2})]\mathrm{d}u\mathrm{d}v. \quad (14.16)$$

Finally, by 2D inverse Fourier transforming Eq. (14.16), the spectrum of the observed complex amplitude at $z = R$ is given by

$$U(u, v)|_{z=R} = \frac{\mathrm{i}}{4\pi} \frac{O(u, v, \sqrt{1/\lambda^2 - u^2 - v^2})}{\sqrt{1/\lambda^2 - u^2 - v^2}}$$

$$\times \exp\ (\mathrm{i}2\pi R\sqrt{1/\lambda^2 - u^2 - v^2}). \quad (14.17)$$

In Eq. (14.17), $O(u, v, \sqrt{1/\lambda^2 - u^2 - v^2})$ implies a hemispherical surface component with the diameter of $1/\lambda$ in the 3D spectrum $O(u, v, w)$ of the 3D object $o(x, y, z)$. This means that the 2D spectrum $U(u, v)$ of the diffracted wave is given by the hemispherical surface component $O(u, v, \sqrt{1/\lambda^2 - u^2 - v^2})$ of the 3D spectrum with the weight of $\mathrm{i}/(4\pi\sqrt{1/\lambda^2 - u^2 - v^2})$ multiplied with a phase component $\exp\ (\mathrm{i}2\pi R\sqrt{1/\lambda^2 - u^2 - v^2})$.

Next, the correspondence relation between the diffracted wavefront and the hemispherical Fourier spectrum equivalent to it is illustrated in Fig. 14.3. Here, the diffracted wavefront to the $+z_0$ direction corresponds to the hemispherical spectrum with the $+w$ axis of symmetry. Similarly, the directions $-z_0$, $+y_0$, and $-y_0$ correspond to the directions $-w$, $+v$, and $-v$, respectively. Thus, it is possible to express the wavefront diffracting in all directions only by the hemispherical Fourier spectrum with the radius $1/\lambda$.
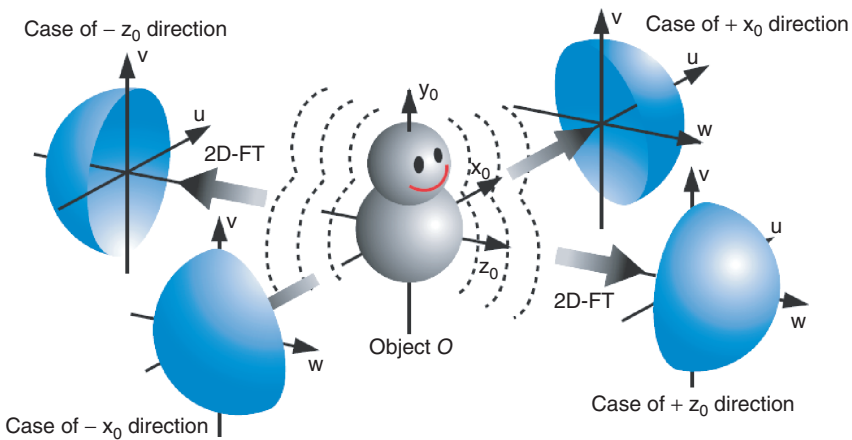

Figure 14.3 Three-dimensional hemispherical Fourier spectrum of diffracted wavefronts in each direction

14.2.1   *Diffracted Waves with Full View Angles*

As an example of the fast calculation method of diffraction, the diffracted images in all directions are demonstrated. Once the 3D Fourier spectrum is obtained, the diffracted images to the arbitrary direction can be readily calculated by extracting the hemispherical spectrum, multiplying the weight factor and the phase components, and running the 2D inverse FFT. To verify this method, we have simulated the diffraction under the configuration where the 2D asterism image shown in Fig. 14.4 is distributed cylindrically as shown in Fig. 14.5.

θ is the angle between the observation direction and the z axis and the diffraction distance R is set to the radius of the cylinder of the object. The results calculated with the observation angle θ varying are shown in Fig. 14.6.

As shown in Fig. 14.6, only the object corresponding to the angle θ is focused on the center of the observation plane, while both ends of each image are out of focus. These results are very natural considering the configuration shown in Fig. 14.5 and the validity of the derived principle has been verified at once.
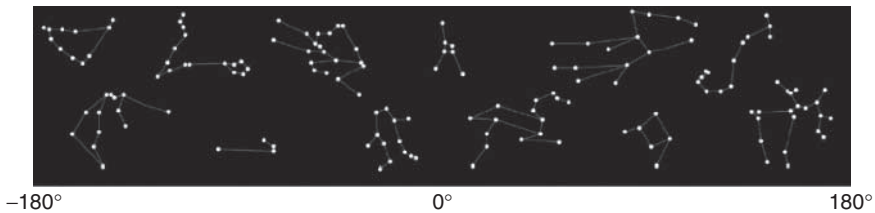

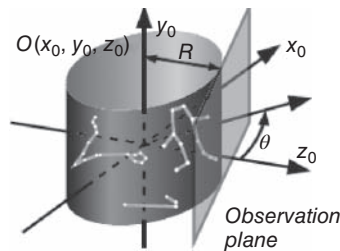−180°   0°   180°

Figure 14.4   Asterism map for the object

Figure 14.5   Schematic for the simulation. Perspective view of the observation plane

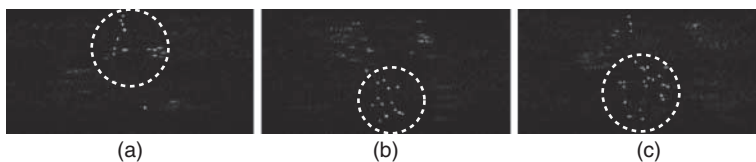(a)   (b)   (c)

Figure 14.6   Diffracted images observed at $\theta = -109°$ (a), $-23°$ (b), and $152°$ (c)

14.3 Point-Spread Function Approach for Spherical Holography

14.3.1 Spherical Object and Spherical Hologram

As far as we know, diffraction to a spherical surface cannot be directly expressed as a convolution. The pixel resolution for a grid on a spherical surface in spherical coordinates is not constant, so the PSF is not spatially invariant. To facilitate fast calculation, an approximation to the integral is proposed, which forces the PSF to be spatially invariant and allows the diffraction integral to be expressed as a convolution.

In this method, to obtain a convolution form, the object surface must also be spherical and concentric with the hologram surface. The geometries of the object and CGH are shown in Fig. 14.7, where a spherical coordinate system is used. Note that the azimuthal axis is defined as ϕ, and the latitudinal axis is defined as θ. The object and hologram surfaces are denoted as $f_o(r_o, \phi_o, \theta_o)$ and $f_h(r_h, \phi_h, \theta_h)$. In this paper we limit the discussion to the case of a two-spherical dimensional object with hologram radius r_h constant, so we can rewrite the distributions as $f_o(\phi_o, \theta_o)$ and $f_h(\phi_h, \theta_h)$. A CGH of a 3D object can be created by taking the superposition of multiple holograms, f_h, computed with successive object radii to an arbitrary resolution. In this discussion we also assume that the object surface radius is smaller than the hologram radius, although this is not a necessary condition mathematically.

Treating the object as a composite of many point light sources, the hologram amplitude distribution $f_h(\phi_h, \theta_h)$ is given as the integral of the spherical wavefronts emerging from all the point light components on the object surface:

$$f_h(\phi_h, \theta_h) = C \iint \frac{f_o(\phi_o, \theta_o) \exp(ikd)}{d} d\phi_o d\theta_o, \qquad (14.18)$$
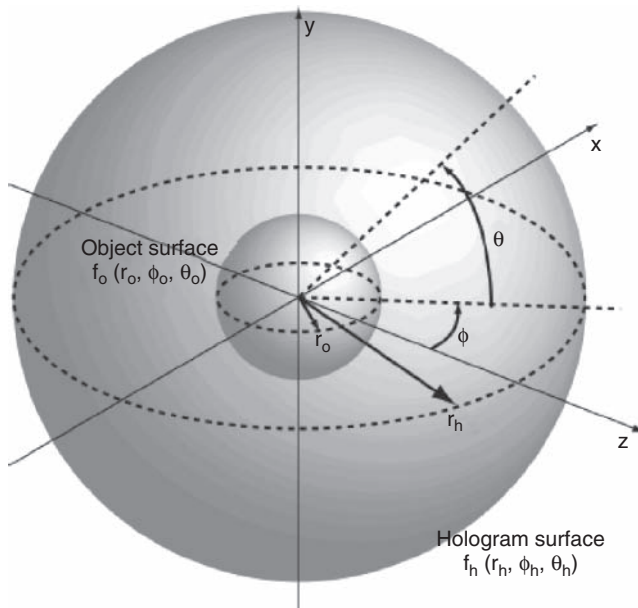

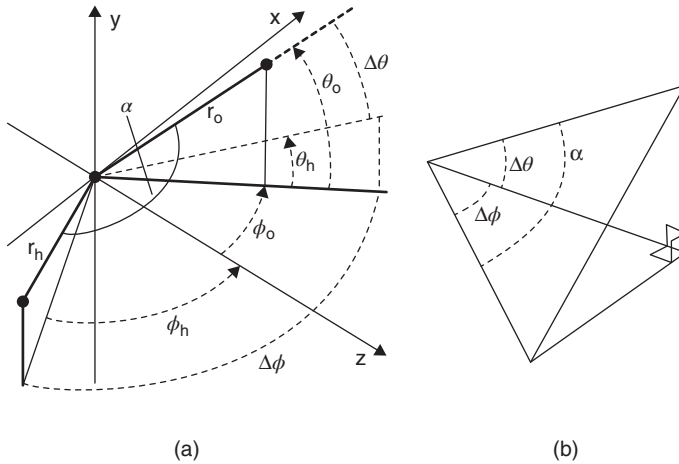**Figure 14.7**   Spherical coordinate system

Figure 14.8   Geometry of the angular approximation (a) and geometry to solve the angle α(b)

where k denotes the wavenumber of the incident light, C denotes a constant, and d, which is the distance between a point on the object and a point on the hologram, is given by

$$d = \{r_o^2 + r_h^2 - 2r_o r_h[\sin(\theta_h)\sin(\theta_o) + \cos(\theta_h)\cos(\theta_o)\cos(\phi_h - \phi_o)]\}^{1/2}. \tag{14.19}$$

To allow the integral to be expressed as a convolution, the distance formula must be expressed as a function of $(\theta_h - \theta_o)$ and $(\phi_h - \phi_o)$.

We propose an approximation to the distance using the geometry of Fig. 14.8(a). The angle α between the two points can be approximated from the angles $\Delta\phi$ and $\Delta\theta$ through the law of cosines using the geometry of Fig. 14.8(b). Then the distance can be calculated using the lengths r_o and r_h, and the angle α with another application of the law of cosines. The resulting equation for distance is

$$d = \{r_o^2 + r_h^2 - 2r_o r_h[\cos(\theta_h - \theta_o)\cos(\phi_h - \phi_o)]\}^{1/2}. \tag{14.20}$$

The approximated PSF can then be defined as

$$h(\phi, \theta) = \frac{\exp\{ik[r_o^2 + r_h^2 - 2r_o r_h \cos(\theta)\cos(\phi)]^{1/2}\}}{[r_o^2 + r_h^2 - 2r_o r_h \cos(\theta)\cos(\phi)]^{1/2}}. \tag{14.21}$$

We substitute the PSF into Eq. (14.18) and obtain

$$f_h(\phi_h, \theta_h) = C \iint f_o(\phi_o, \theta_o) h(\phi_h - \phi_o, \theta_h - \theta_o) d\phi_o d\theta_o, \tag{14.22}$$

which takes the form of a convolution integral,

$$f_h = C f_o * h, \tag{14.23}$$

where * denotes the convolution operation. Thus, the calculation of the diffracted wavefront on the spherical hologram surface can be performed using the FFT, according to the convolution theorem.

14.3.2   Approximation Error

We examine the error analytically and numerically. The exact formula for the distance between a point on the object and a point on the hologram is given by Eq. (14.2). If we insert a $\cos(\phi_h - \phi_o)$ term next to the sine terms,

$$d = \{r_o^2 + r_h^2 - 2r_o r_h [\sin(\theta_h)\sin(\theta_o)\cos(\phi_h - \phi_o) + \cos(\theta_h)\cos(\theta_o)\cos(\phi_h - \phi_o)]\}^{1/2}$$

(14.24)

the equation reduces to the approximation form of Eq. (14.20). It is clear that the error is minimal when the $\cos(\phi_h - \phi_o)$ term approaches a value of one, or $(\phi_h - \phi_o)$ approaches zero. Thus, contributions to the hologram from object points far in the azimuthal axis have larger errors. In addition, when θ_o approaches zero, the approximated distance equation becomes

$$d = \{r_o^2 + r_h^2 - 2r_o r_h [\cos(\theta_h)\cos(\phi_h - \phi_o)]\}^{1/2},$$

(14.25)

which is equivalent to the exact distance. Thus, the contributions to the hologram from object points near $\theta_o = 0$ have a smaller error. The same is true of θ_h. That is, points near $\theta_h = 0$ on the hologram have a smaller error, regardless of the area of the object image observed. Also, as r_o approaches zero, the error decreases. As can be seen from Eq. (14.14), as r_o becomes very small compared with r_h, the r_o^2 and $-2r_o r_h \cos(\phi_h - \phi_o)\cos(\theta_h - \theta_o)$ terms become very small compared with the r_h^2 term, and the distance approaches that given by the exact equation.

14.3.3   Computer Simulation for Spherical Holography

To demonstrate the application of the fast calculation method to holograms of more realistic object distributions, a hologram of the image in Fig. 14.9(a) was generated with the $r_o = 1$ cm
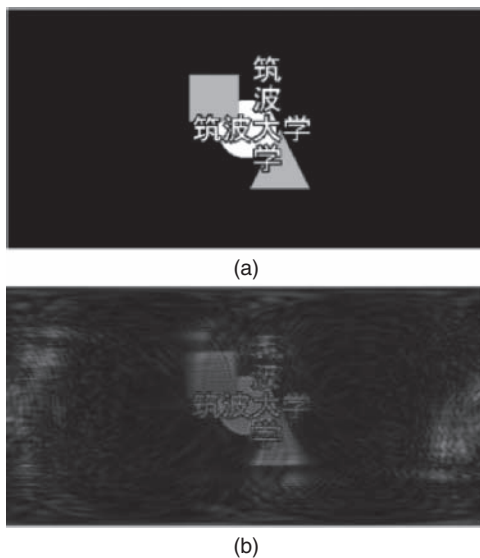

(a)

(b)

Figure 14.9   Object (a) and reconstructed image (b)

and $r_h = 10$ cm, with a wavelength of 300 μm. The reconstructed image on the same spherical surface is shown in Fig. 14.9(b), reconstructed from the hologram region of $|\phi_h| < \pi/4$ and $|\theta_h| < \pi/3$. We note that a reconstruction from a hologram computed through direct calculation of the diffraction integral would yield a perfect reconstruction of the original image in Fig. 14.9(a). Although there is distortion throughout the image, particularly as θ increases, the image is clearly recognizable. This implies that images can be reconstructed from holograms created using this method to a reasonable degree of practical usefulness.

14.4   Rigorous Point-Spread Function Approach

We will describe points on the sphere by their latitude $\theta \in [-\pi/2, \pi/2]$ and longitude $\phi \in [0, 2\pi]$ as shown in Fig. 14.7. Throughout this work we will be dealing with square-integrable functions that span a space $L^2(S)$ on the unit sphere S. Fast computation algorithms take advantage of the cyclic and periodic properties of the transformation kernel for fast calculations. The cyclic and periodic properties exists only if the system is shift invariant between the transformation planes for that operation. Which means that the object and hologram surface should be shift invariant in order to devise a fast computation scheme. To achieve this the object and hologram surface are chosen to be concentric spherical surfaces, so that they can remain shift invariant in rotation along ϕ and θ directions. But these surfaces are defined on a spherical grid, where the sampling points are more dense at the poles than at the equator. Hence, shift invariance is not satisfied. However, in the hologram generation process, the hologram and object are band-limited functions on $L^2(S)$ space. The band-limited functions on $L^2(S)$ space have a very useful and important property that *any rotated version of a band-limited function is also a band-limited function with the same bandwidth* [13–15]. Thus, they are refereed to as having *uniform resolution* at all points on the sphere, meaning that they are shift invariant. The triangular truncation and Gauss–Legendre quadrature method that occurs in the transformation operation is responsible for this property (explained in the next section). Hence the system shown in Fig. 14.7 does have a shift invariance relationship between object and hologram surfaces, and confirms the possibility of a fast computation formula. With this assurance we proceed to develop the fast calculation method for computer generated spherical holography starting from the basic equations of electromagnetism.

An electromagnetic field is defined by Maxwell's equations and its propagation by the Helmholtz wave equation. Hence for any particular system the complex amplitude of a propagating wave at any instance of time and anywhere in space can be found by solving the wave equation, applying its constraints and conditions. Accordingly, for the system shown in Fig. 14.7 the solution can be derived starting from the wave equation as follows. The time independent vector wave equation $u(r, \theta, \phi)$ is expressed the by the Helmholtz equation as

$$\nabla^2 u + k^2 u = 0, \tag{14.26}$$

where r is the radius of the spherical surface of interest and θ and ϕ represent the azimuthal and polar angles in the surface, respectively. The Laplacian operator ∇^2 in spherical coordinates is defined as

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial \phi^2}. \tag{14.27}$$

The scalar wave equation given in spherical coordinates becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial u}{\partial r}) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial u}{\partial\theta}) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2} + k^2 u = 0. \tag{14.28}$$

The solution of Eq. (14.28) can be found by separation of variables [16–18], which can be expressed as shown next

$$u(r, \theta, \phi, t) = R(r)\Theta(\theta)\Phi(\phi). \tag{14.29}$$

The separable variables obey the following four ordinary differential equations:

$$\frac{d^2\Phi}{d\phi^2} + m^2\Phi = 0 \tag{14.30}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}(\sin\theta\frac{d\Theta}{d\theta}) + [n(n+1) - \frac{m^2}{\sin^2\theta}]\Theta = 0 \tag{14.31}$$

$$\frac{1}{r^2}\frac{d}{dr}(r^2\frac{dR}{dr}) + k^2 R - \frac{n(n+1)}{r^2}R = 0. \tag{14.32}$$

The solution to all these equations are derived in by Arfken [18]. Only the final results are used in this research work. The solution to azimuthal Eq. (14.30) is

$$\Phi(\phi) = \Phi_1 e^{im\phi} + \Phi_2 e^{-im\phi}, \tag{14.33}$$

where m must be an integer so that there is continuity and periodicity in $\Phi(\phi)$. Φ_1 and Φ_2 are constants.

The solution to polar Eq. (14.31) is

$$\Theta(\theta) = \Theta_1 P_n^m(\cos\theta) + \Theta_2 Q_n^m(\cos\theta) \tag{14.34}$$

where, P_n^m and Q_n^m are the associated Legendre polynomials of first and second kind, respectively, and Θ_1 and Θ_2 are constants. Q_n^m is not finite at the poles where $\cos(\theta) = \pm 1$, so this solution is discarded ($\Theta_2 = 0$).

For the radial differential equation Eq. (14.32) the solutions are

$$R(r) = R_1 h_n^{(1)}(kr) + R_2 h_n^{(2)}(kr) \tag{14.35}$$

where, $h_n^{(1)}$ and $h_n^{(2)}$ are the spherical Hankel functions of the first and second kind, respectively. Since we are only interested in the outgoing wave, we can neglect the second term ($R_2 = 0$).

The angle functions in Eq. (14.33) and Eq. (14.34) are conveniently combined into a single function called a *spherical harmonic* Y_n^m [17, 18] defined by

$$Y_n^m(\theta, \phi) \equiv (-1)^m \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m \cos(\theta) e^{im\phi} \tag{14.36}$$

where the quantity

$$\overline{P}_n^m = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos\theta) \tag{14.37}$$

is known as the orthonormalized associated Legendre polynomial. The term $(-1)^m$ is called the Condon–Shortley phase. Hence the spherical harmonics can also be represented in short form as shown next (neglecting the Condon–Shortley phase)

$$Y_n^m(\theta, \phi) = \overline{P}_n^m(\cos \theta)e^{im\phi}. \tag{14.38}$$

Combining all these equations, the traveling wave solution to Eq. (14.28) can be represented as

$$u(r, \theta, \phi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{mn}(\omega)h_n(kr)Y_n^m(\theta, \phi). \tag{14.39}$$

The radiated field is completely defined when the coefficient A_{mn} is determined. This is achieved by using the orthonormal property of spherical harmonics. Assume that the wave field $u(r, \theta, \phi)$ is known on a sphere of radius $r = a$. We also drop the time variable (which is not important) for simplicity. Now multiplying each side of Eq. (14.39) (evaluated at $r = a$) by $Y_n^m(\theta, \phi)^*$ and integrating over the sphere, gives

$$A_{mn} = \frac{1}{h_n(ka)} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} u(a, \theta, \phi)Y_n^m(\theta, \phi)^* \sin(\theta)d\theta d\phi \tag{14.40}$$

where, $d\Omega = \sin(\theta)d\theta d\phi'$, is the solid angle for integrating on a sphere. Inserting the expression for A_{mn} back into Eq. (14.39), we get,

$$u(r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\theta, \phi) \left(\left[\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} u(a, \theta', \phi')Y_n^m(\theta', \phi')^* d\Omega' \right] \frac{h_n(kr)}{h_n(ka)} \right). \tag{14.41}$$

Hence the wavefield at any spherical surface $u(r, \theta, \phi)$ can be calculated knowing the wavefield at $u(a, \theta', \phi')$.

For easy understanding and interpretation of Eq. (14.41), it is compared with the well-known angular spectrum of plane waves equation [19], which is given by

$$u(x, y, z) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(k_x x + k_y y)} dk_x dk_y$$

$$\left(\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x', y', 0)e^{-i(k_x x' + k_y y')} dx' dy' \right] e^{ik_z z} \right). \tag{14.42}$$

It is well known that, in Eq. (14.42), the quantity within square brackets represents the source wave field decomposed into spectrum of plane waves in (k_x, k_y). The propagation of the spectrum is defined by the quantity $e^{ik_z z}$, which is known as the transfer function. The propagated spectrum is recomposed into the wave field at destination by the integral over k_x, k_y. Thus this equation defines wave propagation from one plane surface $u(x', y', 0)$ to another surface $u(x, y, z)$. When comparing Eq. (14.41) with Eq. (14.42), the following can be deduced:

- The wavefield in (θ, ϕ) at a radiating spherical surface of radius $r = a$ is decomposed into its wave spectra in (m, n) defined by

$$U_{mn}(a) = \int u(a, \theta, \phi)Y_n^m(\theta, \phi)^* d\Omega. \tag{14.43}$$

- The decomposed wave components (spectra) are expressed by Eq. (14.36), which is composed of a traveling wave component in ϕ, and defined by $e^{im\phi}$ and a standing wave component in θ given by $P_n^m(\cos\theta)$.
- The decomposed wave components in this system can be named spherical wave components in analogy to plane wave components for planar systems. Similarly Eq. (14.43) can be termed a spherical wave spectrum as analogous to the angular spectrum of plane waves.
- The wavenumbers k_x and k_y are imitated by the quantities m/a and n/a. Hence, we can refer to the spherical wave spectrum as a k-space spectrum, due to this analogy.
- Equation (14.43) can be viewed as a forward Fourier transform using $Y_n^m(\theta,\phi)$ as the basis function. In other words, the spectral space for spherical system is spanned by spherical harmonic functions ($Y_m^n(\theta,\phi)$). This is also termed *spherical harmonic transform* (SHT).
- The propagation of a spherical wave spectrum from one spherical surface of radius a to another of radius r is given by

$$U_{mn}(r) = \frac{h_n(kr)}{h_n(ka)} U_{mn}(a).\tag{14.44}$$

- Hence the quantity $h_n(kr)/h_n(ka)$ can be referred to the transfer function (TF) for a spherical system, as opposed to the quantity $e^{ik_z z}$ for a planar system.
- The inverse spherical harmonic transform (ISHT) that recomposes the wave field back from the spectrum is given by

$$u(r,\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} U_{nm}(r)Y_n^m(\theta,\phi).\tag{14.45}$$

The transfer function is crucial since it completely defines propagation and hence it is worth discussing some of its properties. During propagation, amplitude and phase of the spectral components change with distance as defined by the transfer function. What is most important is the rate of change of phase of the transfer function, which determines the sampling requirements. Accordingly the plot shown in Fig. 14.10 reveals that the phase change increases with increasing orders "n" of the transfer function (the plot was generated for wavelength 100 μm,
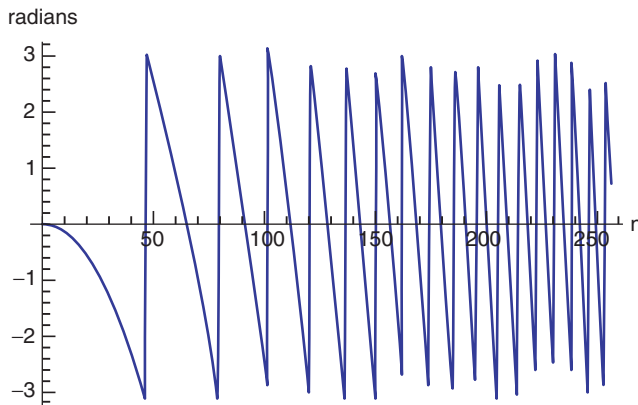

Figure 14.10   Plot of phase (in radians) of the transfer function for increasing order (n)

radius 10 and 0.5 cm, and up to 256 orders). Hence, sampling requirements will be satisfied if highest order "n" of the transfer function is sampled according to Nyquist criteria. It can also be understood that the rate of phase change becomes higher as the distance of propagation increases. This will demand a lot of sampling and also increase numerical errors. It is worth noting here that the spherical Hankel functions are asymptotic in nature. In the far field, the spherical Hankel functions can be approximated by their asymptotic expressions, which is as shown in Eq. (14.46).

$$h_n^{(1)}(x) = (-i)^{n+1} \frac{e^{ix}}{ix}. \tag{14.46}$$

This could yield a formula analogous to the far field Fresnel diffraction formula. However, it requires systematic development of theory with proper analysis of approximations and is not discussed in this chapter. But this can be considered as future work to the proposed method. Therefore, the devised formula, which is analogous to the angular spectrum of plane waves (AS) formula, defines wave propagation between spherical surfaces. The numerical procedure to implement the devised formula is discussed in the next section.

14.4.1   Numerical Computation

The numerical computation of angular spectrum (AS) method heavily depends on the FFT operations for which a lot of tools and methods are available. But the numerical computation of the proposed method heavily depends on the SHT operations. Fast computation was guaranteed from the theory and geometry of the system. Now a numerical procedure that takes advantage of this is required. Fortunately, lot of fast computation numerical methods have been reported for SHT and this research work could make use of it. Since FFT and numerical computation of AS method are well understood, this section intends to introduce SHT and numerical computation of the proposed method in close analogy to the former.

Continuing with the comparison from the previous section, the numerical computation of wave propagation for planar and spherical systems according to Eq. (14.41) and Eq. (14.42) can be represented respectively as

$$u(r, \theta, \phi) = \text{ISHT}[\text{SHT}(u(a, \theta, \phi)) \times TF_s] \tag{14.47}$$

$$u(x, y, z) = \text{IFFT}[\text{FFT}(u(x, y, 0)) \times TF_c] \tag{14.48}$$

where, FFT[...] and IFFT[...] denote the forward and inverse fast Fourier transform operations, while SHT[...] and ISHT[...] denote the forward and inverse spherical harmonic transform operations. TF_s and TF_c are the transfer functions for spherical and planar wave propagations as defined by Eq. (14.41) and Eq. (14.42), respectively. Comparison reveals that the computation method (for spherical systems) is analogous to the angular spectrum of plane waves method (for planar systems) except for the fact that the Fourier transform is replaced by the spherical harmonic transform. Hence evaluation of SHT becomes key, the others being only basic mathematical evaluations. SHT operations are fundamentally different from the FFT operations. In FFT operation the transform is done with a sinusoid and its higher harmonics as the basis function, whereas in the case of SHT, spherical harmonics defined by Eq. (14.36) are

the basis. This also implies that since the transformation kernel for the FFT is a sinusoid, it is an operation defined on a circle. But spherical harmonics are functions on a sphere and hence the transformation is an operation defined on a sphere. Spherical harmonic transforms have been studied extensively, and fast computation algorithms and optimization methods have been proposed. A method that requires only $O(N^2 \log N)$ operations for N grid points, as opposed to the standard N^3 operations, was proposed by Chien *et al.* [13]. They imposed truncations on the spectral components, and used a fast multipole method and fast Fourier transform for evaluation. Their method is called the "spectral truncation method" and the errors due to truncation were well within acceptable limits. Later on, Healy *et al.* [14] achieved the same using $O(N(\log N))^2$ operations. They took advantage of the recursive properties of the associated Legendre polynomial for fast computation. This method was found to be more efficient and hence was chosen for numerical evaluation of the spherical harmonic transforms in this work. A brief outline of the numerical evaluation is presented. For more details please refer to Healy [14] and Driscoll [15].

Though FFT and SHT are fundamentally different, they both are variable separable. Which means a 2D FFT is computed by separating the transformation kernel into its variables and evaluating it as 1D column transform followed by a 1D row transformation. Similarly the spherical harmonic transform kernel Y_n^m given by Eq. (14.38) is also variable separable and can be separated into a ϕ component and θ component. Then it is evaluated as a 1D transform along the ϕ direction followed by another 1D transform along the θ direction. Accordingly, the transformation given by Eq. (14.43) can be represented as shown next

$$U_{mn}(r) = \int_{-\pi/2}^{\pi/2} \left(\int_{-\pi}^{\pi} u(r, \theta, \phi) e^{im\phi} \mathrm{d}\phi \right) \overline{P}_n^m(\cos \theta) \mathrm{d}\theta. \qquad (14.49)$$

First, the quantity within the round brackets alone is to be computed, which is nothing but a Fourier transform operation. The Fourier coefficients $U^m(\theta)$ are evaluated for $m = -N, ..., N$ as shown next

$$U_m(\theta) = \int_{-\pi}^{\pi} u(r, \theta, \phi) e^{im\phi} \mathrm{d}\theta \qquad (14.50)$$

$$= \frac{1}{I} \sum_{i=1}^{I} u(r, \theta, \phi_i) e^{im\phi_i} \qquad (14.51)$$

where $\phi_i = 2\pi i / I$ for $i = 1, ..., I$. The equispaced longitudes ϕ_i enables the use of fast Fourier transform.

Second, the Legendre transform of the Fourier coefficients $U_m(\theta)$ is to be evaluated for $|m| \leq n \leq N$. This is done using the Gaussian–Legendre quadrature as shown next,

$$U_{nm} = \int_{-\pi/2}^{\pi/2} U_m(\theta_j) \overline{P}_n^m(\cos \theta) \sin \theta \mathrm{d}\theta \qquad (14.52)$$

$$= \sum_{j=|m|}^{N} U_m(\theta_j) \overline{P}_n^m(\cos \theta_j) w_j, \qquad (14.53)$$

where θ_j and w_j are, respectively, the Gauss nodes and weights, and are calculated using the Fourier–Newton method as described by Swarztrauber [16]. The Gauss–Legendre quadrature replaces the integral with the sum. The fact that the summation runs only from $|m|$ to N is referred to as triangular truncation. The use of the Gauss–Legendre quadrature method redistributes θ into Gaussian nodes θ_j. This along with the triangular truncation are responsible for uniform resolution on the latitudinal points. Again, the triangular truncation, along with the recurrence property of the Legendre polynomial, helps to achieve fast computation.

In a similar way the inverse spherical harmonic transform can be represented as

$$u(\theta, \phi) = \sum_{m=-N}^{N} \left(\sum_{n=|m|}^{N} U_{nm} P_n^m(\cos \theta) \right) e^{\mathrm{i} m \phi}. \tag{14.54}$$

The inverse transform also follows the same procedure and is computed in two steps but in the reverse order (i.e., Legendre transform first and Fourier transform next) as shown here

$$U_m(\theta) = \sum_{n=|m|}^{N} U_{nm} \overline{P}_n^m(\cos \theta) \tag{14.55}$$

$$u(\theta, \phi) = \sum_{m=-N}^{N} U_m(\theta) e^{\mathrm{i} m \phi}. \tag{14.56}$$

Thus using this numerical procedure fast computation of wave propagation in spherical computer generated holograms is achieved, which uses only $O(N(\log N)^2)$ operations for SHT computation.

There are a lot of software tools available on the Internet to do the SHT operation. Most of them are tuned and dedicated for geophysical processes, which only requires real SHT but holography requires complex SHT. However, the package of SHTools by Wieczorek [20] could do a complex SHT operation in the Fortran language. Though not tested by us, this is the best one available that suits the work related to this paper. Hence, we recommend using this if it is required to quickly reproduce the work in this paper. The next section describes the testing and verification of this numerical procedure using simulations.

14.4.2 Simulation Results on Rigorous Theory

The system considered for simulation experiments is shown in Fig. 14.7. The object $(O(a, \theta, \phi))$ is a spherical surface of radius 1 cm and the hologram $(H(r, \theta, \phi))$ is another concentric spherical surface of radius 10 cm. The reference is considered to be a virtual source emitting spherical waves from the center, that is, the wavefield, due to reference has same phase and amplitude throughout the hologram plane. This is similar to using a plane reference wave with normal incidence in plane holography.

14.4.3 Verification through Comparison

Since this the first occurrence of such a formula in computer generated holography first it is required to test it to see if it obeys the basic diffraction laws. For this, the proposed method by

expecting it to reproduce the already known diffraction results. To achieve this, the proposed method is made to generate already reported diffraction patterns for spherical surfaces. For this the diffraction pattern reported by Tachiki *et al.* [12] for spherical surfaces is used as a reference. Accordingly, a simple object was chosen, which is a spherical surface with two irradiating points at $\phi = -\pi/2$ and $\phi = \pi/2$, as shown in Fig. 14.11. The object and hologram are composed of 256 pixels in the longitudinal (north-south) direction and 512 pixels in the latitudinal (east-west) direction. The wavelength was chosen to be $\lambda = 100\,\mu m$, in order to reduce sampling requirements and visibility of fringes. The procedure for numerical generation of hologram using the proposed method is expressed in an abstract form as shown next.

$$AmplitudeHologram = |ISHT[SHT(Object) \times TF]$$
$$+ ISHT[SHT(Reference) \times TF]|^2.$$

Then a hologram for the same object was simulated using the the well-known direct integration formula defined in Eq. (14.57).

$$H(r, \theta, \phi) = \iint \frac{O(\theta', \phi') \exp(ikL)}{L} dx dy \tag{14.57}$$

where,

$$L = \sqrt{r^2 + a^2 - 2ra[\sin(\theta)\sin(\theta') + \cos(\theta)\cos(\theta')\cos(\phi - \phi')]} \tag{14.58}$$

$$Hologram = |H_{object}(r, \theta, \phi) + H_{reference}(r, \theta, \phi)|^2.$$

The simulation results are shown in Fig. 14.12. The pattern generated by the proposed method matches the one generated by the direct integration method. However, the distribution of brightness and contrast across the pattern is constant for the direct integration method, while it decreases gradually from the center for our proposed method. This inconsistency can be explained as follows. The direct integration formula Eq. (14.57) is the Rayleigh–Sommerfeld diffraction formula [12] without the obliquity factor. The obliquity factor is the cosine of the angle between the normal of the radiating surface to the direction of the observation point. This is responsible for the distribution of light intensity based on the angle (i.e., more bright at the center, gradually decreases outward, and no radiation backwards). However, the spectral method, which is the solution to the boundary value problem of the wave equation, incorporates the obliquity factor, and hence the brightness and contrast varies radially. Moreover, the

Figure 14.11 Two point source object

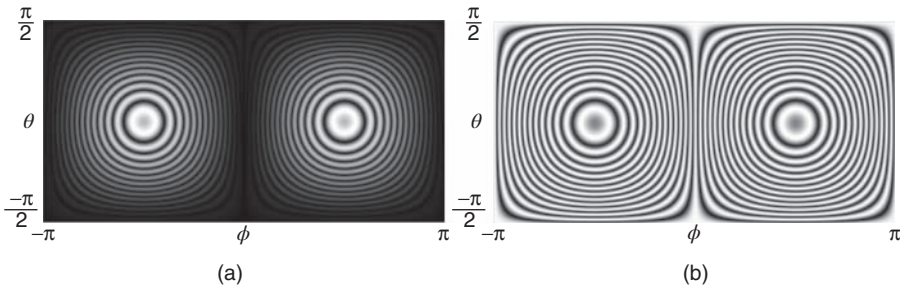(a)   (b)

Figure 14.12   Computed hologram (intensity) using (a) proposed method and (b) direct integration

obliquity factor does not alter the phase of the traveling wave, which in turn does not affect the interference pattern, and hence guarantees a fair comparison.

14.4.4   Hologram Generation

Since the theory is developed in the context of computer generated holography, it is mandatory to verify its applicability to the same. For this, it was decided to perform spherical hologram generation and then reconstruction from the hologram on the computer using our proposed method. The object was assumed to be a single spherical surface with some images inscribed on it. The spherical object for which the hologram was to be made is shown in Fig. 14.13. The object and hologram were composed of 256×512 pixels. The wavelength for simulation was chosen to be $\lambda = 30\,\mu m$. Again, here the wavelength was assumed to be large in order to reduce sampling requirements. Now, using the developed formula, wave propagation was simulated from the object surface to the hologram surface. Since the reference was assumed to be a spherical wave emanating from the center, it contained the same phase and amplitude at the hologram surface. So we have the complex amplitudes of the object and reference as a matrix of complex numbers at the hologram surface. By adding these two complex amplitudes and calculating the intensity, the hologram is produced and is shown in Fig. 14.14.
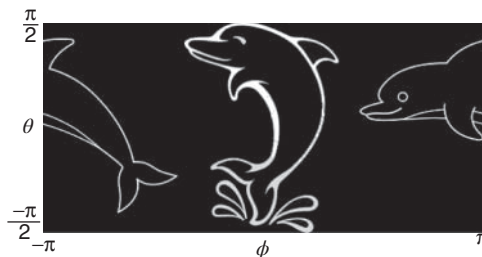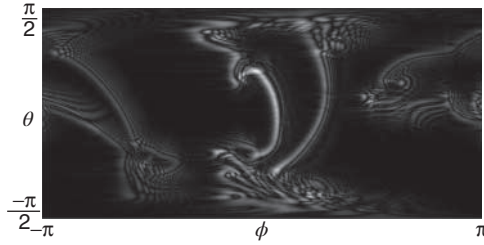

Figure 14.13   Object
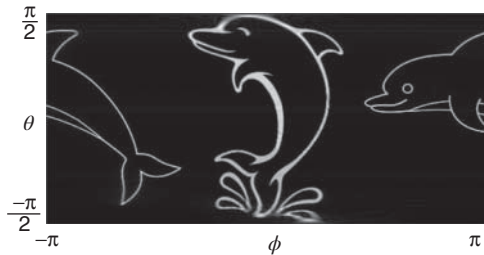
Figure 14.14   Hologram (intensity)

Figure 14.15   Reconstruction

From this hologram the object was reconstructed back onto the original spherical surface using Eq. 14.41. Reconstruction with the original reference will only produce a virtual image at the location of the object. In order to obtain a real reconstructed image at the original object location, the hologram should be illuminated (or reconstructed) using the conjugate of the reference. This means that we are attempting to reconstruct a real image on the spherical surface where the object was earlier located. The conjugate of the reference was produced by taking the complex conjugate of the reference wave field matrix. Accordingly, the numerical procedure of the for reconstruction is expressed in an abstract form as shown next.

$$Reconstruction = |ISHT[SHT(Hologram \times Conjugate[Reference]) \times TF]|^2.$$

The reconstructed real image is shown in Fig. 14.15. The reconstruction matches exactly with the object chosen. The reconstruction is crisp and is also free from any noise. As mentioned earlier, the object and hologram are square integrable band-limited functions on a closed surface. Hence, a rotated (shifted in theta or phi) version of the object or hologram will produce a rotated version of the reconstruction. The wave propagation calculation requires $O(N(\log N))^2$ operations for N sampling points, and hence it is a fast computation formula. The calculations were executed using a scripted language-python on a Dell Precision T7400 machine with 12 GB of RAM memory. A calculation time for the direct integration, convolution, and spectral methods are 3730 s, 0.057 s, and 0.039 s, respectively. This means that the proposed method took the least time for calculation, and hence is the fastest.

14.5   Conclusion

We have revealed the information inherent to the diffracted wavefront by the derivation of the relation with the 3D Fourier spectrum based on the Helmholtz equation and its Green function. The diffracted wavefront propagating in all directions can be expressed completely by just the spherical Fourier spectrum. Moreover, as an example of the application of this basic principle, the simulation of the diffracted images to all directions has been demonstrated and the validity of this principle has been also verified. This principle gives us the basis of the diffraction calculation based on the Fourier spectrum and it is flexibly applicable to other various diffractions, such as the rotation of wavefronts, cylindrical observation, and so on.

References

[1] Lohmann A. W. and D. P. Paris, "Binary Fraunhofer holograms, generated by computer," *Appl. Opt.,* **6**, 1739–1748 (1967).

[2] Leseberg D., "Sizable Fresnel-type hologram generated by computer," *J. Opt. Soc. Amer.*, **6**, 229–233 (1989).

[3] Yoshikawa N., M. Itoh, and T. Yatagai, "Interpolation of reconstructed image in Fourier transform computer-generated hologram," *Opt. Commun.*, **119**, 33–40 (1995).

[4] Leseberg D. and C. Frére, "Computer-generated holograms of 3-D objects composed of tilted planar segments," *Appl. Opt.*, **27**, 3020–3024 (1988).

[5] Matsushima K., H. Schimmel, and F. Wyrowski, "Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves," *J. Opt. Soc. Am. A,* **20**, 1755–1762 (2003).

[6] Yu L., Y. An, and L. Cai, "Numerical reconstruction of digital holograms with variable viewing angles," *Optics Express*, **10**, 1250–1257 (2002).

[7] Soares O. D. D. and J. C. A. Fernandes, "Cylindrical hologram of 360° field of view," *Appl. Opt.*, **21**, 3194–3196 (1982).

[8] Rosen J., "Computer-generated holograms of images reconstructed on curved surfaces," *Appl. Opt.*, **38**, 6136–6140 (1999).

[9] Sando Y., M. Itoh and T. Yatagai, "Fast calculation method for cylindrical computer-generated holograms," *Opt. Express,* **13**, 1418 (2005).

[10] Jackin B. J. and T. Yatagai, "Fast calculation method for computer-generated cylindrical hologram based on wave propagation in spectral domain," *Opt. Express,* **18**, 25546–25555 (2010).

[11] Kak A. C. and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, New York, 1988).

[12] Tachiki M. L., Y. Sando, M. Itoh, and T. Yatagai, "Fast calculation method for spherical computer-generated holograms," *Appl. Opt.*, **45**, 3527–3533 (2006).

[13] Chien R. J. and B. K. Alpert. "A fast spherical filter with uniform resolution." *J. Comput. Phys.*, **136**, 580–584 (1997).

[14] Healy Jr D. M., D. Rockmore, P. J. Kostelec, and S. Moore. "FFTs for the 2-sphere–improvements and variations" *J. Fourier. Anal. Appl.* **9**, 341–385 (1998).

[15] Driscoll J. R. and D. M. Healy. "Computing Fourier transforms and convolutions on the sphere," *Adv. Appl. Math.*, **15**, 201–250 (1994).

[16] Swarztrauber P. N. "On computing the points and weights for Gauss–Legendre quadrature," *SIAM. J. Sci. Computing,* **24**, 945–954 (2002).

[17] Lebedev N.N. *Special Functions and their Applications*. Prentice Hall (1965).

[18] Arfken G.B. *Mathematical Method for Physicisst*. Academic Press, p. 702 (2001).

[19] Goodman J.W., *Introduction to Fourier Optics*, 3rd edn. Roberts and Company Publishers, (2004).

[20] Wieczorek M. SHTools. Website, available at: URL http://shtools.ipgp.fr/ (accessed December 6, 2013).