D. G. Falconer

Noise and Distortion in Photographic Data Storage

Abstract: Noise and distortion limit the usefulness of the photographic transparency as a data-storage medium. The noise, which tends to be multiplicative, derives from the random distribution of silver grains in the photographic image, as well as from thickness variations in the developed emulsion. Distortion, on the other hand, results from 1) the nonlinear relation between transmittance and exposure, 2) the finite width of the emulsion's point-spread function and 3) the existence of an adjacency-enhancement function. Although grain noise remains intrinsic and untreatable, nonlinear distortion—both global and local—may be treated by lowering the contrast of the exposure pattern or, preferably, by recording the data in the form of a phase-modulated carrier wave, as in holography. A solution to the remaining difficulty, namely, linear-global distortion, is obtained through the use of high-resolution, Lippmann-type emulsions.

Introduction

Noise and distortion limit the usefulness of photographic film as a data-storage medium.[1–9] The noise arises from the statistical nature of the photographic process—that is, from natural fluctuations in the number of silver-halide grains exposed and developed per unit area. These fluctuations produce similar variations in the transparency's optical density and emulsion thickness, and hence also in its amplitude and intensity transmittances. Because of the square-root-of-the-sample-size rule, such variations increase approximately as the square root of the mean density. Since amplitude transmittance depends exponentially on both optical density and emulsion thickness, additive variations in these quantities lead to multiplicative fluctuations in the mean transmittance level. Like the grain noise, the usable signal intensity also increases with the mean transmittance. Consequently, optimum signal-to-noise ratios are achieved with low-level, high-contrast exposure patterns.

Distortion, on the other hand, results because the incoming photon stream exposes the emulsion's silver-halide crystals on a roughly one-for-ϵ basis, where ϵ denotes the quantum efficiency of the emulsion. According to Nutting's formula, the optical density increases in proportion to the silver-grain density, and hence also in proportion to the exposure. Consequently, the transmittance, being an exponential function of the optical density, also tends to depend exponentially on the exposure. Such nonlinearity appears distinctly on all T_i vs E curves and generates quadratic and higher-order deviations from strict linearity. Though nonstatistical, quadratic and higher-order distortion degrade the recorded data much as grain noise, unless such deviations remain small. The usual technique for suppressing the effects of nonlinearity is to combine the exposing signal with a large bias, so that the quadratic term diminishes in proportion to the signal-to-bias ratio. Consequently, linear data recording is achieved by using a low-contrast emulsion in conjunction with a low-contrast exposure.

According to the above discussion, the need for high signal-to-noise ratios (SNR) and high signal-to-distortion ratios (SDR) leads to conflicting requirements for the exposure pattern—the former requiring high-contrast exposures and the latter, low-contrast ones. Although the grain noise remains intrinsic and untreatable, one may suppress the various distortion terms by trading off bandwidth for linearity. One such technique, originally noted in holography, phase-modulates the signal pattern onto a carrier wave whose frequency is at least three times the signal bandwidth. In this case the linear, quadratic and higher-order effects can be physically separated by ordinary diffraction techniques. Moreover, once modulated on the carrier, linearity restrictions disappear, so that one may employ both high-contrast film and

The author is with the Stanford Research Institute, Menlo Park, California 94025.

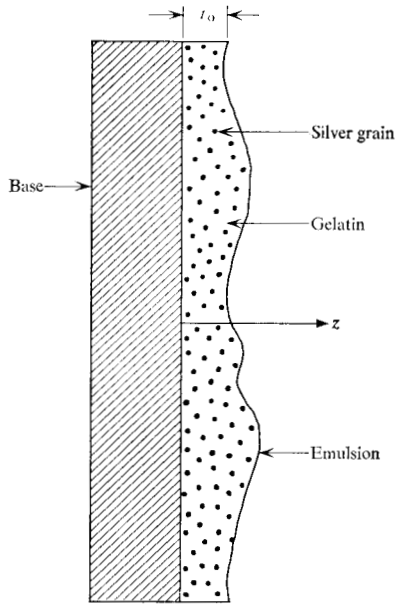

Figure 1 Emulsion geometry. (Dimensions have been exaggerated for clarity.)

high-contrast exposures to increase both the signal-to-noise ratio and absolute signal strength.

The photographic transparency also suffers the problems of finite resolution capability and adjacency-enhancement effects. Loss of resolution, a linear phenomenon, results from the finite diameter and/or the finite number of silver grains in the developed image. However, one can purchase emulsions with grain sizes and grain densities sufficient to resolve the detail of any optical image (provided that exposure levels remain above threshold), thus making the wavelength of light the limiting factor in resolution. The adjacency effect, on the other hand, is a nonlinear phenomenon that arises from the emulsion's tendency to draw unused developer from low-exposure areas of the emulsion to high-exposure ones. Because of this movement, low-exposure regions underdevelop, and high-exposure ones overdevelop, exaggerating thereby any sharp changes or discontinuities in the exposure pattern. Like other nonlinearities, adjacency effects are reduced by using either low-contrast or phase-modulated exposure patterns.

Grain noise

The grain noise discussed above results from the unavoidable statistical fluctuations in the number of silver grains remaining per unit area in the developed emulsion. The exposure process, being a quantum-mechanical interaction between photons and silver-halide crystals, follows, approximately, Poisson statistics. Thus the rms fluctuation ΔN expected in the mean grain density N increases as the square root of N:

$$N \pm \Delta N = N \pm \sqrt{N}.$$

This relation indicates that higher grain densities N lead to larger absolute fluctuations ΔN, but smaller fractional variations $\Delta N/N$. Thus, fine-grain emulsions typically generate less statistical noise than coarse-grain ones.

The connection between grain density N and optical density D follows from the geometry of Fig. 1. According to this figure the light dI lost in the emulsion layer situated between z and $z + dz$ depends on the light intensity at the layer $I(z)$, the number of grains per unit volume (N/t), the mean absorption cross section per grain a and the thickness of the emulsion layer dz:

$$dI = -a(N/t)I(z)dz,$$

where t denotes the nominal thickness of the emulsion. Integration of this expression then yields the light intensity as a function of z:

$$I(z) = I(0) \exp(-aNz/t),$$

where $I(0)$ represents the incident light intensity. The intensity transmittance T_i, defined as the ratio of the emerging intensity to the incident one, thus depends exponentially on the mean grain density N:

$$T_i = I(t)/I(0) = \exp(-aN).$$

On the other hand, the optical density $D = -\log T_i$, increases linearly with N:

$$T_i = 10^{-D} = \exp(-2.3D) = \exp(-aN).$$

Thus, $2.3D = aN$, a relation known as Nutting's formula.[10]

The rms fluctuation in the optical density follows from the proportionality between D and N:

$$D \pm \Delta D = (a/2.3)(N \pm \sqrt{N}).$$

This equation implies that the observed variations in D depend on whether N is measured in grains per square meter or grains per square micrometer. To eliminate this ambiguity, let d represent the dimension of the finest detail to be recorded by the emulsion. Then the mean number of grains n expected in an element with dimension $d \times d$ is just Nd^2, and

$$D \pm \sigma = (a/2.3d^2)(n \pm \sqrt{n}),$$

where σ denotes the *granularity* of the emulsion. Substitution for n then yields

$$\sigma = \sqrt{aD/2.3}/d.$$

This equation indicates that the observed granularity increases as the square root of the mean density (the

$\sqrt{D}$ law[11]) and decreases inversely as the image detail d (Selwyn's law[12]).

In practice, film manufacturers measure the rms granularity of their emulsions with a scanning densitometer, and then use the $\sqrt{D}$ law and Selwyn's law to convert them for use at other densities and other resolutions. In particular, if σ_0 denotes the rms granularity measured at a film density D_0 and with a scanning aperture d_0, then the granularity σ expected at an optical density D and with image detail d is given by

$$\sigma = \sigma_0(d_0/d)\sqrt{D/D_0}.$$

For example, Kodak 649F plate and film—comparatively "quiet" storage media—have $\sigma_0 = 0.5\%$, $D_0 = 0.8$ and $d_0 = 42\ \mu$m.

Fluctuations in optical density, of course, produce corresponding variations in intensity transmittance. If one neglects the $\sqrt{D}$ law—a good approximation if the exposure pattern has moderate-to-low contrast—then the noise-like fluctuations in density look additive:

$$D(x, y) = \text{bias} + \text{signal} + \text{noise}$$

$$= D_b + D_s(x, y) + D_n(x, y).$$

Written this way, the intensity transmittance takes the form

$$T_i = 10^{-D_b - D_s - D_n} = \text{bias} \times \text{signal} \times \text{noise}.$$

In other words, additive fluctuations in grain or optical density produce multiplicative variations in the intensity transmittance.

When D_s and D_n are small relative to unity, both the signal and noise portions of the transmittance become additive. That is,

$$T_i \approx 10^{-D_b}(1 - 2.3\ D_s)(1 - 2.3\ D_n)$$

$$\approx T_b - 2.3T_b\ D_s - 2.3T_b\ D_n$$

$$\equiv T_b + T_s + T_n.$$

In this case both the signal and the noise increase in proportion to the bias transmittance; in addition, T_n increases as the square root of the bias density (the $\sqrt{D}$ law). Thus the rms fluctuation in T_n has the form

$$\text{rms } T_n = 2.3T_b\sigma = 2.3T_b\sigma_0(d_0/d)\sqrt{D_b/D_0}.$$

Consequently, lower exposure levels (that is, small values of D_b) produce the better signal-to-noise ratios.

The amplitude transmittance, defined as the ratio of the amplitude emerging from the emulsion to the amplitude incident upon it, depends in addition on any thickness variations in the transparency. Thickness variations, both systematic and random, arise because the silver grains making up the developed image cause the emulsion to bulge out in proportion to the volume of residual silver. Thus, if t_0 denotes the thickness when no silver is present, and t, the thickness when N grains per unit area develop, then

$$t = t_0 + Nv,$$

where v equals the mean volume of the various silver grains. Substitution from Nutting's formula then yields

$$t = t_0 + 2.3D(v/a).$$

Thus, the height of the relief image varies as the optical density.[7,8]

According to the above, the relative phase shift suffered by the radiation on its trip through the emulsion is just

$$\phi = k(t - t_0)(n - 1) = 2.3kD(v/a)(n - 1) \equiv 2.3\kappa D,$$

where n denotes the refractive index of the gelatin and κ, an empirically determined constant. Consequently, the amplitude transmittance takes the basic form:

$$T_a = T_i^{\frac{1}{2}} \exp(i\phi) = T_i^{\frac{1}{2}} 10^{i\kappa D} = T_i^{\frac{1}{2} + i\kappa}.$$

In other words, intensity and amplitude transmittance are linked by a complex power law. (In practice, the fixing, bleaching and tanning operations also affect the magnitude of the observed phase shift ϕ.)

Transmittance versus exposure characteristics

The emulsion's transmittance versus exposure curve specifies the intensity transmittance T_i expected with an exposure level E. Qualitatively, the T_i vs E characteristic resembles an exponential-decay function since increasing exposure levels lead to ever-lower transmittances. Exponential parameterization also produces the correct asymptotic behavior, since this function tends toward constant transmittance as E tends to zero and toward zero transmittance as E goes to infinity. Exponential dependence is also expected on theoretical grounds. That is, the incoming photons tend to expose the emulsion's silver-halide crystals on a one-for-ϵ basis, where ϵ denotes the quantum efficiency of the emulsion. Consequently, the number of grains exposed per unit area increases more or less linearly with E:

$$N = N_0 + \epsilon E/h\nu,$$

where $h\nu$ represents the energy carried per photon, and N_0 is a number related to the emulsion's fog level. Nutting's formula then yields the exponential behavior

$$T_i = \exp(-2.3D) = \exp(-aN)$$

$$= \exp(-aN_0 - a\epsilon E/h\nu)$$

$$\equiv T_0 \exp(-E/E').$$

Here, T_0 denotes the maximum intensity transmittance of the emulsion, and E' the exposure level that drops T_i to $1/e$ of its maximum value.

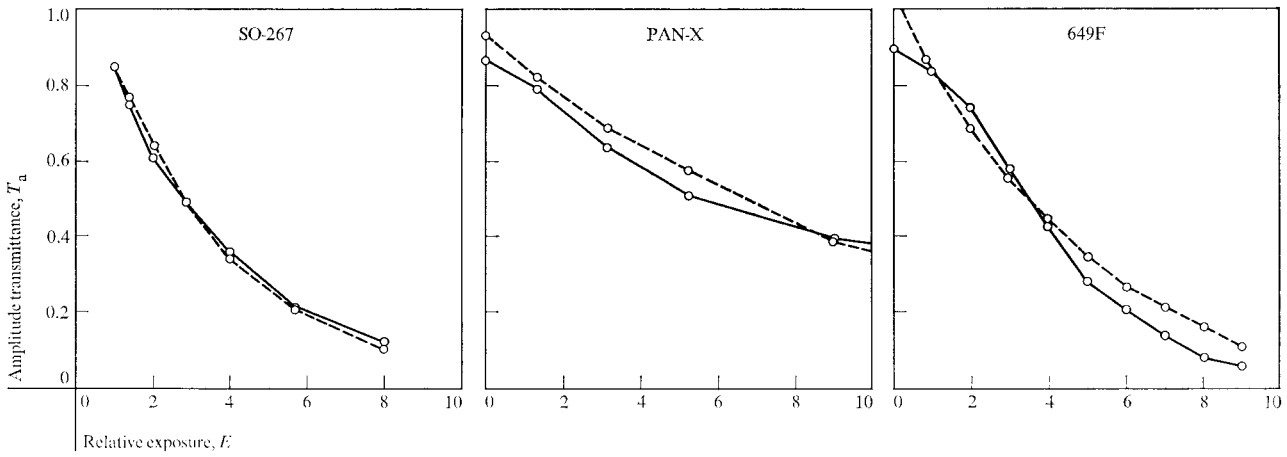

Figure 2 Exponential fits to T_a vs E curves. The dashed line represents a best chi-square fit of the exponential form $T_a = T_0 \exp(-E/E')$ to the experimental points of Horner.[16] (a) SO-267, dev. 3 min. in DK-76; (b) Pan-X, dev. 3 min. in DK-76; (c) 649F, dev. 5 min. in DK-19

Exponentiality also results from linearity on the D vs log E curve. The D vs log E curve, it will be recalled, is popular with the portrait photographer because the human eye responds more or less logarithmically to incident light levels. Thus, while data-storage applications require linearity on the T_i vs E curve, the portrait photographer prefers linearity between log T_i and log E or, more conventionally, between D and log E:

$$D = \gamma \log (E/E_0),$$

where γ and E_0 denote the usual empirical constants. The exposure E, of course, consists of a bias term E_b and a signal term E_s, so that linearity between D and log E means that

$$\begin{aligned} T_i &= (E/E_0)^{-\gamma} \\ &= (E_b/E_0)^{-\gamma}(1 + E_s/E_b)^{-\gamma} \\ &\approx T_b(1 - \gamma E_s/E_b) \\ &\approx T_b \exp(-\gamma E_s/E_b). \end{aligned}$$

Thus, to first order, linearity on the portrait photographer's D vs log E curve implies exponentiality on the data processor's T_i vs E curve.[13] (See also Fig. 2.)

Using exponential parameterization one can look both quantitatively and qualitatively at the distortion generated by the nonlinearity in the T_i vs E curve. In particular, since $E_s/E_b = I_s/I_b$,

$$\begin{aligned} T_i &= T_b \exp(-\gamma I_s/I_b) \\ &\approx T_b[1 - \gamma(I_s/I_b) + \tfrac{1}{2}\gamma^2(I_s/I_b)^2] \\ &\equiv T_b + T_s + T_d . \end{aligned}$$

The first term, T_b, represents the bias transmittance; the second, T_s, the signal transmittance; and the last, T_d, a quadratic distortion term. For accurate data recording the distortion should appear small relative to the signal:

$$\text{SDR} = T_s/T_d = 2/\gamma(I_s/I_b).$$

This formula indicates that both low-contrast emulsions (small γ) and low-contrast exposures (small rms I_s/I_b) are required for good signal-to-distortion ratios (SDR). On the other hand, for good signal-to-noise ratios,

$$\text{SNR} = T_s/T_n = \gamma(I_s/I_b)/2.3\sigma_0(d_0/d)\sqrt{D_b/D_0},$$

both high-contrast emulsions and high-contrast, low-level (small D_b) exposures are required. Consequently, the need for both good SDR and SNR leads to conflicting exposure and emulsion conditions.

To alleviate the conflict, one can phase-modulate the signal (or an equivalent thereof, such as its Fourier or Fresnel transform) onto a carrier wave of spatial frequency ν. The exposure pattern then has the basic form

$$I(x, y) = I_b + I_0 \cos[2\pi\nu x + \phi(x, y)],$$

where I_0 denotes a constant, and $\phi(x, y)$ the phase modulation. The intensity transmittance generated by this pattern contains harmonics of all orders:

$$\begin{aligned} T_i &= T_b \exp[-\gamma(I_0/I_b)\cos(2\pi\nu x + \phi)] \\ &= T_b \sum_n (-\gamma)^n (I_0/I_b)^n \cos^n(2\pi\nu x + \phi)/n! \end{aligned}$$

As is evident from Fig. 3, the power spectrum of the nth order will overlap the first unless

$$\nu + \Omega < n(\nu - \Omega),$$

where Ω denotes the highest spatial frequency in the *phase-factor* $\exp(i\phi)$. Inspection of the inequality indicates that the spectrum of no order will overlap that of the first if $\nu > 3\Omega$. Consequently, by reducing the signal bandwidth in the x-dimension by one-third and using coherent optical techniques to separate the various orders, one can pass the signal undistorted through the nonlinearity, and thus eliminate the "exposure/emulsion" conflict noted above.

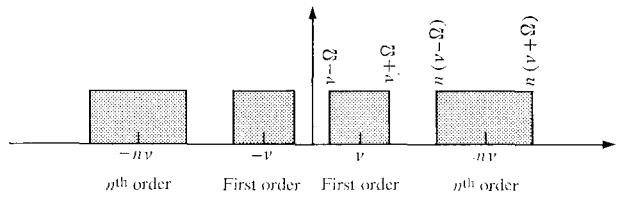

Figure 3 Distorted power spectrum of phase-modulated signal. Note that the low-frequency edge of the nth diffraction order will not overlap the high-frequency edge of the first if $(\nu + \Omega) < n(\nu - \Omega)$.

Resolution and adjacency

The nonlinearity associated with the T_i vs E curve classifies as a local nonlinearity. In photography, local nonlinearity implies that the transmittance observed at the point (x, y) depends only on the exposure at (x, y). In practice, however, global dependencies also arise, so that the transmittance produced at (x, y) depends not only on the exposure at (x, y), but also on the exposure at points nearby (x, y). By restricting consideration to linear- and quadratic-type distortions, one can represent the effects of local and global distortion as products and convolutions, respectively:

$T_i = AE$	local linearity
$T_i = BE^2$	local nonlinearity
$T_i = C * E$	global linearity
$T_i = D * E^2$	global nonlinearity.

Here A and B denote constants; C and D, functions of x and y; and *, two-dimensional convolution.

The finite resolution capability of the photographic transparency classifies, for example, as a global linearity. It is specified quantitatively by defining a point-spread function $S(x, y)$ for the emulsion. The point-spread function, in turn, specifies the image expected when the exposure pattern is a Dirac delta function, and the nonlinear character of the T_i vs E curve has been accounted for. In other words, the observed transmittance has the form

$$T_i = T_b \exp[-\gamma(I_s/I_b) * S].$$

According to the literature[14] the point-spread function looks roughly Gaussian:

$$S(x, y) = (1/2\pi\sigma^2) \exp(-r^2/2\sigma^2),$$

where σ denotes the rms width of the Gaussian distribution, and $r^2 = x^2 + y^2$. Consequently, its Fourier transform $\hat{S}(\xi, \eta)$ also appears Gaussian:

$$\hat{S}(\xi, \eta) = \exp(-\sigma^2\rho^2/2),$$

where ξ and η are measured in radians per unit length, and $\rho^2 = \xi^2 + \eta^2$. For portrait-type films, σ is typically on the order of 10 micrometers; for high-resolution, Lippmann-type emulsions, it may be one micrometer or less.

Adjacency effects, on the other hand, classify as global nonlinearities of the quadratic variety.[15] Adjacency, it will be recalled, tends to enhance sharp changes and discontinuities in the exposure pattern, and is specified quantitatively by defining an *adjacency-enhancement function* $A(x, y)$ for the emulsion. Assuming, as before, an exponential T_i vs E nonlinearity leads to an observed transmittance of the form:

$$T_i = T_b \exp[-\gamma(I_s/I_b) * S - \gamma(I_s/I_b)^2 * A].$$

Horner[16] has suggested that the adjacency-enhancement function has the form of the second derivative of a Gaussian distribution:

$$A(x, y) = -(A_0/\pi)(d^2/dr^2) \exp(-r^2/2\tau^2),$$

where A_0 reflects the strength of the enhancement and τ measures its effective range. Consequently, its Fourier transform takes the form of a Maxwell-Boltzmann distribution:

$$A(\xi, \eta) = A_0(\tau^2\rho^2) \exp(-\tau^2\rho^2/2).$$

Typically, τ is on the order of 10 μm.

The modulation transfer function, as specified and measured by the film manufacturer,[17] represents a mixture of resolution and adjacency effects. To measure the MTF associated with a given emulsion, one exposes the film to a cosinusoidal intensity distribution of the form

$$I(x, y) = I_b + I_s(x, y) = I_0 + I_0 \cos \omega x$$

and then measures the resulting intensity transmittance. The MTF is defined as the ratio of the modulation M' of the developed image to the modulation M of the exposing intensity pattern—after correction for any nonlinearity in the T_i vs E curve. Calling $R = I_s/I_b$ thus leads to the following expression for the MTF:

$$\text{MTF} = M'/M = \text{rms}(R * S + R^2 * A)/\text{rms } R.$$

Substitution for R, S and A, and evaluation of the convolution integrals then yields

$$\text{MTF} = \exp(-\sigma^2\omega^2/2) + A_0(2\tau^2\omega^2) \exp(-2\tau^2\omega^2).$$

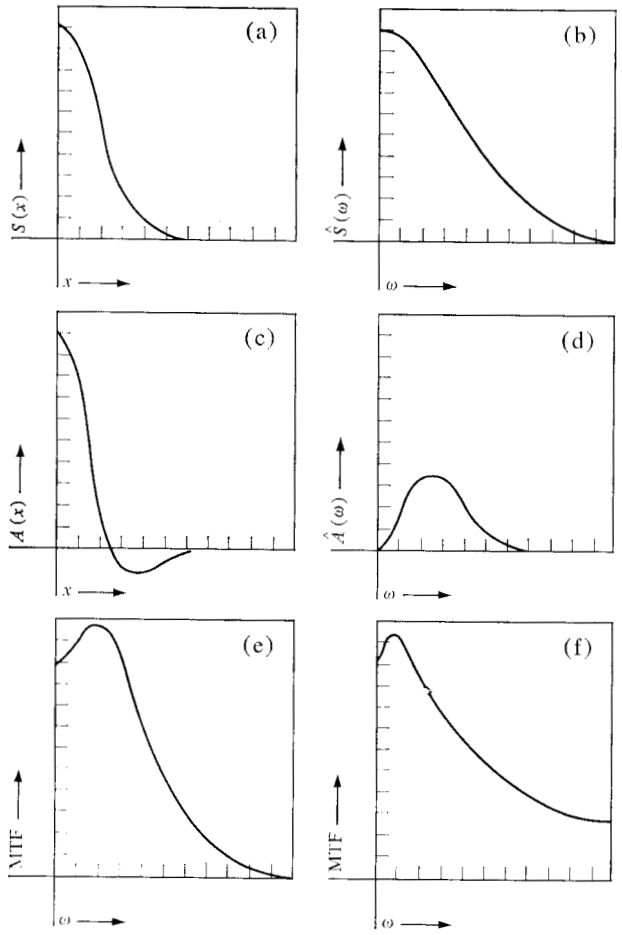

Figure 4 Point-spread, adjacency-enhancement, and MTF curves. (a) Point-spread function; (b) its Fourier transform; (c) adjacency-enhancement function; (d) its Fourier transform; (e) theoretical MTF; and (f) experimental MTF (Royal-X Pan, dev. 5 min. in DK-50[11]).

The first term decreases monotonically with increasing ω; the second rises to a maximum at $\omega = 1/\sqrt{2}\,\tau$, after which it decreases exponentially toward zero. The two terms thus form a typical MTF curve, as shown in Fig. 4. Adjacency effects, it will be noted, generate the experimentally observed bump at the lower spatial frequencies.

Acknowledgments

The author wishes to acknowledge discussions related to this work with L. Cutrona, J. Horner, F. Yu and P. Koch. The author is also indebted to E. Jones for encouraging publication of this research.

References

1. E. Leith, *Photo. Sci. and Eng.* **10,** 75 (1962).
2. R. van Ligten, *J. Opt. Soc. Am.* **56,** 1 (1966).
3. A. Kozma, *J. Opt. Soc. Am.* **56,** 428 (1966); **58,** 436 (1968).
4. C. Helstrom, *J. Opt. Soc. Am.* **56,** 433 (1966).
5. J. Goodman, *J. Opt. Soc. Am.* **57,** 493 (1967).
6. F. Yu, *J. Opt. Soc. Am.* **59,** 342 (1969).
7. J. Altman, *Appl. Opt.* **5,** 1689 (1966).
8. D. Falconer, *Phot. Sci. Eng.* **10,** 133 (1966).
9. C. Burckhardt, *Appl. Opt.* **6,** 1359 (1967).
10. C. Mees, *Theory of the Photographic Process,* The Macmillan Book Co., New York, 1954, p. 810ff.
11. H. Frieser, *Phot. Sci. Eng.* **3,** 164 (1959); J. Eyer, *Appl. Spectr.* **14,** 4 (1960).
12. G. Higgens, *Appl. Opt.* **3,** 1 (1964).
13. G. Zech, private communication (1967).
14. D. Kelly, *J. Opt. Soc. Am.* **51,** 319 (1961).
15. J. Simonds, *Phot. Sci. Eng.* **9,** 294 (1965).
16. J. Horner, private communication (1967).
17. *Modulation Transfer Data,* Eastman Kodak Co., Rochester, New York, Pamphlet No. P-49.

Received December 12, 1969