Journal of Research of the National Bureau of Standards Vol. 56, No. 5, May 1956 Research Paper 2672

Characteristics of an Image-Forming System

Roland V. Shack

Two general approaches to the analysis of an image-forming system are considered. One depends on the image of a point object and the other on the Fourier transform of this image. The two are developed independently and then coordinated, a practical characteristic function being determined for each approach. The relative merits of the two approaches are considered.

1. Introduction

For the past few years considerable energy has been expended in the search for an objective procedure for evaluating the quality of images formed by optical instruments. Existing procedures have been found to be not entirely satisfactory, and much work has been done in measuring previously unused physical parameters, which are objectively determinable, for the purpose of correlating them with the existing quality criteria.

The objection to this is that such an empirical, and therefore much limited, correlation eliminates only one of the faults of the existing criteria, and this is the subjectivity of their determination. Any other weakness is ignored.

A better approach is to analyze the image-forming process as a phenomenon, the aim being to characterize the process in as general and inclusive a way as possible, consistent with practical instrumentation. New criteria of image quality would of course be expected to be developed. Many approaches have been made in this direction also, and the present paper is to be included among them. However, in contrast to some of the published material, the emphasis here is on the practicality and usefulness of the results obtained rather than on mathematical rigor, although the treatment should be rigorous enough to include all essential factors.

Let us consider this matter of practical instrumentation. The heart of the test instrument is the photosensitive detector, for it is this which provides the data by which the tested instrument is evaluated. Three practical photosensitive detectors are available—the eye, the photographic emulsion, and the photocell.

The only test of image quality for which the eye is capable of making quantitative measurements is the resolving-power test. This test is rapid and relatively inexpensive, but the information obtained is incomplete, the precision is low, and the results are subject to variation from individual to individual.

A photographic detector allows quantitative measurements to be made under nonthreshold conditions, but time is required for processing, the processing conditions must be rigidly standardized, the granularity and diffusion in the emulsion affect the results, the response of the film is nonlinear with respect to incident flux, and, in the end, an additional sensing mechanism, such as a microdensitometer, is needed to reduce the emulsion properties to numerical values.

The photocell is probably the most satisfactory photosensitive detector for the test instrument. Within its proper operating range, its characteristics remain reasonably constant, its response is linear with respect to incident flux, its spectral response can be adjusted so as to approximate that of the eye, and its output can very easily produce graphical or numerical results. However, it must be used in conjunction with an aperture that limits the spatial integration of the detail in the image being examined, and there must be provision for relative displacement between the aperture and the image so that various portions of the image may be sampled.

It should also be pointed out that the report is illustrated throughout by the characteristics of an aberration-free system with a circular aperture in monochromatic light, diffraction being the sole source of image degradation, and the light from various points in the object space being noncoherent. This has been selected as an interesting and informative type of system, which real systems tend to approximate as their quality improves. It must be emphasized,

379949—56——1

however, that this is used for illustration only. The material covered applies to all types of images, assuming noncoherence.

There are two viewpoints from which image evaluation can be approached. One, the classical viewpoint, considers the point image to be the most fundamental element in an image process. Any object can be regarded as a summation of points, and its image as the summation of the corresponding point images. An evaluation of the point image would be sufficient to characterize the system.

The other viewpoint involves the concepts of Fourier analysis. Here the object is considered to be the summation of a set of sinusoidal waves distributed in the object plane. These component waves, differing from each other in amplitude, frequency, phase, and direction, are spatially distributed waves. That is, they are spatial, not temporal, sinusoidal variations in brightness throughout the object plane. The image consists of the summation of the images of these component waves. A description of the manner in which the optical system forms images of these component waves would also be sufficient to characterize the system.

In section 2 a way of describing the point image is developed, which can be obtained, at least in principle, from a variety of test objects, namely, a point, a fine line, a variable slit, and a knife edge. Data from any or all of these objects can be represented by a single common curve, which can be interpreted in terms of any of the objects.

Section 3 deals with the Fourier type of approach. An imaging system does not affect the frequency, direction, or sinusoidal character of the component waves. It can only affect their amplitudes and phase relationships. The function that describes these modifications as a function of the frequency and direction of the component waves is also characteristic of the imaging system. The test object required to obtain this information consists of a series of patterns in which the luminance varies sinusoidally, each pattern having a different spatial frequency and all oriented in the same direction. The directional variation can be obtained by rotating the test object with respect to the system being tested.

Section 4 is concerned with the coordination between the two viewpoints. It is shown that it is possible to obtain from either approach the characteristics for a periodic test object consisting of alternate dark and light stripes of equal width, such as is commonly used in resolving-power work. It is also shown that it is possible to transform the characteristics of either approach into those of the other.

Section 5 discusses the application to practice, the interpretation of the results and the relative merits of the two approaches.

The appendix gives the mathematical formulation of the diffraction images used as illustrations.

In the following mathematical statements, constant coefficients are ignored in the integrations unless otherwise indicated. The functions are assumed to be normalized after integration. Also, the object-plane coordinates are reduced to the image plane by application of the magnification.

2. Evaluation of the Point Image

2.1. General Image Formation with the Point Image

The point image is the flux-density distribution in the image plane when the object is a point source (figs. 1, 2). A general object can be considered to consist of a summation of points, and its image the summation of the corresponding point images.

Let $O(x,y)$=general object function,
$\varphi(x,y)$=point image function,
$I(x,y)$=general image function.

Then

$$I(x',y')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} O(x,y)\varphi(x'-x,y'-y)dxdy. \tag{1}$$

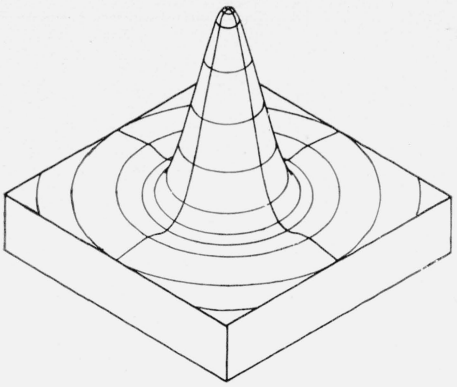

FIGURE 1. *Point-image model.*

In this model the vertical dimension represents flux density (per unit area). It is understood that the rings actually continue indefinitely, whereas only the first bright ring is shown completely.

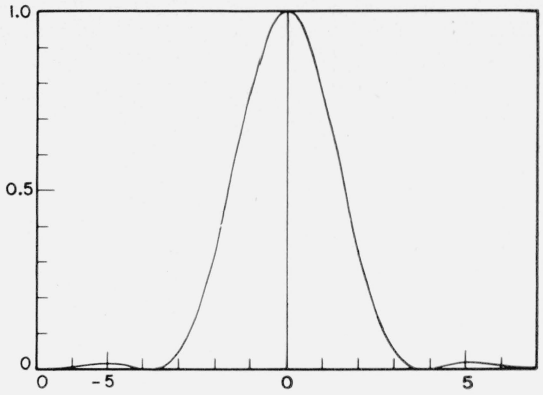

FIGURE 2. *Section of point image.*

The normalized distance in the image plane is measured in z-units, as explained in appendix 1.

The primed variables represent the displacement between the O and φ functions required for the integration. For each displacement, the integral, being a definite integral, establishes a specific value for I. The image function, I, then, is a function of the displacement involved in the integration. The space described by x',y', however, is the same as that described by x,y. O and I can be compared point for point.

The integration can also be written in such a form that the object function is the displaced function, that is,

$$I(x',y')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)O(x'-x,y'-y)dxdy. \tag{2}$$

Either of these forms is perfectly valid, and either may be transformed into the other, provided one recognizes that x,y in eq (1) is not the same as x,y in eq (2). To distinguish them, one might use subscripts on the symbols, but this would make the equations more confusing, and is not necessary if one understands the situation.

The image function is to be sampled with a space-integrating detector in the image plane, that is, a photocell behind an aperture.

Let $A(x,y)$=detector aperture transmission function,
$E(x,y)$=total flux passing through A as a function of the position of A.

Then

$$E(x'',y'')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I(x',y')A(x''-x',y''-y')dx'dy'. \tag{3}$$

Combining this with eq (2) we obtain

$$E(x'',y'')=\int_{-\infty}^{\infty}\int\int\int\varphi(x,y)O(x'-x,y'-y)A(x''-x',y''-y')dxdydx'dy'. \tag{4}$$

From this it can be seen that the functional characteristics of O and A can be interchanged without affecting the measured flux, E. For example, suppose the object were a point source and the aperture a circular hole centered on the point image. The output from the photocell will indicate amount of flux passing through the hole. Then interchange the object and aperture. Now the object is a uniformly luminous disk with a reduced diameter equal to that of the previous aperture, and the new aperture is a minute hole, equal in size to the reduced geometric size of the previous point source. The output from the photocell will be the same as before.

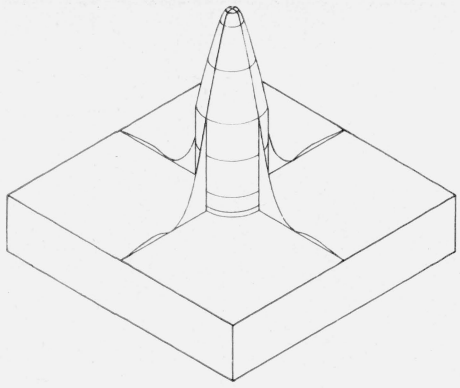

FIGURE 3. *Point-image evaluation—Hopkins' method.*

The solid represents the portion of the total flux that passes through the circular aperture.

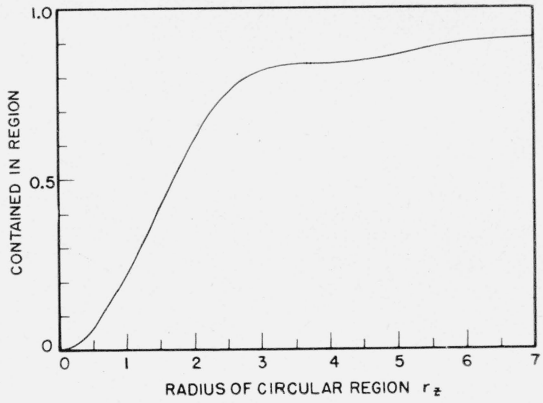

FIGURE 4. *Radial flux distribution of point image.*

This curve shows the normalized volume of the solid of figure 3 as a function of the radius of the limiting cylinder.

Mathematically, and in a more general sense, this situation is as follows: If O is a point source, then

$$E(x'',y'')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(x''-x',y''-y')\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)O(x'-x,y'-y)dxdy\right]dx'dy'$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x',y')A(x''-x',y''-y')dx'dy', \tag{5}$$

and the flux detected depends on the nature of A.

On the other hand, if A is a point aperture, then

$$E(x'',y'')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}O(x'-x,y'-y)A(x''-x',y''-y')dx'dy'\right]dxdy$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)O(x''-x,y''-y)dxdy, \tag{6}$$

which is identical in form with eq (5), except that A is replaced by O. If O in eq (6) had the same functional characteristics as A in eq (5), then they would be mathematically indistinguishable, and the same E will be obtained from either. Also note that eq (6) has the same form as eq (2). As one would expect, the use of a point aperture would produce an undistorted map of the flux-density distribution in the image plane.

2.2. Determination of a Characteristic Function of the Point Image

It should be clear from the above that the function which distinguishes one image-forming system from another is the point-image function φ. The problem is to find some way of describing φ, which can be obtained experimentally and which provides significant information to the user.

The direct mapping of the flux-density distribution in the image of a point object with a point aperture is impractical because of the very low-energy levels involved. A practical method must involve in some manner the integration of the flux over an area.

One method, used by Hopkins [7] [1] is to measure the flux contained within successive concentric circular regions about the center of the point image (fig. 3). The ratio of the flux contained within a circular region to the total flux of the point image is plotted as a function of the radius of the circle (fig. 4).

[1] Figures in brackets indicate the literature references at the end of this paper.

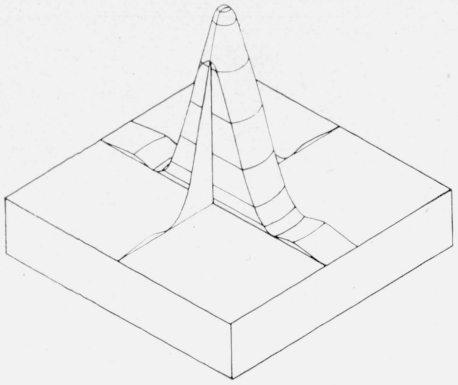

FIGURE 5. *Determination of point-image characteristic.*

The solid represents the portion of the total flux that passes through the receiver slit, the point image being centered on the slit.

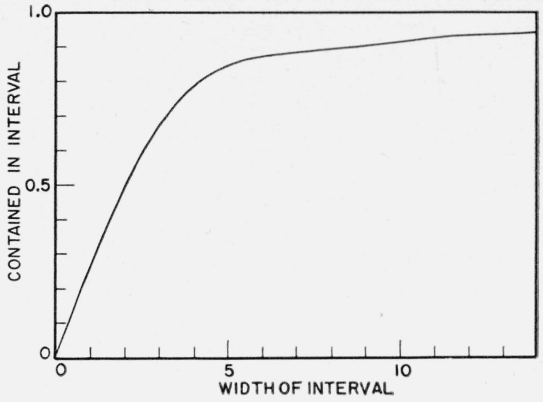

FIGURE 6. *Point-image characteristic.*

This curve shows the normalized volume of the solid in figure 5 as a function of the width of the slab.

This method has several advantages over mapping the flux-density distribution directly. Of course, there is an increase in the energy involved, which increases the signal-to-noise ratio of the measurements. Also, there is a reduction of a three-dimensional function to two dimensions, which is more convenient. Further, this method provides the user with an idea of the contrast with which small detail will be imaged, for, by the interchangeability of the object and aperture function, the resulting curve (fig. 4) can be considered to represent the flux density at the center of the image of a disk as a function of the size of the disk. This disk corresponds to a small object in the scene, later observed by the user.

This method does have disadvantages, however. It is satisfactory for systems in which the point image is radially symmetrical, but is insensitive to the presence of radial asymmetry, such as exists in an astigmatic or comatic image. It also presents the practical difficulty of locating the centers of the apertures or disks in two dimensions with precision.

A different integrative method that allows a considerable increase in the available energy is one in which the integration is limited in one direction only. This is done by measuring the flux contained in successive widths about the center of the point image (fig. 5). The normalized flux contained in a region as a function of the width of the region (fig. 6) is the function that is here proposed as the most useful and practical characteristic function describing the point image. It will hereafter be called the point-image characteristic $K(w)$, where w represents the width of the region. For example, if the integration is limited in the x-direction, then

$$K(w_x)=\int_{-w_x/2}^{w_x/2}\int_{-\infty}^{\infty}\varphi(z,y)\,dy\,dx. \tag{7}$$

It should be noticed that this function does not actually involve a reduction of three dimensions to two, because the curve obtained is a function of the direction in which the integration is taken. This makes the data somewhat less convenient than is true of the previous method, but this is not objectionable because of the additional information obtained. The new method will detect a lack of radial symmetry. However, the inconvenience involved is not too great because most images are bilaterally symmetrical, or nearly so, and two mutually perpendicular orientations are all that are necessary to characterize the image.

The point-image characteristic has other virtues beyond the relatively large amount of available energy and the ability to detect lack of radial symmetry. These arise out of the unidirectional limitation of the integration.

For example, the point-image characteristic is closely connected with the fine-line image. The fine-line image is the image of a line of infinite length but infinitesimal width. The flux density is constant along the length of the line image and varies in a direction perpendicular

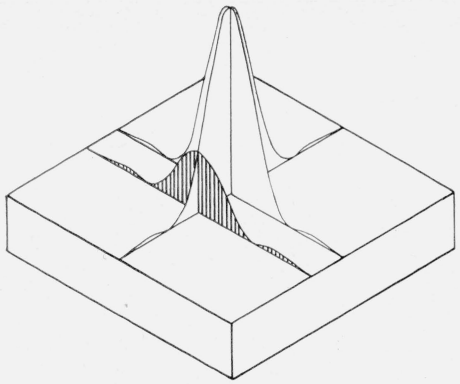

FIGURE 7. *Generation of fine-line image.*

The flux density at a point in a fine-line image is proportional to the area of the section of the point image orientated in the direction of the line.

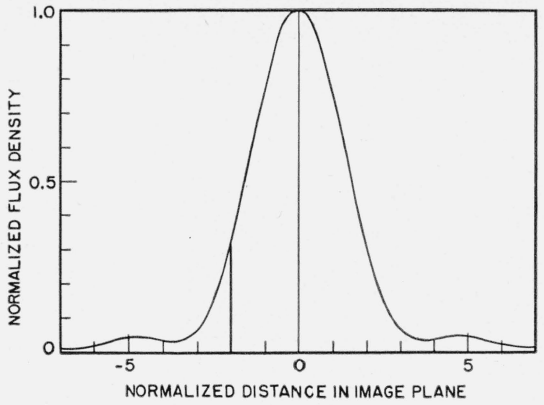

FIGURE 8. *Cross section of fine-line image.*

The vertical line corresponds to the area of the section shown in figure 7.

to the length. It is obtained by integrating the point image in one direction only. For example, the image function for a fine line, the length of which is in the y-direction, is given by

$$L(x)=\int_{-\infty}^{\infty}\varphi(x,y)\,dy. \tag{8}$$

This is illustrated in figures 7 and 8.

The point-image characteristic can be obtained from the fine-line image by

$$K(w_x)=\int_{-w_x/2}^{w_x/2}L(x)\,dx. \tag{9}$$

This is illustrated in figure 9.

Up to this point we have been considering the object to be a point or a fine line and the detector aperture to be a variable slit, the operating mechanism that produces the variation in w. But, because of the interchangeability of the object and aperture, an illuminated variable slit could be used as an object and a fine slit centered on the variable slit image as the detector aperture. This method of obtaining the point-image characteristic is illustrated in figure 10.

The point-image characteristic is also closely connected with the knife-edge image. The latter is related to the point image as follows:

$$S(x')=\int_{-\infty}^{x'}\int_{-\infty}^{\infty}\varphi(x,y)\,dy\,dx. \tag{10}$$

This is shown in figures 11 and 12.

The relationship between the knife-edge image and the fine-line image is given by

$$S(x')=\int_{-\infty}^{x'}L(x)dx. \tag{11}$$

The point-image characteristic is obtainable from $S(x)$ by observing the values of $S(x)$ at $x=-w_x/2$ and $x=+w_x/2$. Then

$$K(w_x)=S(w_x/2)-S(-w_x/2)=\int_{-\infty}^{w_x/2}L(x)dx-\int_{-\infty}^{-w_x/2}L(x)dx=\int_{-w_x/2}^{w_x/2}L(x)dx. \tag{12}$$

These relationships are illustrated in figures 13 and 14.

In summary, it may be pointed out that the point-image characteristic provides a simple yet practical and informative way of evaluating an imaging system. It may be obtained from a variety of test objects, namely, a point, a fine line, a variable-width line, or a knife edge.

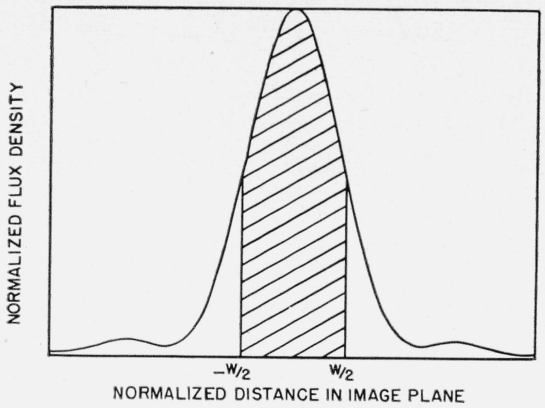

FIGURE 9. *Determination of point-image characteristic from fine-line image profile.*

The cross-hatched area corresponds to the solid in figure 5.

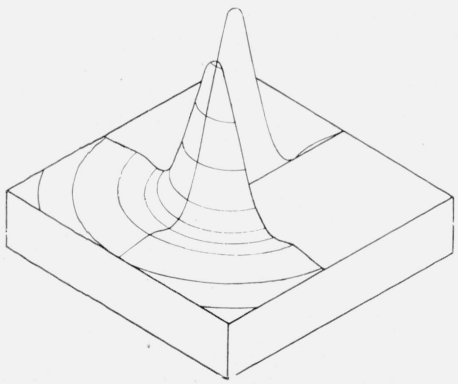

FIGURE 11. *Generation of knife-edge image from point image.*

The flux density at a point in the knife-edge image is proportional to the volume of the solid indicated where the limiting plane has the same orientation as the knife edge and passes through the point in question.

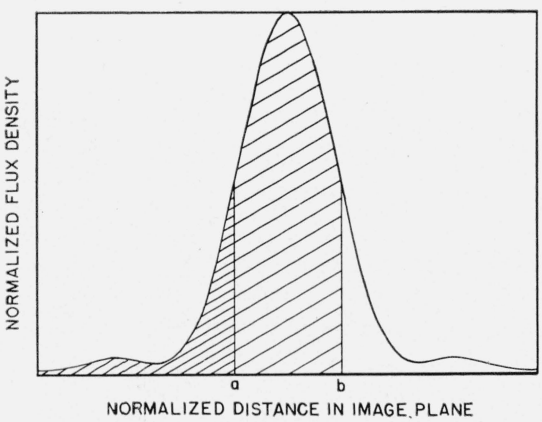

FIGURE 13. *Generation of knife-edge image from fine-line image.*

The flux density at a point in the knife-edge image is proportional to the area under the fine-line image to the left of the corresponding abscissa. The area under the fine-line image between any two abscissa values is proportional to the difference between the corresponding ordinate values of the knife-edge image.

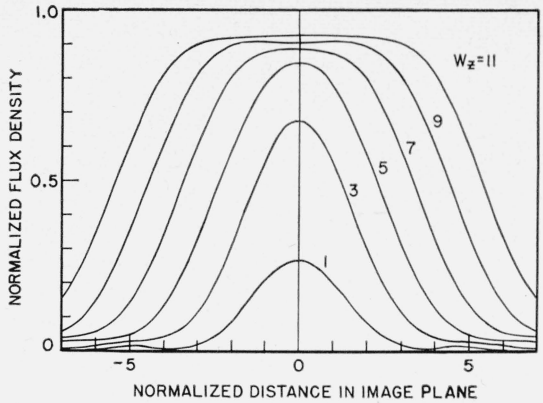

FIGURE 10. *Cross sections of images of finite-width object lines.*

The reduced object line widths are 1, 3, 5, 7, 9, and 11 x-units. Plotting the central flux density as a function of the reduced object line width results in the point-image characteristic curve shown in figure 6.

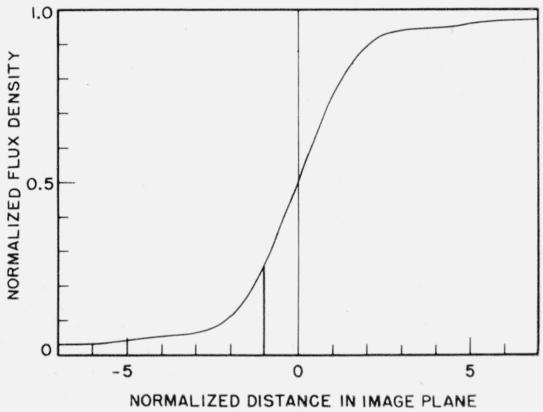

FIGURE 12. *Cross section of knife-edge image.*

The vertical line corresponds to the solid shown in figure 11.

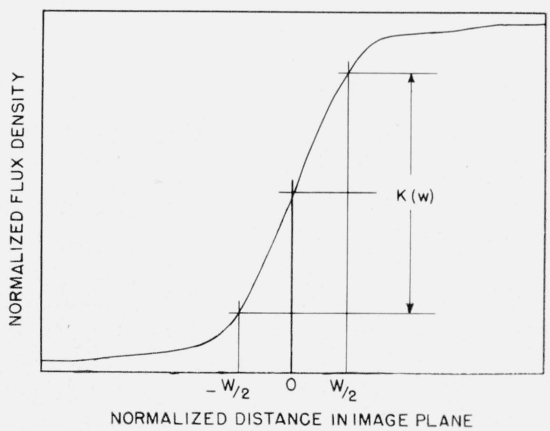

FIGURE 14. *Determination of point-image characteristic from knife-edge image.*

If the abscissa values a and b in figure 13 are made equal to $-w/2$ and $+w/2$ as in this figure, then the ordinate difference shown here is proportional to the corresponding area in figure 13, which itself is proportional to the corresponding value of the point-image characteristic, as indicated in figure 9.

Furthermore, regardless of which way it is obtained, it can be interpreted in terms of any of the objects. This is discussed further in a later section.

3. Evaluation by Fourier Analysis

A general object $O(x,y)$ can be analyzed into a continuum of sinusoidal spatial waves, differing from each other in direction, frequency, amplitude, and phase. The characteristics of these waves are given by the Fourier transform of the object,

$$T_o(\omega_x,\omega_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} O(x,y)\ \exp\ [-i(\omega_x x+\omega_y y)]dxdy, \tag{13}$$

where $T_o(\omega_x,\omega_y)$ is complex, containing both amplitude and phase factors, and ω_x and ω_y are directional frequency components of the component waves.

A component wave itself is represented by

$$W_c=T_o(\omega_x,\omega_y)\ \exp\ [i(\omega_x x+\omega_y y)]. \tag{14}$$

Consider this to be an object. Then by applying eq (2),

$$I_W(x',y')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)T_o(\omega_x,\omega_y)\ \exp\ [i\{\omega_x(x'-x)+\omega_y(y'-y)\}]dxdy$$

$$=T_o(\omega_x,\omega_y)\ \exp\ [i(\omega_x x'+\omega_y y')]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\ \exp\ [-i(\omega_x x+\omega_y y)]dxdy. \tag{15}$$

The integral is the Fourier transform $\Phi(\omega_x,\omega_y)$ of $\varphi(x,y)$, and therefore

$$I_W(x',y')=\Phi(\omega_x,\omega_y)T_o(\omega_x,\omega_y)\ \exp\ [i(\omega_x x'+\omega_y y')]. \tag{16}$$

The function modified by Φ is seen to be simply the component object wave. Each component image wave then will be the product of Φ and the corresponding component object wave. Therefore,

$$T_I(\omega_x,\omega_y)=\Phi(\omega_x,\omega_y)T_o(\omega_x,\omega_y), \tag{17}$$

where T_I is the Fourier transform of the image.

If $I(x,y)$ can be considered to be the object of a second imaging precess, then

$$T_{I_2}(\omega_x,\omega_y)=\Phi_2(\omega_x,\omega_y)T_{I_1}(\omega_x,\omega_y)=\Phi_2\Phi_1 T_o. \tag{18}$$

This can be extended to as many imaging processes as desired.

The transform

$$\Phi(\omega_x,\omega_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\ \exp\ [-i\omega_x x+\omega_y y)]dxdy \tag{19}$$

is seen to be characteristic of the imaging process. Let us examine it in more detail.

The transform is the double integral of the product of two functions, one being the point image $\varphi(x,y)$ and the other a two-dimensional wave $exp[-i(\omega_x x+\omega_y y)]$.

This two-dimensional wave is sinusoidal in one direction and constant in the perpendicular direction, like a corrugated roof. The direction of the sinusoidal variation is inclined to the x-axis by an angle θ, where $\tan\theta=\omega_y/\omega_x$, and the frequency is given by $\omega_\theta=\sqrt{\omega_x^2+\omega_y^2}$ (see fig. 15). The transform can then be considered to be $\Phi(\theta,\omega_\theta)$.

For $\theta=0$, $\omega_y=0$ and the transform becomes

$$\Phi(\omega_x)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\exp[-i(\omega_x x)]dydx=\int_{-\infty}^{\infty}L(x)\exp[-i(\omega_x x)]dx, \tag{20}$$

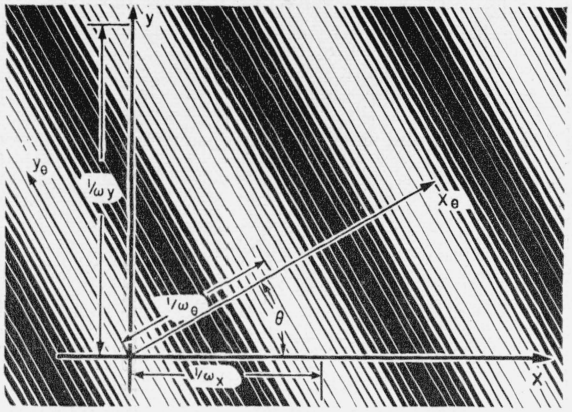

FIGURE 15. *Transformation of coordinates.*

If the direction of the pattern is inclined to the original x-axis by an angle θ, then, in the new coordinate system in line with the pattern, $_\theta=\sqrt{\omega_x{}^2+\omega_y{}^2}$ and $\tan\theta=\omega_y/\omega_x$.

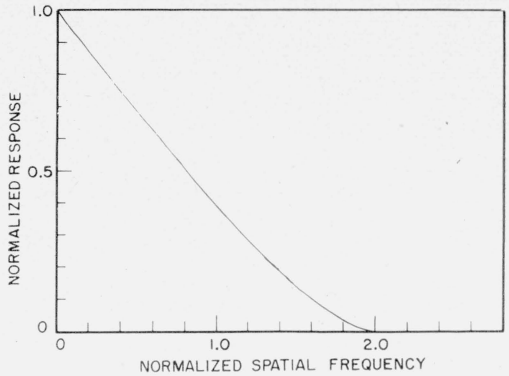

FIGURE 16. *Sine-wave response curve.*

This is the positive side of the cross-section of the Fourier transform of the point image shown in figure 1. It is also the Fourier transform of the fine-line image shown in figure 8.

where $L(x)=\int_{-\infty}^{\infty}\varphi(x,y)dy$ is the fine-line image (fig. 16).

If the x, y coordinates are rotated in x_θ, y_θ through the angle θ (fig. 15), then eq (20) can be generalized into

$$\Phi(\theta,\omega_\theta)=\int_{-\infty}^{\infty}L(x_\theta)\exp[-i(\omega_\theta x_\theta)]dx_\theta. \tag{21}$$

For any θ, the transform Φ of eq (21) is the cross section in the θ-direction of the solid representing the transform $\Phi(\omega_x, \omega_y)$. This is shown in figure 17.

Equation (21) serves as the basis for the experimental determination of Φ, since for each θ it is the transform of the fine-line image oriented with its length perpendicular to the θ-direction. The object can be one in which the flux density varies sinusoidally in one direction, and the detector aperture can be a fine slit.

The object can be represented by

$$O(x_\theta)=A+B\ \exp[i\omega_\theta x_\theta],\quad A\geq B. \tag{22}$$

The constant term is necessary because all flux densities must be positive. The image is given by

$$I(x_\theta)=A+\Phi(\omega_\theta)B\ \exp[i\omega_\theta x_\theta]. \tag{23}$$

Let $M_o=B/A$ be defined as the modulation in the object. The corresponding modulation in the image is $M_I=\Phi B/A$. The transform Φ is obtained by

$$\Phi=\frac{M_I}{M_o}. \tag{24}$$

The use of the modulation automatically compensates for any factor that changes the signal output of the testing device proportionally, such as the transmission factor of the lens, change in brightness of the source, change in gain of the detector amplifier, etc.

It must be remembered that Φ is actually complex, involving both an amplitude and a phase factor. However, if the origin of $\varphi(x,y)$ is properly selected, the phase shift is small. If $\varphi(x,y)$ is symmetrical about its origin, then there is no phase shift involved. In most cases, for evaluation purposes, the phase-shift factor can be neglected.

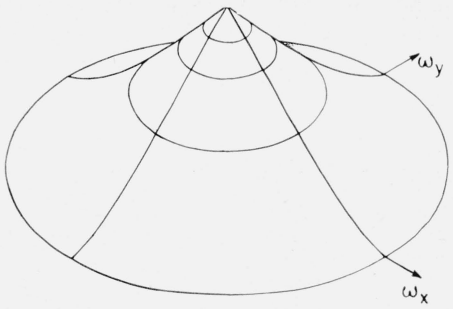

FIGURE 17. *Three-dimensional model of Fourier transform of point image.*

Each radial cross section is the sine-wave response in that direction. Note that the space is frequency space.

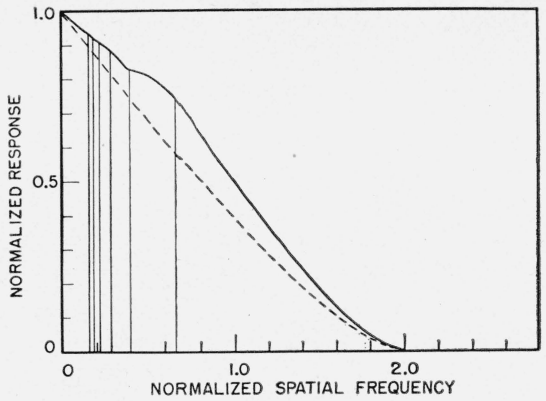

FIGURE 18. *Square-wave response curve.*

The dashed line is the sine-wave response curve from which this was derived. The vertical lines separate regions in which the square-wave image contains different numbers of harmonic components, each region having one more component than its neighbor to the right. In the region farthest to the right, the square-wave image contains only the fundamental.

4. Coordination

If $\Phi(\omega)$ is given, it is possible to predict a similar response curve in which the test object consists of alternate dark and light stripes, a spatial square wave. The square-wave test object can be analyzed into its component harmonics, each of which is attenuated by the value of Φ corresponding to its frequency, and the image is obtained by adding together these attenuated harmonics. For the response curve we are only interested in the peak-to-peak values that are obtained from the values at the centers of the lines and spaces. The square-wave response curve then is given by

$$\psi(\omega)=\frac{4}{\pi}\left[\Phi(\omega)-\frac{1}{3}\Phi(3\omega)+\frac{1}{5}\Phi(5\omega)-\cdots\right]. \tag{25}$$

It should be noticed that there will be only a finite number of terms in the sum because there is a limiting value of ω beyond which Φ remains at zero. This limit exists because of the finite dimensions of the aperture of the system; the larger the aperture the higher the limiting frequency.

Because of the finite range in ω, it is possible to obtain $\Phi(\omega)$ from $\psi(\omega)$ by the inverse process. Here it is necessary to start at the limiting value of ω and work backward. From this limiting value ω_c back to $\omega_c/3$ the sine-wave response is given by

$$\Phi(\omega)=\pi/4\ \psi(\omega) \tag{26}$$

because Φ and ψ are both zero for the odd multiples of ω for ω greater than $\omega_c/3$.

From $\omega_c/3$ to $\omega_c/5$,

$$\Phi(\omega)=\pi/4\ \psi(\omega)+\tfrac{1}{3}\Phi(3\omega), \tag{27}$$

where $\Phi(3\omega)$ may not be zero, and has already been found.

From $\omega_c/5$ to $\omega_c/7$,

$$\Phi(\omega)=\pi/4\ \psi(\omega)+\tfrac{1}{3}\Phi(3\omega)-\tfrac{1}{5}\Phi(5\omega), \tag{28}$$

and so forth.

The relationship between the square-wave response and the sine-wave response is illustrated in figure 18.

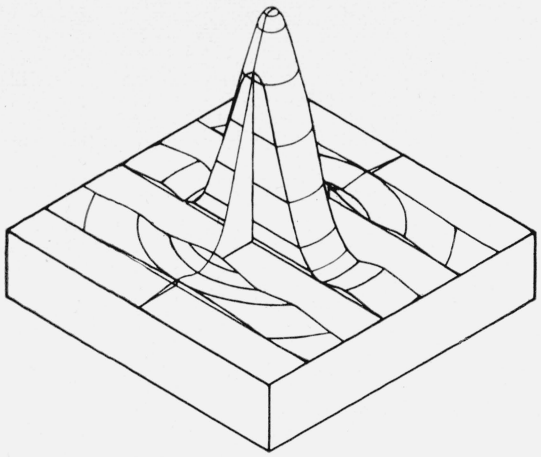

FIGURE 19. *Generation of square-wave response from point image.*

The sum of the volumes of the slabs is proportional to the flux density at the center of a bright line in the image. The sum of the volumes of the slabs that have been removed corresponds to the dark line. The difference between the two is proportional to the square-wave response.

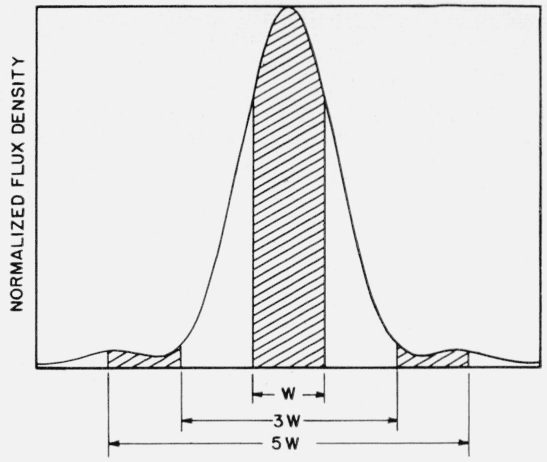

FIGURE 20. *Generation of square-wave response from fine-line image.*

The cross-hatched areas correspond to the slabs in figure 19. The marked widths indicate how the square-wave response can be obtained from the point-image characteristics.

The square-wave response may also be obtained from the point-image characteristic. If we consider the imaging of a square-wave test object as a point-image integration, then the flux density at the center of the image of a bright line in the pattern is proportional to the sum of the volumes of the slabs illustrated in figure 19, or the areas of the stripes in figure 20. The flux density at the center of the image of a dark line is proportional to the sum of the volumes of the slabs (or areas of the stripes) which lie between those illustrated. The sum of these two flux densities is proportional to the total flux in the point or fine-line image, and the square-wave response is given by

$$\Psi(\omega)=E\left(\frac{1}{2\omega}\right)-\left[1-E\left(\frac{1}{2\omega}\right)\right]=2E\left(\frac{1}{2\omega}\right)-1, \tag{29}$$

where $E(1/2\omega)$ is the flux density at the center of the image of a bright line and $1/2\omega=w$ is the width of a line, bright or dark, in the pattern.

It can be seen from figure 20 that

$$E\left(\frac{1}{2\omega}\right)=E(w)=K(w)+[K(5w)-K(3w)]+[K(9w)-K(7w)]+\ldots, \tag{30}$$

where $K(w)$ is the point-image characteristic, so

$$\Psi(\omega)=2\{K(w)+[K(5w)-K(3w)]+\ldots\}-1. \tag{31}$$

It is now apparent that $\Phi(\omega)$, the sine-wave response, may be obtained from $K(w)$, the point-image characteristic. This is done by determining $\psi(\omega)$, the square-wave response, from $K(w)$ and then applying the procedure indicated in eq (26), (27), and (28) to obtain the sine-wave response.[2]

To obtain the point-image characteristic from the sine-wave response is more direct. Consider the test object to be a variable-width line. The transform of the image of this object is obtained by multiplying the transform of the object by the sine-wave response of the system.

[2] This calculation for the case of the unaberrated image was made by the author before he was aware of Steel's equation (see appendix). It was done for 30 different spatial frequencies, using the tabulated values for the knife-edge image given by Struve (see appendix) as well as his approximation for values beyond those tabulated. The agreement between this calculated sine-wave response and Steel's equation is well within the error of calculation.

The transform of this object is of the form

$$T_o = w \frac{\sin \pi \omega w}{\pi \omega w} \tag{32}$$

shown in figure 21, and the transform of the image is given by $T_I = \Phi T_o$. The image itself would be given by the inverse transform

$$I(x_\theta) = \int_{-\infty}^{\infty} \Phi T_o \exp\left[i(\omega_\theta x_\theta)\right] d\omega_\theta, \tag{33}$$

but we are interested only in the value at the center of the line image where $x_\theta = 0$, so

$$K(w_\theta) = I(O) = \int_{-\infty}^{\infty} \Phi T_o d\omega_\theta = 2w \int_0^{\infty} \Phi(\omega_\theta) \frac{\sin \pi \omega w}{\pi \omega w} d\omega_\theta. \tag{34}$$

This equation shows the direct manner in which $K(w)$ may be obtained from $\Phi(\omega)$. For each w selected, the sine-wave response is multiplied by the proper calculated function of the form $\sin x/x$, and the product is integrated graphically or numerically. The result is then multiplied by $2w$, and the value obtained is the desired $K(w)$.

5. Application

It has been shown that an imaging system can be evaluated by means of a description of the point image or, alternatively, a description of its Fourier transform. Each of them has its advantages.

The Fourier approach allows us to combine several systems or to analyze a system into its components, under certain conditions. The principal condition is that each intermediate image formed by each component must be equivalent to a real luminous object having the same flux-density distribution in the image plane. Thus the combining or analyzing process can be applied to a photographic process, a television system, or a system in which each intermediate image is formed on a diffusing surface. It cannot be applied to a telescopic system because the light emerging from the primary image still contains the wave-front deformations produced by the objective and is subject to aberration correction by the ocular.

This does not mean that the Fourier approach cannot be applied to the entire telescopic system. It is just that the whole telescope cannot be precisely evaluated by simply evaluating the components.

Using the Fourier approach also enables one to apply the concepts developed in information and communication theory to imaging processes.

Another virtue of the Fourier approach is that the transform determined for any orientation of the sine-wave pattern in the image plane is a true section of the solid representing the entire transform in the image plane. A similar property does not hold for the point-image characteristic.

It should be noticed also that the sine-wave response of a lens has a finite boundary determined by the aperture of the system, whereas the point-image characteristic is unbounded.

The point-image characteristic, on the other hand, directly provides two basic types of information about the performance of the system, apart from the distribution of flux in the point image. These are the contrast at which an isolated object will be imaged as a function of the object size, and the gradient in the image of the edge of an extended object. The first holds because the point-image characteristic can be obtained by measuring the flux densities at the centers of finite-width line images. The second holds because the characteristic can also be obtained from the image of a knife-edge object.

However, if one knows either the point-image characteristic or the sine-wave response, he can calculate the other, as has been shown. It would probably be preferable to have a research instrument that would determine the sine-wave response, because the transformation to the point-image characteristic is more direct and more suitable for calculating machines than the reverse transformation.

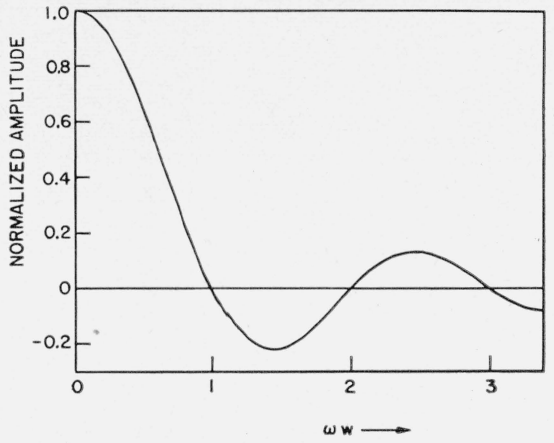

FIGURE 21. *Transform of finite-width line object.*

ω is the frequency of the component wave, and w is the reduced object-line width.

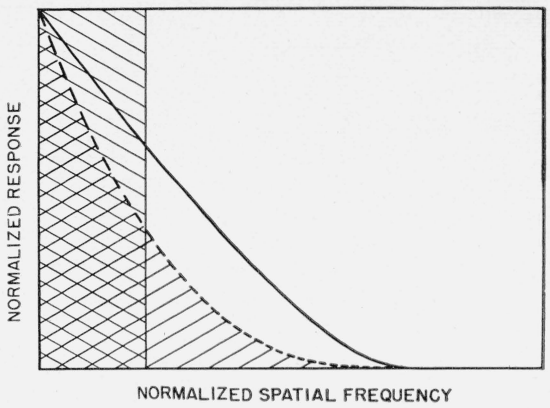

FIGURE 22. *Schade's equivalent bandwidth.*

The solid curve is the sine-wave response curve; the dashed curve is the square of the solid curve; and the rectangular area is equal to the area under the dashed curve. This rectangular area is measured by a single number, its limiting frequency, thereby providing a single number to describe the sine-wave-response curve.

Another problem that comes up is one that is involved in routine testing. It would be desirable to reduce the evaluating curve to a single number with as little loss of significant information as possible, and it would be desirable that this reduction be done automatically in the test procedure.

With respect to the sine-wave response curve, Schade has suggested a reduction in which the curve is squared, ordinate by ordinate, and then integrated. The resulting number is equal to the limiting frequency of a rectangular "response" curve having the same area as the squared sine-wave response curve. This establishes an equivalent bandwidth, shown in figure 22.

The mechanism that would permit this determination directly would be similar to the sine-wave response mechanism, except that a "noise" pattern instead of a sinusoidal pattern would be used. The ideal noise pattern contains all frequencies at equal amplitude but with random phase relations. The fluctuations in the photoelectric output produced by this pattern are fed through a squaring circuit and then integrated. The resulting steady current, indicated on a meter, is proportional to Schade's equivalent bandwidth. One trouble with this system is the difficulty involved in producing an acceptable noise pattern.

Equation (34) indicates another approach to the problem of representing the sine-wave response curve with a single number. The object transform T_0 in this equation can be considered to be a weighting factor for the sine-wave response curve, and the integral to represent the equivalent bandwidth Ω. Then this equivalent bandwidth can be determined by the use of a variable slit or a knife edge, for

$$\Omega=\frac{K(w)}{2w}. \tag{35}$$

For this to be single-valued, w must be fixed, and two convenient possible values appear evident. One is to set $w=1/\omega_c$, where ω_c is the limiting frequency for a theoretically perfect lens having the same aperture as the lens under test. Then the weighting function goes to its first zero at ω_c. This method is illustrated in figure 23.

The other convenient value for w is zero, for the limit of $K(w)/w$ as w goes to zero is the slope at the center of the knife-edge image, as can be seen with the aid of figure 14. As indicated in eq (35), this equivalent bandwidth is given by one-half the slope at the center of the knife-edge image. The weighting function approaches unity for small values of ω as w approaches zero, so this equivalent bandwidth is given by the area under the sine-wave response curve itself. This is illustrated in figure 24. It might be emphasized that this latter relationship is an important and fundamental condition. The slope at the center of a knife-edge image is exactly proportional to the area under the sine-wave response curve.

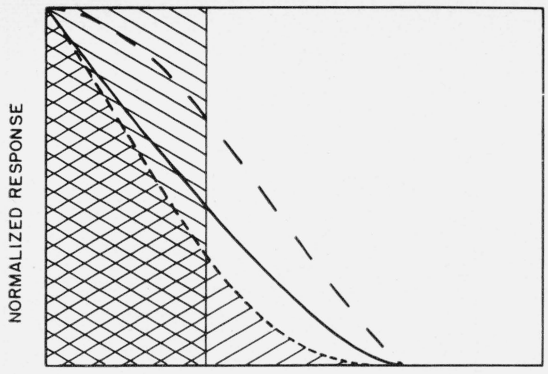

FIGURE 23. *Another equivalent bandwidth.*

The solid curve is the sine-wave response curve; the long-dashed curve is the transform of the finite-width-line object; the short-dashed curve is the product of the other two; and the rectangular area is equal to the area under the short-dashed curve. This provides a different single characteristic number, which can be obtained more easily than Schade's equivalent bandwidth.

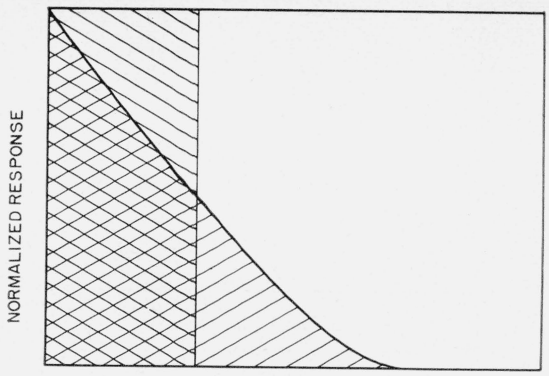

FIGURE 24. *A third equivalent bandwidth.*

This bandwidth is measured by the area under the sine-wave-response curve itself. The area is exactly proportional to the slope at the center of a knife-edge image.

The idea of measuring an equivalent bandwidth may be a good way of reducing the sine-wave response curve to a single number, but this process eliminates one of the advantages of the sine-wave approach, and that is the ability to combine directly a sequence of imaging processes. For most telescopic systems this may be unimportant, but if such a combination is desirable, then perhaps the sine-wave response curve can be characterized well enough by some particular frequency. If the response to some such frequency were established as a measure of quality, then the measure of the quality of the combination is simply the product of the measures of quality of the components.

To summarize, the Fourier approach seems to be more desirable for research and detailed testing, but the determination of the point-image characteristic lends itself to rapid routine testing.

6. Appendix

6.1. Images Produced by an Aberration-Free System With a Circular Aperture in Monochromatic Light

The extended objects are assumed to be illuminated noncoherently. In the following expressions, the unit of displacement in the image is:

$$z=\frac{2\pi ax}{\lambda d}, \tag{36}$$

where a is the radius of the circular aperture, x is the distance in the image plane from the center of the image, d is the distance from the image plane to the aperture, and λ is the wavelength of light.

a. Point Image

This is well known, and its section is given by

$$\varphi(z)=4\left(\frac{J_1(z)}{z}\right)^2. \tag{37}$$

It is illustrated in figures 1 and 2.

b. Fine-Line Image

An expression for the cross section of the fine-line image, implicit in the original work of Struve [1], is explicitly given by Steel [2] as

$$L(z)=\frac{3\pi}{8}\frac{H_1(2z)}{(2z)^2}, \tag{38}$$

where H_1 is what is known as the Struve function, and has been tabulated.[3] The cross section of the fine-line image is illustrated in figure 8.

c. Knife-Edge Image

Struve [1] developed this in a series expansion and tabulated it for z up to 15. If the edge is oriented so that the gradient for the image is positive, then the image is given by

$$S(z)=\frac{1}{2}+\frac{2}{\pi^2}\left\{\frac{3}{1}\frac{2^2z}{1^2\cdot 3^2}-\frac{5}{3}\frac{2^4z^3}{1^2\cdot 3^2\cdot 5^2}+\frac{7}{5}\frac{2^6z^5}{1^2\cdot 3^2\cdot 5^2\cdot 7^2}-\cdots\right\}. \tag{39}$$

Struve also gave a simple approximation, which is in error by less than 0.1 percent for $Z>7$ and 1.0 percent for $Z>3$.

$$S(z)\cong 1-\frac{2}{\pi^2 z}. \tag{40}$$

The knife-edge image is illustrated in figure 12.

d. Finite-Width Line Image

If the width of the Gaussian image is w_z, then

$$\overline{L(z)}=S\left(z+\frac{w_z}{2}\right)-S\left(z-\frac{w_z}{2}\right). \tag{41}$$

This family of images is illustrated in figure 10.

e. Point-Image Characteristic

This is given by

$$K(w_z)=S\left(\frac{w_z}{2}\right)-S\left(-\frac{w_z}{2}\right)=2S\left(\frac{w_z}{2}\right)-1. \tag{42}$$

Combining this with eq (39) gives

$$K(w_z)=\frac{4}{\pi^2}\left\{\frac{3}{1}\frac{2w_z}{1^2\cdot 3^2}-\frac{5}{3}\frac{(2w_z)^3}{1^2\cdot 3^2\cdot 5^2}+\frac{7}{5}\frac{(2w_z)^5}{1^2\cdot 3^2\cdot 5^2\cdot 7^2}-\cdots\right\}. \tag{43}$$

This is illustrated in figure 6.

f. Sine-Wave Response

Steel [2] gives the following expression for the sine-wave response:

$$\Phi(\omega)\begin{cases}=\frac{2}{\pi}\left(\text{arc cos}\frac{\omega}{2}-\frac{\omega}{2}\sqrt{1-\frac{\omega^2}{4}}\right), & 0<\omega<2\\ =0 & ;\quad \omega>2,\end{cases} \tag{44}$$

where $\omega=\pi/w_z$, and w_z is the wavelength of the sinusoid. Notice that there is an absolute limit to the frequency of the pattern that a lens can form, this limit being where the wavelength of the pattern in z units is equal to π.

The frequency, ν, in the image plane is given by

$$\nu=\frac{\omega}{2N\lambda}, \tag{45}$$

where $N=d/2a$ and λ is the wavelength of light. If λ is in millimeters, ν is in cycles per millimeter.

Figures 16 and 17 show the sine-wave response.

[3] E. Jahnke and F. Emde, Tables of functions with formulas and curves, 4th ed. (Dover Publications, New York, N. Y., 1945).

7. References

The data for the curves in the illustrations were calculated from the information contained in the following references:

[1] H. Struve, Beitrag zur theorie der diffraction an fernrohren, Ann. Physik Chemie **17,** 1008 (1882).

[2] W. H. Steel, Étude des effets combinés des aberrations et d'une obturation centrale de la pupille sur le contraste des images optiques. Premiére Partie, Rev. opt. **32,** 4 (1953).

Steel provides information on the point, disk, fine-line, knife-edge, sine wave, and square wave images. Material on the sine wave or Fourier transform approach is contained in the following references.

[3] O. H. Schade, A new system of measuring the specifying image definition, NBS Circular 526 (April 29, 1954).

[4] O. H. Schade, Image gradation, graininess, and sharpness in television and motion picture systems. Part II, J. Soc. Motion Picture Television Engrs. **58,** 181 (1952).

[5] P. Elias, D. S. Grey, and D. Z. Robinson, Fourier treatment of optical processes, J. Opt. Soc. Am. **42,** 127 (1952).

[6] E. W. H. Selwyn, The photographic and visual resolving power of lenses. I. Visual resolving power, Phot. J. **88B,** 6 (1948).

Schade is the first to build a practical machine for obtaining a sine-wave analysis of an optical system. Selwyn applies the Fourier treatment to the prediction of resolving power.

Work on the photoelectric examination of point images through circular apertures is reported in the following paper:

[7] R. E. Hopkins, Measurements of the energy distribution of optical images, NBS Circular 526 (April 29, 1954).

Interest in the knife edge is shown in the following references:

[8] W. Weinstein, Criteria of image-forming quality in photographic objectives, Phot. J. **91B,** 138 (1951).

[9] R. V. Shack, A proposed approach to image evaluation, NBS Circular 526 (April 29, 1954).

Shack also discusses the variable slit. Hopkins is now using knife edges in his work.

WASHINGTON, June 1, 1955.