Trans. [RE-PG[T, IT-2 (June 1956), 56-65

Spatial Filtering in Optics*

EDWARD L. O'NEILL†

Summary—**Starting with the formulation of H. H. Hopkins for the image forming properties of an optical system in terms of a coherence factor over the object plane, the two extreme cases of complete coherence and incoherence are considered. The incoherent case is treated briefly as a low-pass spatial frequency filter.**

In the case of coherent illumination, it is shown that the optical analog of such well-known electrical concepts as equalization [17], edge-sharpening, and the detection of periodic and isolated signals in the presence of noise can be carried out with relative ease. A detailed theoretical treatment of the problem together with illustrations emphasizes the analogy between optical and electrical filtering.

List of Symbols and Fourier Relations[1]

$\hat{t}(x_i, y_i)$ = point image amplitude distribution.
$t(x_i, y_i)$ = point image intensity distribution.
$\hat{o}(x_0, y_0)$ = complex amplitude transmission of object.
$o(x_0, y_0)$ = object intensity distribution.
$\hat{\imath}(x_i, y_i)$ = image amplitude distribution.
$i(x_i, y_i)$ = image intensity distribution.
$\gamma(x_0, y_0; x_0', y_0')$ = "partial coherence factor" in object plane.
$\hat{\tau}(\mu, \nu)$ = complex transmission of aperture.
$\tau(\mu, \nu)$ = transfer function.
$\hat{O}(\mu, \nu)$ = object amplitude spatial spectrum.
$O(\mu, \nu)$ = object intensity spatial spectrum.
$\hat{I}(\mu, \nu)$ = image amplitude spatial spectrum.
$I(\mu, \nu)$ = image intensity spatial spectrum.
$\Gamma(\mu, \nu)$ = intensity variation over the source.

All functions denoted by the same small and capital letter are Fourier Transform pairs, *e.g.*,

$$\hat{o}(x_0, y_0) = \iint_{-\infty}^{\infty} \hat{O}(\mu, \nu) e^{-i(\mu x_0 + \nu y_0)} \, d\mu \, d\nu.$$

Introduction

There has been sufficient evidence over the past few years to indicate that the concepts of electrical communication theory have become a permanent fixture in the field of optics [1], [2], [4], [6], [12], [21-23], so that a detailed description of the electrical-optical analog is superfluous at this point. Nevertheless, it can be pointed out, in retrospect, that most of the emphasis thus far has been confined to the passive role of determining the performance of the optical system [8], [14-16], [23-25], namely, the evaluation of the transfer function or sine wave response curve. A study of the exceptions to this statement shows that, even in these cases, strictly speaking, a communication theory approach has not been employed [10], [13].

With the publication of an article by Maréchal and Croce [17] on increasing the contrast of photographs, it became apparent that the more general problem of detecting and recognizing signals in the presence of noise might be approached through the concepts of electrical communication theory. Not only was the general synthesis of an optical system possible, but it appeared immediately to involve some simplifications as compared to its electrical counterpart.

Rather than consider each optical system as a special problem, it was decided to treat the general problem of image formation by including a factor which describes the illumination over the object plane. For this reason the formulation of Hopkins [8], [9] was adopted to evaluate the final intensity distribution in terms of the "partial coherence factor." It was then possible to discuss the two extreme cases of completely coherent and incoherent systems. The former, being linear in amplitude, could be handled from the Fourier standpoint. The latter, being linear in intensity, was also subject to Fourier analysis and synthesis along with certain basic limitations. However, since in the coherent case, control over the Fourier components that make up the image could be exercised with relative ease, it was possible to carry out a series of simple experiments that bore a one-to-one correspondence to known filtering operations in electrical communication theory. The paper presents both a theoretical and experimental treatment of the general formulation of the problem together with illustrations of the results.

The Fourier Analysis and Synthesis of Coherent and Incoherent Optical Systems

Generalized Image Formation

In any comprehensive treatment of image formation in an optical system, it is convenient to incorporate a factor which describes the mode of illumination over the object plane. For monochromatic systems Hopkins [8], [9] has succeeded in doing this in terms of a "partial coherence factor." Because of the comprehensive character of such a formulation, this term will be employed in the analysis to be presented in this paper. However, only the two extremes of complete incoherence and coherence will be considered here.

As a starting point, consider an element of the source $d\sigma$ (Fig. 1) producing a complex disturbance $\hat{u}(x_0, y_0)$ at the point (x_0, y_0) in the object space where $\hat{u}(x_0, y_0)$ is a

* This work was supported by the Aerial Reconnaissance Lab., Wright Air Development Center, under Air Force Contract No. AF 33(616)-432.

† Boston Univ. Phys. Res. Labs., Boston, Mass.

[1] This notation follows that of Linfoot and Fellgett [11] in which the "hooked" (^) variables denote complex amplitude and the "unhooked" variables denote intensity.

solution of the scalar wave equation. The resultant complex transmitted amplitude is now given by $\hat{u}(x_0, y_0)\hat{o}(x_0, y_0)$. To determine the image distribution $\hat{\imath}(x_i, y_i)$, it is necessary to integrate the point image distribution $\hat{t}(x_i, y_i)$ over the object plane in the following manner:

$$\hat{\imath}(x_i, y_i) = \iint_{-\infty}^{\infty} \hat{t}(x_i - x_0, y_i - y_0)\hat{u}(x_0, y_0)\hat{o}(x_0, y_0)\, dx_0\, dy_0. \tag{1}$$

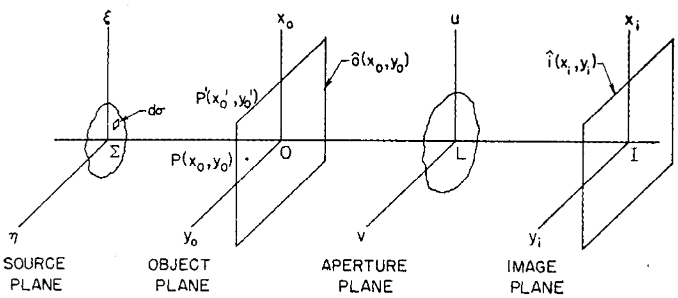


Fig. 1—Generalized image formation.

However, the eye or any physical instrument is sensitive to intensity variations, hence we must determine the image distribution from an independent point in object space (x_0', y_0') and perform a similar integration over the complex conjugate of the functions in (1). We have then

$$di(x_i, y_i) = \hat{\imath}(x_i, y_i)\hat{\imath}^*(x_i, y_i)\, d\sigma \tag{2a}$$

or

$$di(x_i, y_i) = d\sigma \iiiint_{-\infty}^{\infty} \hat{t}(x_i - x_0, y_i - y_0)\hat{t}^*(x_i - x_0', y_i - y_0')$$
$$\cdot \hat{u}(x_0, y_0)\hat{u}^*(x_0', y_0')\hat{o}(x_0, y_0)\hat{o}^*(x_0', y_0')\, dS_0\, dS_0' \tag{2b}$$

where

$$dS_0 = dx_0\, dy_0 \quad \text{and} \quad dS_0' = dx_0'\, dy_0' .$$

Finally, adding up the contributions from the source in the final integral gives

$$i(x_i, y_i) = \iiiint_{-\infty}^{\infty} \left[\iint_{-\infty}^{\infty} \hat{u}(x_0, y_0)\hat{u}^*(x_0', y_0')\, d\sigma \right]$$
$$\cdot \hat{t}(x_i - x_0, y_i - y_0)\hat{t}^*(x_i - x_0', y_i - y_0')$$
$$\cdot \hat{o}(x_0, y_0)\hat{o}^*(x_0', y_0')\, dS_0\, dS_0' . \tag{3}$$

Now, the term inside the brackets is Hopkins' partial coherence factor $\gamma(x_0, y_0; x_0', y_0')$ which, for a plane source, is given by

$$\gamma(x_0, y_0; x_0', y_0') = \iint_{-\infty}^{\infty} \Gamma(\mu, \nu) e^{i[\mu(x_0 - x'_0) + \nu(y_0 - y'_0)]}\, d\mu\, d\nu \tag{4}$$

where $\Gamma(\mu, \nu)$ describes the intensity variation and geometry of the source, and is zero outside the source. γ can be defined in terms of the contrast of the Young's fringes in the (μ, ν)[2] plane formed by pinholes at (x_0, y_0) and (x_0', y_0') when illuminated by a source whose geometry and intensity variation is given by $\Gamma(\mu, \nu)$. Substitution of (4) into (3) results in the intensity distribution over the image plane in the compact form

$$i(x_i, y_i) = \iiiint_{-\infty}^{\infty} \gamma(x_0, y_0; x_0', y_0')$$
$$\cdot \hat{t}(x_i - x_0, y_i - y_0)\hat{t}^*(x_i - x_0', y_i - y_0')$$
$$\cdot \hat{o}(x_0, y_0)\hat{o}^*(x_0', y_0')\, dS_0\, dS_0' . \tag{5}$$

Though (5) appears as a rather formidable formulation of the final image intensity variation, the problem simplifies considerably when either of the two extremes of complete coherence and incoherence are considered.

Frequently it is convenient to carry out the analysis of a system, not in terms of the independent space variables, but rather in the corresponding spatial frequency domain. Hence, proper manipulation with Fourier integrals, similar to Hopkins' treatment of Fourier series, yields

$$i(x_i, y_i) = \iiiint_{-\infty}^{\infty} \tau(\mu, \nu; \mu', \nu')\hat{O}(\mu, \nu)\hat{O}^*(\mu', \nu')$$
$$\cdot e^{i[(\mu - \mu')x_i + (\nu - \nu')y_i]}\, d\mu\, d\nu\, d\mu'\, d\nu' \tag{6}$$

where

$$\tau(\mu, \nu; \mu', \nu') = \iint_{-\infty}^{\infty} \Gamma(s, t)$$
$$\cdot \hat{\tau}(s + \mu, t + \nu)\hat{\tau}^*(s + \mu', t + \nu')\, ds\, dt \tag{7a}$$

and

$$\Gamma(s, t) = \iint_{-\infty}^{\infty} \gamma(x_0, y_0) e^{-i(sx_0 + ty_0)}\, dx_0\, dy_0. \tag{7b}$$

With this generalized formulation in mind, we will now focus our attention on the problems of the Fourier analysis and synthesis of optical systems which are linear in intensity (incoherent) and linear in amplitude (coherent). Before distinguishing between the analysis and synthesis of incoherent systems, it seems appropriate at this point to discuss the physical significance of completely incoherent illumination with respect to (5), (6), and (7). To begin with, when dealing with a self-luminous object, we should expect only point-to-point coherence or incoherence since every point in object space radiates independently of every other point. Mathematically, such a situation can be described in terms of the well-known Dirac "delta-

[2] The relation between the cartesian coordinates (u, v) in Fig. 1 and the reduced coordinates (μ, ν) is

$$\mu = \frac{ku}{p}, \qquad \nu = \frac{kv}{p}$$

where $k = 2\pi/\lambda$ is the wave number of the radiation and p is the distance between the object and aperture plane in Fig. 1.

function." Thus describing the partial coherence factor by the delta function and making use of (7b), we can, conceptually at least, replace the self-luminous object by a transparent one illuminated by an "effective source," infinite in extent. In a similar fashion, for a transparent object illuminated by a large extended source, we can consider such an object self-luminous for all practical purposes, provided the source is large enough [8], [9]. Since this type of illumination is the one most frequently encountered, most of the effort thus far in optics has been directed toward the analysis of incoherent systems.

Incoherent Illumination

Analysis: In a broad sense, the analysis of a linear system can be defined as the determination of the Fourier frequency components that the system passes to make up the output. In the optical case, the sine wave response of the system is determined as a function of the spatial frequency from a knowledge of the geometry and complex transmission of the aperture; *e.g.*, coating, aberrations, etc.

To illustrate this, consider (5) where, for the incoherent case, the partial coherence factor $\gamma(x_0, y_0; x_0', y_0')$ is a "delta" function. This gives

$$i(x_i, y_i) = \iiiint_{-\infty}^{\infty} \delta(x_0 - x_0', y_0 - y_0') \cdot \hat{t}(x_i - x_0, y_i - y_0)\hat{t}^*(x_i - x_0', y_i - y_0') \cdot \hat{o}(x_0, y_0)\hat{o}^*(x_0', y_0')\, dx_0\, dy_0\, dx_0'\, dy_0'. \tag{8}$$

Integrating over $dx_0'dy_0'$ and making use of the sifting property of the "delta" function, we have

$$i(x_i, y_i) = \iint_{-\infty}^{\infty} |\hat{t}(x_i - x_0, y_i - y_0)|^2\, |\hat{o}(x_0, y_0)|^2\, dx_0\, dy_0 \tag{9}$$

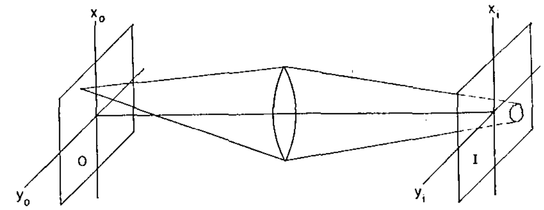
or

$$i(x_i, y_i) = \iint_{-\infty}^{\infty} t(x_i - x_0, y_i - y_0)o(x_0, y_0)\, dx_0\, dy_0 \tag{10}$$

where we see that the system is linear in intensity. (See Fig. 2.) The convenience of the Fourier approach becomes apparent when we perform a Fourier transformation of (10) to the spatial frequency domain and in the process replace the rather cumbersome convolution integral by a multiplication process. This gives

$$I(\mu, \nu) = \tau(\mu, \nu)O(\mu, \nu). \tag{11}$$

Thus we may consider the transfer or system function $\tau(\mu, \nu)$ operating on the spatial spectrum of the object intensity distribution to produce the spatial spectrum of the image intensity distribution. The term $\tau(\mu, \nu)$ has been assigned a variety of names in recent years, and, in fact, may be defined in a number of ways. For the sake of clarity and simplicity, we will define it in this paper as a measure of the reduction in contrast, and relative phase shift in passing from object to image space, of a sinusoidal intensity variation of increasing spatial frequency. Its units along the abscissa will be in reciprocal length (lines/mm), and it will be called the "transfer function."



$$i(x_i, y_i) = \iint_{-\infty}^{\infty} t(x_i - x_0, y_i - y_0)o(x_0, y_0)\, dx_0\, dy_0$$

$$I(\mu, \nu) = \tau(\mu, \nu)O(\mu, \nu)$$

Fig. 2—A linear incoherent optical system.

A review of the literature [16], [18], [23-25] shows that most of the attention thus far in recent years has been focussed upon the evaluation of $\tau(\mu, \nu)$. The reason for this is apparent when we consider the various Fourier relations that exist in systems that are diffraction limited. For example, one way of defining $\tau(\mu, \nu)$ mathematically is in terms of the Fourier transform of the point image intensity distribution; *i.e.*,

$$\tau(\mu, \nu) = \iint_{-\infty}^{\infty} t(x_i, y_i)e^{i(\mu x_i + \nu y_i)}\, dx_i\, dy_i \tag{12a}$$

or

$$\tau(\mu, \nu) = \iint_{-\infty}^{\infty} \hat{t}(x_i, y_i)\hat{t}^*(x_i, y_i)e^{i(\mu x_i + \nu y_i)}\, dx_i\, dy_i \tag{12b}$$

which, after application of a well-known theorem for the Fourier transform of a product, gives

$$\Gamma(\mu, \nu) = \iint_{-\infty}^{\infty} \hat{\tau}(\mu', \nu')\hat{\tau}^*(\mu' + \mu, \nu' + \nu)\, d\mu'\, d\nu' \tag{13}$$

where $\hat{\tau}(\mu', \nu')$ describes the complex transmission and geometry of the aperture and is zero outside the aperture. For a uniform circular aperture with aberrations, we have the following:

$$\hat{\tau}(\mu', \nu') = \begin{cases} e^{ik\Delta(\mu', \nu')} & \mu', \nu' \subset A \\ 0 & \mu', \nu' \not\subset A \end{cases} \tag{14}$$

where A denotes the area of the aperture and where $\Delta(\mu', \nu')$ is the deviation of the actual wave front from a reference sphere (the aberrations). Because of this apparent simplification, it should be possible from lens design data to substitute the proper coefficients (*e.g.*, Seidel aberration coefficients) into the expression for $\Delta(\mu', \nu')$ and perform the convolution operation described by (13). In practice, of course, because of the form of (14), the integral can rarely be evaluated in closed form. Nevertheless, despite the practical difficulties associated with this approach the treatment of an incoherent optical system as a two-dimensional low-pass spatial frequency filter has strong conceptual appeal.

One advantage of such an approach, (13) and (14), is that in theory at least, one can evaluate the transfer function from a knowledge of the aberration coefficients without ever examining in detail the complicated physical diffraction pattern.

Synthesis and Limitations: In place of the passive role one assumes in analyzing linear systems, one might wish to take a more active part in altering the system's performance by changing the image forming properties of the system and, hence, the transfer function. One method of doing this is to apply complex amplitude coatings over the aperture [10], [13], and [19] and hence alter the point image distribution. An extreme in this direction would be to place a central obstruction in the aperture so that the transparent area is annular [18], [25]. In fact, it might be possible to optimize the frequency response within a given prespecified spatial bandwidth of particular interest.[3]

Although several of these schemes have met with some success for special problems, the flexibility afforded to spatial filtering operations through generalized synthesis is quite limited with incoherent systems. A little reflection will show that this must be the case. Basically, we are dealing with the addition of nonnegative intensity variations everywhere in the image plane which will always give rise to background or constant levels, so that all incoherent systems behave as low-pass spatial frequency filters.[4] Bearing this in mind, we are led to the obvious conclusion that general spatial filtering can be performed only in systems in which destructive as well as constructive interference can occur. That is, we must be able to control amplitude and phase and, therefore, we are led to a consideration of the image forming properties of coherent systems.

Coherent Illumination

Analysis: It might be well to preface this section with the statement that the optical system to be discussed here would be described by the microscopist as one employing "point source Kohler illumination." In terms of the equations already derived for the more general case, it is now necessary to impose the condition that, for a point source, $\Gamma(\mu, \nu)$ is given by the Dirac delta function. Under this condition (5) becomes

$$i(x_i, y_i) = \iint_{-\infty}^{\infty} \hat{t}(x_i - x_0, y_i - y_0)\hat{o}(x_0, y_0)\, dx_0\, dy_0 \cdot \iint_{-\infty}^{\infty} \hat{t}^*(x_i - x_0', y_i - y_0')\hat{o}^*(x_0', y_0')\, dx_0'\, dy_0' \qquad (15a)$$

or

$$i(x_i, y_i) = \hat{\imath}(x_i, y_i)\hat{\imath}^*(x_i, y_i). \qquad (15b)$$

It is apparent that in terms of intensity, the system is nonlinear. Therefore, for convenience, the communication theory principles will be applied to a system linear in amplitude, tacitly assuming that (15b) will be needed to evaluate the resultant intensity variation.

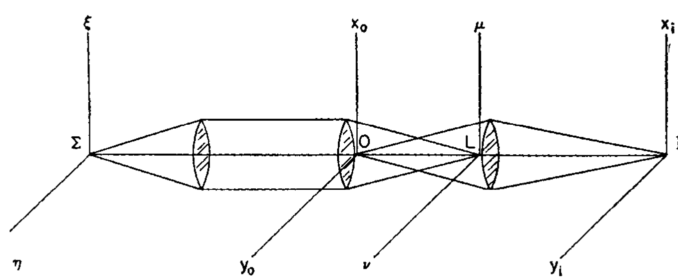
With this in mind, we can give the image amplitude distribution by the convolution integral of the point image amplitude distribution over the complex object amplitude distribution as follows:

$$\hat{\imath}(x_i, y_i) = \iint_{-\infty}^{\infty} \hat{t}(x_i - x_0, y_i - y_0)\hat{o}(x_0, y_0)\, dx_0\, dy_0. \qquad (16)$$

Here again it is possible to replace the convolution integral by a multiplication process by transforming to the spatial frequency domain. This gives[5]

$$\hat{I}(\mu, \nu) = \hat{\tau}(\mu, \nu)\hat{O}(\mu, \nu). \qquad (17)$$

Again this equation may be interpreted as an operation on the object amplitude spectrum to give the image amplitude spectrum. Now, however, we see that the amplitude transfer function is identically given by $\hat{\tau}(\mu, \nu)$ which describes the geometry and complex transmission of the aperture in the (μ, ν) plane. Thus it becomes apparent that the Fourier components that make up the image can be controlled by inserting the proper mask in this plane. (See Fig. 3.)



$$\hat{\imath}(x_i, y_i) = \iint_{-\infty}^{\infty} \hat{t}(x_i - x_0, y_i - y_0)\hat{o}(x_0, y_0)\, dx_0\, dy_0$$

$$\hat{I}(\mu, \nu) = \hat{\tau}(\mu, \nu)\hat{O}(\mu, \nu)$$

Fig. 3—A linear coherent optical system.

[3] For example, one possible criterion might be to minimize the mean square difference between the aberrated system and an ideal one over a given spatial frequency range, *i.e.*,

$$\epsilon = \frac{1}{\omega_2 - \omega_1} \int_{\omega_1}^{\omega_2} (\tau_0 - \tau)^2\, d\omega$$

where τ_0 = transfer function in the ideal case (no aberrations) and τ = actual transfer function. One might then determine the correct balancing of aberration coefficients to make ϵ a minimum. The frequency range $\omega_1 \leq \omega \leq \omega_2$ would of course be determined by the application.

[4] This of course is no longer true for electro-optical systems in which the constant term can easily be eliminated [11].

[5] This formulation can be arrived at by successively applying Huyghen's principle over the object and aperture planes as shown by Rhodes [21], [22] who in turn discusses the conditions under which the Fourier transform relations are applicable.

Before investigating these possibilities, it seems proper to verify the statements given above with a few simple examples. Consider first of all a clear unaberrated circular aperture. Then $\hat{\tau}(\mu, \nu)$ is the familiar cylindrical "top-hat" distribution, which "filter-wise" signifies an ideal low-pass, two-dimensional filter. There exist several basic limitations to perfect imagery in the system under discussion. The first is due to the wave character of radiation itself [6]. If the condition is imposed that the incident and emergent waves in object space must both satisfy the scalar wave equation, there results an inequality in terms of the wavelength of the light and the spacing between detail in the object. For detail closer together than the wavelength of the light, the coefficient of the space variable in the exponent becomes real and negative, instead of imaginary, which results in waves that die out within a few wavelengths of the object. This limitation exemplifies a well-known principle in communication theory which deals with the inability of a carrier to transmit information concerning details closer together than the wavelength of the carrier.

The second limitation imposed is due to the finite dimensions of the aperture. That is, while all diffracted angles between zero and ninety degrees are possible, only those Fraunhofer or Fourier orders will enter into the formation of the image that intercept the aperture in the (μ, ν) plane. Since this point of view is merely a generalized Abbé theory of image formation in the microscope, it is a simple matter to verify Abbé's claim, that for periodic targets the resolution limit of the system can be doubled by employing oblique illumination.

In order to do this, the point source must be moved off axis, which results in a tipped wave passing through the object plane. However, this is equivalent to having axial illumination with a linear phase plate over the object; hence the complex transmission of the object becomes

$$\hat{o}(x_0, y_0) = \hat{o}'(x_0, y_0)e^{ik(\alpha_0 x_0 + \beta_0 y_0)} \tag{18}$$

where α_0, β_0 are the direction cosines of the incident tipped wave. The Fraunhofer spectrum in the μ, ν plane is given by

$$\hat{O}(\mu, \nu) = \iint_{-\infty}^{\infty} \hat{o}'(x_0, y_0)e^{i[(\mu + k\alpha_0)x_0 + (\nu + k\beta_0)y_0]}\, dx_0\, dy_0, \tag{19}$$

and, if we define the terms $\mu_0 = k\alpha_0$ and $\nu_0 = k\beta_0$ (19) becomes

$$\hat{O}(\mu + \mu_0, \nu + \nu_0) = \iint_{-\infty}^{\infty} \hat{o}'(x_0, y_0)e^{i[(\mu+\mu_0)x_0 + (\nu+\nu_0)y_0]}\, dx_0\, dy_0. \tag{20}$$

That is, the spectrum of the object remains the same but is shifted to a new center which is, of course, the geometrical image of the displaced point source.

The resultant amplitude in the image plane is now given by

$$\hat{\imath}(x_i, y_i) = \iint_{-\infty}^{\infty} \hat{\tau}(\mu, \nu)\hat{O}(\mu + \mu_0, \nu + \nu_0)e^{-i(\mu x_i + \nu y_i)}\, d\mu\, d\nu. \tag{21}$$

Changing variables, we obtain

$$\hat{\imath}(x_i, y_i) = e^{i(\mu_0 x_i + \nu_0 y_i)} \iint_{-\infty}^{\infty} \hat{\tau}(\mu' - \mu_0, \nu' - \nu_0) \cdot \hat{O}(\mu', \nu')e^{-i(\mu' x_i + \nu' y_i)}\, d\mu'\, d\nu'. \tag{22}$$

Hence, the transfer function in amplitude is given by the complex transmission of the aperture centered about the shifted position as shown in Fig. 4(a).

To illustrate this point further, consider a crossed sinusoidal target with spatial frequencies ω_x, ω_y in the x and y directions respectively, such that

$$\hat{o}'(x_0, y_0) = e^{-i(\omega_x x_0 + \omega_y y_0)}. \tag{23}$$

Eq. (20) becomes

$$\hat{O}(\mu + \mu_0, \nu + \nu_0) = \iint_{-\infty}^{\infty} e^{i[(\mu+\mu_0-\omega_x)x_0 + (\nu+\nu_0-\omega_y)y_0]}\, dx_0\, dy_0 \tag{24}$$

which, except for a constant, is the integral representation of a delta function, so that

$$\hat{O}(\mu, \nu) = c^2\delta[\mu - (\omega_x - \mu_0), \nu - (\omega_y - \nu_0)] \tag{25}$$

where c is a constant. Substituting this expression into (21) and making use of the sifting property of the delta function, we obtain

$$\hat{\imath}(x_i, y_i) = c^2\hat{\tau}(\omega_x - \mu_0, \omega_y - \nu_0)e^{-i(\omega_x x_i + \omega_y y_i)}e^{i(\mu_0 x_i + \nu_0 y_i)} \tag{26}$$

which becomes, after normalizing,

$$\hat{\imath}(x_i, y_i) = \tau(\omega_x - \mu_0, \omega_y - \nu_0)\hat{o}'(x_i, y_i)e^{i(\mu_0 x_i + \nu_0 y_i)}. \tag{27}$$

Now noting that the last term merely describes the obliquity of the wave, we see that the object structure is reproduced in the image but that the amplitude is modulated by $\hat{\tau}(\omega_x - \mu_0, \omega_y - \nu_0)$. This implies that if the Fourier or Fraunhofer orders do not fall within the region of the shifted aperture function $\hat{\tau}(\mu, \nu)$, the periodicity will not be reproduced in the image. (See Fig. 4.) Now, since the spectrum is shifted by an amount corresponding to the direction cosines of the incident illumination, the final contrast and resolution limit are functions of the complex transmission of the system, the angular size of the aperture, and the direction of illumination. In terms of the sine wave resolution limit of the system, it is necessary that ω_x satisfy the following inequality in order for the periodicity to be detected in the image:

$$\omega_x \leq \mu_0 + \mu_1 \tag{28}$$

where μ/k represents the direction cosine of incident illumination and μ_1/k the maximum angular dimension of the aperture. Then,

$$\frac{2\pi}{P_x} \leq k(\sin\theta_0 + \sin\theta_1) \tag{29}$$

or, at the resolution limit,

$$P_x = \frac{\lambda}{(NA)_{\text{obj.}} + (NA)_{\text{coll.}}}. \tag{30}$$

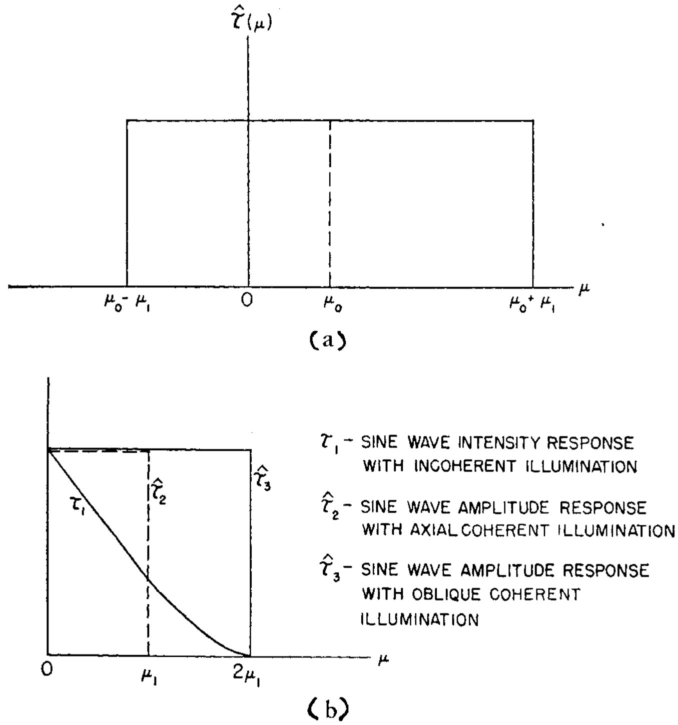


Fig. 4—Comparison of transfer functions.

For axial illumination

$$P_x = \frac{\lambda}{(NA)_{\text{obj.}}}. \tag{31}$$

For oblique illumination with the numerical aperture (NA) of the collimator equal to the numerical aperture of the objective,

$$P_x = \frac{\lambda}{2(NA)_{\text{obj.}}} \tag{32}$$

thus verifying Abbé's claim that the resolution limit for periodic structures can be doubled by employing the proper oblique illumination.

We have illustrated rather sketchily what might be termed the analysis of a linear coherent optical system. It now becomes clear that the synthesis of such a system is more promising than the synthesis of the previously discussed incoherent case. This is evident from an examination of (17) and Fig. 3, where it can be seen that the Fourier spectrum that constitutes the image can be controlled by inserting the proper mask in the (μ, ν) plane and hence controlling $\hat{\tau}(\mu, \nu)$. It is upon this aspect of the system that attention will be focussed in the remainder of the paper.

Before formulation of the communication theory approach to the problem of spatial filtering, it should be pointed out that this scheme of controlling the Fourier spectrum of the image is not new. The obvious illustration of this is the phase contrast microscope; however, the method first came to the attention of the author in an article by Maréchal and Croce [17] on increasing the contrast of photographs by a method known in communication theory as equalization [4]. Further investigation showed that Porter [20] in 1906 had performed a series of then remarkable spatial filtering experiments in verification of the Abbé theory of the microscope.

Synthesis: The purpose of the preceding section was to illustrate how a linear system can be analyzed by Fourier methods. It has been shown that the important contribution of such an analysis is that the rather cumbersome convolution process in the space domain can be replaced by the more practical multiplication process in the spatial frequency domain. It is now evident that the key to such an analysis is the transfer function $\hat{\tau}(\mu, \nu)$, which operates on the input spectrum to give the output spectrum, in this case the object and image respectively. In the synthesis processes to be discussed, it will be demanded that $\hat{\tau}(\mu, \nu)$ assume special characteristics in order to perform certain predescribed operations.

In the selection of $\hat{\tau}(\mu, \nu)$ in the remainder of this paper, the terminology and experience of electrical communication theory will be drawn on wherever applicable so that, to a large extent, the analogy between the electrical and optical systems will be emphasized.

A. Equalization: It is well-known in communication theory that if we have n linear systems in cascade, the final output spectrum is given by

$$F_{\text{out}}(\omega) = \hat{\tau}_1(\omega)\hat{\tau}_2(\omega) \cdots \hat{\tau}_n(\omega)F_{\text{in}}(\omega) \tag{33}$$

where $\hat{\tau}_k(\omega)$ is the transfer function for the k'th system. It should be noted in passing that Fourier methods really demonstrate their versatility in analyses of this kind since the n-fold multiplication process in the frequency domain is the counterpart of a series of n-fold convolution integrals in the space or time domain.

In the process of equalization we would like, in theory at least, to control $\hat{\tau}_n(\omega)$ in such a way that it becomes the reciprocal of the product of $\hat{\tau}_1(\omega)\hat{\tau}_2(\omega) \cdots \hat{\tau}_{n-1}(\omega)$; that is, we would like to design a $\hat{\tau}_n(\omega)$ such that

$$\hat{\tau}_n(\omega) = \frac{1}{\prod_{k=1}^{n-1} \hat{\tau}_k(\omega)}. \tag{34}$$

The reason for this is obvious when we examine (33). Of course in practice such control is not always possible. It was in this respect that Maréchal and Croce were able to increase the contrast of photographs by inserting in the (μ, ν) plane a mask that was dense at the center and gradually became transparent at the edge in such a way that the product of $\hat{\tau}(\omega)$ with the primary imaging transfer function gave a flat response. The result of inserting such a mask is shown in Fig. 5.

B. Edge Sharpening: The next step in altering the spatial spectrum of a photograph is the deletion of the constant term. This is similar to inserting a condenser in an electrical network and represents a high-pass filter. In terms of $\hat{\tau}(\mu, \nu)$ it merely means the insertion of a small opaque spot on axis in the (μ, ν) plane. The result of this operation

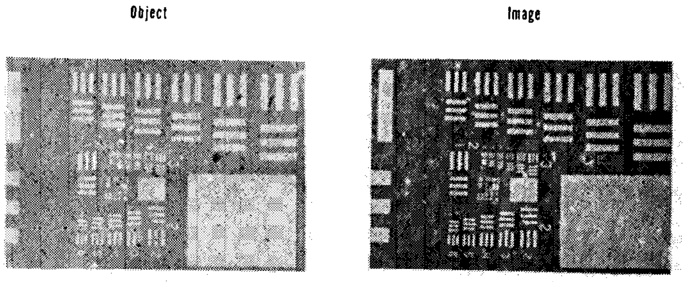


Fig. 5—Increase of contrast.

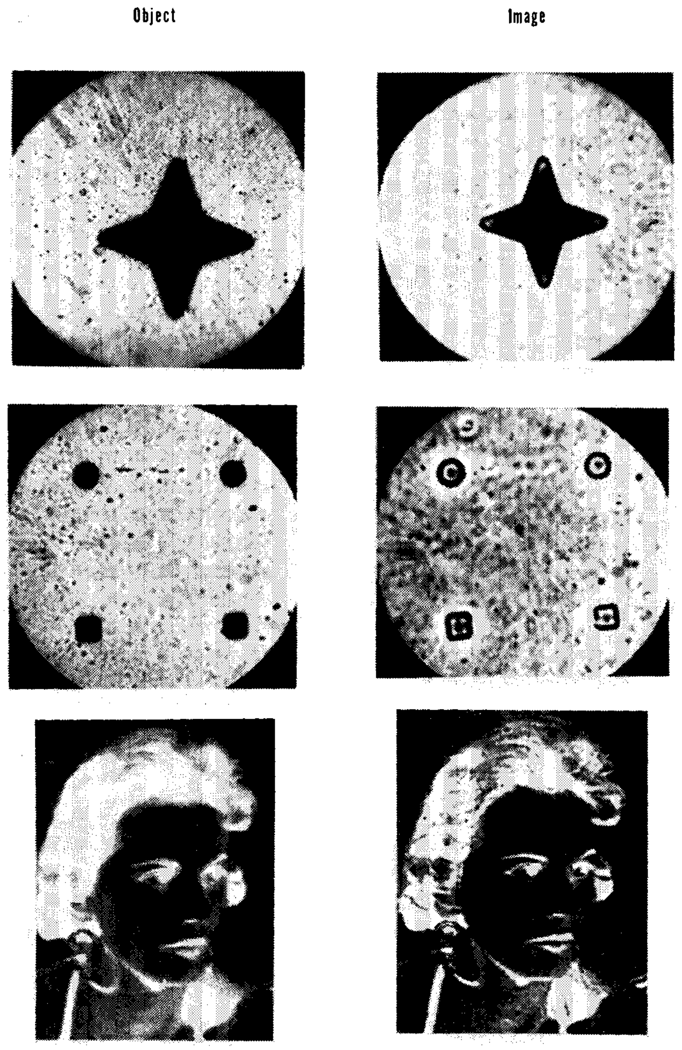


Fig. 6—Edge sharpening.

is shown in Fig. 6. Since, from an electrical standpoint the process is similar to differentiating the input, it would appear possible to add the edge-sharpened photograph to the original, optically or photographically. It should be noted that any operation such as edge sharpening or increasing signal to noise ratio for given geometrical shapes affects similar distributions anywhere in the field of view, since the Fraunhofer spectrum is formed on axis in the (μ, ν) plane. In addition to this favorable quality, the method has an additional property of performing filtering operations in two-dimensions at once, as opposed to electrical schemes.

C. Grain: In the special case where the distracting element in a photograph is the grain, these rapid fluctuations can be smoothed out as long as the size of the desired detail is not of the same order of magnitude. There are, of course, a number of other ways of doing this since all incoherent systems behave as low-pass spatial frequency filters. For example, the photograph could be blurred slightly or viewed out of focus. For the system under discussion, several photographs were chosen that appeared objectionably grainy to the eye. The aperture in the (μ, ν) plane was stopped down by a diaphragm until the visual image appeared noticeably smoother, and the image was then photographed. The result of rejecting these higher spatial frequency components due to grain is shown in Fig. 7. The effect is more pronounced when viewed with some magnification.

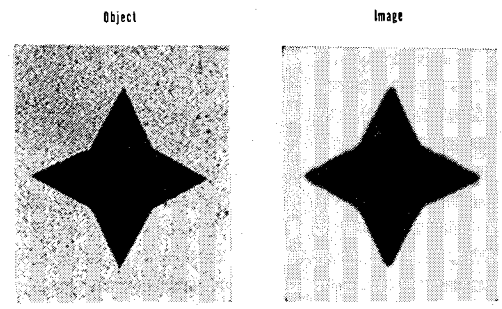


Fig. 7—Isolated detail with grain.

Detection and Recognition of Signals in the Presence of Noise: For convenience, in the remainder of this paper it will be assumed that the object function can be broken up into the sum of two separate functions, o_S and o_N, the signal and noise functions respectively. The reason for this is to maintain the electrical communication theory analog as far as possible. Bearing this analogy in mind, we will now subdivide the types of signals to be discussed into the three main categories of periodic, isolated (transient), and random detail.

A. Periodic Signals: It is a well-known principle in modern communication theory that an ideal method for recovering a periodic signal in noise is to cross correlate the signal plus noise distribution with a "Dirac comb;" *i.e.*, a series of evenly spaced sharp pulses at intervals corresponding to the fundamental period of the signal to be detected. This scheme has obvious applications in radar systems, for example. The optical equivalent of such a cross correlation procedure is to insert an opaque mask in the (μ, ν) plane with pinholes at positions corresponding to the spectral orders of the periodic portion of the target. The image forming lens recombines these spectral orders to produce the periodic disturbance in the image. (See Fig. 8.)

Loosely speaking, the opaque mask with the properly positioned pinholes can be represented as

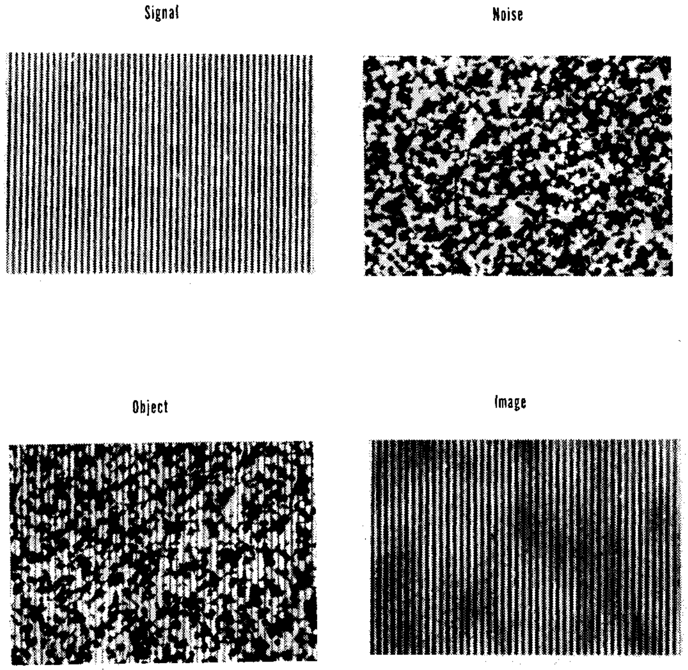


Fig. 8—Periodic detail in the presence of "noise".

$$\hat{r}(\mu, \nu) = \sum_m \sum_n \delta(\mu - \omega_m, \nu - \omega_n) \qquad (35)$$

where $\omega_m = m\omega_x$, and $\omega_n = n\omega_y$.

Multiplication of $\hat{r}(\mu, \nu)$ with the Fourier spectrum of signal and noise has the effect of passing the discrete spectrum of the periodic signal while sampling the continuous noise spectrum at discrete intervals. The result in the image plane is a reproduction of the periodic signal plus a small slowly varying term that oscillates about the average noise level. The calculation of the image distribution can actually be carried out for the case when there is little or no correlation between the noise and filter distributions [7]. Defining the total object distribution as a sum of signal and noise we have

$$\hat{o}(x_0, y_0) = \hat{o}_S(x_0, y_0) + \hat{o}_N(x_0, y_0) \qquad (36)$$

where

$$\hat{o}_S(x_0, y_0) = \sum_m \sum_n c_{mn} e^{-i(\omega_m x_0 + \omega_n y_0)}.$$

The resultant image distribution is a convolution integral of the impulse response with the $\hat{o}(x_0, y_0)$ in (36). Because a Dirac comb is its own Fourier transform, the total integral is the sum of a convolution integral of a Dirac comb function with the periodic signal plus the convolution of the comb function with the noise. Because of the periodicity one may deal with average values of the integrals in terms of the cross correlation function. It can be shown quite readily that the cross correlation of a periodic signal with a Dirac comb function of the same period is, except for a constant, a reproduction of the periodic signal. By contrast, if the noise and comb function are uncorrelated, the result is a constant related to the average noise level. With this in mind, the image distribution becomes

$$\hat{\imath}(x_i, y_i) = K\hat{o}_S(x_i, y_i) + \epsilon. \qquad (37)$$

B. Isolated Signals: The solution to this problem is not new to the communication engineer [7]; for this reason only an outline of the method will be presented here. The photointerpretive equivalent of maximizing the peak signal to rms noise ratio would be maximizing the probability of detecting an isolated piece of detail in unwanted background. The more selective recognition problem would seem to involve a sharpening or outlining operation. For the peak axial signal we may write

$$\text{peak axial signal} \sim \iint_{-\infty}^{\infty} \hat{r}(\mu, \nu)\hat{o}_S(\mu, \nu)\, d\mu\, d\nu \qquad (38)$$

and for the rms noise

$$\text{rms noise} \sim \left[\iint_{-\infty}^{\infty} |\, \hat{r}(\mu, \nu) \,|^2\, d\mu\, d\nu \right]^{1/2} \qquad (39)$$

so that the ratio becomes

$$\frac{\text{peak axial signal}}{\text{rms noise}} \sim \frac{\displaystyle\iint_{-\infty}^{\infty} \hat{r}(\mu, \nu)\hat{o}_S(\mu, \nu)\, d\mu\, d\nu}{\left[\displaystyle\iint_{-\infty}^{\infty} |\, \hat{r}(\mu, \nu) \,|^2\, d\mu\, d\nu \right]^{1/2}} \qquad (40)$$

where again the subscripts S and N denote signal and noise respectively. Several independent investigators have shown that this ratio is a maximum when

$$\hat{r}(\mu, \nu) = \hat{o}_S^*(\mu, \nu) \qquad (41)$$

that is, when the filter function is the complex conjugate of the desired signal spectrum. With such a filter, it can be readily shown that the image distribution is merely the autocorrelation function of the signal alone when signal and noise are uncorrelated.

An approximation to this filter can be achieved in the optical case by photographing the spectrum of the signal alone. The negative transparency can then be used as a mask when the signal and noise combination are viewed. (See Fig. 9). Note, however, that such a filter does not necessarily outline the signal as does the edge sharpening operation, but rather maximizes (40) and hence enhances the probability of detecting the signal. Finally, as Goldman [7] points out, for a given noise level, the ratio in (40) depends only upon the signal energy in the object. This would therefore appear to be a starting point for determining thresholds in terms of input (object) signal to noise ratio.

C. Random Signals: A detailed treatment of this problem is beyond the scope of this paper and present considerations. Nevertheless, for the sake of completeness, it proves of interest to examine how far we can go in ex-

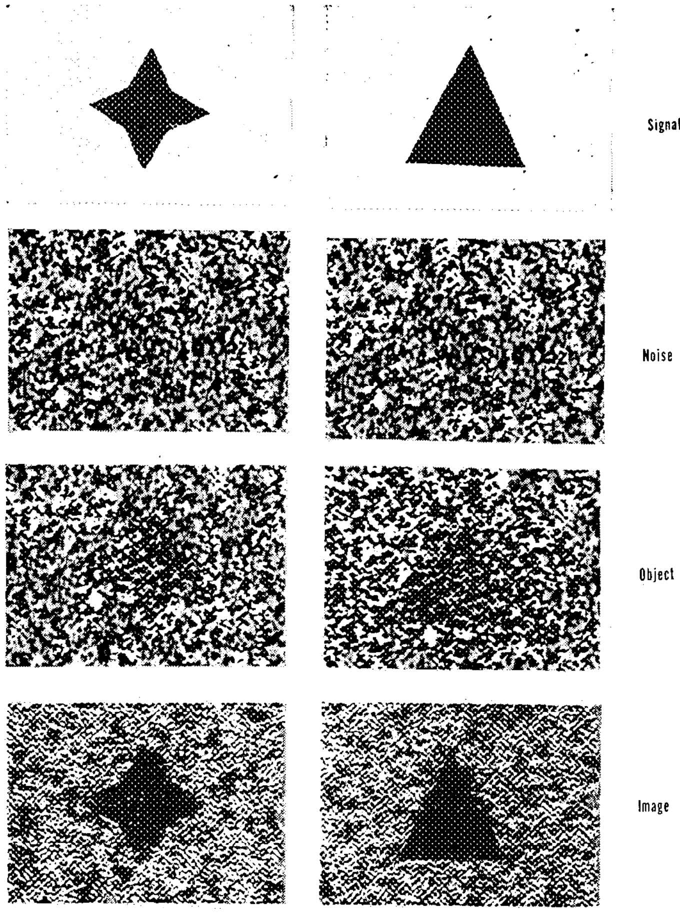


Fig. 9—Isolated signal in the presence of "noise".

tending the one-dimensional voltage vs time electrical considerations to the two-dimensional optical case. In such an extension, a suprising simplification is at once obvious in the optical problem. As pointed out by Cheatham and Kohlenberg [2], the fundamental difference between optical and electrical filters lies in the fact that once a filter function has been determined in the electrical case, one must then impose the condition that its response vanish for $t < 0$. This restriction is basic in that no amount of equipment or ingenuity can circumvent it. Such is not the case in the optical system. The impulse or point source response in optical systems lies on both sides of the space axes and in many cases possesses symmetry. From this point of view the design of an optimum filter presents some simplification in the optical case.

Conclusion

As pointed out recently by Linfoot and Fellgett [12], an optical system can be assessed from two distinct aspects. First of all, in terms of the primary imaging operation, it may be demanded that the intensity distribution in object space be transformed into an intensity distribution in image space with as high a degree of fidelity as possible. Secondly, assuming that it is impossible to exercise control over the primary imaging system, it may be demanded of a secondary imaging system that it extract as much useful information as possible from the photograph at the expense of unwanted background detail.[6] It is upon this latter aspect that attention has been focussed in this paper. Because the problem is not far removed from the problem of detecting signals in the presence of noise in electrical systems, the terminology of electrical network theory has been adopted wherever feasible.

A cursory examination was made of the image-forming properties of an optical system in general; the specialized cases of the analysis and synthesis of incoherent systems were discussed; and the limitations for such systems were mentioned, so far as spatial filtering is concerned. From an investigation of the coherent case, it became apparent that a method for performing spatial filtering could be carried out. To test this hypothesis, several experiments were attempted, in which the determination of the proper filter could be predicted from the equivalent electrical network problem. In fact, for many of the specialized problems of detecting certain geometrical patterns in the presence of unwanted background, the determination of the proper mask (filter) to insert becomes almost self-evident. Actually, for search problems such as this, the amount of information needed is analogous to the electrical case; that is, one must know either the desired pulse shape approximately or the statistical character of the noise (the power spectrum). In brief, if overlapping signal and noise are to be separated at all, they will be separated in their spectra, and the advantage of the optical over the electrical system is that such a separation takes place automatically in the optical case through the phenomenon of diffraction.

Finally, the same caution that the electrical engineer must exercise is warranted here except possibly to a larger extent. The insertion of a filter can often times give rise to false information in the image and unless properly interpreted may, in fact, negate the purpose of the filtering operation. While this information in the electrical case is most often put on a recording device and then studied objectively, it may happen that the optical image is examined visually which creates, as any microscopist knows, a number of psycho-physical problems. Whether or not this represents a basic limitation to the ideas discussed here remains to be determined.

Acknowledgment

The author is indebted to W. Wong for the bulk of the experimental work mentioned in this report and in particular for the production of the figures and diagrams used throughout the paper.

[6] This second aspect may imply reconstruction methods such as increasing contrast or edge sharpening, or it may possibly imply separation of usable and unwanted background detail in the photograph through spatial filtering.

Bibliography

[1] Blanc-Lapierre, A., "Upon Some Analogies Between Optics and Information Theory." Symposium on Microwave Optics, McGill University, Montreal, Canada, June, 1953.

[2] Cheatham, T. P., Jr., and Kohlenberg, A., "Optical Filters—Their Equivalence to and Difference from Electrical Networks." 1954 IRE Convention Record, part 4, "Electronic Computers and Information Theory," pp. 6–12.

[3] Duffieux, P. M., "L'Integrale de Fourier et ses Applications a l'Optique." Besancon, Faculte des Sciences, 1946.

[4] Elias, P., Grey, D. S., and Robinson, Z., "Fourier Treatment of Optical Processes." *Journal of the Optical Society of America*, Vol. 42 (February, 1952), p. 127.

[5] Elias, P., "Optics and Communication Theory." *Journal of the Optical Society of America*, Vol. 43 (April, 1953), p. 229.

[6] Gabor, D., "Lectures on Communication Theory." Technical Report No. 238, Research Laboratory of Electronics, Massachusetts Institute of Technology, April, 1952.

[7] Goldman, S., *Information Theory.* New York, Prentice-Hall, 1953.

[8] Hopkins, H. H., "The Concept of Partial Coherence in Optics." *Proceedings of the Royal Society A*, Vol. 208 (1951), pp. 263–277.

[9] Hopkins, H. H., "On the Diffraction Theory of Optical Images." *Proceedings of the Royal Society A*, Vol. 217 (1953), pp. 408–432.

[10] Jacquinot, P., Boughton, P., and Dossier, B., *Contributions in La Theorie des Images Optiques*. Paris, Editions de la Revue d'Optique, 1949, p. 183.

[11] Kovasznay, S. G., and Joseph, H. M., "Image Processing." Proceedings of the IRE, Vol. 43 (May, 1955), pp. 560–570.

[12] Linfoot, E. H., and Fellgett, P. B., "On the Assessment of Optical Images." *Royal Society of London Philosophical Transactions*, Series A, Vol. 247, p. 931.

[13] Luneberg, R. K., *Mathematical Theory of Optics.* Providence, R. I., Brown University, 1944.

[14] Macdonald, D. E., "Air Photography." *Journal of the Optical Society of America*, Vol. 43 (April, 1953), p. 290.

[15] Macdonald, D. E., "Quality Aspects of the Aerial Photographic System." National Bureau of Standards Circular 526, Proceedings of the N.B.S. Semicentennial Symposium on Optical Image Evaluations, 1951.

[16] Maréchal, A., "The Contrast of Optical Images and the Influence of Aberrations." National Bureau of Standards Circular 526, Proceedings of the N.B.S. Semicentennial Symposium on Optical Image Evaluations, 1951.

[17] Maréchal, A., and Croce, P., "Un Filtre de Frequences Spatiales pour l'Amelioration du Contraste des Images Optiques." *Comptes Rendus de Academie des Sciences*, Vol. 237 (September, 1953), pp. 607–609.

[18] O'Neill, E., "Transfer Function for an Annular Aperture." *Journal of the Optical Society of America*, Vol. 46 (1956), p. 285.

[19] Osterberg, H., and Wilkins, J. E., Jr., "The Resolving Power of a Coated Objective." *Journal of the Optical Society of America*, Vol. 39 (1949), p. 553.

[20] Porter, A. B., "On the Diffraction Theory of Microscopic Vision." *Philosophical Magazine*, Vol. 11 (1906), p. 154.

[21] Rhodes, J. E., Jr., "Analysis and Synthesis of Optical Images." *American Journal of Physics*, Vol. 21 (May, 1953), pp. 5, 337.

[22] Rhodes, J. E., Jr., "Microscope Imagery as Carrier Communication." *Journal of the Optical Society of America*, Vol. 43 (October, 1953), p. 848.

[23] Schade, O. H., "Electro-Optical Characteristics of Television Systems." *RCA Review*, Vol. 9, Nos. 1, 2, 3, and 4 (1948).

[24] Selwyn, E. W. H., "Theory of Resolving Power." National Bureau of Standards Circular 526, Proceedings of N.B.S. Semicentennial Symposium on Optical Image Evaluation, 1951.

[25] Steel, W. H., "Etude des Effets Combines des Aberrations et d'une Obturation Centrale de la Pupille Sur le Contraste des Images Optiques." *Revue d'Optique*, Vol. 32 (1953), pp. 4–26, 143–178, 269–306.