

Structural-information Storage in Holograms

Abstract: The number of degrees of freedom, or structural-information content, of the object wave field recorded in a Leith-Upatnieks hologram is expressed in terms of the resolving power and dimensions of the recording medium, the coherence properties of the primary illumination and the position of the point reference source. In contrast with previous studies, the calculation does not involve the paraxial approximation. It is shown that of all holograms, the Fourier-transform hologram makes the most efficient use of available resolving power and coherence length.

Introduction

An important question in the communication theory of holography concerns the number of degrees of freedom, or structural-information content,¹ of the object wave field recorded in a hologram. Several authors have dealt with the problem,²⁻⁵ but so far only the recording of paraxial wave fields has been considered. The present paper deals with the holographic recording of fields of extended angular aperture. Its aim is to relate the structural-information content of the recorded wave field to the 1) resolving power and dimensions of the recording medium, 2) temporal bandwidth of the primary illumination and 3) location of an off-axis reference source. The effects of the recording medium and reference-source position on the amount of structural information recorded have been described previously in the context of the paraxial approximation.²⁻⁵ To the author's knowledge, the effect of the temporal bandwidth has not yet been discussed.

We do not account for the effects of noise. Indeed, the statistical or metrical¹ aspect of holographic information storage can be considered as separate from the structural aspect, and has been so treated by Smith.⁶ The total information content of the holographic record can be found by attaching a signal-to-noise ratio to each degree of freedom of the recorded wave field.

The holographic system considered in this paper is shown in Fig. 1. The object and plate are arranged symmetrically about a horizontal axis, and their separation is z . The reference source R is located to the left of the object and is constrained to lie on the extension of a line AD connecting opposite edges of the object and plate. The recording of both two- and three-dimensional wave fields is considered. In the two-dimensional case (cylindrical wavefronts) the object and plate can be considered to be rectilinear elements of lengths $2L_S$ and $2L_H$, respectively. In the three-dimensional case (spherical wavefronts) the object and plate are assumed to be circular in cross section with diameters $2L_S$ and $2L_H$.

We allow for the possibility that each element of the object may send out light in all directions. Of course only a part of the total complex wave field actually falls on the plate and, in general, even this part may not be fully recorded for lack of adequate film resolving power and coherence length. For analytical purposes we simplify the situation by restricting the dimensions of the object and plate to such extent that all parts of the complex wave field incident on the plate are recorded. Thus each element of the processed hologram contributes to the reconstruction of each element of the object. It is through the restriction on L_S and L_H that the resolving power and temporal bandwidth limit the number of degrees of freedom of the recorded wave field.

The author is with the Research Division of the American Optical Corporation, P.O. Box 187, Framingham Centre, Mass. 01701.

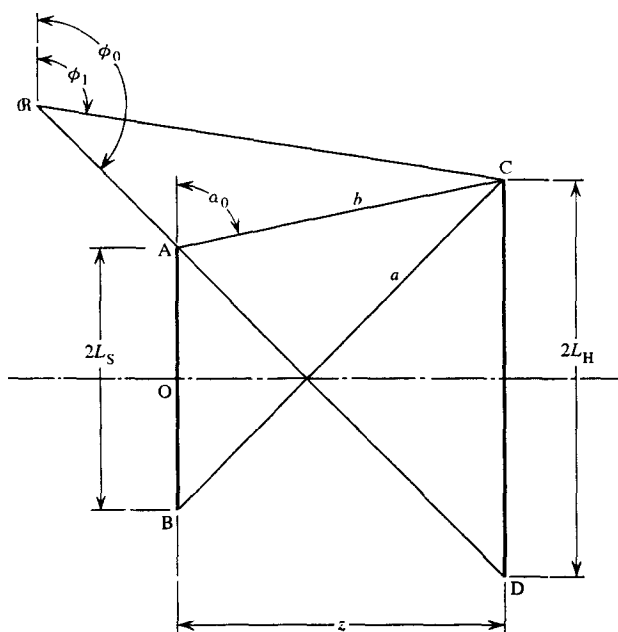


Figure 1 Hologram-recording geometry. The symbol S denotes reference point source, AB the rectilinear or circular object domain, and CD the rectilinear or circular holographic plate.

The remainder of the paper consists of three main parts. The first part reviews an integral representation of the number of degrees of freedom of a coherent wave field. In the second part the number of degrees of freedom of the object wave field recorded in the holographic system of Fig. 1 is expressed in terms of the resolving power and dimensions of the recording medium. The third part describes the limit on the degree of freedom arising from the finite coherence length.

The structural-information content of a coherent wave field

The number of degrees of freedom of a wave field can be defined as the number of independent real parameters needed to describe the field completely.^{5,7} In the communications context this number specifies the structural-information content of the field.¹ M. von Laue's original treatment⁷ of the subject was limited to those cases in which the radiation field suffers the same angular limitation at each element of a given cross section. In this paper we make use of a somewhat more general definition of the degree of freedom due to Gabor.¹ Gabor's definition allows for spatial variations in angular aperture, and is thus suited to the study of wide-angle (space-variant) holographic imagery.

Consider first the two-dimensional case, that of a monochromatic (scalar) wave field propagating in the plane of the paper. The sources of the field, radiating at wavelength λ , are distributed along a plane curve S as shown in Fig. 2. According to Gabor,¹ the number F of degrees of freedom of the field emerging from S and incident on a second plane curve $\mathcal{H}$ is given by

$$F = 2 \iint d\left(\frac{\cos \alpha}{\lambda}\right) ds, \quad (1)$$

where the integrand represents the differential space-bandwidth product associated with an elementary pencil of radiation incident on an element of arc ds in $\mathcal{H}$ and confined in direction to an element of plane angle $d\alpha$. Here α is the angle that the direction considered makes with the tangent vector T at ds . The factor of 2 in this formula accounts for the fact that the coherent wave field has both an amplitude and a phase description.

An expression appropriate to three-dimensional wave fields can also be given. In this case S denotes a surface distribution of sources and $\mathcal{H}$ refers to a surface of observation (see Fig. 3). The number of degrees of freedom of the (scalar) wave field emerging from S and incident on $\mathcal{H}$ can be written¹

$$F = 2 \iint \frac{\cos \theta}{\lambda^2} d\Omega dA, \quad (2)$$

where the integrand represents the differential space-bandwidth product associated with an elementary pencil of radiation incident on an element of area dA in $\mathcal{H}$ and confined in direction to an element of solid angle $d\Omega$. Here θ is the angle between the mean direction considered and the normal vector n at dA . The factor of 2 in this formula has the same meaning as in Eq. (1).

In general, the angle and space integrals appearing in Eqs. (1) and (2) are not independent. When they are independent, the corresponding optical system can be described as space-invariant. The factor $\frac{1}{2}F$ is then known as the space-bandwidth product.⁴

The significance of the degree of freedom for holography can be illustrated as follows. Suppose that, in Fig. 2 (or Fig. 3), a hologram in position S produces a real image at $\mathcal{H}$. The local spatial bandwidth of radiation incident on a given element of $\mathcal{H}$ is $\int d(\cos \alpha)/\lambda$ (or $\int \cos \theta d\Omega/\lambda^2$). Now the reciprocal of this factor determines the smallest resolved interval of complex-amplitude structure at the given element. Thus the quantity $\frac{1}{2}F$ measures the number of resolution cells or, more precisely, the number of independent data points^{1,8} contained in the total complex-amplitude image. This is the significance of the degree of freedom for microscopy.

Note that F also has an interpretation for visual display holography, where the smallest resolved interval of the image is determined not by the aperture of the plate but

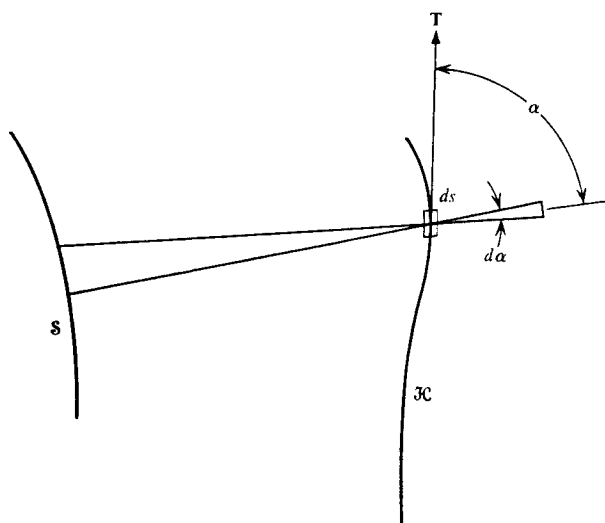
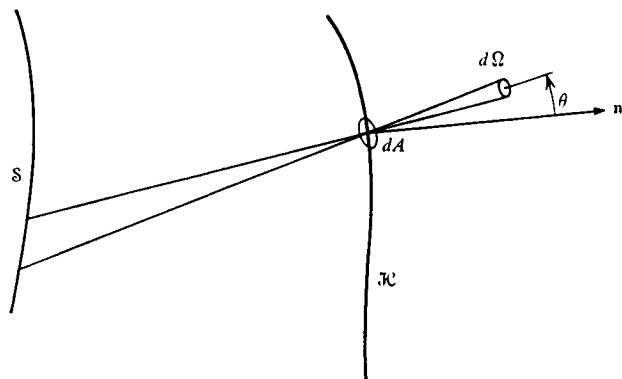


Figure 2 Illustration of Gabor's representation of the number of degrees of freedom of a two-dimensional wave field. The field originates at a plane curve S and is incident on a second plane curve H .

Figure 3 Illustration of Gabor's representation of the number of degrees of freedom of a three-dimensional wave field. The field originates at a surface S and is incident on a second surface H .



by the pupil of the viewer's eye. Suppose that a hologram in position H produces a virtual image in the location S . In the display application two factors are of interest, the differential angular field of view $d\alpha$ (or $d\Omega$) and the available increment of motion $\sin \alpha ds$ (or $\cos \theta dA$) of the viewer's eye in a direction normal to the principal ray in $d\alpha$ (or $d\Omega$). The product of these factors, when integrated over the total field of view and hologram aperture, is proportional to the degree of freedom F . Thus, of two competing display systems operating at the same wavelength, the one for which F is greater may

be said to have the greater "visual effectiveness" (more parallax for given angular field of view, or vice versa).

Let us now return to the holographic system of Fig. 1. When the illumination is monochromatic, and the film is perfect, the number of degrees of freedom of the recorded object wave field can be computed from Eqs. (1) and (2). In the two-dimensional case (rectilinear object and plate) integration of Eq. (1) gives

$$F = 2(2/\lambda)(a - b), \quad (3)$$

and in the three-dimensional case (circular object and plate), integration of Eq. (2) gives

$$F = 2(\pi/\lambda^2)[\pi(a - b)^2/4], \quad (4)$$

where, in both cases, a and b are the path lengths indicated in Fig. 1. In these expressions the factor 2 is that already described in connection with Eqs. (1) and (2), and the second factor is the maximum spatial bandpass available at the wavelength λ . The third factor depends only on the path difference $a - b$. Thus, as should be expected, the degree of freedom is closely related to the number of half-wavelengths contained in the interval $a - b$. When the plate dimensions are small compared with the separation z , in both cases $\frac{1}{2}F$ reduces to the ordinary space-bandwidth product.

Equations (3) and (4) apply when the object is permitted to occupy the complete space between A and B (Fig. 1). With this arrangement the reconstructed wave field has superimposed on it an unwanted background (intermodulation image) that arises as a result of the object field interfering with itself. The effect of the background is largely removed by restricting the object to the lower half OB of the original domain AB. As only one-half the available field of view is used, the wave field recorded in the hologram has $\frac{1}{2}F$ degrees of freedom, where F is given by Eq. (3) or Eq. (4).

We have derived Eqs. (3) and (4) on the assumptions of perfect film and monochromatic illumination. The same expressions can be applied to the case of imperfect film and nonmonochromatic illumination, provided that all parts of the complex disturbance incident on the plate are recorded (see the Introduction). It remains to determine the largest possible value of $a - b$ consistent with this condition. We consider in order the following three cases: imperfect film and monochromatic light, perfect film and nonmonochromatic light, and imperfect film and nonmonochromatic light.

Structural information and film resolving power

Since the path difference $a - b$ is determined by the separation and dimensions of the object and plate, our first task is to relate the latter quantities to the resolving power of the recording medium. The following treatment,

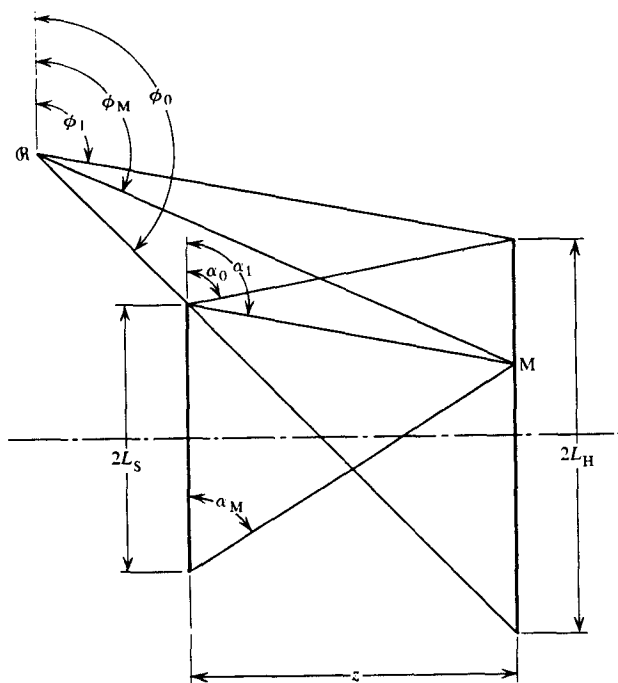


Figure 4 The holographic system of Fig. 1, with notation appropriate to deriving the resolution criterion, Eq. (6).

which assumes monochromatic light, can be considered an extension of the work of Leith et al.⁹ and of Stigliani et al.¹⁰

The recording medium may be assumed to have finite depth. In general, the medium records a three-dimensional system of standing waves, and the resolving power required is determined by the maximum standing-wave frequency produced within the medium.

The standing wave of maximum spatial frequency is formed within a small region of the plate, indicated by M in Fig. 4. The location of the region M clearly depends on the position of the reference source. However, because of the symmetry of the recording system, the nodal surfaces of this particular standing wave are approximately at right angles to the surface of the plate, regardless of where the reference source is located. Thus, the basic restriction imposed on the recording geometry by the resolving power N (lines/mm) can always be written,

$$N\lambda \geq \cos \alpha_M - \cos \phi_M, \quad (5)$$

where, as shown in Fig. 4, α_M is the inclination from the vertical of a ray extending from the lower edge of the plate through the region M, and ϕ_M is the inclination of a reference ray also extending through M.

Equation (5) is to be solved for the object dimension L_S . A somewhat involved calculation (see the Appendix) gives

$$L_S + \beta L_H \leq \mu z / (1 - \mu^2)^{1/2}, \quad (6)$$

where

$$\mu = N\lambda/2 \quad (7)$$

and

$$\beta = \frac{\cos \alpha_0 - \cos \phi_1}{\cos \alpha_0 - \cos \phi_0}, \quad (8)$$

where the angles α_0 , ϕ_0 and ϕ_1 are defined as in Fig. 1. In this formula the parameter μ expresses the resolving power of the recording medium in reciprocal half wavelengths. The parameter β expresses the position of the reference source and can take any value between zero and one. For example, the value $\beta = 1$ corresponds to plane reference waves ($\phi_1 = \phi_0$) and $\beta = 0$ corresponds to spherical reference waves emanating from a point at the upper edge of the object ($\phi_1 = \alpha_0$; lensless Fourier-transform holography). Equation (6) shows that, in general, the resolving power limits both the plate size and the object size (for given z).⁹ However, in the special case $\beta = 0$, only the object size is limited.^{11,12} Of course, if the film is perfect ($\mu = 1$), there is in principle no restriction on either L_S or L_H .¹⁰ It should be noted that Eq. (6) is independent of the depth of the recording medium, a direct consequence of the symmetrical recording geometry.

The amount of structural information contained in the recorded monochromatic object wave field can now be calculated. The general result is obtained by eliminating the object dimension L_S in Eqs. (3) and (4) by means of Eq. (6).

An approximate version of the general result enables us to see more clearly the effect of the resolving power. We write

$$a - b = (2L_S L_H / D) [1 + \frac{1}{8}(2L_S L_H / D^2)^2 + \dots], \quad (9)$$

where

$$D^2 = L_S^2 + L_H^2 + z^2. \quad (10)$$

If

$$\frac{1}{8}(2L_S L_H / D^2)^2 \ll 1, \quad (11)$$

then only the leading term in Eq. (9) need be retained. Note that since $2L_S L_H / D^2 \leq 1$ for all values of L_S and L_H , the approximation is a good one for all but the most extreme angles of diffraction. Within this approximation, Eqs. (9), (6) and (3) give for the rectilinear hologram

$$F \leq 4NL_H \left\{ \frac{1 - \beta x / \mu}{[1 - 2\beta \mu x + (1 + \beta^2)x^2]^{\frac{1}{2}}} \right\}, \quad (12)$$

and Eqs. (9), (6) and (4) give for the circular hologram

$$F \leq (\pi N L_H)^2 \left\{ \frac{1}{2} \frac{(1 - \beta x / \mu)^2}{1 - 2\beta \mu x + (1 + \beta^2)x^2} \right\}, \quad (13)$$

where

$$x = (1 - \mu^2)^{1/2} (L_H / z). \quad (14)$$

These results can be interpreted as follows.

1) The factor $4NL_H$ appearing in Eq. (12) is the space-bandwidth product associated with the one-dimensional recording medium. Similarly, the factor $(\pi L_H N)^2$ in Eq. (13) is the space-bandwidth product associated with the circular recording medium.¹³ Now the number of independent real data in a photographic image cannot exceed the space-bandwidth product of the recording medium.²⁻⁵ Thus, in each expression, the factor in brackets $\{\dots\}$ represents the efficiency⁵ with which the corresponding holographic medium is exploited for the storage of structural information.

2) From Eq. (12) we see that the system consisting of rectilinear object and hologram has a potential storage efficiency of one. In a strict sense, the maximum efficiency can be attained only with perfect film ($\mu = 1$), thus permitting an object of indefinitely large width. However, when the film is not perfect, an approximation to unit efficiency can be obtained by imposing the condition $x/\mu \ll 1$. A result similar to this was found by Lukosz⁵ for "single-sideband" holograms produced with plane-wave reference illumination and paraxial wave fronts. When the film is not perfect, and when the ratio x/μ cannot be made small, the efficiency is maximized by choosing $\beta = 0$. That is, in general, the lensless Fourier-transform arrangement makes the most efficient use of the given storage medium. Note, however, that the storage efficiency of this arrangement decreases with increasing numerical aperture (or increasing x), for the resolving power is greater than is necessary to record the interference fringes formed at the edges of the plate. In the paraxial approximation ($x \ll 1$), and with $\beta = 0$, Eq. (12) reduces to an expression given by Parrent and Reynolds² for the amount of structural information stored in a one-dimensional Fourier-transform hologram.

3) From Eq. (13) we see that the system consisting of circular object and plate also has the greatest potential storage efficiency when the film is perfect or, barring perfect film, when $x/\mu \ll 1$. When the film is not perfect, and when x/μ is not small, the Fourier transform system ($\beta = 0$) once again provides the most efficient means of storing structural information.

4) However, the maximum storage efficiency of the three-dimensional system is not one but one-half. The reason for this is that the resolving power of the plate is not fully utilized in the direction normal to the direction of

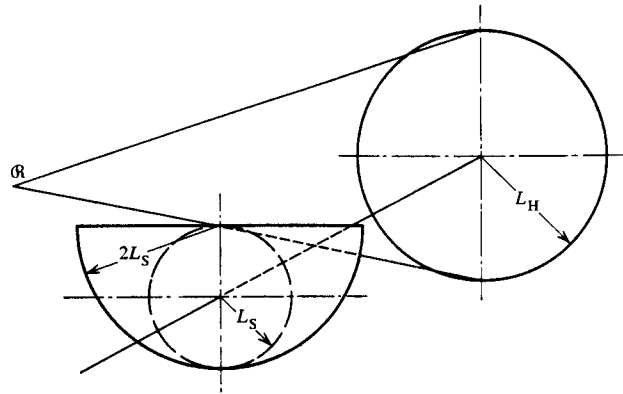


Figure 5 A modified three-dimensional holographic system in which the circular object domain, the dashed circle, is replaced by a 180° sector of twice the area.

sideband shift (i.e., in the direction normal to the plane of Fig. 1).¹⁴ The efficiency of the system can be improved by replacing the circular object domain by a 180° sector of twice the area, as shown in Fig. 5. When the object dimension L_S is restricted by a medium of moderate resolving power, the efficiency of the modified system is nearly double that given by Eq. (13) for the system with circular object domain. However, as $\mu \rightarrow 1$, and the allowable object dimension increases, the efficiency of the modified system decreases in the limit to the previous value of one-half. For when the ratio $L_S/z \rightarrow \infty$, the solid angle subtended at the plate by the 180° sector approaches that subtended by the circular domain.

The above considerations apply equally well when the object occupying the interval AB in Fig. 1 is replaced by a lens of the same aperture. This lens can be used to record the focused image of a coherent object (image-plane hologram),⁴ or to record its Fraunhofer diffraction image (Fraunhofer hologram). Indeed, the lens can be removed altogether, leaving an aperture in an opaque screen; the hologram records the incident wave field arising from any three-dimensional distribution of coherent objects located to the left of the aperture. The maximum allowable dimensions of the aperture and hologram are determined from Eq. (6), and the maximum number of degrees of freedom in the recorded wave field is given by Eq. (12) or (13), depending on the number of spatial dimensions considered.

We note finally that if one requires angular separation of the intermodulation background from the first-order image, then the storage efficiencies will amount to about one-half those given by Eqs. (12) and (13).

a medium of given dimensions and resolving power. Thus, for the two-dimensional system, we can equate the right-hand members of Eqs. (12) and (19) [the former with $\beta = 0$, the latter with $(r - s) - (a - b) = 0$] to obtain

$$\Delta l = \frac{2\mu L_H}{[1 + (1 - \mu^2)(L_H/z)^2]^{\frac{1}{2}}}. \quad (21)$$

That is, for given values of μ , L_H and z , the coherence length Δl need not exceed the value given by Eq. (21). Note the two limiting cases

$$\Delta l = 2\mu L_H \quad (\mu = 1 \text{ or } L_H/z \ll 1) \quad (22)$$

and

$$\Delta l = 2\mu z \quad (\mu \ll 1 \text{ and } L_H/z \gg 1). \quad (23)$$

In all cases the allowable object dimension is given by Eq. (6) with $\beta = 0$.

These equations apply as well to the three-dimensional Fourier-transform system. The parameter L_H in this case refers to the radius of the circular hologram.

Acknowledgments

This work was begun under the auspices of a fellowship granted for the academic year 1966-67 by the Institute of Science and Technology of the University of Michigan. The author thanks C. R. Worthington of the Mellon Institute, Carnegie-Mellon University for his interest and guidance in the pursuit of the solution to this problem.

Appendix

The resolution criterion, Eq. (6), can be derived with the aid of Fig. 4. The region M denotes the location of the interference fringes of highest spatial frequency, and the resolving power needed to record them is given by Eq. (5). Now by adding and subtracting identical quantities to and from the right-hand side of Eq. (5), we can write this condition in the form

$$\begin{aligned} N\lambda \geq & (\cos \alpha_0 - \cos \phi_0) - (\cos \phi_1 - \cos \phi_0) \\ & + (\cos \alpha_M - \cos \alpha_1) \\ & + (\cos \phi_1 - \cos \phi_M) - (\cos \alpha_0 - \cos \alpha_1). \end{aligned} \quad (A1)$$

In Eq. (A1) the quantity $(\cos \alpha_0 - \cos \phi_0)$ measures the numerical aperture of the system as seen from the upper edge of the object, and the quantity $(\cos \phi_1 - \cos \phi_0)$ measures the spatial-frequency content of the reference wave. The quantity $(\cos \alpha_M - \cos \alpha_1)$ measures the "numerical field of view" as seen from M.

As for the final two terms in Eq. (A1), we can make the following observations. In the case of plane-wave reference illumination the region M occurs at the top of the plate, and the two terms in question each vanish identically. When the reference illumination issues from

a point at the upper edge of the object, the region M occurs at the center of the plate; the two terms are in this case equal but opposite in sign, and therefore cancel. On geometrical grounds it is clear that in any intermediate case the combination of these two terms can be neglected. Hence, to a good approximation, Eq. (A1) can be written

$$N\lambda \geq \beta(\cos \alpha_0 - \cos \phi_0) + (\cos \alpha_M - \cos \alpha_1), \quad (A2)$$

where β is given by Eq. (8). We have thus expressed the resolution criterion in terms of the numerical aperture, the numerical field of view and the parameter β , which specifies the position of the reference source.

The problem of relating the structural-information content of the recorded wave field to the resolving power N requires that we solve Eq. (A2) for the object dimension L_s . An exact solution is not available for arbitrary values of β . However, an approximate solution can be obtained as follows.

When $\beta = 1$ (plane-wave reference illumination) an exact solution for L_s is readily obtained:

$$L_s \leq \frac{\mu z}{(1 - \mu^2)^{\frac{1}{2}}} - L_H, \quad (A3)$$

where $\mu = N\lambda/2$. An exact solution for L_s can also be found when $\beta = 0$ (Fourier-transform arrangement):

$$L_s \leq \frac{\mu z}{(1 - \mu^2)^{\frac{1}{2}}}. \quad (A4)$$

Consider now the paraxial case in which $(\cos \alpha_0 - \cos \phi_0) \approx 2L_H/z$ and $(\cos \alpha_M - \cos \alpha_0) \approx 2L_s/z$. This gives, for arbitrary values of β ,

$$L_s \leq \mu z - \beta L_H. \quad (A5)$$

A comparison of Eq. (A5) with Eqs. (A3) and (A4) suggests the following generalization of Eq. (A5) to the wide-angle case:

$$L_s \leq \frac{\mu z}{(1 - \mu^2)^{\frac{1}{2}}} - \beta L_H, \quad (A6)$$

which is equivalent to Eq. (6). This result is exact in the limiting cases $\beta = 1$ and $\beta = 0$. It is not exact for intermediate values of β , but the expression no doubt provides an adequate estimate of the resolving power needed for given values of L_H , L_s and z .

References and notes

1. D. Gabor, *Progress in Optics* **1**, 109 (1961).
2. G. B. Parrent and G. O. Reynolds, *J. Opt. Soc. Am.* **56**, 1400 (1966).
3. J. B. DeVelis and G. O. Reynolds, *Theory and Applications of Holography*, Addison-Wesley Publishing Co., Reading, Massachusetts, 1967, pp. 102-104.
4. A. W. Lohmann, IBM Research Paper RJ-438 (May 9, 1967).

5. W. Lukosz, *J. Opt. Soc. Am.* **57**, 932 (1967).
6. H. M. Smith, *Principles of Holography*, Interscience Publishers (division of John Wiley and Sons, Inc.), New York, 1969, pp. 217-226.
7. M. von Laue, *Ann. d. Physik* **44**, 1197 (1914).
8. G. Toraldo di Francia, *J. Opt. Soc. Am.* **45**, 497 (1955).
9. E. N. Leith, J. Upatnieks, B. P. Hildebrand and K. Haines, *J. Motion Picture Televis. Engrs.* **74**, 893 (1965).
10. D. J. Stigliani, R. G. Semonin and R. Mittra, *Proc. IEEE* **55**, 1509 (1967).
11. J. T. Winthrop and C. R. Worthington, *Phys. Letters* **15**, 124 (1965).
12. R. F. van Ligten, *J. Opt. Soc. Am.* **56**, 1009 (1966).
13. These are the "transverse" space-bandwidth products associated with the recording media. As already mentioned, the recording geometry of Fig. 1 precludes the appearance of the added dimension of emulsion depth.
14. An explanation of this effect in terms of the hologram spectrum is given in Ref. 4.
15. The value of R required to realize condition (17) is obtained by subtracting Eq. (16) from Eq. (15) (both expressions taken with equality signs):

$$(r - R) + (s - R) = a + b.$$
16. Note the formal correspondence,

$$\begin{aligned} \lambda &\leftrightarrow L_H \\ 1/\Delta\lambda &\leftrightarrow N. \end{aligned}$$

Received December 12, 1969