OPTICAL HOLOGRAPHY
Student Edition

Robert J. Collier
Christoph B. Burckhardt
Lawrence H. Lin

Bell Telephone Laboratories
Murray Hill, New Jersey

ACADEMIC PRESS  New York  San Francisco  London
A Subsidiary of Harcourt Brace Jovanovich, Publishers
Chapter 3

GEOMETRIC ANALYSIS OF POINT-SOURCE HOLOGRAMS

In Chapter 2 the Gabor in-line hologram and the Leith–Upatnieks off-axis hologram were described in terms of the interference of light coming from two point sources. Despite the simplifications involved, such point-source holograms clearly illustrate many of the basic features of holography. From them much can be learned about the spacing of the interference fringes to be recorded, about properties of the virtual and real images that are generated, and about the magnification obtainable in the reconstruction process.

There are no point sources of light in the physical world, but the extended sources and illuminated subjects we do encounter can be thought of as collections of point sources. Let us suppose that \( a_1, a_2, \ldots, \) etc. represent the complex amplitudes of light waves arriving at the hologram plane from one such collection of subject point sources. If \( r \) represents the reference light complex amplitude arriving at the hologram, then the total complex amplitude at the hologram is

\[
a_1 + a_2 + \cdots + r.
\]

The quantity important to the holographic recording process is the total intensity

\[
I = (a_1 + a_2 + \cdots + r)(a_1^* + a_2^* + \cdots + r^*)
\]

\[
= a_1a_1^* + a_2a_2^* + \cdots + rr^* + (a_1a_2^* + a_2a_1^* + \cdots)
\]

\[
+ (a_1^*a_2 + a_2^*a_1 + \cdots) + r^*(a_1 + a_2 + \cdots).
\]

Apart from the intermodulation terms (the terms \( a_1a_2^* + a_2a_1^* + \cdots \), which indicate interference among the subject wave components, the subject point sources behave independently. Concern with effects of the intermodulation terms on reconstruction can be removed in the ways touched on in Chapter 2. For the case of in-line holography, the reference wave amplitude is made much larger than the subject wave amplitude. The intermodulation terms then become negligible. In the case of off-axis holography, the angle between subject and reference waves is chosen large enough to angularly separate the image waves from the waves diffracted by the intermodulation terms. (The latter waves propagate in directions close to that of the illuminating beam.)

Since the subject point sources can for our purposes be regarded as independent, we restrict our attention here to a single subject point source and to the holograms formed with a reference point source. We assume all waves travel from left to right. We further assume that (1) all illumination is perfectly coherent, (2) the holograms are exposed and developed to produce an amplitude transmittance proportional to the intensity of the interference pattern, and (3) the holograms behave as plane diffraction gratings.

The reader who anticipates making his own holograms and seeks to observe the results and properties described in this chapter should be forewarned regarding the choice of recording media. Many of the effects to be discussed, e.g., simultaneous observation of real and virtual images and the results of illuminating the hologram at a different angle or with a different wavelength than that used to form it, can best be observed when the hologram truly behaves as a plane diffraction grating. The high-resolution photographic emulsions normally used in holography range from 6 to 15 \( \mu m \) in thickness; therefore small reference-to-subject beam angles must be employed to avoid the angular and spectral selective properties of volumegrating behavior. One recording material thin enough to behave essentially as a plane diffraction grating is thermoplastic ([3.1], see also Chapter 10). Holograms formed in it will exhibit the properties we shall analyze.

### 3.1 Computation of Subject–Reference Phase Differences

In the analysis to follow, a subject, reference, or illuminating wave arriving at any point \( Q \) on the hologram plane in Fig. 3.1 is to be represented by the difference in its phase at \( Q \) over its phase at a fixed point of origin \( O \). (It is assumed that the amplitude of the spherical wave arising
from each point source is approximately uniform across the hologram plane and therefore plays no part in the essential process.) If we assume that the space on either side of the hologram has the same index of refraction and if the hologram can be regarded as very thin, then relative phase can be computed from geometrical light-path differences. We shall employ the basic analysis introduced by Meier [3.2]. It involves making paraxial approximations. (For a nonparaxial analysis see Champagne [3.3].)

Let \( a = a_0 \exp(i \varphi_a) \) be the complex amplitude of the light arriving at the hologram plane from the subject point source and let \( r = r_0 \exp(i \varphi_r) \) be the complex amplitude of the reference wave at the hologram. Then, as in Section 2.5 (but without restriction to the in-line configuration), the intensity recorded at the hologram plane is

\[
I = a_0^2 + r_0^2 + r^* a + r a^*
\]

We are primarily interested in the interference terms \( r a^* + r^* a \)

which describe the periodic spatial variation of intensity in the interference fringes. The periodicity of the hologram fringes is governed by the argument of the cosine, i.e., by the phase difference \( \varphi_r - \varphi_a \).

Consider now hologram formation as indicated in Fig. 3.1. The subject point source \( P \) is in the \( x_1 y_1 \) plane, \( z_1 = -d \) units from the origin \( O \) of the \( z \) axis. The origin \( O \) is located in the hologram \( x_2 y_2 z_2 \) plane. (The primes will be carried until we reach that stage of the analysis where enlargement of the hologram plane is considered.) A reference point source \( R \) is located in some arbitrary plane \( x_r y_r \), a distance \( z_r \) from the hologram plane. If \( R \) is to the left of the hologram plane and the reference wave diverges from \( R \), \( z_r \) is negative (as shown); if \( R \) is to the right of the hologram plane and the reference wave converges to \( R \), \( z_r \) is positive. We wish to compute \( \varphi_r - \varphi_a \) at an arbitrary point \( Q \) in the hologram plane. We can differentiate \((1/2\pi)(\varphi_r - \varphi_a)\) with respect to a spatial coordinate to determine the number of cycles of intensity variation per unit distance along the coordinate axis. We then know how many fringes per unit distance in that direction must be recorded by the photographic emulsion as a function of the arrangement of \( P, R \), and the recording plate.

The starting phases of the waves emanating from \( P \) and \( R \) are perfectly arbitrary. Let us assume that these have been adjusted so that each wave has the same phase at the point \( O \) in the hologram plane. We can call that phase value zero. Since \( P \) and \( R \) are point sources, they each emit a spherical wave whose phase at any point in space is proportional to the radial distance of that point from the source. Then, by computing the path difference \( PQ - PO \), we obtain the phase \( \varphi_a \) of the light complex amplitude at \( Q \) coming from \( P \). By similar computation we obtain the phase \( \varphi_r \) of the light complex amplitude at \( Q \) coming from \( R \). The sign of the phase at \( Q \) relative to that at \( O \) must be carefully considered. The two diagrams of Fig. 3.2 are useful in discussing this point. The magnitude of the phase difference \( \varphi_a \), corresponding to the path difference \( PQ - PO \), is \( |\varphi_a| = (2\pi/\lambda) |PQ - PO| \)

where \( \lambda \) represents the wavelength. If \( P \) is a real point source of diverging spherical waves and if \( PQ > PO \), then the wavefront arriving at \( Q \) was emitted at an earlier time than that simultaneously arriving at \( O \) (see Fig.
3. Geometric Analysis of Point-Source Holograms

3.2a). Therefore its phase at \( Q \) must be less than that of the front at \( O \) (assuming phase to increase with time), and consequently \( \varphi_a = -(2\pi/\lambda) \times (PQ - PO) \). In Fig. 3.2b a converging reference wave is indicated. Here \( R \) exists as an effective point focus on the opposite side of the hologram from \( P \). For \( RQ > RO \), the phase of the wavefront at \( Q \) is greater than that at \( O \) since the wavefront at \( Q \) was emitted later. Thus, for a converging reference wave, \( \varphi_a = -(2\pi/\lambda)(RQ - RO) \), while for the usual case of a diverging reference wave, \( \varphi_a = -(2\pi/\lambda)(RO - RQ) \).

We can now return to the computation of \( \varphi_a - \varphi_s \) for the case where \( P \) and \( R \) are each sources of diverging light on the same side of the hologram. For the phase of the subject wave at \( Q \), \( \varphi_s(x'_s, y'_s) \), the reference phase at \( Q \), obtained in a fashion analogous to that used for \( \varphi_s \), is

\[ \varphi_a = \frac{-2\pi}{\lambda_1} (PQ - PO) \]

\[ = \frac{-2\pi}{\lambda_1} \left\{ [ (x'_s - x_i)^2 + (y'_s - y_i)^2 + z_i^2 \right\}^{1/2} - \left\{ [x_i^2 + y_i^2 + z_i^2] \right\}^{1/2} \]

\[ = \frac{-2\pi}{\lambda_1} \left\{ \left[ 1 + \frac{(x'_s - x_i)^2 + (y'_s - y_i)^2}{z_i^2} \right]^{1/2} \right\} \]

\[ - \left\{ \left[ 1 + x_i^2 + y_i^2 \right]^{1/2} \right\} \]

where \( \lambda_1 \) is the wavelength of the light used to form the interference pattern and where we understand \( z_i \) to have a negative value so that the sign of \( \varphi_a \) remains negative. (Thus the character of the wave, i.e., diverging or converging, is carried implicitly in \( z_i \).) If both \( P \) and \( Q \) are not far off the \( z \) axis, and if \( z_i \) is large enough, \( \varphi_a \) can be approximated to the first order in \( 1/z_i \) by

\[ \varphi_a = \frac{2\pi}{\lambda_1} \left[ \frac{1}{2z_i} (x'_s^2 + y'_s^2 - 2x'_s x_i - 2y'_s y_i) \right]. \]  (3.3)

The next higher-order terms in the binomial expansion are third-order in \( 1/z_i \). The first-order approximation will be satisfactory for most of the purposes of this chapter. We shall point out when an expression derived in this chapter with the aid of Eq. (3.3) differs, because of the approximation, from that obtained through other considerations. \( \varphi_s(x'_s, y'_s) \), the reference phase at \( Q \), obtained in a fashion analogous to that used for \( \varphi_s \), is

\[ \varphi_s = \frac{2\pi}{\lambda_1} \left[ \frac{1}{2z_i} (x'_s^2 + y'_s^2 - 2x'_s x_i - 2y'_s y_i) \right]. \]  (3.4)

The subject-reference phase difference at \( Q \) is then given by

\[ \varphi_s - \varphi_a = \frac{2\pi}{\lambda_1} \left[ (x'_s^2 + y'_s^2) \left( \frac{1}{2z_i} - \frac{1}{2z_i} \right) \right] \]

\[ - x'_s \left[ \frac{x_i}{z_i} - \frac{x_i}{z_i} \right] - y'_s \left[ \frac{y_i}{z_i} - \frac{y_i}{z_i} \right] \]

\[ = \frac{2\pi}{\lambda_1} \Delta l. \]  (3.5)

The quantity in the brackets is the path difference \( \Delta l \) for light traveling to \( Q \) from \( P \) as against that traveling to \( Q \) from \( R \).

3.1.1 THE IN-LINE HOLOGRAM

Both subject and reference point sources are on-axis in the case of the in-line hologram so that in Eq. (3.5), \( x_i, y_i, x'_s, \) and \( y'_s \) are all zero. If we relabel \( z_i = -u \) and \( z_i = -v \) to correspond with the notation of Section 2.4.1, then the path difference in Eq. (3.5) becomes

\[ \Delta l = (x'_s^2 + y'_s^2) \left( \frac{1}{2z_i} - \frac{1}{z_i} \right) \]

\[ = (x'_s^2 + y'_s^2) \left( \frac{1}{u} - \frac{1}{v} \right) = \frac{x'_s^2 + y'_s^2}{2f} = \frac{\theta^2}{2f} \]  (3.6)

where we have used Eq. (2.1), \( f^{-1} = u^{-1} - v^{-1} \), and where \( \theta \) is the radial distance from the origin \( O \) at the hologram plane. The bright fringes of the interference pattern occur whenever \( \Delta l = n\lambda_1 \) where \( n \) is an integer. Since \( \Delta l \) is symmetric about the origin, the fringes are circular and form the zone plate pattern given by

\[ \Delta l = \frac{x'_s^2 + y'_s^2}{2} \left( \frac{1}{u} - \frac{1}{v} \right) = \frac{\theta^2}{2f} = n\lambda_1. \]  (3.7)

As stated in Eq. (3.2) the spatial variation of the interference pattern intensity is governed by a cosine function, \( \cos(\varphi_s - \varphi_a) = \cos(2\pi \Delta l / \lambda_1) \). If \( \Delta l \) were linearly dependent on the spatial variable, then a constant frequency could be ascribed to the cosinusoidal intensity variation. This is generally not the case, but a local spatial frequency, the fringe frequency \( \nu(\varphi) \), can be defined. (Here \( \varphi \) is the spatial variable measured along a direction perpendicular to the crests of the interference pattern, and \( \nu \) is considered to be a function of \( \varphi \).) We define \( \nu \) as the spatial rate of change of
the phase of the intensity pattern at $Q$ divided by $2\pi$ radians:
\[ r(\varphi) = \frac{\partial (\varphi_2 - \varphi_1)}{\partial \varphi} \cdot \frac{1}{2\pi} \frac{\partial (A_l)}{\partial \varphi}. \]  
(3.8)

For $A_l$ given in Eq. (3.7),
\[ r(\varphi) = \rho |\beta_1|. \]  
(3.9)

Thus, proceeding radially outward from the hologram center, the fringe frequency increases linearly with $\varphi$. At some value of $\varphi$, $v$ may exceed the resolution capability $v_m$ of the photosensitive medium. Such a radius defines the limiting aperture and the image resolution of the hologram.

When comparing the recording resolution requirements of in-line holograms with other hologram-forming configurations to be analyzed in this chapter, we shall find it convenient and sufficient to consider only $v_r$ the fringe frequency component of $v$ along the $x_2$ direction. As a further simplification we consider $R$ to be at infinity (a plane wave reference with $z_r = \infty$). For this case [see Eq. (3.6)]
\[ \xi' = \frac{\partial (\varphi_2 - \varphi_1)}{2\pi \partial \varphi} = -\frac{x_{2}'}{z_1 \lambda_1}. \]  
(3.10)

The farther the subject is from the hologram, the coarser and the more easily recorded are the fringes. Gabor attempted to apply this fact in his "projection method" (Section 2.2). Unfortunately image resolution decreases as well if $x_2'$ is limited.

### 3.1.2 The Off-Axis Hologram

The relation obtained by setting $A_l = n\lambda_1$ in Eq. (3.5),
\[ A_l = (x_2'^2 + y_2'^2)\left(\frac{1}{2}\right)\left(\frac{1}{z_1} - \frac{1}{z_2}\right) - x_2\left(\frac{x_{1}'z_{2}' - x_1z_{2}}{z_1 - z_2}\right) - y_2\left(\frac{y_{1}'z_{2}' - y_1z_{2}}{z_1 - z_2}\right) = n\lambda_1, \]  
(3.11)

is the general expression for a circle whose center has the coordinates
\[ x_2' = \frac{z_2 x_{2}' - z_1 x_1}{z_1 - z_2}, \quad y_2' = \frac{z_2 y_{2}' - z_1 y_1}{z_1 - z_2}, \]  
(3.12)
and whose radius $\rho$ is given by
\[ \rho^2 = \left(\frac{z_2 x_{2}' - z_1 x_1}{z_1 - z_2}\right)^2 + \left(\frac{z_2 y_{2}' - z_1 y_1}{z_1 - z_2}\right)^2 + \frac{2n\lambda_1 z_1 z_2}{z_1 - z_2}. \]  
(3.13)

Suppose we consider the off-axis intensity pattern formed by the interference of an axial plane wave reference ($x_r = y_r = 0, z_r = \infty$) with a spherical subject wave diverging from an off-axis point $(x_1, y_1, z_1)$. The center of the set of circular fringes, whose radii correspond to integral values of $n$ in Eq. (3.13), is given by Eq. (3.12) as $x_2' = x_1$ and $y_2' = 0$. A zone-plate pattern is thus centered at the foot of the perpendicular dropped from $P$ to the hologram plane. The situation is indicated in Fig. 3.3. If the photographic plate is centered at $O$ so that it can record only off-center portions of the interference pattern, a Leith–Upatnieks hologram is obtained. The fringe frequency $\xi'$ in the $x_2'$ direction can be found by differentiating $A_l/\lambda_1$ in Eq. (3.11) under the conditions $x_1 = y_1 = z_1 = 0$ and $z_2 = \infty$. The result is
\[ \xi' = -\frac{x_{2}'}{z_1 \lambda_1} + \frac{x_1}{z_1 \lambda_1}. \]  
(3.14)
Assuming the hologram plate to be centered at \( O \), let us compare \( \varepsilon' \) in Eq. (3.14) with that found for the in-line configuration, Eq. (3.10). At the hologram center \( (x'_0 = 0) \) the in-line fringe frequency is zero while the off-axis fringe frequency is \( x'_{0}/a_{1} \). As one proceeds outward in the negative \( x' \) direction (see Fig. 3.3), the frequency of each fringe system increases linearly with \( x' \) and this frequency difference is maintained. At the edge of the hologram are generated the highest frequencies to be recorded. If the photosensitive medium on the hologram plate is to record the fringes in the off-axis case, it must have a resolution capability in addition to that required for the in-line case. Examination of Eq. (3.11) reveals that had we considered an off-axis reference plane wave \( x' \) of subject, the last term \( \pi_{0} \) in Eq. (3.15) becomes \( \pi_{1}/a_{1} \), which is the mean angle to the axis made by the subject wave, i.e., the angle to the \( z \) axis made by a ray passing from \( P \) to the center of the hologram at \( O \). Thus the mean angle between subject and reference beam provides the difference in the maximum fringe frequency generated in an off-axis hologram as against the in-line hologram.

In the practical case of an extended subject, either the width of the subject or the width of the recording plate can cause the fringe frequency \( \varepsilon' \) to exceed the plate resolution \( v_{m} \). When the plate is small compared to the subject, the last term in Eq. (3.14) \( [\pi_{1} - \pi_{0}]a_{1} \) in Eq. (3.15) is dominant. Point sources at the extreme dimension of the subject produce the maximum fringe frequency at the plate. If \( \varepsilon' > v_{m} \) for the extreme portions of the subject, then such portions are not recorded. On the other hand, when the plate is much larger than the subject, the dominant term in Eq. (3.14) might be the first. Beyond some value of \( x'_{0} \) all subject points produce zone-plate fringes whose frequency \( \varepsilon' > v_{m} \). That value of \( x'_{0} \) defines the practical extent of the hologram record.

### 3.1.3 Lensless Fourier Transform Hologram

We consider now the arrangement shown in Fig. 3.4 where subject and reference points are in the same plane. The coordinates of the subject point \( P \) are \( (x_{1}, y_{1} = 0, z_{1}) \) and those of the reference point \( R \) are \( (x_{e}, y_{e} = 0, z_{e} = z_{1}) \). The phase difference \( \varphi_{r} - \varphi_{a} \) in Eq. (3.5) becomes

\[
\varphi_{r} - \varphi_{a} = \frac{2\pi}{a_{1}} \left( \frac{x_{1}}{z_{1}} - \frac{x_{e}}{z_{e}} \right) x'_{0}.
\]

Differentiation of \( \varphi_{r} - \varphi_{a} = 2\pi x'_{0}/a_{1} \) with respect to \( x'_{0} \) yields the constant fringe frequency

\[
\varepsilon' = \frac{x_{1} - x_{e}}{z_{1} a_{1}}.
\]

Since the intensity of the interference pattern is independent of \( y' \), as is apparent from Eq. (3.16), the fringes are vertical, uniformly spaced, linear fringes. Their intensity varies cosinusoidally in the \( x' \) direction. (The physical arrangement and results are equivalent to those of Thomas Young's experiment. The method was suggested by Winthrop and Worthington for X-ray holography [3.4] and by Stroke for optical holography [3.5].)

As may be seen in Fig. 3.4, the ratio \( x_{1}/z_{1} = \tan \theta_{f} \approx \theta_{f} \) in our first-order approximation, and similarly the ratio \( x_{e}/z_{e} \approx \theta_{e} \). We can write \( \varphi_{r} - \varphi_{a} \) in Eq. (3.16) as

\[
\varphi_{r} - \varphi_{a} = \frac{2\pi}{a_{1}} (\theta_{f} - \theta_{e}) x'_{0},
\]

an expression depending only on the angle subtended at the hologram by the distance separating \( P \) and \( R \). If points \( P \) and \( R \) are at an infinite distance from the hologram \( (z_{e} = z_{1} = \infty) \), while \( x_{1}/z_{1} \approx \theta_{f} \) and \( x_{e}/z_{e} \approx \theta_{e} \) are still finite, Eq. (3.18) continues to hold. The waves arriving at the hologram from point sources at infinity are plane waves. They are the far-field pattern or Fourier transforms of the point sources. Hence the linear fringe system of Eq. (3.18) can be regarded as the interference of a plane wave reference with the Fourier transform of the subject point source \( P \). Not only can the hologram formed as in Fig. 3.4 be illuminated with the original reference point \( R \) to produce an image of \( P \), but it can equally well be illuminated by a plane wave to reconstruct a plane wave which is the Fourier transform of the point source \( P \).
68 3. Geometric Analysis of Point-Source Holograms

of \( P \). In the latter case the reconstructed wave must be observed in the far field to obtain the image of \( P \). The required second Fourier transformation can be performed optically by placing a lens adjacent to the hologram and observing the focal pattern in the back focal plane of the lens. A further discussion of the lensless Fourier transform hologram is given in Chapter 8.

By placing the reference source close to the subject, \( x_1 - x_i \) in Eq. (3.17) can be kept small and the fringe frequency \( \xi' \) kept low. To the extent that the approximations leading to Eq. (3.3) hold, \( \xi' \) is constant across the hologram plate, and holograms can be formed on plates of low resolution. For extended subjects, \( x_1 - x_i \) is a function of the width of the subject. Fringes produced by extreme portions of the subject may yet exceed the plate resolution, thus preventing these portions from being recorded. However, for small subjects the lensless Fourier transform hologram configuration produces a uniform, low-frequency fringe system over a large plate area. The result is a wide-aperture hologram which images with high resolution.

Equation (3.18) may be used to express the fringe separation, \( d = 1/\xi' \), for the interference pattern formed by the intersection of two plane waves:

\[
d = \lambda_i/(\theta_i - \theta_f). \tag{3.19}
\]

Suppose that as in Section 1.3.1, \( \theta_f = -\theta_i \). Substituting into Eq. (3.19) we obtain

\[
20_i d = \lambda_i \tag{3.20}
\]

which is a small-angle approximation to Eq. (1.10)

\[
2d \sin \theta = \lambda.
\]

That the lensless Fourier transform hologram is equivalent to the hologram record of two intersecting plane waves is understandable when one notes that in \( (\varphi_p - \varphi_s) \) are phase expressions for two spherical waves of equal but opposite curvature. Phase contributions due to curvature of the wavefronts cancel leaving only those contributions due to the difference in mean directions of the waves.

3.2 Reconstruction with a Point Source

Having considered various hologram-forming arrangements, let us now investigate the reconstruction process. We assume that it is possible to magnify or demagnify the hologram after formation and before reconstruction. To take account of this, the hologram plane coordinates are relabeled \( x_2 = mx_2' \) and \( y_2 = my_2' \), where \( m \) is the linear magnification. We also assume that the reconstructing wavelength \( \lambda_r \) need not be the same as the forming wavelength \( \lambda_i \); their ratio is given by \( \mu = \lambda_r/\lambda_i \). The reconstructing or illuminating wave originates from a point source \( C(x_c, y_c, z_c) \) as in Fig. 3.5. We do not require \( C \) to be the original reference source; it may be the source of a diverging wave or the focus of a converging wave.

When the hologram is properly recorded in photographic emulsion, its amplitude transmittance \( t \) is proportional to the intensity \( I \) given by Eq. (3.1) (see also Section 1.8), where the spherical wave intensities \( I_1 = a_0^2 \) and \( I_2 = r_0^2 \) are approximately uniform over the hologram plane. Hence for holograms of point sources, diffraction results only from illumination of the spatially varying interference terms in the transmittance proportional to

\[
a_0^* + ra^*.
\]

The complex amplitudes of the diffracted waves at the hologram plane are proportional to the products of the illumination complex amplitude \( e \) times the above transmittance terms,

\[
er_0^* + er^*a,
\]

where \( e = e_0 \exp(i\varphi_c) \). In Section 1.8 it is said that the first term above, containing \( a^* \), yields a real image while the second, containing \( a \), produces a virtual image. As we shall see, this is not always the case. Nevertheless the phase of the diffracted wave

\[
er_0^* = e_0 r_0 a_0 \exp[(i\varphi_0 + \varphi_2 - \varphi_a)]
\]
3. Geometric Analysis of Point-Source Holograms

is to be labeled
\[ \varphi_R = \varphi_c + \varphi_r - \varphi_a, \]  
(3.21)

while the phase of the wave \( e^{i \alpha} \) is to be labeled
\[ \varphi_V = \varphi_c - \varphi_r + \varphi_a. \]  
(3.22)

As in the computation of \( \varphi_a \), we can cause the phase \( \varphi_c \) of wave \( e^c \) to be zero at the origin \( O \) and calculate the relative phase at some arbitrary point \((x_2, y_2)\) on the hologram plane. Thus
\[ \varphi_c(x_2, y_2) = \frac{2\pi}{\lambda_2} \left[ \frac{1}{2x_2} (x_c^2 + y_c^2 - 2x_c x_2 - 2y_c y_2) \right]. \]  
(3.23)

The axial distance \( z_c \) can be either positive or negative corresponding to illumination by a converging or diverging wave respectively. We can now substitute into Eq. (3.22) the values of \( \varphi_c, \varphi_r \), and \( \varphi_a \) given in Eqs. (3.23), (3.3), and (3.4), respectively, to obtain
\[ \varphi_V = \frac{2\pi}{\lambda_2} \left[ \frac{1}{2x_2} (x_c^2 + y_c^2 - 2x_c x_2 - 2y_c y_2) \right]. \]  
(3.24)

Similarly
\[ \varphi_B(x_2, y_2) = \frac{\pi}{\lambda_2} \left[ (x_2^2 + y_2^2) \left( \frac{1}{z_2} + \frac{\mu x_2}{m z_2} - \frac{\mu y_2}{m z_2} \right) - 2 \left( \frac{x_2}{z_2} + \frac{\mu x_2}{m z_2} - \frac{\mu y_2}{m z_2} \right) \right]. \]  
(3.25)

If the hologram is indeed to image the point-source subject \( P \), the phase \( \varphi_V \) and \( \varphi_B \) must correspond to those of spherical waves. A first-order approximation to a spherical-wave phase distribution over the hologram can be written, as in Eq. (3.3),
\[ \varphi(x_2, y_2) = \frac{2\pi}{\lambda_2} \left[ \frac{1}{2x_2} (x_c^2 + y_c^2 - 2x_c x_2 - 2y_c y_2) \right]. \]  
(3.26)

In the above equation, \( z_2 \) is the hologram-to-image plane separation while \( x_2 \) and \( y_2 \) represent the coordinates of the image point \( P \) in the image plane (see Fig. 3.5). We must try to arrange \( \varphi_V \) and \( \varphi_B \) to have the same form as \( \varphi \). If this can be done, then the image waves are, to first order, spherical and converge or diverge according to the signs of \( \varphi_V \) and \( \varphi_B \). They represent the case of perfect first-order imaging of the point source. The higher-order terms, neglected in the expansion of \( \varphi, \varphi_V, \) and \( \varphi_B \), however, may differ and so represent aberrations (see Section 3.4).

3.3 Characteristics of the Images

By factoring out the coefficient of \((x_2^2 + y_2^2)\) in \( \varphi_V \) and \( \varphi_B \), one can produce the desired form, indicating perfect first-order imaging. The image coordinates \((x_{2v}, y_{2v}, z_{2v})\) for \( \varphi_V \) and \((x_{2b}, y_{2b}, z_{2b})\) for \( \varphi_B \) can be identified as
\[ z_{2v} = \left( \frac{1}{z_2} + \frac{\mu}{m z_2} \right)^{-1} \left( \frac{m^2 x_2 z_2}{z_2} + \frac{\mu}{m z_2} \right), \]  
(3.27)

\[ x_{2v} = \frac{m^2 x_2 z_2}{z_2} + \frac{\mu}{m z_2} + \frac{\mu x_2}{m z_2}, \]  
(3.28)

\[ y_{2v} = \frac{m^2 y_2 z_2}{z_2} + \frac{\mu}{m z_2} + \frac{\mu y_2}{m z_2}, \]  

\[ z_{2b} = \left( \frac{1}{z_2} + \frac{\mu}{m z_2} \right)^{-1} \left( \frac{m^2 x_2 z_2}{z_2} + \frac{\mu}{m z_2} \right), \]  

\[ x_{2b} = \frac{m^2 x_2 z_2}{z_2} + \frac{\mu}{m z_2} + \frac{\mu x_2}{m z_2}, \]  

\[ y_{2b} = \frac{m^2 y_2 z_2}{z_2} + \frac{\mu}{m z_2} + \frac{\mu y_2}{m z_2}. \]  

Along with the relations defining the image location, we can define expressions for the lateral magnification \( M_{lb} \) as
\[ M_{lb} = \frac{dx_2}{dx_1} = \frac{dy_2}{dy_1}. \]
from which we obtain

\[ M_{\text{lat},V} = m \left( 1 + \frac{m^2 z_1}{\mu z_0} - \frac{z_1}{z_0} \right)^{-1} \]

\[ M_{\text{lat},R} = m \left( 1 - \frac{m^2 z_1}{\mu z_0} - \frac{z_1}{z_0} \right)^{-1} \]

and the angular magnification \( M_{\text{ang}} \) as

\[ M_{\text{ang}} = \frac{d(x_0/z_0)}{d(x_1/z_1)} \]

from which we obtain

\[ |M_{\text{ang}}| = \mu/m. \]

### 3.3.1 In-Line Hologram Images

The letters \( V \) and \( R \) refer to the reconstructed waves \( \mathbf{e}^* \mathbf{a} \) and \( \mathbf{e}^* \mathbf{a}^* \), respectively. Whether the images produced by these waves are actually virtual or real depends upon their divergence or convergence, i.e., upon the sign of \( z_V \) and \( z_R \). A negative value implies a diverging wave and a virtual image; a positive value implies a converging wave and a real image. (Note: \( z_1 \), the subject-to-hologram distance, has a negative value if the subject is a real object.) We shall begin our analysis of image characteristics with Gabor’s in-line holography where reference, subject, and illuminating sources are all on axis so that

\[ x_v = x_1 = z_0 = 0. \]

Only the \( x \) and \( z \) components of the image coordinates will be analyzed, since no new information is obtained by considering the \( \mu \) component.

Gabor’s “projection method” required that the subject be placed close to the source, i.e., \( z_0 = z_1 + \Delta \) where \( \Delta \) is a negative distance and where \( \Delta/z_1 \ll 1 \) (see Section 2.2). Suppose \( \mu = m = 1 \) and \( z_0 = z_1 \). (This corresponds to Gabor’s all-optical verification of his invention.) With the above \( x \) and \( z \) values Eqs. (3.27) and (3.28) give

\[ x_{SV} = 0, \quad x_{SV} = z_1 = z_0 - \Delta, \]

\[ x_{SR} = 0, \quad z_{SR} = \left( \frac{2}{z_0} - \frac{1}{z_1} \right)^{-1} = z_0 + \Delta. \]

The images lie close to and symmetric about the illuminating source. Since \( z_0 \) is negative, both images are virtual. To photograph these, Gabor needed a lens to form real images at a photographic plate.

For \( \mu = m = 1 \), the lateral magnification in Eq. (3.29) becomes

\[ M_{\text{lat},V} = \left[ 1 + z_1 \left( \frac{1}{z_0} - \frac{1}{z_1} \right) \right]^{-1}, \]

\[ M_{\text{lat},R} = \left[ 1 - z_1 \left( \frac{1}{z_0} + \frac{1}{z_1} \right) \right]^{-1}. \]

When the illuminating source is located at the original reference source position \( z_0 = z_1 \), then \( M_{\text{lat},V} = 1 \) and \( M_{\text{lat},R} = -1 \), remembering that \( z_0/z_1 \approx 1 \). (The \(-1\) magnification in the case of the virtual image produced by \( \varphi_R \) indicates an inverted image.) When the illuminating wavefront has less curvature at the hologram than the reference wave, i.e., \( |z_0| > |z_1| \), the lateral magnification increases. It becomes large, approaching \( z_0/\lambda \) as \( z_0 \to \infty \) and the illumination becomes a plane wave. However the corresponding value of the image plane distance \( z_0 \) becomes large as well. The angular magnification, \( M_{\text{ang}} = \mu/m = 1 \), remains constant.

Gabor’s original plan was to form the hologram at electron wavelengths and illuminate it at optical wavelengths. For this case \( \mu = \lambda_0/\lambda_e \approx 10^6 \). To avoid aberrations, he planned to scale the hologram by a factor \( \mu = m \) and to place the illuminating source at distance \( z_0 = mz_1 \) from the hologram. The lateral magnification under these conditions becomes

\[ M_{\text{lat}} = \pm m = \pm \mu, \]

but the angular magnification remains unity. Consequently the distance from hologram to image plane is \( \mu \) times the distance between object and hologram.

The essential feature of the “transmission method” of Haines and Dyson (see Section 2.2) was to place the subject close to the hologram so that \( |z_1| \ll |z_1| \). Again with \( x_v = x_1 = z_0 = 0, \mu = m = 1, \) and \( z_0 = z_1 \), we obtain from Eqs. (3.27) and (3.28)

\[ x_{SV} = 0, \quad x_{SV} = z_1, \]

\[ x_{SR} = 0, \quad z_{SR} = \left( \frac{2}{z_0} - \frac{1}{z_1} \right)^{-1}. \]

If the reference is a plane wave so that \( z_0 \to \infty \), then a virtual image is found at \( z_1 \) and a real image at \(-z_1\), the images symmetric this time about the hologram. Both images are upright since in this case \( M_{\text{lat},V} = M_{\text{lat},R} = +1 \).
3. Geometric Analysis of Point-Source Holograms

A result of general consequence, not confined to on-axis holography, can be obtained by considering the hologram to be illuminated by a plane wave \((z_e = \infty)\) of wavelength different from that used to form it \((\mu > 1)\).

With no scaling \((m = 1)\), the lateral magnification \(M_{lad} = (1 - Z_t/Z_s)^{-1}\) depends on the ratio \(Z_t/Z_s\) but is independent of the wavelength change. If, in addition, the reference wave had been a plane wave \((z_r = \infty)\), then no magnification whatsoever is obtainable in the reconstruction process. Of course scaling up the dimensions of the hologram can produce large lateral magnification even with plane reference and illuminating waves. However, optical enlargement of the hologram is an impractical and undesirable step in an otherwise lensless imaging process. It also places the image plane at a considerable distance from the hologram. For example, when \(z_e = z_r = \infty\) and \(\mu = m\), the axial distance of the real image is \(Z_{2R} = -mZ_1 = -\mu Z_1\).

3.3.2 LEITH–UPATNIKES OFF-AXIS HOLOGRAM IMAGES

Leith's and Upatnieks' method allows off-axis positions for the subject, reference, and illuminating sources, and there is no need to restrict the reference source to the \(z\) axis as we did in Section 3.1.2. Effects of illuminating the hologram with a wavelength and mean angle different from those of the reference beam can be illustrated easily by assuming plane waves for both reference and illuminating beams. In this section and throughout the remainder of the chapter we shall consider that the hologram dimensions are kept constant so that \(m = 1\). Again we confine our attention to the \(x\) and \(z\) image coordinates. With these simplifications Eqs. (3.27) and (3.28) reduce to

\[
x_{3y} = x_1 + \left(\frac{x_e}{z_e}\right)\frac{Z_1}{\mu} - \left(\frac{x_r}{z_r}\right)Z_1 = x_1 + z_1\left(\frac{\theta_e}{\mu} - \theta_r\right), \quad z_{3y} = \frac{Z_1}{\mu},
\]

\[
x_{3R} = x_1 - z_1\left(\frac{\theta_e}{\mu} + \theta_r\right), \quad z_{3R} = -\frac{Z_1}{\mu}
\]

where \(\theta_e = \tan \theta_e = x_e/z_e\) and \(\theta_r = \tan \theta_r = x_r/z_r\) are the angles that the illuminating and reference beams make with the positive direction of the \(z\) axis (see Fig. 3.6).

When the illuminating wave is identical to the reference, \(\mu = 1\) and \(\theta_e = \theta_r\). A virtual image then appears at the location of the original subject source \((x_1, Z_1)\), and a real image appears in a plane \(Z_1\) distant on the other side of the hologram from the illumination sources. Both images are upright.

As in Section 1.4, it is customary to describe the response of a diffraction grating to incident illumination in terms of the angles of incidence and diffraction. Since holograms discussed in this chapter are similar to plane diffraction gratings, it is appropriate to calculate the diffraction angles corresponding to illumination of a hologram with a plane wave making an angle \(\theta_i\) to the hologram plane. We can simplify matters further by considering \(\theta_e = \theta_r = 0\) as in Fig. 3.7. Then from Eq. (3.35) the diffraction angle corresponding to the \(\varphi_v\) diffracted wave is

\[
\theta_{3y} = x_{3y}z_{3y} = \frac{x_1}{z_1} = \theta_1 \quad \text{(a negative angle for} \ Z_1 < 0),
\]

\[
\theta_{3R} = x_{3R}z_{3R} = \frac{x_1}{z_1} = \theta_1
\]

Fig. 3.6. Illumination of Leith–Upatnieks hologram with plane wave \(c\). Hologram had been made with spherical wave \(a\) and plane wave reference \(r\). (Note: \(\theta_a\) and \(\theta_e\) are positive angles while \(\theta_r\) is negative.)

Fig. 3.7. Illumination of off-axis hologram with a plane wave \(c\) where \(\theta_e = \theta_r = 0\).
and the angle corresponding to $\theta_R$ is
\[ \theta_R = \frac{x_{3R}}{z_{3R}} = -\frac{x_1}{z_1} = -\varphi_i \quad (\text{a positive angle for } z_1 < 0). \]

The image waves are as shown in Fig. 3.7. The effect of a wavelength change in the illuminating beam is to multiply the diffraction angles by the wavelength ratio $\mu$. On the other hand for $\mu = 1$ but $\theta_y \neq \theta_i = 0$, the effect is to add $\theta_y$ to both $\theta_{3V}$ and $\theta_{3R}$, essentially rotating the emerging beams about the $y$ axis.

For the general case of an off-axis reference plane wave and a similar illuminating wave, the diffraction angles obtained from Eq. (3.35) are
\[ \begin{align*}
\theta_{3V} &= \mu \theta_i + \theta_y - \mu \theta_r \\
\theta_{3R} &= -\mu \theta_i + \theta_y + \mu \theta_r.
\end{align*} \tag{3.36} \]

A choice of angles often used in making holograms is $\theta_y = -\alpha$, $\theta_r = +\alpha'$ and $\theta_i = +\alpha$. If $\mu$ is kept equal to 1, the diffraction angles $\theta_{3V} = -\alpha$ and $\theta_{3R} = +3\alpha$ are as shown in Fig. 3.8.

Thus far, all of our results have been derived from a first-order approximation. Our analysis, when applied to elementary hologram formation and illumination with plane waves, should yield the familiar plane diffraction grating response, Eq. (1.11)
\[ d(\sin i + \sin \delta) = \lambda. \]

However the first-order analysis allows only a small angle approximation to Eq. (1.11) to be obtained. To illustrate, let us first restate Eq. (1.11) in a form pertinent to holographically formed gratings. Consider a grating to have been formed by interference of two plane waves on a photographic emulsion as in Fig. 1.4. For the case shown there, each forming angle can be set equal to $\theta_1$ so that $d = \lambda_1/(2 \sin \theta_1)$. Substituting $d$ into Eq. (1.11) and assuming that an illuminating plane wave is incident at an angle $i = \theta_1$, we obtain for the angle of diffraction $\delta$ (which is also the angle $\theta_y$ in the notation of this chapter)
\[ \sin \delta = \sin \theta_y = 2(\lambda_0/\lambda_1) \sin \theta_1 - \sin \theta_1 = \sin \theta_1(2\mu - 1). \tag{3.37} \]

Now let us evaluate the diffraction angle $\theta_{3V}$, in Eq. (3.36), under similar conditions. The conditions are met by setting $\theta_i = -\theta_i$ and $\theta_r = \theta_i$. We obtain
\[ \theta_{3V} = \theta_1(2\mu - 1), \tag{3.38} \]
the first-order approximation to Eq. (3.37). Thus, results obtained through first-order geometric considerations hold only so long as $\sin \theta = \tan \theta = \theta$.

### 3.3.3 Images When All Sources Are Equidistant from the Hologram

Suppose the subject and reference point sources lie in the same plane so that $z_1 = z_3$, and suppose that the illuminating wave is identical to the reference $(x = x_3, z = z_3)$. Equations (3.27) and (3.28) reduce to
\[ \begin{align*}
x_{3V} &= x_1(1 - \mu) + \mu x_1, \quad z_{3V} = z_1, \\
x_{3R} &= x_1(1 + \mu) - \mu x_1, \quad z_{3R} = z_1. \tag{3.39}
\end{align*} \]

The configuration forms the lensless Fourier transform hologram, and in this case both images are virtual, lying in the original subject plane. If the reference source lies on the $x$ axis, then the images are symmetric about the $z$ axis. The image at $(x_{3R}, z_{3R})$ is inverted.

Any of the waves can be converging to a point which is a positive distance from the hologram plane. Suppose the subject beam is such a converging beam so that $z_1$ is positive and suppose $z_r = z_o = -z_1$. Then
\[ z_{3V} = z_1/(2\mu - 1), \]
and the image produced by $q_v$ is real, for $2\mu > 1$. On the other hand the image produced by $q_R$ is characterized by
\[ z_{3R} = -z_1/(2\mu + 1) \]
and is virtual. If, instead, it is the illuminating wave that is chosen to be converging to a point a positive distance from the hologram, then inspection of $z_{3V}$ and $z_{3R}$ in Eqs. (3.27) and (3.28) reveals that both images are real; in this case the image at $(x_{3V}, z_{3V})$ is inverted.
3.4 Third-Order Aberrations

Equation (3.26) representing a spherical-wave phase distribution over the hologram is only a first-order approximation in 1/z. The next order of approximation consists of a number of terms all multiplying the factor (1/z)^3. This is also true, of course, of the expressions for \( \theta_{k} \), \( \theta_{r} \), and \( \theta_{b} \) in Eqs. (3.3), (3.4), and (3.23), respectively. The third-order terms of Eqs. (3.3), (3.4), and (3.23) are to be added together in accordance with Eq. (3.21) or (3.22) to give the third-order terms in \( \theta_{k} \) or \( \theta_{b} \). The phase differences between the third-order expansion of Eq. (3.26) and the third-order expansion of \( \theta_{k} \) (or \( \theta_{b} \)) are the aberrations.

Meier [3.2] has calculated hologram aberrations in terms of their usual classifications: spherical aberration, coma, astigmatism, field curvature, and distortion. He finds that one or the other of the waves diffracted by the hologram yields an image free of all aberrations providing the illuminating wave is identical with the reference. For this case the magnification is unity. Magnification is achieved (1) by illuminating the hologram with a spherical wave whose curvature differs from the reference, keeping \( \mu = m = 1 \); (2) by illuminating with a wavelength differing from that used to form the hologram, \( \mu \neq 1 \); and (3) by scaling the hologram, \( m \neq 1 \). The first method cannot be employed without producing aberrations. If plane waves are used for reference and illumination, then the condition \( \mu = m \) and \( \theta_{k} = \theta_{b} \) produces an aberration-free image from \( \phi_{k} \), while \( \mu = m \), \( \theta_{k} = -\theta_{b} \) produces an aberration-free image from \( \phi_{b} \). (The scaling, however, requires a lens which may degrade the image.) When reference source and subject sources are equidistant from the hologram (the lensless Fourier transform configuration, \( z_{r} = z_{s} \)), magnification can be achieved without optical scaling and with zero spherical aberration. The magnification is obtained by making \( \mu > 1 \). However at least one of the other aberrations will be present.

REFERENCES