coherent spread function, and its incoherent transfer function is equal to the autocorrelation of its coherent transfer function.

The fundamental relations derived in this chapter are stated as theorems and used in the succeeding chapters to obtain some practical results for imaging systems with *circular*, *annular*, and *Gaussian pupils*.

1.2 RAYLEIGH-SOMMERFELD THEORY OF DIFFRACTION AND THE HUYGENS-FRESNEL PRINCIPLE

Many scalar (as opposed to vector) theories of diffraction have been proposed over the years to explain what is and might be observed when an optical system forms the image of an object.[2] We discuss only one of them, namely, the Rayleigh-Sommerfeld theory,[3] and obtain the Fresnel-Huygens principle from it. The concept of aberrations of a diffracted wave is introduced, their Fresnel and Fraunhofer approximations are discussed, and the transfer function of free space is obtained. This theory is adequate for what we need to discuss in terms of the diffraction theory of image formation, including the effects of aberrations of an imaging system. The following two theorems are derived in this section.

Theorem 1. *As a wave propagates, its disturbance according to the Huygens-Fresnel principle is given by the superposition of secondary spherical wavelets weighted by the amplitudes at the points where they originate on the wave.*

Theorem 2. *Under certain approximations, the propagation of a wave is described by a Fourier transform of its complex amplitude modified by a quadratic phase factor in the Fresnel approximation, or without modification in the Fraunhofer approximation.*

1.2.1 Rayleigh-Sommerfeld Formula

Consider an optical wave of wavelength λ and wavenumber $k = 2\pi/\lambda$ propagating in the z direction, as illustrated in Figure 1-1. Suppose we are given the complex amplitude $U(\vec{r};0)$ of the wave in the plane $z = 0$, where $\vec{r} = (x, y)$ is the 2D position vector of a point in a plane defined by the z value. The complex amplitude $U(\vec{r};z)$ in a plane at a distance z can be determined by solving the Helmholtz scalar wave equation for free-space propagation, namely,

$$\left(\nabla^2 + k^2\right)U(\vec{r};z) = 0 \quad , \tag{1-1}$$

where ∇^2 is the 3D Laplace operator, i.e.,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad . \tag{1-2}$$

In Figure 1-1, $\vec{r}\,'$ is used in the $z = 0$ plane to distinguish it from the $\vec{r}$ in the z plane, as the two planes become the source and observation planes, respectively.

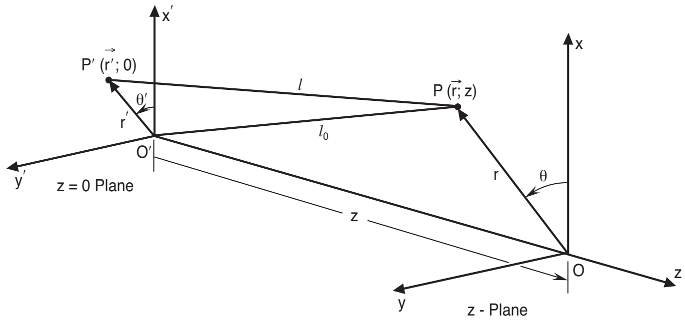

Figure 1-1. Geometry of wave propagation for determining the complex amplitude $U(\vec{r}; z)$ in a z plane from its knowledge $U(\vec{r}\,'; 0)$ in the $z = 0$ plane. $O'P = l_0$, $P'P = l$ and $O'O = z$.

To relate $U(\vec{r}; z)$ to $U(\vec{r}; 0)$, we decompose both of them into plane waves. Thus, for example, we write

$$U(\vec{r}; 0) = \int A(\vec{v}; 0) \exp(-2\pi i \vec{r} \cdot \vec{v})\, d\vec{v} \quad , \tag{1-3}$$

where $A(\vec{v}; 0)$ is the amplitude of a plane wave propagating with direction cosines (α, β, γ) such that the spatial frequency $\vec{v}$ is given by

$$\vec{v} = \frac{1}{\lambda}(\alpha, \beta) \tag{1-4}$$

and

$$\alpha^2 + \beta^2 + \gamma^2 = 1 \quad . \tag{1-5}$$

$A(\vec{v}; 0)$ is also referred to as the spectral component of $U(\vec{r}; 0)$ with a spatial frequency $\vec{v}$. Written in the form $A(\alpha/\lambda, \beta/\lambda; 0)$, it is also called the *angular spectrum* of $U(\vec{r}; 0)$. [Equation (1-4) may be obtained by comparing the exponential in Eq. (1-3) with the standard form of a plane wave $\exp\left[i\left(\omega t - \vec{k} \cdot \vec{r}\right)\right]$, where ω and $\vec{k}$ are its angular frequency and propagation vector, and t is time.]

Multiplying both sides of Eq. (1-3) by $\exp(2\pi i \vec{r} \cdot \vec{v}\,')$ and integrating, we obtain

$$\begin{aligned}\int U(\vec{r}; 0) \exp(2\pi i \vec{r} \cdot \vec{v}\,')\, d\vec{r} &= \int A(\vec{v}; 0)\, d\vec{v} \int \exp\left[2\pi i \vec{r} \cdot (\vec{v}\,' - \vec{v})\right] d\vec{r} \\ &= \int A(\vec{v}; 0)\, \delta(\vec{v} - \vec{v}\,')\, d\vec{v} \\ &= A(\vec{v}\,'; 0) \quad ,\end{aligned}$$

or

$$A(\vec{v}; 0) = \int U(\vec{r}; 0) \exp(2\pi i \vec{r} \cdot \vec{v})\, d\vec{r} \quad , \tag{1-6}$$

where $\delta(\cdot)$ is a Dirac delta function. It is clear from Eqs. (1-3) and (1-6) that $U(\vec{r};0)$ and $A(\vec{v};0)$ form a 2-D Fourier-transform pair; $U(\vec{r};0)$ is the inverse Fourier transform of $A(\vec{v};0)$ and $A(\vec{v};0)$ is the Fourier transform of $U(\vec{r};0)$. (For a definition of the Fourier transform, see the Appendix.) Similarly, we may write

$$U(\vec{r};z) = \int A(\vec{v};z)\exp(-2\pi i\vec{r}\cdot\vec{v})d\vec{v} \quad , \tag{1-7}$$

where $A(\vec{v};z)$ is the amplitude of a plane-wave component of spatial frequency $\vec{v}$ in the z plane given by

$$A(\vec{v};z) = \int U(\vec{r};z)\exp(2\pi i\vec{r}\cdot\vec{v})d\vec{r} \quad . \tag{1-8}$$

Since $U(\vec{r};z)$ satisfies the wave equation, which is linear, each component of the plane waves that comprise it must also satisfy it. Hence, substituting $A(\vec{v};z)\exp(-2\pi i\vec{r}\cdot\vec{v})$ for it into Eq. (1-1), we obtain

$$\gamma^2 k^2 A(\vec{v};z) + \frac{\partial^2 A(\vec{v};z)}{\partial z^2} = 0 \quad . \tag{1-9}$$

Solving Eq. (1-9), we find that

$$A(\vec{v};z) = A(\vec{v};0)\exp(ik\gamma z) \quad , \tag{1-10}$$

showing how the amplitude of a plane wave component changes as it propagates. We note that only its phase changes. Substituting Eq. (1-10) into Eq. (1-7), we obtain

$$U(\vec{r};z) = \int A(\vec{v};0)\exp(ik\gamma z)\exp(-2\pi i\vec{r}\cdot\vec{v})d\vec{v} \quad . \tag{1-11}$$

Let $h(\vec{r};z)$ be the inverse Fourier transform of $\exp(ik\gamma z)$, i.e.,

$$h(\vec{r};z) = \int \exp(ik\gamma z)\exp(-2\pi i\vec{r}\cdot\vec{v})d\vec{v} \quad , \tag{1-12}$$

so that

$$\exp(ik\gamma z) = \int h(\vec{r};z)\exp(2\pi i\vec{r}\cdot\vec{v})d\vec{r} \quad . \tag{1-13}$$

Substituting Eq. (1-13) into Eq. (1-11), we obtain

$$\begin{aligned} U(\vec{r};z) &= \int h(\vec{r}\,';z)d\vec{r}\,'\int A(\vec{v};0)\exp\left[-2\pi i(\vec{r}-\vec{r}\,')\cdot\vec{v}\right]d\vec{v} \\ &= \int h(\vec{r}\,';z)U(\vec{r}-\vec{r}\,';0)d\vec{r}\,' \quad , \end{aligned} \tag{1-14}$$

where we have used Eq. (1-3). By change of variables, Eq. (1-14) containing the convolution integral can also be written

$$U(\vec{r};z) = \int h(\vec{r}-\vec{r}\,';z)U(\vec{r}\,';0)d\vec{r}\,' \quad . \tag{1-15}$$

The integrals in Eqs. (1-14) and (1-15) giving the wave field $U(\vec{r};z)$ atthe z plane represent a convolution of $h(\vec{r}; z)$ and the wave field $U(\vec{r}; 0)$ in the $z = 0$ plane.

If a point source of unit amplitude is placed at a point $\vec{r}_o$ in the $z = 0$ plane, i.e., if

$$U(\vec{r}\,'; 0) = \delta(\vec{r}\,' - \vec{r}_o) \quad , \tag{1-16}$$

then Eq. (1-15) shows that the complex amplitude at a point $\vec{r}$ in the z plane is given by

$$U(\vec{r}; z) = h(\vec{r} - \vec{r}_o; z) \quad . \tag{1-17}$$

Thus, $h(\vec{r}; z)$ represents the complex amplitude at a point $\vec{r}$ in the z plane due to a point source of unit amplitude located at $\vec{r} = 0$ in the $z = 0$ plane. It is called the complex amplitude *point-spread function* (or *impulse response*) *of free space*. Carrying out the integration in Eq. (1-12), it can be shown that[4]

$$h(\vec{r}; z) = \frac{1}{\lambda}\left(\frac{1}{kl_0} - i\right)\frac{z}{l_0}\frac{\exp(ikl_0)}{l_0} \quad , \tag{1-18}$$

where

$$l_0 = \left(z^2 + r^2\right)^{1/2} \tag{1-19}$$

is the distance $O'P$ between the origin in the $z = 0$ plane and the observation point $(\vec{r}; z)$. Equation (1-18) is a mathematical expression for a Huygens' spherical wavelet diverging from the point source, and it describes the complex amplitude point-spread function of free space. We note from Eq. (1-12) that $h(\vec{r}; z) \to \delta(\vec{r})$ as $z \to 0$, i.e., it becomes the point source, as expected. Accordingly, Eq. (1-15) is a mathematical description of Huygens-Fresnel principle, namely, that *the complex amplitude in the z plane is a linear superposition of Huygens*' *secondary spherical wavelets* $h(\vec{r} - \vec{r}\,'; z)$ *weighted by the amplitudes* $U(\vec{r}\,'; 0)$ *of the wave where they originate* (Theorem 1). The diffracted wave field described by Eq. (1-15) is shift invariant (or isoplanatic) in that a spherical wavelet at a point $\vec{r}$ due to a source point at $\vec{r}\,'$ depends on $\vec{r} - \vec{r}\,'$, i.e., the form of a spherical wavelet is independent of the location of its origin in the $z = 0$ plane, except for a shift in the center of the distribution.

Substituting Eq. (1-18) into Eq. (1-15), we obtain

$$U(\vec{r}; z) = \frac{1}{\lambda}\int U(\vec{r}\,'; 0)\left(\frac{1}{kl} - i\right)\frac{z}{l}\frac{\exp(ikl)}{l}\,d\vec{r}\,' \quad , \tag{1-20}$$

where (see Figure 1-1)

$$l = \left[z^2 + |\vec{r} - \vec{r}\,'|^2\right]^{1/2} \tag{1-21a}$$

$$= z + \frac{1}{2z}\left(r^2 + r'^2 - 2\vec{r}\cdot\vec{r}\,'\right) - \frac{1}{8z^3}|\vec{r} - \vec{r}\,'|^4 + \ldots \tag{1-21b}$$

is the distance $P'P$ between a source point $(\vec{r}\,';0)$ and the observation point $(\vec{r};z)$. Equation (1-20) is the *Rayleigh-Sommerfeld formula* describing propagation of a wave from one plane to another.

1.2.2 Fresnel and Fraunhofer Approximations

For large values of z, $kl >> 1$, $l \simeq z$ (so that the obliquity factor z/l, representing the cosine of the angle $P'P$ makes with the z axis, may be assumed to be unity) except in the exponent where we retain additional terms according to Eq. (1-21b) to within a fraction of a wavelength (since they are multiplied by k). Hence, Eq. (1-20) may be written

$$U(\vec{r};z) = \frac{-i}{\lambda z}\exp\left[ik\left(z+\frac{r^2}{2z}\right)\right]\int \exp\left(\frac{ikr'^2}{2z}\right)U(\vec{r}\,';0)\exp\left(-\frac{2\pi i}{\lambda z}\vec{r}\cdot\vec{r}\,'\right)d\vec{r}\,' \quad ,$$

$$\text{for } z^3 >> k\left|\vec{r}-\vec{r}\,'\right|^4_{\max}\big/8 \quad \text{(Fresnel)} \quad , \qquad \text{(1-22a)}$$

and

$$U(\vec{r};z) = \frac{-i}{\lambda z}\exp\left[ik\left(z+\frac{r^2}{2z}\right)\right]\int U(\vec{r}\,';0)\exp\left(-\frac{2\pi i}{\lambda z}\vec{r}\cdot\vec{r}\,'\right)d\vec{r}\,' \quad ,$$

$$\text{for } z >> kr'^2_{\max}\big/2 \quad \text{(Fraunhofer)} \quad . \qquad \text{(1-23a)}$$

Thus, depending on the value of the distance z relative to the extent of the regions of the source field and observation, *the complex amplitude in a z plane is proportional to the inverse Fourier transform of the complex amplitude in the* $z = 0$ *plane with or without modification by a quadratic phase factor* $kr'^2/2z$ (Theorem 2). Equations (1-22a) and (1-23a) represent diffraction integrals in the *Fresnel* and *Fraunhofer approximations*, respectively. It should be noted that the Fresnel condition of large distance z in Eq. (1-22a) is a sufficient but not a necessary condition. What is necessary is that the sum of the neglected terms be small so that their contribution to the diffraction integral is negligible.

The integrals in Eqs. (1-22a) and (1-23a) are referred to as representing the *Fresnel* and *Fraunhofer diffraction patterns* of the distribution $U(\vec{r}\,';0)$. The region of space satisfying the condition $z >> kr'^2_{\max}/2$ is called the Fraunhofer or the *far-field region of diffraction*. The condition itself is called the Fraunhofer or the *far-field condition*. The region of space satisfying the Fresnel condition $z^3 >> k\left|\vec{r}-\vec{r}\,'\right|^4_{\max}/8$ but not the Fraunhofer condition is often referred to as the region of Fresnel or *near-field diffraction*. The region of very small z values is referred to as the Rayleigh-Sommerfeld region of diffraction. It should be evident, though, that the Rayleigh-Sommerfeld integral in Eq. (1-20) will yield accurate results (within the range of its validity) regardless of the value of z. Similarly, the Fresnel integral will yield accurate results in the Fraunhofer region as well. However, calculations in the Fraunhofer region are simpler and more common in imaging applications (due to cancellation of the quadratic phase factor of free-space propagation by the focusing quadratic phase factor provided by the imaging system).