J. A. Armstrong

Fresnel Holograms: Their Imaging Properties and Aberrations

Abstract: A simple and unified treatment is given of the properties of the magnified or demagnified images reconstructed from Fresnel holograms. The resolution attainable in wavefront reconstruction is discussed with particular attention to the aberrations of reconstructed images. Explicit expressions are given for the five primary wave aberrations, viz., spherical aberration, coma, astigmatism, curvature of field, and distortion.

1. Introduction

Holography is the science of producing images by wavefront reconstruction. In general no lenses are involved. The reconstructed image may be either magnified or demagnified compared to the object. Three-dimensional objects can be reconstructed as three-dimensional images. The wavefronts which are recorded photographically and later reconstructed are those due to diffraction of light by the object; thus, corresponding to the difference between Fresnel and Fraunhofer diffraction, one may produce either Fresnel or Fraunhofer holograms. This paper will deal exclusively with the properties of Fresnel holograms. Although some of the results presented here have been stated previously, they are derived in a simple and unified fashion which leads naturally to the treatment of hologram aberrations.

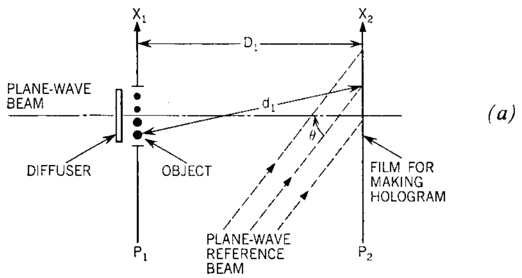
Wavefront reconstruction was invented by Gabor and expounded by him in a series of classic papers.[1–3] After the work of Gabor there were few contributions[4,5] until recently.[6–13] The decline in interest was doubtless due to two difficulties. The first difficulty was that in Gabor's method of image reconstruction the real image was superposed on a field emanating from a virtual image. The second difficulty was that the light sources suitable for producing holograms were of very low brightness 15 years ago. The difficulty of overlapping real and virtual images was removed by the observation[5b,6] that a high-spatial-frequency carrier wave could be used in the hologram construction process in such a way that the real and virtual images would be well separated in the reconstruction. The invention of the laser has overcome the difficulty of working with sources of low brightness.

This paper is organized as follows. First we will show how to calculate in the Fresnel approximation the geometrical properties of the magnified or demagnified images reconstructed from holograms. This will be done for arbitrary illumination of the object and for both two- and three-dimensional objects. A discussion will then be given of the resolution attainable in wavefront reconstruction. Finally we will discuss the aberrations of the images produced by wavefront reconstruction.

2. Geometrical properties

Consider the arrangements shown in Fig. 1a for the construction of a hologram and in Fig. 1b for the reconstruction of the image. In *construction*, a two-dimensional object of transmission $T(x_1)$ (typically, a film transparency) in the plane P_1 is illuminated from behind by monochromatic light of wavelength λ_1; the diffraction field due to the object is superposed on a reference or carrier wave with which it interferes. The carrier wave has plane wavefronts and is derived from the source that illuminates the diffuser. The resulting total field is recorded

Figure 1a Schematic diagram of hologram construction in light of wavelength λ_1. The object (of transmission $T(x_1)$) is illuminated with diffuse light produced by scattering the monochromatic plane wave from a diffusing plate. An optical system is used to derive the plane-wave reference beam from the same source as the plane wave incident on the diffuser.



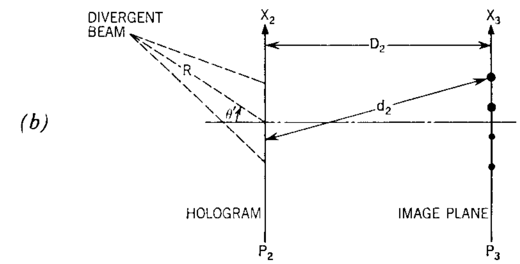


Figure 1b Diagram of image reconstruction using the hologram made as in Fig. 1a. The reconstruction is carried out in light of wavelength λ_2. The hologram is illuminated by a beam of divergent monochromatic light. The radius of curvature of the spherical wavefronts is R.

photographically in plane P_2, producing the hologram. The separation between object and hologram is shown in Fig. 1a as D_1. In *reconstruction*, the developed hologram is illuminated solely by a divergent beam of coherent light of wavelength λ_2, giving rise to a diffraction field containing a real image in the plane P_3. The separation between hologram and real image is shown in Fig. 1b as D_2. For the sake of definiteness we use a plane wave as reference or carrier wave in construction, and a divergent beam of radius of curvature R in reconstruction.

The following properties will be derived:

1) The quantities D_1, D_2, R, λ_1, and λ_2 are related by[4]

$$(\lambda_2 R)^{-1} = (\lambda_1 D_1)^{-1} - (\lambda_2 D_2)^{-1}. \tag{1}$$

2) The image is magnified laterally by an amount $M = (D_2\lambda_2)/(D_1\lambda_1)$ in all directions parallel to the plane of the hologram, and longitudinally by an amount $(D_2/D_1)M$ in the direction perpendicular to the plane of the hologram.[3]

3) The geometrical properties of the reconstructed image are independent of the nature of the illumination of the object provided this illumination is independent of time, and provided the coherence length of the light is longer than any path difference from object to hologram.[8b]

4) The angle of reconstruction θ' is related to the angle of incidence θ of the reference beam on the hologram by the condition $\sin\theta' = (\lambda_2/\lambda_1)\sin\theta$.

Although all calculations to be made here will assume a plane wave reference beam, the methods used can be readily adapted to any other configuration and in fact constitute a general procedure for analyzing holographic problems. Properties (1), (3), (4), and the expression for lateral magnification will be derived first, using two-dimensional objects. The expression for the longitudinal magnification, which applies to three-dimensional objects, is then easily derived from Eq. (1). We will treat explicitly only transparent objects, but it is known that holograms can be made in the light scattered from opaque objects.[8]

- *Two-dimensional objects and images*

Let us consider now the calculation of the field in the plane of the hologram (the P_2 plane in Fig. 1a) when a two-dimensional object of complex amplitude transmittance $T(x_1)$ is illuminated from behind with light of wavelength λ_1. We will take the most general expression for this illumination, namely

$$\int dk_x\, f(k_x) \exp\,[ik_{1x}x_1 + i\phi_k] \exp\,[-i\omega t].$$

Here the quantity ϕ_k refers to x-independent phases; the function $f(k_x)$ describes the decomposition of the field into plane waves. Henceforth the exp $[-i\omega t]$ factor will be omitted. We immediately write down the diffraction field due to the object, illuminated as described, at the hologram plane.

$$E_{\text{diff}}(x_2) = \underbrace{\int dx_1 \int dk_x\, f(k_x) \exp\,[ik_{1x}x_1 + \phi_k(x_1)]}_{\text{illumination}} \times \underbrace{T(x_1)}_{\text{object}} \cdot \underbrace{\frac{\exp\,[ik_1 d_1]}{d_1}}_{\text{diffraction}} \tag{2}$$

In this expression d_1 is the optical path from a point x_1 in the object plane to a point x_2 in the film plane. We have already made use of the so-called small angle approximation in writing Eq. (2) in place of the more general form of the Fresnel-Kirchhoff diffraction formula.[14] In order to proceed we must expand d_1 in powers of x_1 and x_2; the Fresnel approximation consists in cutting off this expansion after the second order terms: $d_1 = D_1 + (x_2 - x_1)^2/2D_1$. The total field at the hologram is the diffraction field plus the field of the reference or carrier beam. The total field may be written

$$E(x_2) = E_{\text{ref}} + E_{\text{diff}} = A_1 \exp\,[ik_1 \sin\theta x_2] + \frac{1}{D_1}\int dx_1 \int dk_x\, f(k_x) \exp\,[ik_{1x}x_1 + i\phi_k]$$

$$\times\ T(x_1)\ \exp\left[ik_1 D_1 + ik_1(x_2 - x_1)^2/(2D_1)\right]. \quad (3)$$

The film records the intensity EE^*; however, only the terms in the intensity which are linear in the diffracted field are of interest to us, since it is these terms which give rise to the reconstructed images. We need assume only that the properties of the film are such that when developed, there are terms in its amplitude transmittance which are linear in $E_{\text{diff}}E^*_{\text{ref}}$ and $E^*_{\text{diff}}E_{\text{ref}}$.

The subsequent calculations will deal with the properties of the reconstructed real image, corresponding to the term $E^*_{\text{diff}}E_{\text{ref}}$ and we will therefore use this latter product as the analytical expression of the transmission of the hologram. In reconstructing the hologram we use the geometrical arrangement shown in Fig. 1b. The hologram is illuminated obliquely at an angle θ' by a divergent beam whose spherical wavefronts have radius of curvature R. Thus the field of the illumination, of wavelength λ_2, in the plane of the hologram P_2 is

$$\frac{A_2}{R}\exp\left[ik_2 r\right] = \frac{A_2}{R}\exp\left[ik_2 R + ik_2 x_2^2/(2R) - ik_2 \sin\theta' x_2\right]. \quad (4)$$

The linear variation of phase with x_2 is due to the non-normal incidence of the wave. Again using the Fresnel-Kirchhoff diffraction formula we calculate the field due to diffraction of the field, Eq. (4), by the hologram $E^*_{\text{diff}}\ E_{\text{ref}}$. This final field, which exhibits the reconstructed image, is

$$E(x_3) = \frac{A_1 A_2}{D_1 R D_2}\exp\left[-ik_1 D_1 + ik_2 R\right]$$
$$\times \iiint dx_2\, dx_1\, dk_x \underbrace{f(k_x)\exp\left[-ik_{1x}x_1 - i\phi_k\right]}_{\text{illumination of object}}$$
$$\times \underbrace{T^*(x_1)\exp\left[-\frac{ik_1(x_2 - x_1)^2}{2D_1} + ik_1 \sin\theta x_2\right]}_{\text{hologram}}$$
$$\times \underbrace{\exp\left[\frac{ik_2 x_2^2}{2R} - ik_2 \sin\theta' x_2\right]}_{\text{illumination of hologram}}$$
$$\times \underbrace{\exp\left[ik_2 D_2 + ik_2\frac{(x_3 - x_2)^2}{2D_2}\right]}_{\text{diffraction}}. \quad (5)$$

To obtain a useful expression we must do the integrations over x_1 and x_2; the integration over x_2 will be done first and will itself produce several useful results. We collect all terms in the exponentials in Eq. (5) that are proportional to x_2^2 and to x_2. The terms depending on x_2^2 are: $(x_2^2/2)[(k_2/R) + (k_2/D_2) - (k_1/D_1)]$. Clearly we will be well advised to make this coefficient of x_2^2 equal 0. For it to be so Eq. (1) must be satisfied; this requirement relates the reconstruction distance to the construction distance, the divergence of the reconstructing beam, and the two wavelengths involved. We can rewrite Eq. (1) in a form which allows a simple interpretation, $(1/R) + (1/D_2) = (\lambda_2/\lambda_1)(1/D_1)$. The distances R and D_2 are related by the simple thin-lens equation for a lens of focal length $(\lambda_1/\lambda_2)D_1$. The hologram is not, of course, a real lens. It is, however, a Fresnel zone plate of equivalent focal length $(\lambda_1/\lambda_2)D_1$. The many similarities between thin lenses and zone plates are helpful in understanding wavefront reconstruction.

By choosing R, D_1, and D_2 to satisfy Eq. (1) we have simplified the integration over x_2 to the point where it can be done explicitly; we have

$$\int_{-\infty}^{\infty}\exp\left[ix_2\left(k_1\sin\theta - k_2\sin\theta' + \frac{k_1}{D_1}x_1 - \frac{k_2}{D_2}x_3\right)\right]dx_2$$
$$= 2\pi\delta\left(k_1\sin\theta - k_2\sin\theta' + \frac{k_1}{D_1}x_1 - \frac{k_2}{D_2}x_3\right). \quad (6)$$

We see immediately that we should choose the angle θ' of reconstruction illumination so that $\sin\theta' = (\lambda_2/\lambda_1)\sin\theta$. With this simplification we can rewrite the delta function as follows:

$$\delta\left(\frac{k_1 x_1}{D_1} - \frac{k_2}{D_2}x_3\right) = \frac{D_1}{k_1}\delta\left(x_1 - \frac{D_1}{D_2}\frac{k_2}{k_1}x_3\right) = \frac{D_1}{k_1}\delta(x_1 - M^{-1}x_3),$$

where we have put $M \equiv (D_2\lambda_2)/(D_1\lambda_1)$. It is now a trivial matter to do the integration over x_1 in Eq. (5); it consists simply in replacing x_1 everywhere it occurs by $(M^{-1}x_3)$. We can therefore write down the final expression for the field in the reconstructed image in the plane P_3:

$$E(x_3) = \frac{\lambda_1 A_1 A_2}{R D_2}\exp\left[\frac{ik_2 x_3^2}{2D_2}(1 - M^{-1})\right]\int dk_x\, f(k_x)$$
$$\times \exp\left[ik_x M^{-1}x_3 + i\phi_k\right] \times T^*(M^{-1}x_3). \quad (7)$$

Note that since $T^*(x_1) = T^*(M^{-1}x_3)$, the scale of the image in the x_3 plane is M times greater than the scale of the object. That is, the factor $M = (D_2\lambda_2)/(D_1\lambda_1)$ is the expression for the magnification achieved during the reconstruction of the image. Note that there is both geometrical and wavelength magnification.[3] Examination of Eq. (7) shows that the image field is essentially the same as the object field, except that it has been magnified, its brightness altered, and phase factor added. This latter factor, of course, in no way affects the intensity of the image, which is what one sees or records; its quadratic

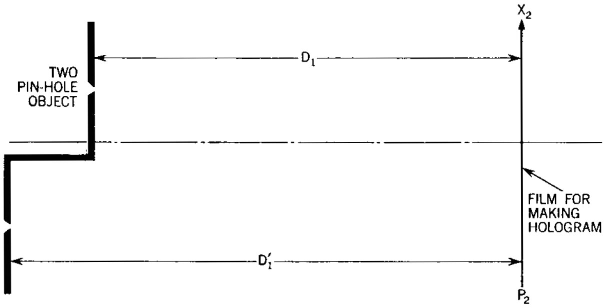


Figure 2 A simple two-pinhole object for illustrating three-dimensional image reconstruction.

dependence on x_3 shows simply that the light is divergent as it traverses the image plane.

Note also that the illumination of the image is the same as that of the object, apart from the fact that the illumination has undergone magnification along with the object. The expression used for the illumination has been completely general and it is clear that the details of the illumination enter in no way into the geometry of reconstruction.[8b]

- *Three-dimensional objects and images*

So far we have been discussing two-dimensional objects only. The extension of the discussion to three-dimensional objects is straightforward and, as we shall see, brings in in general a new magnification factor. Consider the extremely simple three-dimensional object shown in Fig. 2. We have seen in the previous discussion that each point in the object a distance D_1 from the hologram produces in effect a zone plate of focal length $(\lambda_1/\lambda_2)D_1$. Points located a distance D_1' from the hologram produce zone plates of focal length $(\lambda_1/\lambda_2)D_1'$. For each type of point we may write Eq. (1). That is,

$$\frac{1}{R} + \frac{1}{D_2} = \frac{\lambda_2}{\lambda_1}\left(\frac{1}{D_1}\right),$$

and

$$\frac{1}{R} + \frac{1}{D_2'} = \frac{\lambda_2}{\lambda_1}\left(\frac{1}{D_1'}\right).$$

Subtracting these two equations one finds

$$\Delta\left(\frac{1}{D_2}\right) = \frac{\lambda_2}{\lambda_1}\,\Delta\left(\frac{1}{D_1}\right). \tag{8}$$

If D_1 and D_1' are not too different we may express Eq. (8) as follows:

$$\frac{\Delta D_2}{D_2^2} = \frac{\lambda_2}{\lambda_1}\left(\frac{\Delta D_1}{D_1^2}\right).$$

From this expression we find the relation between the axial separation of the points in image space and their separation in the object space:

$$\Delta D_2 = \frac{D_2^2\lambda_2}{D_1^2\lambda_1}\,\Delta D_1 = \left(\frac{D_2}{D_1}\,M\right)\Delta D_1.$$

Thus the axial magnification is D_2/D_1 times the transverse magnification.[3] This property is exactly analogous to the situation which exists with lenses, where if the object and image distances are O and I, the transverse magnification is I/O, whereas the axial magnification is $(I/O)^2$. Hence, faithful three-dimensional magnification is clearly possible, but it must be achieved solely by a change in wavelength between construction and reconstruction. With this latter restriction it becomes necessary to recall that the process of changing wavelengths also involves changing the angle θ' which, since $|\sin\,\theta'| \leq 1$, may limit the degree of three-dimensional magnification that is practically achievable.

3. Resolution

Another hologram property of great interest is the resolution attainable in the reconstructed image. This resolution is determined by three factors:

1) Ideally, the ultimate resolution is determined by the hologram aperture, not by any property of the film used. We will therefore indicate explicitly the effects of hologram size on resolution.

2) In practical cases the limit of resolution may be set by the fact that presently available films cannot record spatial frequencies above a definite limit. This is the problem treated in Refs. 11 and 13. We will discuss this calculation but briefly.

3) Finally, the resolution in the reconstructed image depends on the aberrations of the hologram itself. Wave aberrations of holograms are to be treated in Section 4.

First, we consider the limit imposed on resolution by hologram size. We return to Eq. (6), omitting the irrelevant sine terms, and allow the integration over the hologram coordinate x_2 to run only from $+L$ to $-L$, where $2L$ is the extent of the hologram in the x_2 direction. We find

$$\int_{-L}^{+L} \exp\left[ix_2\left(\frac{k_1x_1}{D_1} - \frac{k_2x_3}{D_2}\right)\right]dx_2 = 2\sin\left[\left(\frac{k_1x_1}{D_1} - \frac{k_2x_3}{D_2}\right)L\right]\left[\left(\frac{k_1x_1}{D_1} - \frac{k_2x_3}{D_2}\right)\right]^{-1}. \tag{9}$$

It will be recalled that for the hologram of unlimited aperture the integration over x_2 gave a delta function. Consider the simple but sufficient illustration of an object $T(x_1)$ which is a single point: $T(x_1) = \delta(x_1 - a)$. The integration over x_1 in Eq. (5) merely replaces x_1 in Eq. (9) by a. That is, the image of a point object is

$$2\sin\left[\left(\frac{k_1a}{D_1} - \frac{k_2x_3}{D_2}\right)L\right]\left[\frac{k_1a}{D_1} - \frac{k_2x_3}{D_2}\right]^{-1}.$$

This is the familiar sin $(qx)/x$ function. The interval in the x_3 plane between the maximum and the first zero in intensity of the image of a point is $\Delta x_3 = (\lambda_2 D_2/2L)$. Now two object points separated by a distance Δx_1 will, from the discussion following Eq. (7), be reconstructed at a separation $M\Delta x_1$. For this separation to be resolved we require that it be greater than the spread in the image of a point. That is, we require $(D_2\lambda_2/D_1\lambda_1)\Delta x_1 > (D_2\lambda_2/2L)$. Clearly this is true only if $\Delta x_1 > (D_1\lambda_1/2L)$, which is the condition that the object detail be resolved under the conditions of the construction. For the idealized holograms we are considering, the processes of magnification or demagnification do not result in loss of detail provided the aperture of the hologram is large enough to resolve the object detail in the first place. It is again clear that there is a deep analogy between lenses and holograms. As far as *resolution* is concerned, the hologram is equivalent, in the construction process, to a lens of f-number $D_1/2L$, and in the reconstruction process, to a lens of f-number $D_2/2L$.

Second, we consider the problem of the limited capacity of photographic film to record high spatial frequencies. It is the most elementary result of diffraction theory that the smaller the object causing diffraction, the greater is the angular spread of the diffracted light. Diffracted light impinging on the hologram at large angles with respect to the reference beam gives rise to interference fringes whose spacing may be far smaller than the resolution limit of the film used to make the hologram; if this is the case information concerning the small detail in the object will be lost.

A method for sidestepping this limitation of the film is given in Refs. 11 and 13. (It is called "Fourier-transform holography" since a lens must be used during reconstruction to take the Fourier transform of the field distribution emerging from the illuminated hologram). It is shown that information about small details can be recorded even though the spatial frequencies involved are too high to be recorded. Fourier-transform holography is especially designed to function in the x-ray region, where the position of the diffracted beams is uniquely determined by the object being x-rayed and cannot be altered by the experimenter. Thus in x-ray diffraction much of the information about the structure of the object is contained in beams which make very large angles indeed with respect to the hologram. The fact that these directions are fixed by the object is due to the periodic structure of the object in the z-dimension (perpendicular to the hologram plane) as well as in the x and y dimensions. There are objects (e.g., film transparencies) whose entire information content resides in the x-y variation of the transparency. Although photographic images have finite thickness in the z-dimension, the location of developed grains in this dimension is purely random. Hence, transparencies and other objects whose diffracting centers in the z-direction are randomly distributed may be considered to be two-dimensional objects. When such an object has x-y details so fine as to give rise to spatial frequencies that are not recordable on film, a simple remedy can be used: the object is illuminated in diffuse light,[8,9,12] causing diffracted beams to leave the object in all directions. (This is the property that is lost when the object has non-random structure in the z-direction). Now the small details will diffract light into a wide range of spatial frequencies; among these will be some that can be recorded on film. Thus high-resolution holograms of two-dimensional objects can be constructed using Fresnel diffraction alone; however, as shown in Ref. 11, one must use a lens—and hence Fraunhofer diffraction during reconstruction—to make high-resolution holograms of three-dimensionally periodic objects using x-rays.

4. Hologram aberrations

In Section 2 we presented a general and simple method for calculating the geometrical properties of holographically reconstructed magnified or demagnified images. The calculations were made in the Fresnel approximation, in which the reconstructed images exhibit no aberrations. Since in real circumstances, however, magnified or demagnified images will suffer from aberrations, it now becomes our aim to evaluate the aberrations of holograms and to discuss the circumstances under which they can be minimized. Clearly an understanding of the aberrations of holograms is vital to the achievement of the high resolution we discussed in Section 3.

Gabor, in one of his papers,[3] considered aberrations from a point of view suggested by his interest in the imaging properties of holograms made with illuminating beams suffering from various aberrations, and his interest in the correction of aberrations as a means for distinguishing between the real and the virtual images. Instead, we consider the aberrations of the *hologram itself*, quite apart from these essentially unrelated problems.

It will be shown that, except for the case of unit magnification, images reconstructed from holograms exhibit the five primary wave aberrations of geometrical optics: viz., spherical aberration, coma, astigmatism, curvature of field, and distortion. It will be seen that by suitable choice of the constructing and reconstructing wavelengths, and of the constructing and reconstructing distances, any one of these five may be eliminated (in one case two can be eliminated simultaneously: astigmatism and curvature of field). Further, it will be shown explicitly that the spherical aberration of the hologram can also be corrected by other means during the construction and reconstruction process. We will give explicit expressions for all of the hologram aberrations in the uncorrected case; the treatment of hologram aberrations will be seen

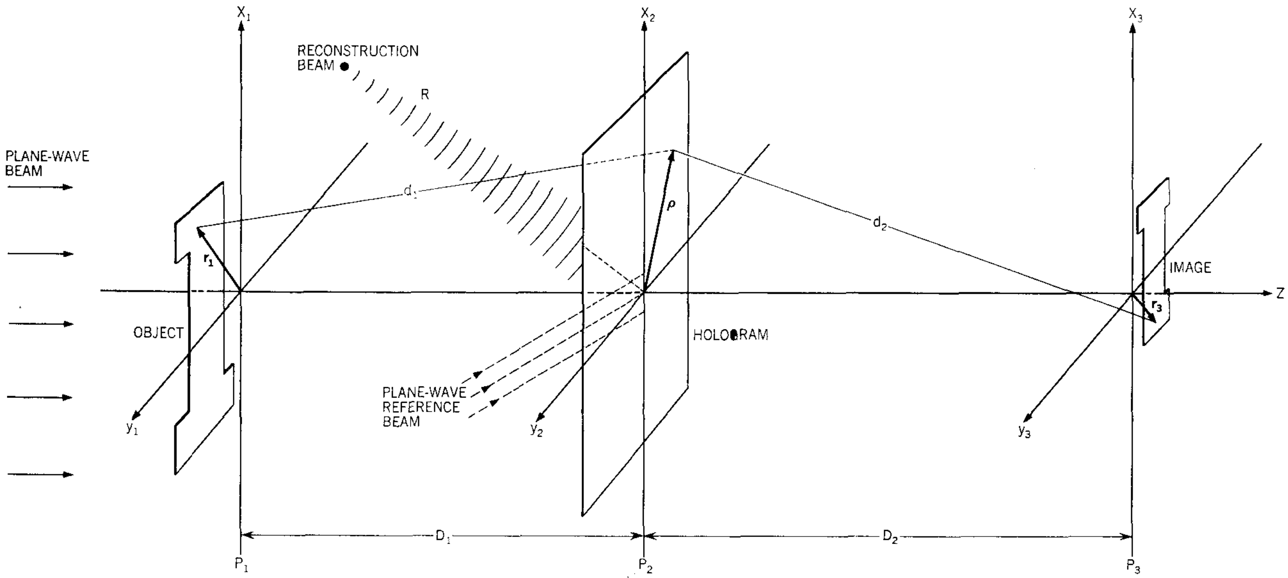


Figure 3 Diagram of the wavefront reconstruction process for use in discussing the aberrations of holographically produced images. Plane-wave illumination, plane-wave reference, and divergent reconstruction beams are shown. The first two are used to produce the hologram; only the third is used in reconstructing the image.

to be very similar to the treatment of lens aberrations, except that the holographic case is vastly more simple. Explicit conditions will be given for the simultaneous elimination of the two most serious aberrations, spherical aberration and coma.

Figure 3, like Figure 1 on page 172, identifies the essential elements of hologram construction and reconstruction, but includes additional details needed for the discussion of aberrations. The optical path from an arbitrary object point $\mathbf{r}_1$ to an arbitrary hologram point ρ is d_1; the distance from that hologram point to an arbitrary image point is d_2.

The starting point for a calculation of the hologram aberrations is Eq. (5), which is the analytic expression for the light field in the reconstructed image in plane P_3. We write it now explicitly for two-dimensional objects.

$$E(x_3, y_3) = \iint d\mathbf{r}_1 d\rho \times \underbrace{T^*(\mathbf{r}_1) \exp\left[-ik_1 d_1 + ik_1 \sin\theta x_2\right]}_{\text{hologram}} \times \underbrace{\exp\left[(ik_2\rho^2)/(2R) - ik_2 \sin\theta' x_2\right]}_{\text{hologram illumination}} \times \underbrace{\exp\left[ik_2 d_2\right]}_{\text{diffraction from hologram}} \qquad (10)$$

We have written this equation for plane-wave object illumination and have omitted unimportant constant amplitude factors. The distances d_1 and d_2 are given by

$$d_1 = D_1\left[1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{D_1^2}\right]^{1/2} = D_1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2D_1} - \frac{1}{8D_1^3}\left[(x_2 - x_1)^2 + (y_2 - y_1)^2\right]^2 + \cdots \qquad (11a)$$

$$d_2 = D_2\left[1 + \frac{(x_3 - x_2)^2 + (y_3 - y_2)}{D_2^2}\right]^{1/2} = D_2 + \frac{(x_3 - x_2)^2 + (y_3 - y_2)^2}{2D_2} + -\frac{1}{8D_2^3}\left[(x_3 - x_2)^2 + (y_3 - y_2)\right]^2 + \cdots \qquad (11b)$$

When only terms of order D and D^{-1} are retained in the expansions, one has the Fresnel approximation and Eq. (10) takes the explicit form of Eq. (5). It should come as no surprise that in discussing the aberrations of the system we retain terms up to order D^{-3}.

The aberrations of the holographic process are found solely by considering the integrations of x_2 and y_2 in Eq. (10). It was shown in Section 2 that if the argument of the exponential is *linear* in x_2 (and y_2) the reconstruction is a faithful magnified image of the object. In the Fresnel approximation, terms quadratic in x_2 or y_2 were eliminated by imposing the condition

$$\frac{1}{R} + \frac{1}{D_2} = \frac{\lambda_2}{\lambda_1}\frac{1}{D_1}. \qquad (12)$$

Here, as before, R is the radius of curvature of the spherical wave which illuminates the hologram. We of course continue to impose this condition, but it is no longer sufficient to cause all terms nonlinear in x_2 or y_2 to drop out of the argument of the exponentials in Eq. (10). *These nonvanishing terms which are nonlinear in the hologram coordinates (or linear in the hologram coordinates and nonlinear in the object coordinates) constitute the wave aberrations of the hologram.* Clearly the quantity we must investigate is $k_2d_2 - k_1d_1 = \Delta$. A tedious but straightforward algebraic computation gives

$$\Delta = -\frac{1}{8}\left(\frac{k_2}{D_2^3} - \frac{k_1}{D_1^3}\right)\rho^4 + \frac{1}{2}\left(\frac{Mk_2}{D_2^3} - \frac{k_1}{D_1^3}\right)\rho^2 K_1^2$$
$$-\frac{1}{2}\left(\frac{M^2k_2}{D_2^3} - \frac{k_1}{D_1^3}\right)K_1^4$$
$$-\frac{1}{4}\left(\frac{M^2k_2}{D_2^3} - \frac{k_1}{D_1^3}\right)\rho^2 r_1^2$$
$$+\frac{1}{2}\left(\frac{M^3k_2}{D_2^3} - \frac{k_1}{D_1^3}\right)r_1^2 K_1^2. \qquad (13)$$

In obtaining this expression we have used Eq. (12) and have dropped all terms in $k_2d_2 - k_1d_1$ that are independent of x_2 or y_2 as well as the terms linear in both $\mathbf{r}_1$ and ρ that produce the delta function in the treatment of Section 2. We have introduced the notation $r_1^2 = x_1^2 + y_1^2$, $\rho^2 = x_2^2 + y_2^2$, and $K_1^2 = x_1x_2 + y_1y_2$; and we have also made use of the fact that a magnification M occurs between object and image.

The first term in Eq. (13) is the spherical aberration; the second is the coma; the third is the astigmatism; the fourth is the curvature of field; and the last is the distortion. This identification is facilitated by comparison of Eq. (13) with the generalized treatment of lens aberrations given in Section 5.3 of Ref. 14. It is clear that all the aberrations which are familiar in lens theory also occur in holography and that these five primary aberrations account for all of the aberrations of a hologram up to third order (terms in Δ up to order D^{-3}). Although the expressions in Eq. (13) for the aberrations are written in terms of the coordinates of the object and the hologram, they could easily be written in terms of the coordinates of the hologram and the image.

Consider now the problem of correcting for these aberrations of the reconstruction process. If we set each expression equal to zero, in turn, we obtain the conditions under which each will be corrected. Thus, using $M = (D_2\lambda_2/D_1\lambda_1)$, we have Table 1.

Table 1 Aberrations and conditions for their correction.

Aberration	*Condition*
• Spherical aberration	$\left(\frac{D_2}{D_1}\right)^3 \frac{\lambda_2}{\lambda_1} = 1$
• Coma	$\left(\frac{D_2}{D_1}\right)^3 \frac{\lambda_2}{\lambda_1} = M;\ D_1 = D_2$
• Astigmatism, and • Curvature of field	$\left(\frac{D_2}{D_1}\right)^3 \frac{\lambda_2}{\lambda_1} = M^2;\ \frac{\lambda_2}{\lambda_1} = \frac{D_2}{D_1}$
• Distortion	$\left(\frac{D_2}{D_1}\right)^3 \frac{\lambda_2}{\lambda_1} = M^3;\ \lambda_1 = \lambda_2$

From these relationships we see that if one aberration is made to vanish, the others generally cannot. (There are two exceptions: if $M = 1$, all the aberrations vanish together; if $M \neq 1$ astigmatism and curvature of field can be simultaneously corrected.) If $D_1 = D_2$ the image will not suffer from coma; if $\lambda_1 = \lambda_2$, there will be no distortion. Astigmatism and curvature of field are eliminated by scaling wavelength and distance the same way.

It is useful to characterize the aberrations in terms of their dependence on the dimensions of the hologram. If the linear dimensions of the hologram are of order L, then according to Eq. (13) the spherical aberration varies as L^4, the coma as L^3, the astigmatism and curvature of field as L^2, and the distortion as L. Clearly the spherical aberration will under most circumstances be the most serious aberration and the one for which it is most important to correct. The spherical aberration is the only hologram aberration that is independent of the object or image coordinates. This feature makes it possible to correct for spherical aberration by illuminating the hologram in a divergent beam that has built into it the amount of spherical aberration necessary to compensate that of the hologram. This can be done by producing the reconstruction beam with a simple lens of known spherical aberration. The amount of this spherical aberration is calculated as follows. To eliminate coma we take $D_1 = D_2$. The lens forming the illuminating beam must introduce a spherical aberration expressed as a phase retardation at the hologram of magnitude $-(1/D_1^3)(k_2 - k_1)\rho^4$. This aberration can be related to the particular choice of lens by using standard formulae for the spherical aberrations of simple lenses. If desirable in a particular application, the correction for spherical aberration can be wholly or partially accomplished by predistorting the plane-wave reference beam used in hologram construction so that it will have all or part of the required spherical aberration. Note that the choice $D_1 = D_2$ is also the one which assures faithful three-dimensional magnification.

Acknowledgments

The author has had fruitful discussions on holography with Drs. G. J. Lasher and W. V. Smith.

References

1. D. Gabor, *Nature* **161**, 777 (1948).
2. D. Gabor, *Proc. Roy. Soc.* (London) **197A**, 454 (1949).
3. D. Gabor, *Proc. Phys. Soc.* **B64**, 449 (1951).
4. G. L. Rogers, *Proc. Roy. Soc.* (Edinburgh) **A63**, 193 (1952).

5a. P. Kirkpatrick and H. M. A. El Sum, *J. Opt. Soc. Am.* **46**, 825 (1956).

5b. A. Lohmann, *Optica Acta* **3**, 97 (1956).

6. E. N. Leith and J. Upatnieks, *J. Opt. Soc. Am.* **52**, 1123 (1962).
7. E. N. Leith and J. Upatnieks, *J. Opt. Soc. Am.* **53**, 1377 (1963).

8a. J. Upatnieks and E. N. Leith, *J. Opt. Soc. Am.* **54**, 579A (1964).

8b. E. N. Leith and J. Upatnieks, *J. Opt. Soc. Am.* **54**, 1295 (1964).

9. G. W. Stroke, "An introduction to optics of coherent and non-coherent electromagnetic radiations," The Univ. of Michigan Engineering Summer Conferences on Lasers, Ann Arbor, May 1964.
10. G. W. Stroke and D. G. Falconer, in *Symposium on Optical and Electro Optical Information Processing Technology*, ed. by J. T. Tippett and L. C. Clapp, D. Berkowitz and C. J. Koester, MIT Press, Cambridge 1964.
11. G. W. Stroke and D. G. Falconer, *Phys. Lett.* **13**, 306 (1964).
12. G. W. Stroke and D. G. Falconer, *Phys. Lett.* **15**, 238 (1965).
13. G. W. Stroke, *Appl. Phys. Lett.* **6**, No. 10, 201 (1965).
14. Born and Wolf, *Principles of Optics*, Pergamon Press, London and New York, 1959, p. 379.

Received March 23, 1965.

Note added in proof

Since submission of this manuscript the author has learned of treatments of hologram aberrations carried out by R. W. Meier and, independently, by E. N. Leith; their papers are expected to appear shortly in the *Journal of the Optical Society of America.*

It should be emphasized in connection with Eq. (10) that the beam used to illuminate the hologram during reconstruction is chosen to have parabolic rather than spherical wavefronts. This has the result of removing any image aberrations that are due to the illumination rather than to the hologram itself. As discussed by Meier in his forthcoming paper, the use of a truly spherical illuminating beam gives rise to additional contributions to each type of aberration. These aberrations are, of course, independent of the image or object coordinates (as is the hologram spherical aberration discussed in the present paper). Thus all the aberrations due to spherical illumination can be corrected by choosing the parabolic wavefronts for which we have made our calculations. Since the illuminating beam is derived effectively from a point source, it should not be difficult to predistort the beam to have parabolic shape modified by the amount of spherical aberration calculated in the present text. The author is grateful to Dr. Meier for a discussion of these latter points.