RESTORATION OF ATMOSPHERICALLY DEGRADED IMAGES

VOLUME 2

WOODS HOLE SUMMER STUDY

JULY 1966

Advisory Committee to the
Air Force Systems Command

NATIONAL ACADEMY OF SCIENCES
NATIONAL RESEARCH COUNCIL

(6) # RESTORATION OF ATMOSPHERICALLY DEGRADED IMAGES.

VOLUME 2.

(9) Report on

WOODS HOLE SUMMER STUDY

JULY 1966.

(11) 1966 (15) AF 18(600)-2891

(12) 255 p.

Advisory Committee to the
Air Force Systems Command

NATIONAL ACADEMY OF SCIENCES
NATIONAL RESEARCH COUNCIL

(402 299)

ADVISORY COMMITTEE
TO THE
AIR FORCE SYSTEMS COMMAND

Brian O'Brien, Pomfret, Connecticut, Chairman
Richard G. Folsom, Rensselaer Polytechnic Institute
William R. Sears, Cornell University
Oswald G. Villard, Jr., Stanford University
Willis H. Ware, The RAND Corporation
Kenneth S. McAlpine, Executive Secretary

OPTICAL MASER PANEL

Stephen M. MacNeille, American Optical Company, Chairman
John K. Galt, Bell Telephone Laboratories, Inc.
Brian O'Brien, Pomfret, Connecticut

Carroll O. Alley, Jr., University of Maryland, Consultant

Copies available from

Defense Documentation Center
Cameron Station
Alexandria, Virginia

CONTENTS

APPENDIX 1

THE VISIBILITY LABORATORY PROGRAM OF RESEARCH ON RESTORATION OF ATMOSPHERICALLY DEGRADED IMAGES

James L. Harris, Sr.

The purpose of this paper is to present a brief description of the program of research which is in progress at the Visibility Laboratory of the University of California, San Diego, and to summarize the essential results of this effort to date. This research has been supported by the Advanced Research Projects Agency.

RESEARCH OBJECTIVES

Atmospheric turbulence imposes a degradation on images of objects in space recorded by means of large ground-based optical systems. The human visual system is relatively inefficient in the extraction of information from badly degraded images. The goal of image processing is to operate on the degraded images in such a way as to produce a processed image which the human visual system can efficiently interpret. While the processing is usually in the direction of correcting the image for the degradation, the emphasis is on producing processed images from which object information can be extracted, a goal which is not necessarily consistent with concepts of image fidelity or subjective impression of the quality of the processed images.

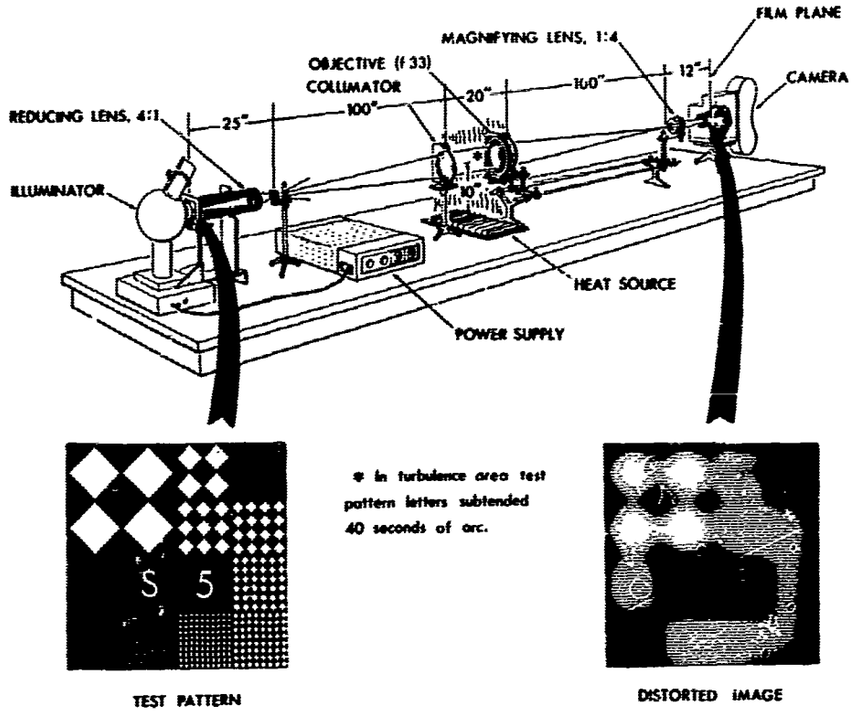


FIGURE 1 Laboratory turbulence generation system.

ARTIFICIAL TURBULENCE GENERATION

The image-processing research has centered upon experiments performed on images which have been degraded by heat-induced turbulence in a laboratory optical bench setup. A sketch of the system is shown in Figure 1.

The light source and projector optics serve to illuminate a film transparency which is the "object" in the experiment. The diffuser is inserted to insure that the object is incoherently illuminated. The collimating lens is used to place the object at infinity from the point of view of the telescope. The heat source is an array of resistance wire and is controlled by means of a variac. The image formed by the telescope objective is magnified and recorded on photographic film. The light level in the system allows photographic exposure periods as short as approximately 1 msec. The dimensions shown in the figure are typical but are varied from one experiment to the next.

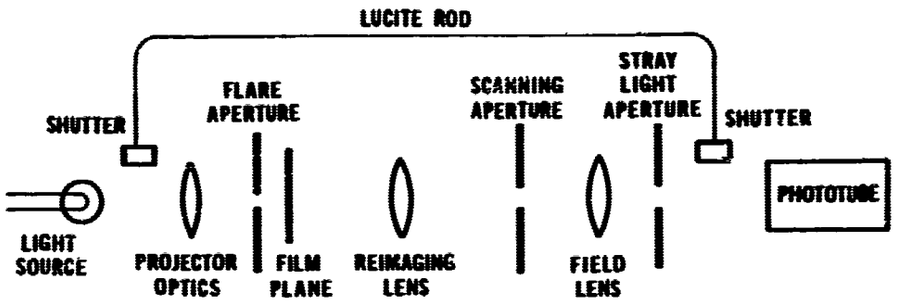


FIGURE 2 Sketch of film scanner.

IMPLEMENTATION OF PROCESSING OPERATIONS

Film Scanning

The degraded film images obtained from the optical bench turbulence generation system are read with a film scanner designed and built by the Visibility Laboratory. A sketch of the scanner is shown in Figure 2. The light source and projector optics illuminate the film, which is then reimaged into the plane of the scanning aperture. The flux passing through the scanning aperture is deposited on the photocathode of the multiplier phototube. The flare aperture located just behind the film is removed when the film is first inserted, to allow a full-field-of-view projection of the image onto the scanning aperture plane. This facilitates such problems as registration and determination of width of scan. During scanning, the flare aperture is in the position shown, so that only the small region of film associated with the reading is illuminated. This greatly reduces the total flux incident on the reimaging lens and hence reduces the flare from this lens. The field lens at the far end of the system images the reimaging lens on the photocathode of the multiplier phototube. By placing an aperture just in front of the photocathode, all stray light coming from regions outside the solid angle of the reimaging lens is eliminated.

A Lucite rod runs from the light source to the photocathode with a solenoid-operated shutter controlling the path. A similar shutter is in the main path of the scanner. The Lucite rod and shutter arrangement makes it possible to sample the auxiliary path at any desired time, as a means of achieving calibration information which can be used to correct the film scan data for long-term drift of the system. Considerable care has been taken to use maximum voltage stabilization for both the light source and the multiplier phototube. It has been determined by experience that the stability of the system is such that the use of the auxiliary Lucite path is not required for most applications.

The scanning operation is achieved by discrete stepping of the film using stepping motors. The number of steps taken in a horizontal line is adjustable from 1 to 90 by means of switch settings on a series of binary counters. At the end of each line the scanner quickly resteps and is stepped vertically one step by the vertical stepping motor. The number of vertical steps in the scan is adjustable from 1 to 128 in the same manner as the horizontal case. When the scanner has completed the specified number of vertical steps it resteps both horizontally and vertically, returning to its starting position. The stepping speed is variable up to a maximum mechanical limitation of approximately 30 steps per second.

The output from the multiplier phototube goes to a voltage-to-frequency converter and is then displayed on a counter. Before inserting film into the scanner, the system is adjusted so that the counter reads 1.0000.

The counter output is sequentially sampled and used to drive an IBM card punch. The card punch used limits the scanning rate to approximately four 4-digit picture elements per second. An incremental magnetic tape recording unit is presently being built into the system, which will allow scanning rates up to the 30-step-per-second mechanical limitation of the present mechanical configuration.

In performing a scan for an image-processing experiment, a gray scale on the same film strip is also scanned and recorded as a card deck.

Computer Operations

The computer used is a CDC 3600 located at the Computer Center of the University of California, San Diego. The first step in the computer processing is that of constructing a characteristic curve for the film from the gray scale card deck. Each image scan data point is then corrected to a relative exposure value by means of the characteristic curve.

In general, the next processing operation is that of calculating a two-dimensional complex Fourier series representation of the image. The series is described in trigonometric form so that for each spatial frequency there are four coefficients, corresponding to terms of the type $\sin k_1x \sin k_2y$, $\sin k_1x \cos k_2y$, $\cos k_1x \cos k_2y$, and $\cos k_1x \sin k_2y$. For a typical 60 x 60 image scan there are 30 x 30 spatial frequencies, each with four coefficients, or a total of 3,600 coefficient values. The programs also furnish the information in terms of modulus and phase values for each spatial frequency.

Normally, the next computer operation consists of operating on this image spectrum with a processing optical transfer function, which consists of a modulus and phase shift for each spatial frequency component. This corrected Fourier series is then summed to obtain a numerical description of the processed image, usually with a set of instructions for truncating the series so as to control the noise level in the processed image.

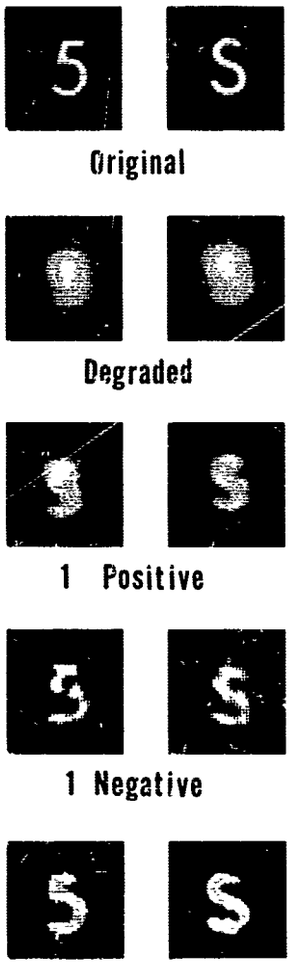


FIGURE 3 Laboratory turbulence degradations and restorations.

Output Display

For research purposes many steps in the mathematical processing operations are recorded, and the output from the computer consists of a very large quantity of printout material. Pictorial information is also provided in the form of a deck of "picture" cards (soon to be replaced by the magnetic tape system). The card deck defines the relative irradiance of each of a matrix of points on the restored image on the basis of a linear 100-step gray scale. The computer has normalized the output image so that the darkest point on the image is zero and the brightest point is 100.

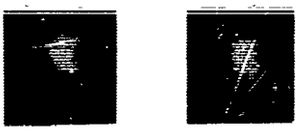
The card deck is read and used to time gate a cathode-ray tube spot, which is made to perform a discrete line scan similar to that used in the film scanner. The cathode-ray tube is photographed with a Polaroid scope camera. The lens in the camera has been masked to a square aperture and defocused so that the square image of the CRT spot has approximately the dimensions of the distance between spot positions. A Polaroid positive is used for quick inspection, with the negative stored for record and for future prints as required.

EXAMPLES OF EXPERIMENTAL ACHIEVEMENTS

A few selected examples of processing results will serve to illustrate the principal results of this research program.

Early work in the program was directed toward processing experiments on images degraded by long time exposures in the optical bench turbulence generation system. A typical result is shown in Figure 3.

In this series an "S" and a "5" were processed using a simple analytic model of the optical transfer function. In the case of long time exposures the point-spread function tends to become a smooth symmetric function so that the optical transfer function is real, i.e., no phase shifts. The processing shown in Figure 3 consists of correcting the modulus of the image spectrum for the attenuation caused by turbulence.



ORIGINAL DISTORTION

$$R = \exp\left\{\frac{i^2 + {}^2}{ai^2 + bj^2}\left[1 - \exp\left(\frac{i^2 + j^2}{(ci^2 + dj^2)^{1/2}}\right)\right]^2\right\}$$

$$f = (i^2 + j^2)^{1/2}$$

$$R = e^{5.56(1-e^{-.27f})^2}$$

$$R = e^{.5f}$$

$$R = 1 + .15f^3$$

$$R = 1 + f^2$$

FIGURE 4 Restorations with various analytic processing modulation transfer functions.

In the early processing experiments the computer programs had not yet been developed for correcting the film transmission readings with the characteristic curve of the film. It was therefore necessary to scan positive transparencies of reasonable linearity. The results shown for "1 positive" were obtained in this way. The results for "1 negative" are from identical processing except that the negative was scanned along with gray scales on the same film strip, and the computer made a "positive" prior to performing the processing operations. The results shown as "4 negatives" were obtained by scanning 4 negatives and averaging to enhance the signal-to-grain noise ratio prior to processing.

An important finding with respect to the long-time-exposure experiments is that precise information about the optical transfer function is not required in order to achieve useful processing results. Figure 4 shows a series of processing results for the degraded images of Figure 3 with various analytic processing optical transfer functions as shown. This insensitivity to the choice of processing function holds for the long exposure case only and is due to the fact that there is only small phase shifting at any spatial frequency.

In the case of short time exposures, the point-spread function is far from symmetric and the optical transfer function has phase shifts of great significance. An interesting comparison of long and short time exposure images is shown in Figure 5.

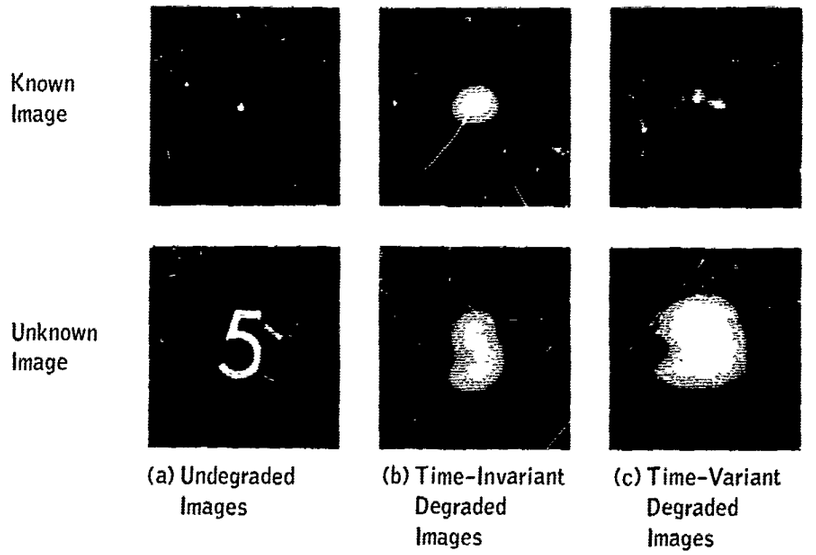


FIGURE 5 Comparison of long and short exposure degradation and point-spread functions.

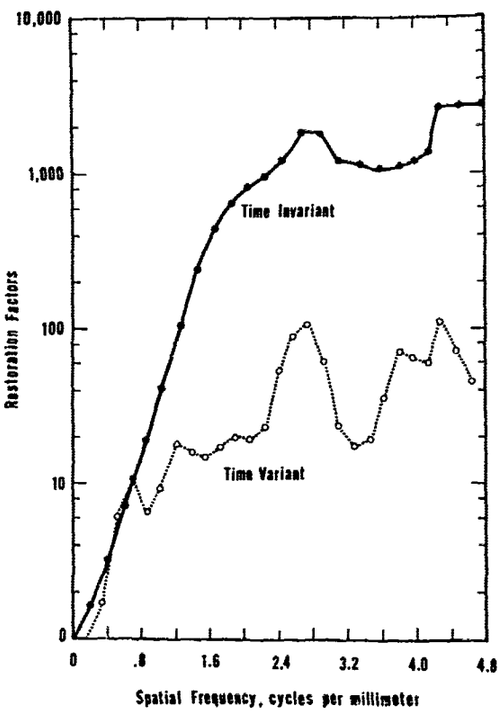


FIGURE 6 Modulus of processing optical transfer functions.

An experiment was performed in which the point images for each case were scanned and Fourier transformed to obtain the optical transfer function, which was then used to process the "unknown" image. The moduli of the processing optical transfer functions are shown in Figure 6. The significant fact is that there is more than an order of magnitude difference in these two functions at the higher frequencies. For any of the higher frequencies, the signal-to-noise (S/N) ratio is therefore significantly greater for the short exposure than for the long exposure.

The counterpart to Figure 6 is the plot shown in Figure 7 of the phase corrections associated with the optical transfer functions for the two cases. Here we see that the phase shifts for the short exposure are very pronounced

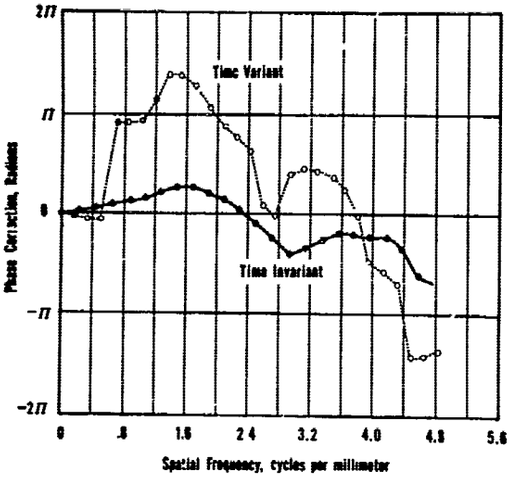


FIGURE 7 Phase of processing optical transfer functions.

compared with those for the long exposure. This is the result which we should expect, since it is the addition of vectors of variable phase for a series of short exposures which generates the attenuation associated with the long exposure.

It is important to realize conceptually that in the case of an exposure sufficiently short to freeze the atmosphere, the degree of spatial integration which takes place is inversely proportional to spatial frequency. The highest spatial frequencies are generated by small zones at either edge of the entrance pupil of the optical system. When the characteristic dimension of the turbulence is large compared with the area of each of the zones, then the spatial integration involved will not produce large attenuation of the modulus of the optical transfer function, but between the two zones there may be large phase shifts. It is therefore characteristic of the short exposure that the highest frequencies will suffer small attenuations and large phase shifts, while the midfrequencies will suffer large attenuations and smaller phase shifts.

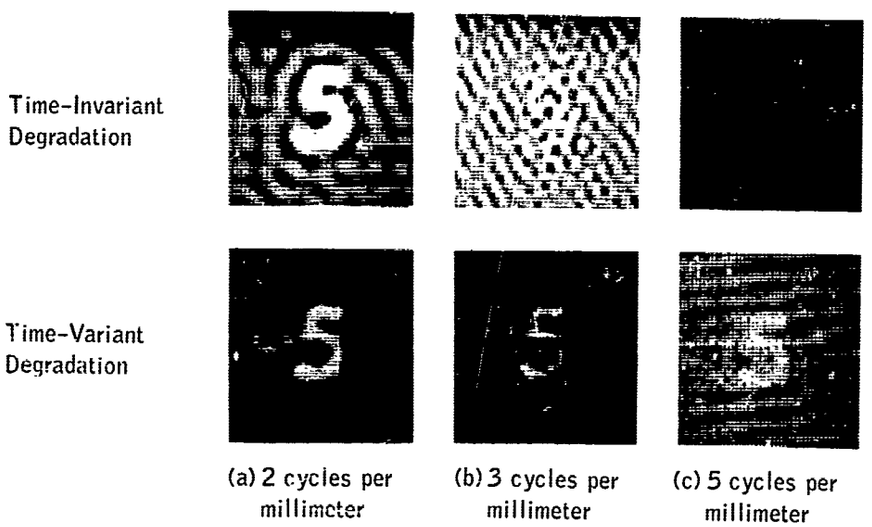
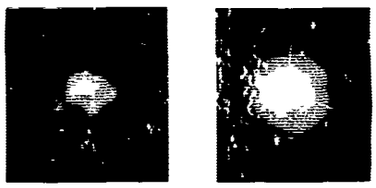


FIGURE 8 Restorations of long and short exposure degradations.



(a) (b)

FIGURE 9 (a) Point-spread function and (b) degraded numeral 5 for short-time-exposure experiment.

FIGURE 10 Restoration of degraded image of Figure 9(b).

Restorations of the two degradations are shown in Figure 8. The Fourier series representation was truncated at the spatial frequencies indicated. It was observed that restorations of the long-time-exposure case were unsuccessful beyond about 2 cycles/mm, whereas the short time exposure was successfully restored up to the full 5 cycles/mm allowable with our 0.1-mm scanning aperture.

Figure 9 shows a more recent example of a short-time-exposure degradation with a higher turbulence level than that previously used. As before, the point image was scanned and used to compute the processing optical transfer function for the correction of the spectrum of the degraded numeral 5. The final processing result is shown in Figure 10.

PRESENT AND FUTURE ACTIVITIES

A set of 500 short-exposure images of both a point source and an unknown object have been recorded in the turbulence generation system. These images are in the process of being scanned.

The first step in the analysis of the point image data will be to generate statistics of the optical transfer function, which will initially consist of independent frequency-of-occurence functions for both amplitude and phase, and may at a later date include the generation of two-dimensional (amplitude and phase) frequency-of-occurrence functions if required. This information will be carefully studied to determine how many parameters are required to give a reasonable description of the statistics involved.

The set of 500 images of the unknown object will then be used to perform a wide variety of restoration experiments, using a number of techniques and various assumptions about the extent of a priori information. These experiments will include: (a) transfer function known for each exposure, with multiple exposures used to improve the S/N ratio; (b) transfer function known statistically in the form of frequency-of-occurrence functions, with maximum likelihood and similar statistical tools used to make a best estimate of the correct image; and (c) transfer

function unknown, with various numbers of degraded images added to obtain a long-time-exposure transfer function, and with restoration accomplished by "guessing" at a smooth modulus correction. Many additional experiments will undoubtedly develop from the experiences encountered in performing these first experiments.

On the conclusion of these experiments, it is hoped that a similar set of data can be acquired from a real telescope facility.

APPENDIX 2

HOW MANY PICTURES MUST WE TAKE TO GET A GOOD ONE?

R. E. Hufnagel

Let $E(x, y)$ represent the complex field amplitude in the focal plane of a diffraction-limited telescope of focal length F and aperture radius R, which is viewing a point source of light through an intervening turbulent atmosphere. Let $I(x, y) = \left|E(x, y)\right|^2$ be the intensity distribution in the image.

Let $M(\xi, \zeta)$ be the normalized lateral coherence function for the light in the entrance pupil, defined in the same manner as was done by Hufnagel and Stanley.[1] With a scaling of the coordinates, M is identical to the average "atmospheric" optical transfer function.

We now consider the special case where the seeing is much worse than the potential diffraction limit of the telescope. This is analogous to saying that $M \approx 0$ when $\xi^2 + \zeta^2 > r_c^2$ for $r_c \ll R$.

When this is the case we may consider that the light which forms the image comes from several effectively uncorrelated sources within the aperture, and by the central limit theorem we may consider the E field in the entrance pupil to be a complex Gaussian random variable.

If this is the case,* then we may use some results obtained by Goodman[2] in connection with light scattering from rough surfaces.

*To be of nontrivial interest we must exclude the possibility that M is limited only because of simple random tilting of planar (flat) wavefronts. For atmospheric turbulence this will require that $r_c \lesssim 10R$.

First the probability density function for I(x, y) becomes

$$p(I) = \frac{1}{\langle I(x, y)\rangle} \exp\left[-I(x, y)/\langle I(x, y)\rangle \right] , \qquad (1)$$

which is just the chi-squared distribution for two degrees of freedom (one for the real part of E and one for the imaginary part).

Goodman also reports that the normalized autocovariance function for I becomes for our case

$$\frac{\langle I(x, y)\, I(x+\Delta x,\ y+\Delta y)\rangle}{\langle I(x, y)\rangle \langle I(x+\Delta x,\ y+\Delta y)\rangle} - 1 = \left[\frac{2J_1(2\pi R\rho / F\lambda)}{(2\pi R\rho / F\lambda)} \right]^2 , \qquad (2)$$

where $\rho = (\Delta x^2 + \Delta y^2)^{1/2}$.

This implies that the covariance distance is equal to the Airy disk radius of the aperture R. Since the turbulence does not significantly disturb the polarization of the light these formulas hold for both unpolarized and polarized light.

The interpretation of this is that under the stated conditions the point-spread function will be composed of high-contrast "grain" (not unlike the speckle pattern from laser-illuminated diffuse objects), that the grain "size" will be the size of the Airy disk, and that the individual grains will be "uncorrelated."

Examination of short-exposure stellar images from the 82-in.-diameter McDonald Observatory telescope[3] shows these features.

It is of academic (and perhaps practical) interest to compute the probability that any one of these "grains" will contain most of the image flux. When this occurs, the image (for that instant of time) will be nearly diffraction limited. The probability that any single grain will contain over half the total flux is given approximately by

$$P = \sum_j \exp\left[-\frac{\frac{1}{2}\sum_k \langle I_k\rangle}{\langle I_j\rangle} \right] ,$$

where j and k summations are over the uncorrelated "grains."* By replacing the sums by integrations over the blur circle, and assuming the blur to have a Gaussian spatial distribution with a standard deviation r_b, one obtains

*Even it the "grains" are uncorrelated, they are not independent; the approximation partially ignores the dependence.

$$p \approx \int_0^\infty \frac{2\pi r\, dr}{A} \exp\left\{ - \frac{\frac{1}{2} \int_0^\infty \frac{2\pi r'\, dr'}{A} \exp\left[- \frac{1}{2} (r'/r_b)^2 \right]}{\exp\left[- \frac{1}{2} (r/r_b)^2 \right]} \right\} ,$$

where A is the area under the normalized autocovariance function given by Eq. (2). If $r_b^2 >> A$, we may keep only the first two terms in the power series expansion for the inside denominator, to yield (after integrating the inside numerator)

$$p \approx \int_0^\infty \frac{2\pi r\, dr}{A} \exp\left\{ (-\pi r_b^2/A) \left[1 + \frac{1}{2} (r/r_b)^2 \right] \right\}$$

which becomes

$$p \approx 2 \exp(-\pi r_b^2/A) .$$

Noting that $A = (F\lambda)^2/\pi R^2$ and (because of the Fourier transform relationship between M and the blur circle) that $r_b \approx (F\lambda)/\pi r_c$, we get finally

$$p \approx 2 \exp\left[- (R/r_c)^2 \right] \text{ for } R >> r_c .$$

Let us assume that wavefront "samples" become uncorrelated in 0.1 sec of time, and ask what value of R/r_c will give a reasonable probability of getting one sample which yields a diffraction-limited image in a time of 50 sec. For this case p need be approximately 2×10^{-3}, or $R/r_c = \sqrt{7} = 2.6$. This does not satisfy the condition that $r_c < 10R$, but if we reinterpret r_c as a kind of "pseudo coherence radius" corresponding to the pseudo (short-time) transfer function (where wavefront tilts have been removed) then one might just sneak by. In any event, the deviation from effective normality for E will be such that p will be greater than calculated here.

The pseudo coherence radius, r_s, will be a function of both R and r_c. We have estimated that $r_s = R/2.6$ when $r_c \approx R/7$.

To illustrate this with numbers, let the average (long-time) blur be 1 sec of arc, which corresponds to $r_c = 5$ cm. Then if R = 35 cm (30-in. diameter), one

has a reasonable chance of getting a picture with nearly 1/7-sec-of-arc resolution if he makes 500 exposures. The model needs refinement, but this preliminary result does not seem too unreasonable. Note that there is an optimum R associated with each r_c and p. If R is less than this optimum, diffraction-limited pictures will be more common but will have more blur from diffraction. If R is greater than optimum, one will not obtain diffraction-limited pictures at all.

REFERENCES

1. R. E. Hufnagel and N. R. Stanley, J. Opt. Soc. Am. 54, 52 (1964).
2. J. W. Goodman, Proc. IEEE 53, 1688 (1965).
3. J. Texereau, Sky and Telescope 25, 347 (1964).

APPENDIX 3

AN IMPROVED MODEL TURBULENT ATMOSPHERE

R. E. Hufnagel

The purpose of this paper is to report on the distribution of turbulent fluctuations of optical refractive index in the earth's atmosphere. The strength values given here supersede the corresponding estimates reported by Hufnagel and Stanley in 1964.[1]

No attempt will be made here to report on or interpret in detail the hundreds of reports and papers which have been written by many workers in the past few years. However, a few selected references will be given.

The model is based on the Kolmogoroff theory of turbulence. We define the structure function for the index of refraction N in the usual manner[2] as

$$D_N(\vec{r}; h) = \left\langle [N(\vec{r}') - N(\vec{r}' + \vec{r})]^2 \right\rangle , \qquad (1)$$

where $\vec{r}$ is the separation vector and $\vec{r}'$ is a position at an altitude h above local ground.

For small $\vec{r}$ the turbulence will be reasonably isotropic so that in this region we will replace $\vec{r}$ by $r \equiv |\vec{r}|$.

The function $D_N(\vec{r})$ has the following approximate form:

$$D_N(\vec{r}) = \begin{cases} (1/3) C_N^2 \, r^2 \, r_o^{-4/3} & \text{for } r \leq r_o, \qquad (2) \\ C_N^2 \left[r^{2/3} - (2/3) \, r_o^{2/3} \right] & \text{for } r_o \leq r \ll r_k, \qquad (3) \end{cases}$$

where

$C_N(h)$ we call the structure parameter,

$r_o(h)$ we call the inner scale factor.

This particular simplified form for D was chosen because it and its first derivative are continuous at $r = r_o$, and it has the correct limiting properties. Theoretically,

$$r_o \approx 5\nu^{3/4}\epsilon^{-1/4}, \tag{4}$$

where ν is the kinematic viscosity and ϵ is the power per unit mass of air dissipated by viscosity. The inner scale factor is related to the inner scale length given by Tatarski and others by a factor of 5.

The distance r_k in Eq. (3) represents the separation distance r for which the turbulence may be considered strictly isotropic with a Kolmogoroff spectrum. The departure from Eq. (3) usually comes first for $\vec{r}$ in vertical directions. Indeed if $\vec{r}$ is horizontal, Eq. (3) often appears to be valid for r up to an order of magnitude or two larger than r_k. We have purposely avoided here the concept of an "outer scale" of turbulence, since its use seems to cause much (unrecognized) confusion. Part of the difficulty results from its common usage as a scalar, when in fact, it is a vector or (for velocity fluctuations) a third rank tensor. In some cases r_k is associated with the Obukhov scale length (see Lumley and Panofsky, and Webb in the Bibliography).

Figures 1 and 2 show plots of C_N, r_o, and r_k versus altitude. The solid and dashed lines represent "background" conditions for the atmosphere. Quite often, however, there are major disturbances superimposed upon this background, associated with frontal activity or some other special set of circumstances. We have indicated in dotted lines the strengths associated with some typical disturbances. "Typical" should not be interpreted as "average"; there is considerable variation from night to night, indeed from hour to hour. The atmosphere will almost never be exactly as shown.

Direct measurements of temperature or optical index fluctuation statistics have been rather scanty above ~ 30 m elevation. Above 3 km they are nonexistent, and we have been forced to infer the turbulence properties from various ground-based observations, especially scintillation.

In Figure 1, the value of $\int C_N^2\, dh$ integrated over the peak shown at 12 km is 2×10^{-12} in units of $(\text{index})^2 \times \text{cm}^{1/3}$. This value is inferred from near zenith measurements of stellar scintillation, corrected for finite aperture telescope size, and corrected for polychromatic light, using Tatarski's relations relating C_N to

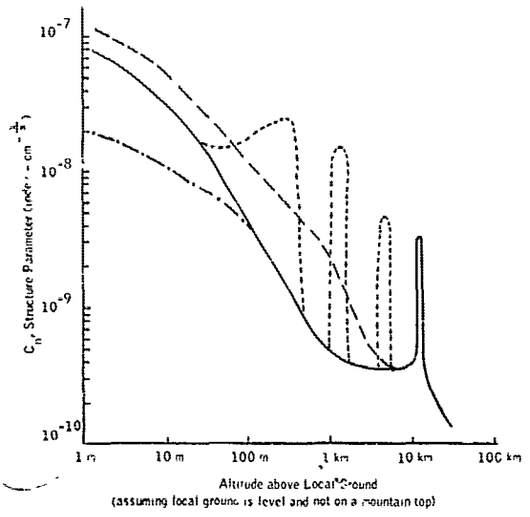


FIGURE 1 Model atmosphere. Structure parameter versus altitude. ---, sunny day; ——, clear night; -·-·-, dawn-dusk minimum; · · · ·, disturbed layers (see text).

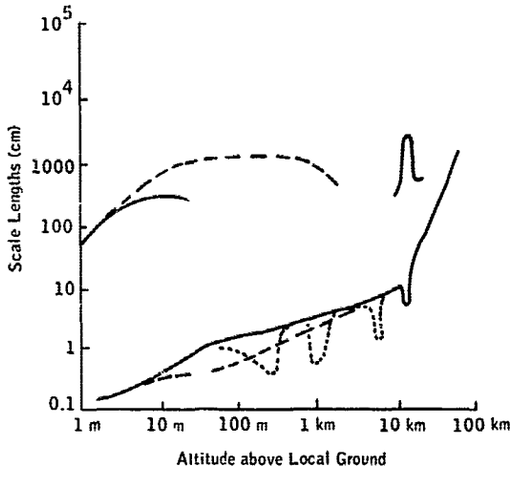


FIGURE 2 Model atmosphere. Inner scale factor r_o, lower curves; distance r_k, upper curves (incomplete for lack of experimental data). ---, sunny day; ——, clear night; · · · ·, disturbed layers (see text).

scintillation. From night to night $\int C_N^2\, dh$ varies with an approximately log normal probability distribution (with a standard deviation of log 2). It has a positive correlation with tropopause wind speed, but we have found no significant dependence on the geographic location of the observatory (from Pulkovo, USSR, to the southwestern United States).

The strength of the "dotted" layers has been estimated from a few scattered measurements at the $h < 3$ km levels and from the author's own observations on certain nights when special circumstances allowed the turbulence strengths and altitudes to be uniquely estimated from the ground.

The turbulence properties near the ground in active convective and near neutral conditions (daytime) are discussed in the first five entries of the bibliography.

Theory and experiments for stable conditions (night time) are in a very sad state – almost as bad as for the upper atmosphere.

REFERENCES

1. R. E. Hufnagel and N. R. Stanley, J. Opt. Soc. Am. 54, 52 (1964), Figure 6.

2. Ibid., Section 5.

SELECTED BIBLIOGRAPHY

Turbulence near the Ground

J. L. Lumley and H. A. Panofsky, The Structure of Atmospheric Turbulence, John Wiley & Sons, Inc., New York (1964); especially Section 5.9.

A. M. Obukhov, "The Structure of the Temperature and Velocity Fields in Free Convection," Bull. Acad. Sci. USSR, Geophys. Ser. No. 9 (1960), translation.

E. K. Webb, "Daytime Thermal Fluctuation in the Lower Atmosphere," Appl. Opt. 3, 1329 (1964).

L. R. Tsvang, "Some Characteristics of the Spectra of Temperature Pulsations in the Boundary Layer of the Atmosphere," Bull. Acad. Sci. USSR, Geophys. Ser. No. 10 (1963), translation.

Upper Troposphere and Lower Stratosphere

R. E. Hufnagel, Section II-B-4 (U) of "Atmospheric Properties for Laser Applications," I. D. A. Rept. S229 (Apr. 1966), Secret.

Reports from the 1965 International Colloquium on the Microstructure of the Atmosphere held in Moscow. The meeting is summarized in Izv. Acad. Sci. USSR, Atmospheric and Oceanic Physics 1, (Oct. 1965), translation. In particular see "The Fine-Scale Structure of the Atmosphere" by E. R. Reiter, Colorado State Univ., and "Empirical Data on the Small-Scale Structure of Atmospheric Turbulence" by A. S. Gurvitch, L. R. Tsvang, and A. M. Yaglom, Academy of Sciences of the USSR, Moscow, 1965.

Relevant material sometimes appears in the journal Atmospheric and Oceanic Physics translated from Russian by the American Geophysical Union.

APPENDIX 4

ON THE MEAN SHORT-TERM BLUR

R. E. Hufnagel

The main purpose of the calculation to follow is to show by example how one proceeds to calculate the mean short-term blur. We shall make some approximations (i.e., square aperture, geometric optics, no intensity scintillation, and monochromatic light) in order to arrive at hand-computable results quickly, but these approximations are not essential to the basic method. The computation of more exact results would make a fine thesis topic.

Without loss of generality we will assume a focal length of unity and a unit sized aperture. The variables $0 \leq x \leq 1$, $0 \leq y \leq 1$ are coordinates in the system pupil. The quantity $W(x,y)$ is the deviation of the wavefront from the reference surface at the pupil.

One may easily[1] show that the center of gravity of the point-spread function lies at a position

$$\vec{\rho} = \int I(x, y)\ \vec{\text{grad}}\ W(x, y)\, dx\, dy \ , \tag{1}$$

where I is the (normalized) intensity of the light entering the system pupil.

Let $P(\vec{r})$ be the point-spread function corresponding to W, where $\vec{r}$ is a two-dimensional vector in the image plane. We are interested in calculating P in a coordinate system shifted by an amount $\vec{\rho}$ which is also a function of W. Then $P_s(\vec{r}) \equiv P(\vec{r} + \vec{\rho})$ is the short-term blur (point-spread function).

Consider the Fourier transform of $P(\vec{r})$,

$$T(\vec{k}) = \iint P(\vec{r}) \exp(-2\pi i \vec{k}\cdot\vec{r})\, d\vec{r} . \quad (2)$$

Then

$$T_s(\vec{k}) \equiv \iint P_s(\vec{r}) \exp(-2\pi i \vec{k}\cdot\vec{r})\, d\vec{r} \quad (3)$$

might be called a short-term optical transfer function. However, there may be no isoplanatic region surrounding P_s and the subscript s should also serve to remind the reader that this may be a pseudo-transfer function, i.e., a Fourier transform of a spread function without an isoplanatic region.

From Eqs. (2) and (3) and the definition of P_s in terms of P it follows that

$$T_s(\vec{k}) = T(\vec{k}) \exp(2\pi i \vec{k}\cdot\vec{\rho}) . \quad (4)$$

We shall now proceed to calculate the mean value of $T_s(\vec{k})$. The brackets $\langle\ \rangle$ will designate an ensemble average. Note that $\langle T(k)\rangle$ is just the long-term optical transfer function, described by Hufnagel and Stanley.[2]

In terms of W, T is given by

$$T(\vec{k}) = \iint G(x,y)\, G^*(x+\lambda k_x,\ y+\lambda k_y) \cdot \left\{ \exp \frac{2\pi i}{\lambda} [W(x,y) - W(x+\lambda k_x,\ y+\lambda k_y)] \right\} dx\, dy , \quad (5)$$

where λ is the wavelength of light and $\vec{k} \equiv (k_x, k_y)$. The function G(x, y) is the system pupil function, and for a diffraction-limited system is unity everywhere inside the pupil and zero outside. For ease of later calculation we will assume a square pupil as stated earlier, rather than a conventional round one. Thus we have

$$T(\vec{k}) = \int_0^{1-\lambda k_y} \int_0^{1-\lambda k_x} \{\ \}\, dx\, dy , \quad (6)$$

where the same quantity is inside the curly brackets as in Eq. (5).

Note that Eqs. (5) and (6) do not admit intensity fluctuations to accompany W fluctuations; this is another simplification.* To further simplify the later calculations, we shall set $k_y = 0$, without loss of generality, thus obtaining

$$T(k_x) = \int_0^1 \int_0^{1-\lambda k_x} \exp \frac{2\pi i}{\lambda} \left[W(x, y) - W(x + \lambda k_x, y) \right] dx\, dy \qquad (7)$$

For most of this paper, we shall pass to the geometrical optics limit (another approximation), which is equivalent to taking the limit as $\lambda \to 0$. Eq. (7) becomes

$$T(k_x) = \int_0^1 \int_0^1 \exp \left[-2\pi i k_x \, (\partial/\partial x)\, W(x, y) \right] dx\, dy \ . \qquad (8)$$

Going back to Eq. (4), we get

$$T_s(k_x) = \int_0^1 \int_0^1 \exp \left\{ 2\pi i k_x \left[\rho_x - (\partial/\partial x)\, W(x, y) \right] \right\} dx\, dy \ , \qquad (9)$$

where $\vec{\rho} \equiv (\rho_x, \rho_y)$. For simplicity of notation we shall define $u(x, y) \equiv \partial W/\partial x$ and

$$\bar{u}(x, y) \equiv \int_0^1 \int_0^1 u(x, y)\, dx\, dy \ . \qquad (10)$$

Note the relationship of Eq. (10) to Eq. (1). Thus

$$T_s(k_x) = \int_0^1 \int_0^1 \exp \left\{ 2\pi k_x \left[\bar{u} - u(x, y) \right] \right\} dx\, dy \ . \qquad (11)$$

If u is a Gaussian random variable, then

---

*Those wishing to include the effects of intensity fluctuations should be warned that, in general, intensity is correlated with W(x, y) and also with $\vec{k} \cdot \vec{\rho}$. This would, of course, complicate the analysis.

$$\langle T_s(k_x) \rangle = \int_0^1 \int_0^1 \exp \left\{ - 2\pi^2 k_x^2 \langle [u(x,y) - \bar{u}]^2 \rangle \right\} dx\, dy \quad . \tag{12}$$

We shall now concentrate for a while on finding the right-hand averaged quantity which we shall call Q:

$$Q(x,y) \equiv \langle [u(x,y) - \bar{u}]^2 \rangle \quad . \tag{13}$$

First,

$$Q = \langle u^2 \rangle - 2 \langle u \bar{u} \rangle + \langle \bar{u} \bar{u} \rangle \, . \tag{14}$$

We shall define the autocovariance function for u as $C_u(x', y') \equiv \langle u(x,y)\, u(x+x', y+y') \rangle$, assuming that u is a statistically stationary (in space) random variable. Also, we shall define the structure function for the same process as

$$D_u(x', y') = \langle [u(x,y) - u(x+x', y+y')]^2 \rangle \quad ,$$

$$D_u(x', y') = 2[C_u(0,0) - C_u(x', y')] \quad . \tag{15}$$

The quantity $\langle \bar{u} \bar{u} \rangle$ is given by

$$\langle \bar{u} \bar{u} \rangle = \int_0^1 \int_0^1 \int_0^1 \int_0^1 \langle u(x,y)\, u(x', y') \rangle \, dx\, dy\, dx'\, dy'$$

$$= \int_0^1 \int_0^1 \int_0^1 \int_0^1 C_u(x-x', y-y')\, dx\, dy\, dx'\, dy' \quad .$$

By rotating the (x, x') and (y, y') coordinate systems 45°, we obtain

$$\langle \bar{u} \bar{u} \rangle = 4 \int_0^1 \int_0^1 C_u(\xi, \zeta)\, (1-\xi)\, (1-\zeta)\, d\xi\, d\zeta \quad . \tag{16}$$

In terms of structure functions we find that

$$Q(x, y) = \int_0^1 \int_0^1 D_u (x-x', y-y') \, dx' \, dy'$$

$$- 2 \int_0^1 \int_0^1 D_u(x', y') (1-x') (1-y') \, dx' \, dy' \quad , \tag{17}$$

where we have substituted $x' = \xi$, $y' = \zeta$ in the right-hand integral since they are just dummy variables.

To evaluate this further we need to know D_u, which we can express in terms of D_W, the structure function for W. In particular we have

$$C_u(\xi, \zeta) = (1/2) \frac{\partial^2}{\partial \xi^2} D_W(\xi, \zeta) \quad , \tag{18}$$

and D_u is related to C_u by Eq. (15). For Kolmogoroff turbulence with a finite inner scale of turbulence, D_W grows initially as $\xi^2 + \zeta^2$, changing later to $(\xi^2 + \zeta^2)^{5/6}$, and still later saturating to a fixed value.

To keep the example uncomplicated, let us assume that

$$D_W(\xi, \zeta) = \alpha^2 (\xi^2 + \zeta^2) - \frac{\beta^2}{12} (\xi^2 + \zeta^2)^2 + \ldots \quad , \tag{19}$$

where α and β are constants. Then it follows that

$$C_u(\xi, \zeta) = \alpha^2 - \frac{\beta^2}{2} (\xi^2 + \frac{1}{3} \zeta^2) + \ldots \tag{20}$$

and

$$D_u = \beta^2 (\xi^2 + \frac{1}{3} \zeta^2) + \ldots \quad . \tag{21}$$

Since C_u in real life is everywhere positive, it follows that this approximation to D_W is reasonable only if $\beta^2 \lesssim \alpha^2$, which does not allow a very good fit to the Kolmogoroff-type statistics. Nevertheless, we can substitute Eq. (21) into Eq. (17) to get the simple result

$$Q(x, y) = \beta^2 [(x - 1/2)^2 + (y - 1/2)^2] + \ldots \quad . \tag{22}$$

Substituting this into Eq. (12) with a change of variables: $\mu = x - 1/2$, $\nu = y - 1/2$, we get

$$\langle T_s(k_x) \rangle = \int_{-1/2}^{+1/2} \int_{-1/2}^{+1/2} \exp\left[- \frac{1}{2} (2\pi\beta k_x)^2 (\mu^2 + \frac{1}{3} \nu^2)\right] d\mu \, d\nu \ ,$$

which integrates to yield,

$$\langle T_s(k_x) \rangle = \frac{\sqrt{3}\,\pi}{2(\pi\beta k_x)} \operatorname{erf}\left(\frac{\pi\beta k_x}{\sqrt{2}} \right) \operatorname{erf}\left(\frac{\pi\beta k_x}{\sqrt{6}} \right) \ . \tag{23}$$

For small βk_x the power series expansion is

$$\langle T_s \rangle = 1 - \frac{2}{9} (\pi\beta k_x)^2 + \frac{1}{27} (\pi\beta k_x)^4 + \cdots \ , \tag{24}$$

and asymptotically for large βk_x it goes as

$$\langle T_s \rangle \approx \frac{\sqrt{3}\,\pi}{2(\pi\beta k_x)^2} \ . \tag{25}$$

Figure 1 shows a plot of $\langle T_s \rangle$. Since it is a geometric optics limit, it is valid only if $\lambda k_x \ll 1$.

For Kolmogoroff-type turbulence, $\langle T_s \rangle$ will be different in detail from what is shown here, but the relatively gradual decrease with k_x should remain. See also the end of Appendix 5.

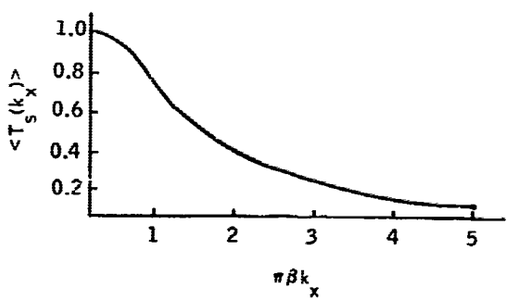


FIGURE 1 The mean short-term optical transfer function $\langle T_s \rangle$.

In another calculation similar to the one shown here, we evaluated $\langle T_s \rangle$ by diffraction theory for the limiting case of $\lambda k_x \to 1$, $k_y = 0$. The result was

$$\lim_{\lambda k_x \to 1} \left[\frac{\langle T_s(k_x, 0) \rangle}{1 - \lambda k_x} \right] = \frac{\sqrt{6}\,\pi}{(2\pi\beta/\lambda)} \operatorname{erf}\left(\frac{\pi\beta}{\sqrt{6}\,\lambda} \right) \tag{26}$$

where we recognize the factor $1 - \lambda k_x$ as the transfer function associated with the lens alone and the quantity on the right as being that part of $\langle T_s \rangle$ associated with the turbulence.

INTERPRETATION OF RESULTS

Let us define $T_m(k) \equiv \exp(2\pi i \vec{k} \cdot \vec{\rho})$.

If there were not short-term blur at all, but just random image displacement, the long-term transfer function would be

$$\langle T_m(k_x) \rangle = \langle \exp(-2\pi i k_x \rho_x) \rangle$$

$$= \exp\left[-2\pi^2 k_x^2 \langle \bar{u} \cdot \bar{u} \rangle \right] \tag{27}$$

$$= \exp\left\{ -2\pi^2 k_x^2 \left[\alpha^2 - (1/9)\beta^2 \right] \right\} , \tag{28}$$

where we used Eqs. (16) and (20) to get Eq. (28). If $T_m(k)$ were uncorrelated with $T_s(k)$, we would then expect that the average long-term transfer function $\langle T(k) \rangle$ would obey the relation

$$\langle T(k) \rangle = \langle T_s(k)\, T_m(k) \rangle \stackrel{?}{=} \langle T_s(k) \rangle \langle T_m(k) \rangle . \tag{29}$$

Using Eq. (28) for $\langle T_m \rangle$ and Eq. (24) for $\langle T_s \rangle$ we obtain

$$\langle T_m \rangle \langle T_s \rangle = \left[\exp\left(-2\pi^2 k_x^2 \alpha^2 \right) \left(1 + \frac{2\pi^2 k_x^2}{9\beta^2} + \ldots \right) \right]$$

$$\times \left[1 - \frac{2}{9} \pi^2 \beta^2 k_x^2 + \ldots \right] ,$$

or to terms in k_x^2,

$$\langle T_m \rangle \langle T_s \rangle \approx \exp\left(-2\pi^2 k_x^2 \alpha^2\right) .$$

In the geometric limit, however,

$$\langle T(k) \rangle = \exp\left(-2\pi^2 k_x^2 \langle u^2 \rangle\right) = \exp\left(-2\pi^2 k_x^2 \alpha^2\right) .$$

Thus $\langle T \rangle \approx \langle T_m \rangle \langle T_s \rangle$ under the conditions stated. For small β, we find that $\langle T \rangle \approx \langle T_m \rangle \langle T_s \rangle$ in the diffraction calculations represented by Eq. (26), as well.

A one-dimensional version of these calculations was done by Hufnagel with similar results and formed the basis of the expression

$$\langle T_s \rangle \approx \langle T(k) \rangle / \langle T_m \rangle = \langle T(k) \rangle \exp\left(2\pi^2 k^2 \langle \rho^2 \rangle\right) \tag{30}$$

given by him in 1963.[3] Since $\langle \rho^2 \rangle$ and $\langle T \rangle$ are easy to calculate, this is a convenient formula to use. It agrees with available experimental data, but is not rigorously correct.

If we apply Eq. (30) blindly to Kolmogoroff-type turbulence (with zero inner scale), we obtain the result

$$\langle T_s \rangle = \exp\left\{ -\frac{2\pi^2}{\lambda^{1/3}} \gamma k^{5/3} \left[1 - (k/k_o)^{1/3} \right] \right\} , \tag{31}$$

where $k_o = \lambda D$ is the aperture diffraction cutoff frequency and γ is a constant depending on the optical path turbulence strength. Here the effect of distortion correction is to add* the extra factor $\left[1 - (k/k_o)^{1/3} \right]$ into the expression for $\langle T \rangle$ which is

$$\langle T \rangle = \exp\left(-2\pi^2 \gamma k^{5/3} \lambda^{-1/3}\right) . \tag{32}$$

A more accurate formula would probably contain a factor of the form

$$\left[1 - \epsilon (k/k_o)^{1/3} \right] ,$$

where $0 < \epsilon < 1$, would be related to the factor given in Eq. (26), and would also depend on the nature of the intensity scintillation present.

*For a nonunity focal length, $k_o = \lambda DF$, where F = focal length, and k is measured in cy/length. Note also that we have generalized back to a two-dimensional $\vec{k}$, where $k = |\vec{k}|$.

One implication of Eq. (29) is that distortion has a small correlation with short-term blur. This raises questions with regard to the idea of looking for "sharp" pictures when the distortion is small. However, if the turbulence is not stationary in time, there would be a benefit since this implies (in this formulation) that β is correlated with α in some larger ensemble.

REFERENCES

1. F. Replogle et al., "Optical Propagation Study," RADC-TR-65-511, Space Surveillance and Instrumentation Branch, RADC, Griffiss AFB (Jan. 1966), Appendix A by R. E. Hufnagel, p. A14.

2. R. E. Hufnagel and N. R. Stanley, J. Opt. Soc. Am. 54, 52 (1964), Eq. (2.5).

3. R. E. Hufnagel, "Understanding the Physics of Seeing through Turbulent Atmospheres," presented at a local SPIE mtg., Patrick AFB, Fla., Nov. 1963.

APPENDIX 5

CORRELATION FUNCTIONS FOR OPTICAL TRANSFER FUNCTIONS

R. E. Hufnagel

As stated in the main body of this report, a possible useful concept in connection with problems of isoplanatism would be that of a correlation function for optical transfer functions. Let us define such a function as

$$H = \langle T_A(\vec{k})\, T_B^*(\vec{k}) \rangle \tag{1}$$

where T_A and T_B will be two transfer functions. If A = B, then H becomes a kind of second moment of T from which one can determine the variance of T.

One could define a quantity H_s analogous to H but for $T_s(\vec{k})$ as defined by Eq. (3) of Appendix 4. In many regards H_s will be a more useful and interesting quantity than H, but it is harder to compute and we will not consider it further at this time.

The difference signified by the subscripts A and B may be from a difference in object field position, or without much added complexity, a difference in the time of the event, or a combination of both. Thus we shall write

$$H = H(\vec{k}; \vec{\Omega}, \tau) , \tag{2}$$

where $\vec{\Omega}$ represents a difference in field position angle and τ represents a difference in time. Further let $\vec{\omega}$ represent the angular rate of target motion and $\vec{V}_n(z)$ represent the projection of the wind velocity at z on a plane normal to the line of sight. Here z will represent slant range along the line of sight.

We now define a modified optical path structure function

$$D(\vec{\rho}) = \left\langle \left\{ \int_0^\infty [N_A(z) - N_B(z)]\, dz \right\} \right\rangle^2 , \qquad (3)$$

where N_A is the index of refraction measured from $\vec{\rho}'$ in the plane of the entrance pupil along a path parallel to z to point A, and N_B is measured from $\vec{\rho} + \vec{\rho}'$ in the entrance pupil to point B on the object at a time τ later. We will assume the object to be far beyond the atmospheric turbulence so that we need not consider path convergence near the object.

We let $\vec{\rho}''$ represent a two-dimensional vector in the directions normal to the slant path axis so that the index of refraction at any point in space can be written as $N(\vec{\rho}'', z)$.

With this notation, $\vec{\rho}''$ for the A situation becomes $\vec{\rho}'' = \vec{\rho}'$, and for the B situation $\vec{\rho}'' = \vec{\rho} + \vec{\rho}' + z(\vec{\Omega} - \vec{\omega}\tau) - \vec{V}_n(z)\tau$. With local statistical stationarity we may subtract $\vec{\rho}'$ from both terms to yield

$$D(\vec{\rho}, \vec{\Omega}, \tau) = \left\langle \left[\int_0^\infty \left\{ N(0, z') - N[\vec{\rho} + z'(\vec{\Omega} - \vec{\omega}\tau) - \vec{V}_n(z')\tau] \right\} dz' \right] \right\rangle , \qquad (4)$$

where both N's are measured at the same time; the turbulence has been assumed to be frozen and moving with the wind.

For the case of Kolmogoroff turbulence, this expression simplifies somewhat to become[1]

$$D(\vec{\rho}, \vec{\Omega}\tau) = 2.91 \int_0^\infty C_N^2(z) \left| \vec{\rho} + z(\vec{\Omega} - \vec{\omega}\tau) - \vec{V}_n(z)\tau \right|^{5/3} dz , \qquad (5)$$

where we have assumed a zero inner scale for turbulence. C_N is defined in Appendix 3.

We do not expect this quantity to be commonly (if ever) completely evaluated except perhaps when $C_N(z)$ assumes a very simple form like a delta function.

Note, however, that if $C_N(z)$ is delta-like or if $\vec{V}_n(z)$ is nearly constant in direction and roughly proportional to z for the z values where C_N is large, then there are $\vec{\Omega}$ such that $z(\vec{\Omega} - \vec{\omega}\tau) - \vec{V}_n(z)\tau \approx 0$, which means that one could measure a $T_s(\vec{k})$ at $\tau = 0$ at one point in the field (e.g., at an edge or at a specular point), and use it when $\tau \neq 0$ at some other point $\vec{\Omega}$ in the field.*

*It is the author's experience that to a reasonable degree, this is a common effect. It represents a valuable piece of a priori knowledge for image restoration of extended objects.

For many practical calculations one will assume that $\vec{\Omega}$, $\vec{\omega}$, τ, or $\vec{V}$, or a combination of them are zero.

Having defined D we must now show how it is used. Let us define:

$$W_A(\vec{p}') = \int_0^\infty N\, dz \tag{6}$$

integrated from $\vec{p}'$ in the entrance pupil to point A at the time corresponding to A, and correspondingly for W_B.

Using a near-field approximation, we have

$$T_A(\vec{k}) = \int \exp\left\{ -\frac{2\pi i}{\lambda} [W_A(\vec{p}') - W_A(\vec{p}' + \lambda\vec{k})] \right\} d\vec{p} \tag{7}$$

and similarly for $T_B(\vec{k})$. The $\vec{p}$ integration is over the common area of pupil and shifted pupil overlap. We have set the focal length to unity. Now

$$\left\langle T_A(\vec{k})\ T_B{}^*(\vec{k}) \right\rangle = H$$

$$= \int d\vec{p}' \int d\vec{p}'' \left\langle \exp \frac{2\pi i}{\lambda} \left[W_A(\vec{p}') - W_A(\vec{p}' + \lambda\vec{k}) - W_B(\vec{p}'') + W_B(\vec{p}'' + \lambda\vec{k}) \right] \right\rangle . \tag{8}$$

Assuming that W is a Gaussian, statistically stationary random variable we obtain from Eq. (8),

$$H = \int d\vec{p}' \int d\vec{p}'' \exp\left[-\frac{2\pi^2}{\lambda^2} F \right] , \tag{9}$$

where F is the expected value of the square of the quantity in square brackets in Eq. (8).

Upon closer examination, one discovers that F can be written as a sum of $D(\vec{p}, \vec{\Omega}, \tau)$ functions, defined in Eq. (4), where $\vec{p}$ is expressed in terms of $\vec{p}' - \vec{p}''$, $\lambda\vec{k}$, their sums and their differences.

The integrals in Eq. (9) are easily evaluated by computer with Eq. (5) as an aid, but to set up the expressions, one needs to lock himself in a room for a few days!

We have had an opportunity to evaluate a much simpler quantity, namely H_A (the case when A = B). The calculations were done for a one-dimensional aperture of unit size. In this case $\vec{p} = p$, $\vec{k} = k$, $\Omega = \omega = \tau = 0$.

$$T(k) = \int_0^{1-\lambda k} \exp \frac{2\pi i}{\lambda} \left[W(p') - W(p' + \lambda k) \right] dp' \tag{10}$$

and

$$H = \int_0^{1-\lambda k} dp' \int_0^{1-\lambda k} dp'' \exp \left[- \frac{2\pi^2}{\lambda^2} \left[2D(\lambda k) + 2D(p) - D(p+k) - D(p-k) \right] \right], \tag{11}$$

where $p = p' - p''$ and $D(p) \equiv D(p, 0, 0)$.

Since the integrand depends on p' and p'' only through p, we switch coordinates and integrate out over the direction orthogonal to p.

$$H = 2 \int_0^{1-\lambda k} (1 - \lambda k - p) \exp \left[\quad \right] dp , \tag{12}$$

where the quantity inside the square brackets is as before.

Substituting for Eq. (5), we obtain

$$H = 2 \int_0^{1-K} (1-K-p) \exp \left\{ -\gamma \left[2K^{5/3} + 2p^{5/3} - (p+K)^{5/3} - (p-K)^{5/3} \right] \right\} dp , \tag{13}$$

where $K \equiv \lambda k$, and

$$\gamma = \frac{2\pi^2}{\lambda^2} (2.91) \int_0^{\infty} C_N^2 (z) \, dz . \tag{14}$$

We have integrated Eq. (13) in a few limiting cases and will present the most interesting result here.

Numerical substitution shows that the quantity in square brackets in Eq. (13) may be approximated by

$$K^{5/3}\left[\,2x^{5/3} - \frac{10}{9}\,x^2\,\right] \cong K^{5/3}\,x^2 \tag{15}$$

for $0 \leq x \leq 0.8$ where $x = p/K$, provided that $\gamma K^{5/3} >> 1$ when $K > 0.55$.

Equations (13) and (15) combine to yield

$$H = 2K \int_0^{1-K} (1 - K - Kx)\, \exp\left(-\gamma K^{5/3} x^2\right) dx \tag{16}$$

or, carrying out the integration,

$$H = \frac{\sqrt{\pi}\,K\,(1-K)}{\sqrt{\gamma K^{5/3}}}\ \mathrm{erf}\left(\frac{1-K}{\sqrt{\gamma K^{5/3}}}\right) - \frac{K^2}{\gamma K^{5/3}}\left\{1 - \exp\left[\frac{(1-K)^2}{\gamma K^{5/3}}\right]\right\}. \tag{17}$$

As $K \to 1$, $H \to 0$ as expected; but for $0.55 < K \lesssim 0.7$ with $\gamma K^{5/3} >> 1$ we have

$$H \cong \frac{\sqrt{\pi}\,K\,(1-K)}{\sqrt{\gamma\,K^{5/3}}} \cong \frac{0.4}{\sqrt{\gamma\,K^{5/3}}}\,. \tag{18}$$

Now

$$\langle T(k) \rangle = \exp\left(-\gamma K^{5/3}\right) \tag{19}$$

and we thus have $\langle T(k) \rangle \cong 0$ for $\gamma K^{5/3} >> 1$. Yet

$$H = \left\langle \left|T\,(k)\right|^2 \right\rangle \cong \frac{0.4}{\sqrt{-\log \langle T \rangle}}\,. \tag{20}$$

The significance of this is that it indicates how strong the effect of phase wander in T_s is in making $\langle T \rangle$ small. Thus $\langle |T| \rangle$ may be of sufficient size to warrent restoration even in the presence of noise even though it would be hopeless to consider a restoration attempt on $\langle T \rangle$.

The key problem of really significant restoration will be one of estimating the phases of the image Fourier components.

REFERENCE

1. R. E. Hufnagel and N. R. Stanley, J. Opt. Soc. Am. 54, 52 (1964), Eqs. (7.1) and (7.2).

APPENDIX 6

SUGGESTED ATMOSPHERIC EXPERIMENTS RELATING TO IMAGE RESTORATION

S. A. Collins, Jr.

The following contains a brief discussion of atmospheric effects applicable to the image reconstruction problem, and a suggested program to provide atmosphere information necessary to improve reconstruction.

It is well known that atmospheric turbulence is caused by differential heating or cooling of the air. Temperature gradients different from the normal adiabatic lapse rate induce vertical air currents which cause refractive index fluctuations. The refractive index spatial variations extend from millimeters in length to tens of meters and sometimes greater. The observed temporal spectrum of the fluctuations is down over 30 dB at 500 cps, so millisecond photographs generally stop the atmospheric motion.

The effects of these refractive index fluctuations are to diffract light traversing the air, causing both amplitude and phase fluctuations of the light waves. As viewed through a telescope, a stellar image will fluctuate in intensity, move, and break up into small pieces. Roughly speaking, turbulent eddies of characteristic dimension D produce only phase fluctuations if they are at a distance from the receiver small compared with $D^2/2\lambda$, and both phase and amplitude fluctuations if they are at a distance comparable with or greater than $D^2/2\lambda$.

In order to use image processing to overcome atmospheric image degradation, it would be sufficient to know the point-spread function of the atmosphere. Indeed the ideal case would be to take a series of fast photographs, knowing the point-spread function at the instant of each photograph. Known image-restoration

techniques could then be used to restore the image. However, it is not feasible to measure the instantaneous point-spread function with every picture. Stars are not sufficiently numerous for this purpose, and the task of stationing an airborne reflector within the telescope field of view to the desired accuracy is insurmountable.

In practice, the best that can be done is to record ensembles of short-exposure point-spread functions for statistical analysis. These would at least allow the use of statistical decision theory; for example, one might ask how many photographs should be analyzed using the average point-spread function before the resulting composite photograph is reliable. The measurements should be made under a variety of atmospheric conditions; e.g., summer, winter, various wind velocities and directions, various temperature gradients, and various velocities of winds in the upper atmosphere. Attempts should be made to correlate the size of the average point-spread function with meteorological data.

At the same time, one should investigate the temporal fluctuations of the point-spread function, or the optical transfer function, of the atmosphere. That is, some of the data should be obtained at a rate comparable with the rate of change of the point-spread function. This will check estimates of the speed of data acquisition necessary to have it regarded as instantaneous.

In addition to the experimental study of point-spread functions, measurements should be made of the isoplanatic patch size at the altitude of usual objects. The isoplanatic patch is defined roughly as the region on the target over which any two points may be placed and still give a good fringe pattern at the receiver. Several techniques are available for a measurement of this region. One is an examination of double stars. The angular separation of stars which no longer move together gives the size of the isoplanatic patch. Another technique involves the use of a high-flying airplane with two rows of corner-cube reflectors on its underside, as discussed below.

In general, measurements of point-spread functions and isoplanatic patches should be made looking through the majority of the atmosphere, i.e., with the source well above the tropopause (which is at about 40,000 ft). However, it is of considerable importance, both to passive image-processing schemes and to Goodman's wavefront-reconstruction scheme, to know whether most of the turbulence is near the ground or whether a significant amount is at the tropopause. For this reason, it would be desirable to try an experiment using an airplane with on-board laser sources or corner-cube reflectors flying at various levels.

We conclude with a rough calculation relating to an experiment which uses a high-flying airplane with a string of equally spaced corner-cube reflectors along the wings and a string along the fuselage. The airplane is illuminated, preferably with laser light, and is photographed using a large telescope. Simultaneously, the

intensity across the pupil is recorded photographically. For precise studies, magnetic tape image recording might be more desirable.

The images are analyzed in several ways. First, the fringe pattern in the aperture illumination is considered. The spatial spectrum should drop off in increments determined by the corner cube spacing, the highest spatial frequencies being due only to interference between retroreflectors with great separation. By examining the drop-off, one should be able to infer directly the size of the isoplanatic patch. This will also give directly the requisite proximity to the object of the holographic reference point required for Goodman's wavefront-reconstruction technique.

The second method of analysis concerns the various corner-cube images. Images within the same isoplanatic patch are expected to have the same point-spread function and will therefore have well-correlated patterns. The decrease in correlation with distance between images then allows a measure of the isoplanatic region. Further, many such images give a good ensemble of point-spread functions to use for the statistical point-spread analysis.

If the airplane is at an altitude of 80,000 ft, a spacing of 60 cm between corner cubes gives 5-sec-of-arc angular separation. These spots should easily be resolved with the Cloudcroft telescope. Further, there should be enough of them to give almost a continuous modulation transfer function.

A state-of-the-art Q-switched laser should be satisfactory. For a single corner-cube reflector at a range R, the relative energy returned to the receiver is

$$E_R/E_t = \frac{\sigma t^2 A g}{4\pi R^4 (\Delta\theta)^2} ,$$

where

$$\sigma = \frac{4\pi b^4}{3\lambda^2}$$

is the cross section of a corner-cube reflector,[1] and

b = corner cube size = 1.44 cm,

R = airplane height = 80,000 ft = 24.4 km,

λ = light wavelength = 10^{-4} cm,

$\Delta\theta$ = transmitter beam angle = 10^{-4} rad,

E_t = transmitter energy = 2 J,

t^2 = path loss $\approx$ 1/2,

A = representative receiver area,

g = intensity gain (area demagnification of receiving optics) = 10^2.

These numbers give

$$E_R/A = 4.0 \times 10^{-8} \text{ J/cm}^2.$$

The intensity gain was written assuming the 48-in. telescope aperture would be imaged on a 4.8-in. plate, giving an area demagnification of 100.

The value E_R/A compares quite favorably with the intensity required to fog Kodak IV-N film. This film requires[2] 10^{-15} J per resolution element and has 4×10^6 resolution elements per sq cm, giving 4×10^{-9} J/cm^2 to fog the film. That is a factor of 10 less than the available light intensity. Hence when the corner cubes are imaged with the telescope, there will be much more than the required light intensity.

The 1.44-cm corner-cube reflectors are just slightly smaller than the telescope resolution size, which is

$$\lambda R/d = \frac{10^{-4} \text{ cm} \times 2.44 \times 10^6 \text{ cm}}{(48 \text{ in.} \times 2.54 \text{ cm/in.})} = 2 \text{ cm}.$$

More work should be done on the method of analysis of results, and the experiments should be tried at ground level first, but this seems like quite a promising experiment to measure isoplanatic patch size.

REFERENCES

1. D. K. Barton, Radar Systems Analysis, Prentice-Hall, Englewood Cliffs, N. J. (1964), p. 71.
2. Kodak Plates and Films for Science and Industry, Pamphlet P-9, Eastman Kodak Company, Rochester, N. Y. (1962), p. 4D.

APPENDIX 7

AN INTERFEROMETRIC METHOD FOR MEASURING THE AVERAGE EFFECTIVE TRANSFER FUNCTION OF THE ATMOSPHERE

R. V. Shack

An instrument is described herein which can replace the eyepiece or camera of an astronomical telescope without further modification of the telescope and which can use a real star as the light source, yet which is capable of measuring the effective transfer function of the atmosphere as a continuous function over the full range of spatial frequencies defined by the diffraction limit of the telescope. The measurements are not affected by the aberrations or the transfer function of the telescope itself.

The heart of the instrument is a Koester prism used as a folding or inverting interferometer. The optical arrangement is shown in Figure 1. Light from the left half of the exit pupil enters the right half of the prism, is totally reflected by the outside face, and is partially reflected by the beam-splitting surface. Light from the right half of the pupil enters the left half of the prism, is totally reflected by the outside face, and is partially transmitted by the beam-splitter surface. Thus the two halves of the wavefront emerging from the pupil are superimposed as if one had been folded over on the other. Fringes which are nominally perpendicular to the folding edge are introduced by rotating the component prisms slightly about an axis perpendicular to the beam-splitting surface during assembly. The light need not be highly monochromatic, and, in fact, white light fringes can be seen because of the equality of the two light paths. In practice a moderately wide-band filter would be used to increase the number of observable fringes.

The inherent stability of the interferometer is very high, because the paths of the two beams between splitting and recombination are contained entirely within the glass of the prism. The spacing and position of the fringes along the dividing edge are fixed and determined by the twist given to the prisms when they are cemented together, and displacement of the star image in the direction of the dividing edge by a moderate amount will have no influence on the fringe pattern.

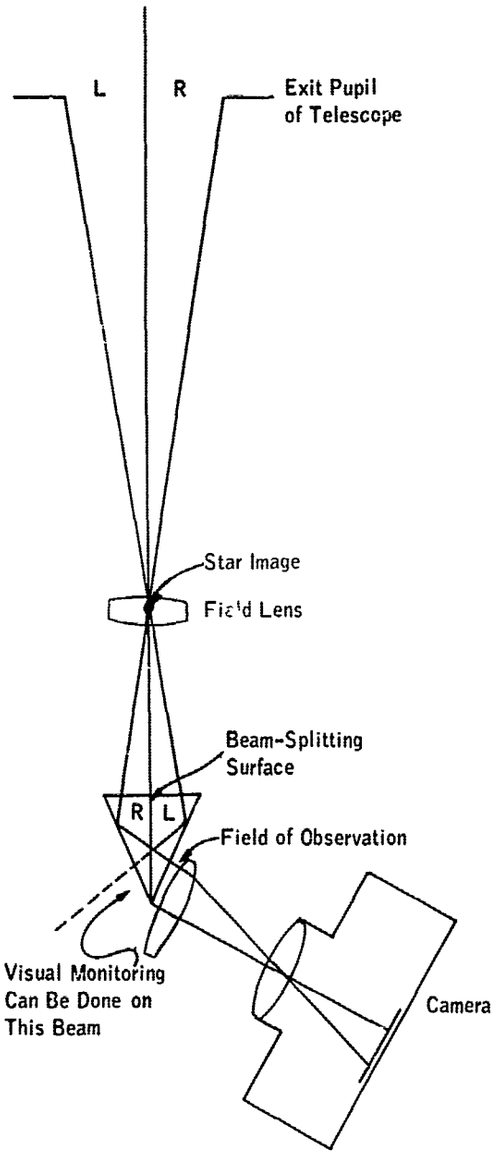


FIGURE 1 A Koester prism used as a folding interferometer.

In operation the fringe pattern would be photographed with an exposure time long in comparison with the fluctuation time of the turbulence and the variation in visibility of the time-averaged fringes as a function of distance from the folding edge determined [Figures 2(a) and 2(b)]. A plot of fringe visibility versus distance from the folding edge has exactly the same shape as the average effective transfer function of the atmosphere.

The proof of the latter statement is as follows. First, consider that the average effective transfer function for the atmosphere is the mutual coherence function of the incident disturbance over the space of the entrance pupil of the optical system, and the process being assumed stationary, the average obtained at any particular pair of points will be the same as the average over all sets of points with the same separation in the same direction.

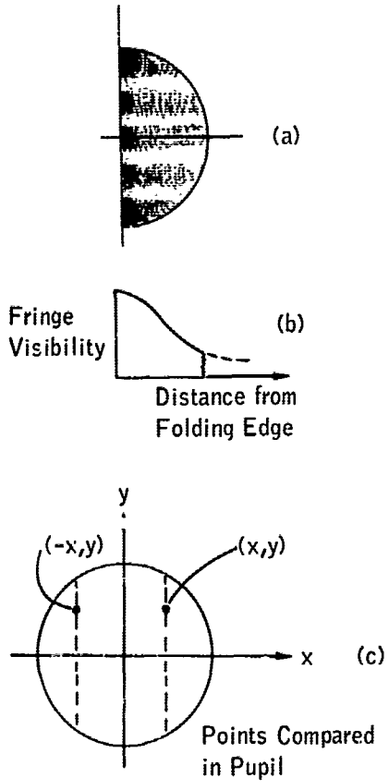


FIGURE 2 Variation of fringe visibility with distance from the folding edge.

Now consider the fringe pattern observed in the interferometer projected back into the pupil. The contribution from one beam is given by

$$A_+ = f(+x, y)e^{iay} = f_+e^{iay},$$

and the other by

$$A_- = f(-x, y)e^{-iay} = f_-e^{-iay},$$

where f(x, y) represents the complex wave disturbance in the exit pupil, and the ay term simply represents the deliberate tip of the wavefront in the y direction which produces the fringe pattern with the fringes nominally perpendicular to the dividing edge, the tipping being assumed equally distributed between the two beams. Then the resultant complex amplitude in the fringe pattern is given by

$$A = A_+ + A_- = f_+e^{iay} + f_-e^{-iay},$$

the complex conjugate of which is

$$A^* = f_+^*e^{-iay} + f_-^*e^{iay},$$

and the intensity of the fringe pattern is given by

$$AA^* = f_+f_+^* + f_-f_-^* + f_+f_-^*e^{i2ay} + f_-f_+^*e^{-i2ay}.$$

The mean value of this (long time integration) is

$$\langle AA^* \rangle = \langle f_+f_+^* \rangle + \langle f_-f_-^* \rangle + \langle f_+f_-^* \rangle e^{i2ay} + \langle f_-f_+^* \rangle e^{-2ay},$$

but if the process is stationary,

$$\langle f_+f_+^* \rangle = \langle f_-f_-^* \rangle = \langle ff^* \rangle = \langle |f|^2 \rangle$$

and

$$\langle f_+f_-^* \rangle = \langle f_-f_+^* \rangle,$$

so

$$\langle AA^* \rangle = 2\langle |f|^2 \rangle + 2\langle f_+ f_-^* \rangle \cos 2ay$$

$$= 2\langle |f|^2 \rangle \left[1 + \frac{\langle f_+ f_-^* \rangle}{\langle |f|^2 \rangle} \cos 2ay \right].$$

The fringe visibility, $\langle f_+ f_-^* \rangle / \langle |f|^2 \rangle$, is seen to be the mutual coherence function for points in the pupil separated by a distance 2x, and therefore a plot of visibility versus distance from the folding edge is identical in shape to the average effective transfer function for the atmosphere.

Other forms of folding interferometer are possible and will give the same results. A different form has already been built by Block Associates, and is soon to be used by James Elliott of Harvard Observatory on a 12-in. -aperture telescope. Plans are presently under way at the University of Arizona to use this type of instrument on large-aperture telescopes. Extensive and continuous use of this type of instrument at an observation site should provide considerable information as to the statistical nature of the seeing which can be expected at the site, possibly through the telescope itself for which this information will be of use.

As for the short-exposure transfer function, this device will not, unfortunately, give direct results, but it will provide information as to the fluctuations in the wavefront which can be of use in estimating the short-term transfer function. Note that it will work on artificial light sources, coherent or incoherent, monochromatic or polychromatic, so that specifically designed experiments using other than natural stars as light sources are perfectly feasible.

APPENDIX 8

A METHOD FOR MEASURING THE SHORT-TERM OPTICAL EFFECTS OF ATMOSPHERIC TURBULENCE

F. Dow Smith

The transfer function between object and image in a telescope is affected both by the refractive characteristics of the turbulent atmosphere and by the instrument itself. In the case of a long-term average, it is possible to assign to the atmosphere a transfer function which may then be incoherently combined with the transfer function of the optical system. Empirical data, consistent with theory, suggest that in all cases whether or not subsequent restoration is planned, short exposures will give the best results. In this short-term case the concept of a transfer function for the atmosphere alone is of limited value, especially when the seeing is good. The over-all transfer function must be obtained from the total wave aberration in the entire optical train between object and image. Not much is known of the detailed character of the wave aberration contributed by the atmosphere. Experimental measurements of this are therefore needed, both as a basis for image restoration studies and for prediction of system performance as a function of design parameters.

Fundamental then is the shape of the wavefront from a distant point as it is incident on the aperture of the telescope. Since the wavefront can be used as the starting point for any optical calculation using geometrical or diffraction optics, I believe that its determination should be given priority over all other measures of the optical effect of the atmosphere. This wavefront can be expected to change in time, but probably not much more rapidly than 100 cps or so. Thus any method of measurement must be fast, and it is desirable that data be obtained in a rapid

time sequence to permit a study of the wavefront as a function of time. Clearly, statistical analysis of the results will be needed to predict the most probable levels of degradation under given conditions and to evaluate the range of degradation.

An ideal method would not only provide data as suggested above to permit studying the nature of atmospheric degradation, but would also (1) duplicate the ray paths traveled by the light for the specific application, (2) permit a study of the size of the isoplanatic region, (3) permit analysis of the effect of tranverse target motion during exposure (i.e., scanning of the atmosphere during exposure as the telescope tracks the moving object), and (4) permit a measurement of the wavefront simultaneously with the main exposure.

It seems likely that a single method capable of achieving all these goals, if indeed one exists, would be complex and expensive. A variety of approaches are possible and several have been suggested by various members of the study group. Interferometric measures rate very high, but these should be capable of yielding the wavefront unambiguously. This is not possible, for example, from a shearing interferogram made at a single value of the shear.

I propose the use of the Hartmann test as a practical way of gathering a substantial amount of the data simply, inexpensively, and using established state of the art. The Hartmann test is a classic method for evaluating the performance of an optical system. It has been popular with astronomers for many years as a practical way for evaluating the over-all performance of a telescope. Its widespread use has been inhibited traditionally by the complexity of data reduction. This is no longer a serious problem since computer programs now exist which compute the shape of the wavefront. A second difficulty is lack of light. With conventional light sources, extremely long exposures are needed. This is not overriding in the case of a classical astronomical telescope with a good sidereal tracking mount which can track a star as long as needed. This limitation can be lifted by use of the laser. The illumination problem for this test is substantially the same as for an interferometer operating with a point source, since in each case, enough light must be available to expose a projected image of the aperture.

The test itself involves placing in the aperture of the telescope a plate perforated with an array of small holes. These holes have diameters characteristically 1/50 the diameter of the large aperture (e.g., perhaps 1-in. diameter for a 48-in. aperture). A glass photographic plate is exposed in positions displaced (say a few centimeters in the case of an f/6 system) in front of and behind the focal plane. The plate record will be an array of spots arranged as a geometrical projection of the aperture plate. If the system is imperfect, errors in this projection can be associated with ray aberrations and ultimately with a wavefront error.

To continue with the example of the 48-in. f/6 system, the spots on the plate would have large diameters, but the centroids can be determined accurately to within 2 μ, sufficient to identify errors of a fraction of a wavelength. In a recent laboratory evaluation of a mirror of about 4-ft diameter, tests were made both by the Hartmann procedure and with an interferometer. The results agreed to something between 1/5 and 1/10 wavelength over the surface.

It should be noted that in the classic use of this test on a telescope, a bright star is used for the light source. A long exposure is made while the telescope tracks the star. If the scintillations due to the atmosphere are statistically symmetrical in terms of beam displacement at the aperture, then the effect of the atmosphere is simply to enlarge the spots without changing their position. Thus the long-term effect of the atmosphere is averaged out. With short exposures, however, the instantaneous displacement of the spot will be recorded, giving a measure of the total aberration due to atmosphere and instrument.

A useful set of experiments using this method could be carried out by flying a laser in an airplane. The telescope would track the airplane so as to collect the light from the on-board laser. Depending on laser power, some aiming of the laser toward the telescope might be needed. It has been estimated that even in the extreme low-power case of a 10-mW gas laser, a beam 1 ° in diameter would give useful exposures. On-board corner cubes with a very high-power laser at ground level should also be considered.

APPENDIX 9

ATMOSPHERIC EFFECTS ON IMAGES - A PHENOMENOLOGICAL APPROACH

James L. Harris, Sr.

This paper contains some random thoughts about atmospheric properties and their measurement. I have great respect for the notion of a long-range program of research directed toward understanding the physics of atmospheric turbulence. Here, however, I adopt the short-range point of view that the atmosphere is a large black box which operates stochastically on the images which I wish to record and process. For this short-range black-box viewpoint, I will be content to obtain experimental evidence about certain critical features of the stochastic process involved.

It is clear that the size of an isoplanatic patch depends in a very complex way on the distribution and characteristics of turbulence along the path between the object and telescope, with the near turbulence tending to produce large isoplanatic regions. It is also intuitively clear that there is a different isoplanatic patch size for every spatial frequency and that the higher the spatial frequency, the smaller the region. This can be visualized geometrically by picturing a cone connecting a point in space with the periphery of the telescope entrance pupil, and a similar cone to a second point in space, the two points separated by an angle α, measured at the telescope. At any distance from the telescope the cross sections of the two cones will have a region of intersection. This region will, of course, decrease with distance, but the point to be made here is that a projection of this common cross section back to the telescope entrance pupil, shows that the very first portions of the wavefront to become uncorrelated with respect to the turbulence will be those zones on the entrance pupil of the telescope which produce the high spatial frequencies.

As has already been pointed out, the isoplanatic region is not some sharply bounded patch. The point-spread function slowly changes shape with position. How then can we define the isoplanatic region, which depends on a continuously varying function of distance as well as of spatial frequency? I would like to try it, strictly from the point of view of an image processer.

Let me start by assuming that I am viewing some extended object in space and that I isolate two small regions separated by a selected angular distance. Suppose now that I take a sequence of photographs, being careful to take out gross image motion by aligning center of gravity for each picture or some similar operation. Let me further assume that I have selected regions in which there is some prominent feature, surrounded by a reasonably uniform background. I now examine the first picture, and by Fourier analysis determine the complex two-dimensional Fourier series representation of each of the two regions. For each spatial frequency component, I take the ratio of the modulus for each region and the difference between phases. I repeat the operation for a second picture of the two regions. If there were strict isoplanatism, both the modulus ratio and the phase difference for each spatial frequency should be the same for each picture. I repeat this procedure for each of my series of pictures, thus generating a body of data which allows me to study the variability of the modulus ratios and phase differences as a function of spatial frequency. It would be well if I took appropriate note of the noise level of my sensor, underscoring those modulus and phase values which may in fact be noise, in order that I do not confuse isoplanatism with lack of correlation due to noise.

Suppose that I have completed the analysis described above, not only for these two regions but also for other selected regions having varied angular separations. How big is an isoplanatic region?

From the point of view of the image processer, I might first venture that isoplanatism exists over those angular regions and those spatial frequencies where the fluctuations in the modulus ratios and the fluctuations in phase differences are within the bounds of modulus and phase variations associated with reasonable picture quality.

Such an answer would be overlooking an important concept. The concept is that I may be able to make partial correction to an image by linear operations over a region in which isoplanatism exists only for that part of the image degradation which results from nearby turbulence. By partial correction I may make a significant improvement in image quality even though I leave uncorrected that portion of the degradation due to distant turbulence.

I can examine this possibility by returning to my body of data and examining, for one region, the variability of the modulus and phase from picture to picture.

So long as the variability for one region from picture to picture is significantly greater than the variability associated with comparing the two regions, I can expect partial spatial frequency correction to give reasonable results.

The experiments described above do not lead to a single numerical definition of isoplanatism, but they do tell me what I would like to know in order to process images. I suppose this is known as the field of Partial Isoplanatism.

PHASE STRUCTURE FUNCTIONS

Consider the case of a wave distorted by phase variation only. If I consider two points on the entrance pupil of my telescope separated by a distance d, then for a wavelength λ, the time averaged optical transfer function for a spatial frequency $f = d/\lambda$ cycles/radian will be determined (assuming spatial and temporal stationarity) by the properties of the phase difference between the two points. If, for example, I assume that the phase difference is normally distributed with a mean equal to the correct phase difference and a variance $\sigma^2(d)$ which is a function of the separation, d, of the two points, then the time averaged optical transfer function will be $e^{-\sigma^2(d)/2} = e^{-\sigma^2(\lambda f)/2}$.

A question about which we were quite vocal the first week of the study was, "Does the optical transfer function reach a plateau?" This is equivalent to asking, "Does $\sigma^2(d)$ reach a plateau for some distance d?" The answer to this question weighs heavily on the feasibility of large-diameter optics or large-baseline interferometer systems. It is important to recognize that the phase structure function, $\sigma^2(d)$, is a function associated with phase difference only, i.e., due to gross climatology there may be substantial fluctuations of absolute phase at the two points, but it is only the phase difference fluctuations which determine the optical transfer function properties.

There are a number of obvious approaches to determining $\sigma^2(d)$. For an image processer, perhaps the most direct approach is to measure optical transfer functions, as, for example, star image analysis. This would be a good solution provided that images of adequate signal-to-noise (S/N) ratio are obtained, since it yields directly the statistics of the optical transfer function which can be analyzed to investigate the existence of a plateau in the optical transfer function. Interferometry offers another direct approach. Perhaps the technique described by Shack in Appendix 7 will provide a clever way of obtaining experimental evidence on this important point.

APPENDIX 10

VISUAL OBSERVATIONS OF SEVERAL SATELLITES AT THE AFAL CLOUDCROFT ELECTRO-OPTICAL SURVEILLANCE RESEARCH FACILITY

Edmund T. Tyson

The observations were made at the Newtonian focus of the 48-in. aperture 312-in. focal length telescope. A Unitron Orthoscopic eyepiece of 6 mm focal length mounted in a rack and pinion focusing tube was used to observe the image. The field stop in the eyepiece is about 4 mm in diameter resulting in a field of view of about 1.5 min of arc. The magnification was 1,320X. On each observing night, a selection of eyepieces was tried and on each occasion the 6-mm eyepiece was found to be most suitable. Since the satellites subtended several seconds of arc, the planet Jupiter was used as the test object for eyepiece selection. (Incidentally, Jupiter was spectacular.) I had always regarded those sketches of the planet showing many bands with some suspicion. I am convinced they exist.

The first satellite observed on 6 March 1966 was Spadat Object No. 744, 64051, an S4 stage launched in January 1964. This rocket body is 80 ft long and 20 ft in diameter. It tumbles with a period of about 1 min. The satellite was acquired at 01 35 04 UT at an elevation of 16°, azimuth of 245°. The slant range was 1,070 km. During the early portion of the track, the image was starlike in appearance and showed the typical short spectrum. As the track progressed, the image assumed the shape observed in previous observations with an image orthicon camera. Minimum slant range was 460 km at 01 37 20 UT at an elevation angle of 53° and azimuth of 172°. The most interesting features were the shadow effects as the

satellite tumbled. The engines and smaller diameter nose would alternately appear and disappear. The black band painted around the base of the smaller cylindrical section was visible. The most surprising feature was the very light blue appearance of the smaller cylindrical section.

The second satellite observed on 6 March 1966 was Spadat Object No. 2070, 6615, Cosmos 110, the Russian vehicle carrying two dogs. The satellite was acquired at 01 55 56 UT at an elevation angle of 16°, azimuth of 240°, and a slant range of 1,060 km. The satellite had a starlike appearance and maintained this appearance throughout the transit. Minimum slant range of 400 km occurred at 01 58 25 UT at an elevation of 54°, azimuth of 319°. The satellite appeared to slowly increase in brightness as the transit progressed. No short-term fluctuations in brightness were noted. The satellite entered the shadow at 02 00 10 UT at a slant range of 890 km at an elevation of 16°, azimuth of 32°.

The third satellite observed on 6 March 1966 was Spadat Object No. 1085, 65091, Pegasus A, a meteorite detection experiment with wing panels 96 ft long and 13 ft wide attached to an S4 stage. Acquisition was accomplished at 02 06 10 UT at a slant range of 1,450 km, elevation angle of 16° and azimuth of 265°. The short spectrum gradually changed into the S4 stage as the satellite moved toward culmination. The meteorite detection panels were invisible until just before culmination which occurred at 02 09 14 UT at a slant range of 690 km with an elevation angle of 49°, azimuth of 194°. The satellite was a brilliant white with no sign of the bluish color noted on 744. Very slow rotation about the longitudinal axis was noted. Past midpoint, the satellite had the appearance we have frequently recorded with an image orthicon camera. The satellite disappeared into the earth's shadow at 02 11 20 UT; slant range was 1,105 km at an elevation angle of 23° and azimuth of 130°.

Four satellites were observed on 8 March 1966. The first was No. 744 acquired at 01 53 47 UT at a slant range of 1,050 km; elevation/azimuth was 16°/263°. The image assumed shape as slant range decreased. Again it was most interesting to observe the shadow variations on the cylindrical and conical elements. In one attitude the satellite appeared to be split in two by a shadow. The nose cylinder showed the same bluish cast noted previously. The paint in the vicinity of the engines was brownish. Minimum slant range was 410 km at 01 56 05 UT; elevation/azimuth was 65°/183°. The direction of the tumble appeared to be aligned approximately with the earth's axis and motion was in the same direction. The satellite disappeared into the earth's shadow at 01 57 52 UT at a slant range of 860 km. No anticipated reddening was noted.

The second satellite observed was Spadat Object No. 049, 60091, Echo I. This object is an aluminized polyethelene balloon 100 ft in diameter. Since its

launch in August 1960, it has lost its shape to some degree. Visual observations at Sulphur Grove (near WPAFB, Ohio) in 1961 using both 5-in. and 12-in. telescopes failed to show other than a starlike image. The purpose in observing Echo I was twofold: to determine if a more powerful telescope could resolve the balloon and to compare the image appearance with Satellite 2070 observed previously. The satellite was acquired at 02 03 00 UT at a slant range of 2,750 km; elevation/azimuth was 17°/240°. The image was starlike in appearance throughout most of the transit. In the region of culmination, brownish blobs appeared alternately on one side then the other of the almost dazzling solar reflection from the aluminized surface. Minimum slant range was 1,450 km at an elevation of 76°, azimuth of 326°. Calculation shows the balloon subtends 4 sec of arc. This dimension is easily resolved with the eyepiece used. [A later satellite (1050) subtended the same angle and was resolved.] But the brilliance of the central image made positive identification of the shape impossible. The iris was varied in an attempt to increase visibility but no improvement was noted. The extrafocal images were starlike. This was the same effect noted with satellite 2070. When this satellite is observed again, higher magnification should be used to reduce the apparent brightness. The transit was terminated shortly past mid-point in order to acquire the next satellite.

The third satellite observed was Spadat Object No. 1088, 65092, Apollo B. This satellite has the same shape and size as the cylindrical/conical nose section of satellite 744. It was observed to determine if it had the bluish appearance of the nose section of 744 and to see if it could be clearly resolved. Acquisition was accomplished at 02 20 00 UT at a slant range of 1,480 km; elevation/azimuth of 16°/254°. Throughout the entire transit, the image was a "soft" rectangle in shape with the atmosphere prism contributing color at each end. The edges were not sharp enough to permit a positive statement that it was resolved. No tumbling motion was detected in accord with previous observations with image orthicon photography. The central portion of the image did not appear bluish. Minimum slant range was 980 km at 02 22 42 UT; elevation/azimuth was 29°/200°. The satellite entered the shadow at 02 24 32 UT at a slant range of 1,250 km; elevation/azimuth was 20°/157°. No reddening was noted.

Satellite 1085 was acquired at 02 36 05 UT at a slant range of 1,460 km; elevation/azimuth was 16°/240°. The image changed from a white blob at acquisition to a T shape as the transit progressed. Maximum image dimension was 4 sec of arc. Minimum slant range was 1,220 km at 02 38 07 UT; elevation/azimuth was 21°/203°. Tracking continued until entry into the earth's shadow at a slant range of 1,330 km at 02 39 25 UT; elevation/azimuth was 18°/178°.

COMMENTS

The tracking performance of the mount was excellent. Tracking rates ranged from 10 sec of arc per sec to 1.2° per sec. Minor deviations about the line of sight were easily corrected by the off-set track capability of the 15-in. guide telescope and tracker. The satellite was always in the field of view after automatic tracking was initiated.

The eyepiece field stop is 3.9 mm in diameter. The focal length of the telescope is 7918.5 mm. The field of view is $3.9/7918.5 = 4.95 \times 10^{-4}$ rad or 100 sec of arc. The magnification is 7918/6 = 1320X. The limiting resolution of the 48-in. telescope is about 0.1 sec of arc. An object element subtending this angle will appear to subtend 0.1 x 1320 = 132 sec of arc or 2 min, a factor of 2 larger than the limiting resolution of the eye. "Seeing" conditions on the two nights were about equal – perhaps 0.7 to 1 sec of arc. Edge sharpness and surface detail observations of satellites as well as the Jupiter observations tend to confirm this.

After the satellite observations were completed on 8 March 1966, several stars, Jupiter, and the Orion nebula were observed. Sirius and its companion were easily resolved. The primary problem here is one of differential brightness – 10.6 magnitudes. The glare from the primary tends to obscure the companion.

Other stars were observed with eyepieces of varying focal length. A "zoom" eyepiece was tested and is considered unsatisfactory for good visual work. Its field is relatively small, it requires readjustment of focus at each focal length and the images did not appear to be of the same quality as with the Unitron eyepieces.

The intensity of the shadow bands in the extra-focal image was weak with relatively long striations in the direction of motion. This would be expected from the quality of "seeing" during satellite observations.

RECOMMENDATIONS

Visual observations, with appropriate precautions, can make full use of the sensitivity and wide dynamic range of the eye. Limiting resolution of the eye can be minimized or, in this experiment, can be matched to the performance of the optical system. Perhaps the most important characteristic of the eye for these tests was the dynamic range capability, since the illumination level was no problem and choice of eyepiece permitted over-matching the eye to the optical system. The next step is to equip the Newtonian position with a reflex camera, framing or cine, that can be operated by an observer on the telescope. The observer must be able to see the image for guiding corrections and to select exposure and frames for photography and to check focus. Photography with a format smaller than

70 mm will require careful guiding. With an observer guiding the telescope, a 16-mm format should prove quite adequate. The Newtonian focus should be re-imaged to f/16 or f/32 depending on the film speed.

APPENDIX 11

PHOTOGRAPHY OF SATELLITES: SOME NOTES ON OPTIMUM IMAGE SIZE AND DENSITY

J. S. Courtney-Pratt

INTRODUCTION

There are several factors to consider. Assume that we hoped to be able to resolve to p sec of arc in an extended image of the satellite (p may be, e.g., 1/10). The image ought to be big enough to avoid undue degradation of resolution in the photographic emulsion. This involves the dropping modulation transfer function (MTF) as spatial frequencies increase and the rise in noise as smaller and smaller areas of the emulsion are considered.

A long effective focal length F_T in the telescope gives a larger image on the emulsion, but it is less bright in proportion to F_T^2 and so longer exposures are necessary. Fast emulsions are coarse and grainy. Long exposures will be less sharp than very short exposures as they include the time averaging of the image fluctuations due to the atmospheric turbulence. They also are smeared because of object rotation and inaccuracies in tracking. If one uses a large effective focal length, the lateral jumping of the image due to inaccuracies in tracking is large and one may need a larger format in the camera.

If in processing the image, one increases the amplitudes of recorded high spatial frequencies, then, at the same time, one increases the over-all granularity noise (but not proportionately). At very low film densities, the rate of change of density with illumination is small. At very high film densities, the rate of change of density with illumination is again small. Between these there is a maximum which, in practice, lies near the low-density end of the range. The optimum

mean density for any chosen film is a function of the spatial frequencies involved as well as of the processing conditions.

PHOTOGRAPHIC DETECTION*

In the paper by R. Clark Jones[2] entitled "On the Minimum Energy Detectable by Photographic Materials," data are given on the energy detectivity and contrast detectivity of four Eastman Kodak films** — Royal-X (emulsion No. 6128), Tri-X (5233), Plus-X (4231), and Panatomic-X (5240, now discontinued). The minimum energy that can be detected by any detector involves not only the properties of the detector but also the reliability of the detection and the number of permitted false alarms. A simple concept is discussed in Clark Jones' paper that involves only the properties of the detector — the Noise Equivalent Energy, E_n, which is defined as the amount of signal energy that produces a density increment equal to the root-mean-square density fluctuation of the granularity. E_n is a function of the mean density to which the emulsion is exposed, and also a function of the size of the region on which the signal falls. E_n is shown to be proportional to the diameter of a circular signal-receiving region of area A_o (or to the square root of the area, if the region on which the signal falls is not circular) provided both lateral dimensions of the region are large compared with ℓ_p, where ℓ_p is a length defined in terms of the point-spread function of the film. (ℓ_p is about equal to the width of the finest point or line image that the emulsion will record or is about equal to the separation of two lines that can just be resolved by the emulsion.)

The minimum detectable signal — or rather the Noise Equivalent Energy, E_n — is obviously a function of the background density. For a very lightly exposed film, the rate of increase of density with exposure is small, and so small-signal exposures near the toe of the density/log exposure curve are not easily detectable. Furthermore, additional exposure when the film has been exposed almost to saturation produces little detectable effect. Between these low limits lies a maximum. The maximum is fairly steeply peaked for some emulsions, but not for others.

---

*This and some subsequent sections follow the general arguments developed in reference 1, though changed where appropriate, as here we are considering satellite images that are resolvable in the telescope whereas there we were considering unresolved or point satellite images against a bright background.

**Although this paper was published some 8 years ago, it contains the only self-consistent published data of this form that I know of. In recent years there have, of course, been advances in photographic emulsions. It would be most advantageous if the emulsion manufacturers would provide data of the kind R. Clark Jones has discussed in his paper for the newest emulsions so that quantitative comparisons, and assessments of the relative advantages of particular new emulsions, could be made using the method described in the present paper.

In the paper cited,[2] curves and tables are given for the four films of density D versus log exposure, $\log_{10} U_a$, where U_a is in erg cm^{-2}; and of the root-mean-square density fluctuation σ for a given signal area versus density D. These may be used directly. However, Clark Jones has also computed and given various other functions in graphical form. Some of these are in reference 2; others are in the two companion papers.[3,4]

These figures were all obtained by exposing the films to a light of a narrow band of wavelengths near 4,300 Å. Graphs are also given of the sensitivity versus wavelength. The films are panchromatic. Thus, to a rough approximation, the sensitivity is independent of wavelength over a reasonable wavelength range. In the calculations which follow, it has been assumed that the sensitivity is constant over the wavelength range 3,500 Å to 6,500 Å, and equal to the value at 4,300 Å.

If further it is assumed that a change of wavelength merely shifts the D versus log U curves sideways without changing their shape, and that the granularity at a given density is independent of the wavelength, then the Noise Equivalent Energy, E_n, will also be independent of wavelength in the range 3,500 Å to 6,500 Å.

For a given signal region of area A, where the root-mean-square density fluctuation is σ, the energy increment that would give a density increment equal to σ is obviously $\sigma A dU/dD$;

$$\therefore \quad E_n = \sigma A \, (dU/dD) . \tag{1}$$

If a film is exposed to background illumination and to a signal image, the signal will increase the density of the developed film in that region. If a photoelectric scanning system is used to scan the developed film, its output will contain a signal pulse as the scan aperture traverses the signal region, and will contain noise due to the granularity. For signal regions of significant lateral dimensions (i.e., greater than ℓ_p), the ratio of peak signal to root-mean-square noise will have a maximum when the dimensions (and azimuth) of the scanning area are equal to those of the signal region. That is, for optimum photometer detection of the signal recorded on a film when the signal region is a circle of diameter i ($i > \ell_p$), the scan aperture of the photometer should also be a circle of the diameter i. Similarly, for a rectangular signal area of dimensions j, k ($j > \ell_p$, $k > \ell_p$) the scan aperture should also be a rectangle of dimensions j, k and should be similarly aligned in azimuth in the film plane.

PROBABILITY OF DETECTION

As stated above, the minimum energy that can be detected by any detector involves not only the properties of the detector but also the reliability of the detection and the number of permitted false alarms. The term used above – the Noise Equivalent Energy, E_n – is defined as the amount of signal energy that produces a density increment equal to the root-mean-square density fluctuation of the granularity, σ. Suppose that the variation of density from the mean is of approximately a Gaussian distribution. Suppose it is known that the signal, if it occurs at all, must fall within a particular region Q'. Suppose a density increment $\Delta D'$ is measured in the region Q', and that for example $\Delta D' = 0.67\sigma$; then the chances are 50:50 that this is a random variation. If, however, $\Delta D' = 2.0\sigma$, the probability is 21:1 against this being a random variation.

It might, therefore, be thought that a criterion could be set, and that it could be said that the chances are 20:1 that the signal is real if the density increment $\Delta D = 2.0\sigma$, and 1,000:1 if $\Delta D = 3.3\sigma$. This would be reasonable enough if the coordinates on the film of the signal region Q' were known exactly.

Suppose now that the position of the signal region Q' is not known and that the area of the exposed film within which a search must be made for a signal is $\overline{A}$, where $\overline{A}/A$ is a large number, e.g., 10^5. Consider these 10^5 regions of area A. The chances are high that in one or more of these a density deviation from the mean could be found that was greater than 2.0σ. To achieve a reliability of detection of 20:1 when the signal may occur in any of 10^5 places requires a density deviation ΔD greater than 5.0σ. Similarly, for a reliability of detection of 1,000:1 when there may be 10^6 places to search, ΔD must be greater than 6σ. It will be noted that the reliability of detection rises rapidly as P increases above 6, where P is the ratio of $\Delta D/\sigma$.

OPTIMUM EXPOSURE DENSITY

Consider a satellite of mean brightness (luminance) B and a small piece, P_s, of that satellite of brightness B + δB and of such a size that it is resolvable with the telescope that we are using.*

*The analysis in this paper aims at optimizing the exposure for detection of small differences in contrast of small neighboring parts of the satellite itself. That is, we are hoping to detect the small piece P_s of the satellite of brightness B + δB against the remainder of the satellite of brightness B.

At other times we might well wish to optimize exposure for detection of a small piece of a satellite P'_s (e.g., a boom or vane) that projected well out from the body of the satellite. Then, the background brightness to be considered would be the brightness of the sky B_{sky} rather than the satellite brightness B.

For a matte white satellite, B will be about 3.5 candle $cm^{-2} = 1.7 \times 10^5$ erg sec^{-1} cm^{-2} per unit solid angle (in the wavelength range 3,500 to 6,500 Å). *

Let

F_T = effective focal length of telescope in cm,

d_T = effective diameter of telescope in cm,

T_T = transmission factor of telescope,

R = range of satellite in cm,

d_Σ = diameter of image of satellite in cm,

d_S = diameter of image of the small piece P_s of the satellite in cm,

t = exposure time in sec,

U = energy per unit area falling in the extended image of the satellite in erg cm^{-2}, and

δU = energy per unit area falling in the extended image of the piece P_s of the satellite in erg cm^{-2}.

The area of the image of the piece P_s of the satellite is

$$A = \frac{\pi}{4} d_S^2 . \qquad (2)$$

Now the root-mean-square fluctuation of density σ varies inversely as $\sqrt{A}$,

$$\therefore \quad \sigma = \sigma_{10} \sqrt{\frac{A_{10}}{A}} \qquad (3)$$

where the subscript 10 refers to a circular signal (and scan) region 10 μ in diameter, i.e.,

$$\sigma = \sigma_{10} \left[\frac{\frac{\pi}{4} \cdot 10^{-6}}{\frac{\pi}{4} \cdot d_S^2} \right]^{\frac{1}{2}} = \sigma_{10} \cdot \frac{10^{-3}}{d_S} . \qquad (4)$$

---

*Note that this figure is considerably greater than the brightness of the full moon as the albedo of the moon is so low. B, for the full moon, is about 0.25 candle $cm^{-2} \doteq 1.2 \times 10^4$ erg sec^{-1} cm^{-2} per unit solid angle.

Many satellites will not be predominantly white and their brightness B will be several times smaller than the figure for white satellites. We will, in the subsequent calculations, take B = 1 candle $cm^{-2} \doteq 5 \times 10^4$ erg sec^{-1} cm^{-2} per unit solid angle.

As shown in Eq. (1), the Noise Equivalent Energy $E_n = A\sigma dU/dD$, i.e.,

$$E_n = \frac{\pi}{4} d_S^2 \frac{10^{-3}}{d_S} \sigma_{10} \frac{dU}{dD}$$

$$= \frac{\pi}{4} \cdot 10^{-3} d_S \cdot \sigma_{10} \frac{dU}{dD} \quad . \tag{5}$$

From geometrical considerations we can write down

$$U = B \cdot \frac{\pi}{4} d_T^2 \cdot \frac{T_T}{F_T^2} \cdot t \qquad \text{erg cm}^{-2} \tag{6}$$

and

$$\delta U = \delta B \cdot \frac{\pi}{4} d_T^2 \cdot \frac{T_T}{F_T^2} \cdot t \qquad \text{erg cm}^{-2}. \tag{7}$$

As stated above, for reasonable certainty of detection, the signal δU must be several times greater than the Noise Equivalent Energy E_n. That is, $\delta U \geq PE_n$. For minimum detectable signal, take $P = P_o$, i.e.,

$$\delta U = P_o E_n \, . \tag{8}$$

Substituting in Eq. (8) from Eqs. (5) and (7):

$$\delta B \frac{\pi}{4} d_T^2 \frac{T_T}{F_T^2} t = P_o \frac{\pi}{4} 10^{-3} d_S \sigma_{10} \frac{dU}{dD} \, . \tag{9}$$

Rewriting this,

$$\delta B = \frac{10^{-3} P_o d_S F_T^2}{T_T d_T^2} \cdot \left(\sigma_{10} \frac{dU}{dD}\right) \cdot \frac{1}{t} \, . \tag{10}$$

So the smallest relative increment in brightness $\delta B/B$ that can be detected is given by:

$$\frac{\delta B}{B} = \frac{\pi}{4} \cdot 10^{-3} P_o d_S \left(\frac{\sigma_{10}}{U} \cdot \frac{dU}{dD} \right) . \tag{11}$$

To minimize $\delta B / B$, i.e., to be able to detect the minimum changes in contrast or reflectance of the satellite, we should minimize the term enclosed in brackets, i.e., maximize J where we define

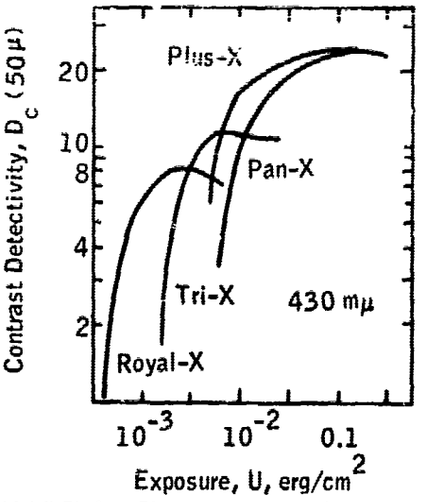


FIGURE 1 The variation of contrast detectivity D_c of the four films versus the background exposure U in erg/cm^2 (for radiation of $\lambda = 4,300$ Å). The curves are for a signal region 50 μ in diameter.

$$J = \left[\frac{U}{\sigma_{10}} \cdot \frac{dD}{dU} \right], \quad (12)$$

Clark Jones[4] has plotted a function he calls "Contrast Detectivity" for a number of emulsions. His function $D_c(50\ \mu)$ is actually just 5 times the term J above; i.e., $D_c(50\ \mu) = 5J$, and so has maxima for the different emulsions at the same values of exposure. Figure 1 is copied from his paper. Plus-X can be seen to be better than Royal-X by about a factor of 3; though the exposure U at which it is best is some 40 times greater.

We should consider for a moment the value of d_S, which also comes into Eq. (11). Clearly we gain by reducing it as far as possible, i.e., by using as small a value of F_T as we can while still keeping d_S above the value of the line spread function, ℓ_p.

Table 1 gives the emulsion parameters corresponding to the exposures that give maximum values of the contrast detectivity, D_c. Approximate values are also given for the line-spread function assuming that it is equal to the reciprocal of the spatial frequency at which the modulation transfer function is 50 percent.

TABLE 1 Emulsion Parameters Corresponding to the Exposures that Give Maximum Values of the Contrast Detectivity, D_c
(Taken from the paper by R. Clark Jones)

Film	$\log_{10} U_a$	U_a (erg cm^{-2})	D_c (50 μ) maximum values	D	σ_{10}	U_a^{-1} (erg^{-1} cm^2)	$\frac{\Delta U}{\Delta D}$	$\frac{\sigma_{10}}{U} \frac{dU}{dD}$ (= $5/D_c$)	ℓ_p approx (mm)
Royal-X	-2.57	$2.49 \cdot 10^{-3}$	8.1	0.410	0.1676	$3.72 \cdot 10^{2}$	$0.99 \cdot 10^{-2}$	$6.17 \cdot 10^{-1}$	1/30
Tri-X	-2.16	$6.92 \cdot 10^{-3}$	11.3	0.247	0.0920	$1.44 \cdot 10^{2}$	$3.44 \cdot 10^{-2}$	$4.57 \cdot 10^{-1}$	——
Plus-X	-1.00	$1.00 \cdot 10^{-1}$	24.3	0.718	0.0703	1.00 10	$29.2 \cdot 10^{-2}$	$2.05 \cdot 10^{-1}$	1/60
Pan-X	-0.75	$1.78 \cdot 10^{-1}$	23.3	0.565	0.0607	5.62	$62.5 \cdot 10^{-2}$	$2.13 \cdot 10^{-1}$	——

We can use these values for any given telescope parameters and find (for the emulsions listed) the optimum exposures, assuming that the image does not degrade as exposure is increased.

Actually of course, the image does increase in blur as we increase the exposure time because of object rotation, because of imperfections in tracking, and because of atmospheric turbulence. For these reasons, it may sometimes be appropriate to use exposure times shorter than would be indicated by the above analysis or to use any one of the methods of electronic image intensification.

Examples

1. Consider, as an example, photographing a satellite (as Tyson has done) with the 48-in. aperture telescope at Cloudcroft, using Plus-X film. The effective focal length in his records was 900 in. The size on the film of an image 1 sec of arc across was 0.11 mm. Most film can resolve 9 cycles/mm easily. Suppose, however, we hope to find detail (at least sometimes) of 0.1 sec of arc (i.e., p = 0.1). We then are hoping to find detail at 90 cycles/mm. The MTF of Plus-X has then fallen to about 40 percent. This is not too serious, though if p were 0.05, the MTF would have dropped to a few percent and detection of low-contrast features in the object would be impossible.

The noise can be a serious problem. The mean density for Plus-X for optimum resolution is about 0.7, and we will assume that the exposure time t was adjusted to give this density. From Eq. (6),

$$t = \frac{U F_T^2}{B \cdot \frac{\pi}{4} d_T^2 T_T} . \qquad (13)$$

So with $U = 1.0 \times 10^{-1}$ erg cm^{-2}, $F_T = 900$ in., $B = 5 \times 10^5$ erg sec^{-1} cm^{-2} per unit solid angle, $d_T = 48$ in. and $T_T = 1/4$, $t = 3.5 \times 10^{-3}$ sec.

For Plus-X at D = 0.7, $\sigma_{10} = 0.07$. Thus for 90 cycles/mm, $\sigma \doteq 0.063$. The signal-to-noise (S/N) ratio is adequate at $D/\sigma = 0.7/0.063 \doteq 11$. However, if we consider processing the image in any way that increases the relative amplitude of the high spatial frequencies, we increase the effective over-all noise and reduce the S/N ratio somewhat.

Thus we see that the choice of effective focal length $F_T = 900$ in., and Plus-X film was in this case reasonable enough if we had hoped eventually to reach a resolution of about 0.1 sec of arc, but the magnification would have been inadequate if we had hoped eventually to reach a resolution of 1/20 sec of arc. The actual value of the exposure time that Tyson used was 1/60 sec, about 5 times greater

than the optimum suggested here, and hence his images were often very dense (though this allowed considerably greater dynamic range). For Plus-X, the maximum of contrast detectivity is not very sharply peaked and so the longer exposure is not too far from optimum.

If very long dynamic range, in particular cases, is important, it would be better to avoid high densities and keep near the optimum exposure. This could be done in one of several ways: (a) One could use one (or more) beam splitters, and take two (or more) adjacent photographs with exposure ratios of, say, 10 or 20 to 1. (b) One could use XR film, or a modified version of this with more overlap between the H-D curves of the various emulsion layers. As an image becomes too dense in a fast layer, the exposure would rise toward optimum in the next slower layer.

If the problem is the considerable variation in the apparent brightness from one frame to the next, rather than small highlights within any one image (as is sometimes the case), one would do better to use an electronically adjusted exposure time rather than set the exposures for all pictures of the sequence at a figure sufficient to record the faint images — as this obviously leads to overexposure for bright members of the set.

2. Consider the use of a fast film such as Royal-X. At the same value of F_T (900 in.) we could have decreased the exposure time by a factor of 40 and might have thought this sufficient justification for the choice of the coarser but faster film. However, the S/N ratio turns out much worse. We would need to increase the effective focal length (working with Royal-X and aiming to reach a resolution of 0.1 sec of arc) to 3,600 in.

Royal-X film and F_T = 900 in.	Royal-X film and F_T = 3,600 in.
MTF for Royal-X at 90 c/mm $\doteq$ 5%	MTF for Royal-X at 22 c/mm $\doteq$ 60%
Optimum density for Royal-X = 0.4	Optimum density = 0.4
Exposure time t for this density = 90 μsec	Exposure time t for this density = 1.5 msec
σ_{10} for Royal-X = 0.168 at D = 0.4	σ_{10} = 0.168 at D = 0.4
$\therefore$ σ at 90 c/mm = 0.15 (for p = 0.1)	$\therefore$ σ at 22 c/mm = 0.038 (also for p = 0.1)
S/N ratio (D/σ) = 2.5, which is too low for distinguishing even high-contrast features	S/N ratio (D/σ) = 11, which is the same as for Plus-X with a focal length of 900 in.

Had we changed from Plus-X and a focal length of 900 in. to Royal-X and a focal length of 3,600 in., we would not have been able to decrease the exposure time by the factor of 40, which is the ratio of their optimum exposure energies,

but only by the ratio 40/16 = 2.5. It is doubtful whether it would be worth changing to the faster film for so small a gain, especially as the problems of keeping the image reasonably steady in the format would have been considerably worsened. Moreover, the sensitivity to detection of contrast is also worse for Royal-X than for Plus-X by almost a factor of 3, as can be seen directly from Figure 1.

GREGORY INTEGRATION OF EXPOSURES

Up to this point in this paper, the analysis applies to any passive system of photography where the magnification and resolution are such that the satellite gives an extended image at the plane of the recording emulsion.

R. L. Gregory has proposed, in the first section of Chapter V of Volume 1 of this report, a system that will, in principle, obviate many of the degradations due to imperfections in tracking and due to the image distortions introduced by atmospheric turbulence. (We will later consider the problem of object rotation.)

His method integrates a number of short exposures chosen automatically at moments of low distortion and low displacement. If there were no reciprocity failure and no intermittency effect, the sum of the exposure times, in his system, should equal the exposure time suggested by the analysis in the preceding pages.

We have measured the frequency spectra of scintillation of stars and planets and have found no significant amplitudes at frequencies above 1,000 cps. It would thus seem that there would be no necessity to reduce the individual exposures in the Gregory system below, say, 10^{-4} (or at most 10^{-5}) sec. Reciprocity failure effects and intermittency effects will thus be small; and we can, to a reasonable approximation, use the analysis above to compute the total integrated time of exposure that would be optimum.

We do not know what the optimum duty cycle should be for Gregory's method; but if, at a guess, it were say 1 percent or 0.1 percent, then the over-all time for which the device should integrate appropriate exposures would be 100 t or 1,000 t, respectively.

These times might well be longer than the time for which one could consider the object to present a stationary aspect. Some satellites rotate slowly and move slowly across the sky — but others have a high rate of change. In these latter cases, we might need either to select the exposures for Gregory integration during similar short periods in each cycle of rotation of the satellite or to alter the duty cycle to a value somewhat worse than the optimum for stationary objects, so that there could be sufficient integration within the available time.

For example, with a satellite rolling at say 1 revolution per second, we might reasonably integrate exposures anywhere within say 1/30 of a revolution, or 33 msec. If the required total exposure were 3.5 msec, we still could work

with a duty cycle of 1:10, which might well give significantly improved images. However, with a satellite rolling four or five times faster, the time available for integration in any one revolution would be too short, and one would need to select similar periods from successive revolutions.

CONCLUSION

We have presented a method for the computation of optimum exposure times for detection of small differences of contrast in small extended images of satellites.

The exposure times with a telescope of 48-in. aperture diameter can be surprisingly short. High framing rates are therefore quite possible. It would seem worthwhile to take some fast sequences of pictures with millisecond exposure times and to examine the results. Some, at least, of the pictures with such short exposures may have avoided the degradation due to atmospheric turbulence.

It should be relatively simple to set up equipment for Gregory Integration of exposures. Exposure times of the duration suggested in this paper could easily be produced by summing exceedingly brief exposures. Even if exposure densities considerably greater than those suggested by the analysis in this paper should prove to be desirable (because, for example, of interest in long dynamic range rather than optimum detection), Gregory Integration could easily achieve the desired values, without needing excessive numbers of exposures and without difficult shuttering requirements.

REFERENCES

1. J.S. Courtney-Pratt, "A Note on the Possibility of Photographing a Satellite Near the Moon," J. Phot. Sci. 9, 36 (1961).

2. R. Clark Jones, "On the Minimum Energy Detectable by Photographic Materials: Part 2 - Results for Four Current Kodak Films," Phot. Sci. Eng. 2 (4), 191 (Dec. 1958).

3. R. Clark Jones, "On the Quantum Efficiency of Photographic Negatives," Phot. Sci. Eng. 2 (2), 57 (Aug. 1958).

4. R. Clark Jones, "On the Minimum Energy Detectable by Photographic Materials: Part 3 - Energy Incident on a Microscopic Area of Film," Phot. Sci. Eng. 2, (4) 198 (Dec. 1958).

APPENDIX 12

ON STATISTICAL ESTIMATION OF INCOHERENTLY ILLUMINATED OBJECTS

Robert S. Kennedy

The relevance of statistical estimation theory to the restoration of incoherent images degraded by the atmosphere is the subject of this paper. Although the scope, and more importantly, the depth of the remarks reflect the limited time that has been available for the study, they do suggest that statistical techniques may be quite relevant and valuable when the task is either to restore images of many different objects or is to obtain a single image from a large quantity of raw data, e. g., from many degraded images of the same object. It also appears that the applicability of the techniques to orbital objects will be principally limited by the uncertainties introduced by object motion; were it not for this motion, there would be no theoretical limit to the quality of the restored image. The principal conclusion is that the techniques of statistical estimation theory are sufficiently promising to warrant a further analysis employing a more realistic model of atmospheric degradation and focusing more upon the combined limitations imposed by the effects of object motion and of the atmosphere. In the present summer study, these effects have only been considered separately.

Our approach has been to first determine the class of image restoration problems to which statistical theories are conceptually and philosophically applicable, and then to examine the structure and performance of the processors that

result from these theories. A variety of simplifying assumptions are invoked in the quantitative phases of the inquiry, but the conclusions do not appear to be critically dependent upon them.

For example, if one supposes that the instantaneous point-spread function of the atmosphere is known, and that the processor noise is Gaussian of mean zero, then the maximum likelihood estimator of the object is identical to that linear estimator which yields a zero average error and a minimum mean-square error. Moreover, this estimator does not differ substantially from that linear estimator which minimizes the mean-square error without constraining the average error to be zero. The processors that emerge from a preliminary consideration of unknown atmospheric point-spread functions also possess similarities, but the results are too incomplete to be conclusive.

In the interest of simplicity, the analysis is restricted to objects contained in an isoplanatic patch, but the extension to larger objects or to less isoplanatic atmospheres is straightforward. This does not imply that the difference between isoplanatic and nonisoplanatic problems is trivial; rather it implies that the difficulties lie in the implementation and performance of the resulting system rather than in their analysis.

THE ROLE OF STATISTICAL TECHNIQUES

It appears that restoration problems can be grouped into three classes to which statistical techniques apply in varying degrees. At one extreme are those problems wherein the degraded images of a large number of different low-priority objects are to be restored. In this instance, highly structured processing systems which require a minimum of human intervention are in order. Moreover, in such quantity operations, one is prone to tolerate some variability in the quality of the restored images and is willing to speak of the average performance of the system. Statistical estimation theory is of course tailored to just this class of problems.

At the other extreme is a class of problems wherein one possesses a limited amount of raw data about a very high-priority object. In this instance, it is both reasonable and feasible to have the entire processing system under human supervision and there is no need to introduce any particular structure into the processor until the data have been collected. Moreover one is probably unwilling to speak of "average performance" or "average quality" of the restored image. Rather, one seeks to extract from the data all of the intelligence that is "humanly possible." Statistical theory has little to offer in such situations other than to suggest some very general and obvious processors that should be placed at the disposal of the human supervisor.

Between the extremes of the preceding paragraphs lies the situation wherein a great deal of redundant data is available about a very high-priority object. As the amount of such data increases, a point is reached at which the human mind boggles and some of the human supervision and control must be sacrificed to a more structured processor. Statistical estimation theory provides one means of suggesting structures in such applications. Although the effectiveness of these structures depends upon the character of the redundancy in the raw data, they will tend to become rather efficient in both relative and absolute terms whenever the data consist of many "independent" degraded images of a single fixed object.

Stated alternately, they will be effective if the data exhibit the kinds of "statistical regularity" that is typified by games of chance. This regularity, which appears as one considers longer and longer sequences of roughly independent events subject to the same conditions, permits one to make very strong statements as to some of the conditions to which the events are subject, e.g., the bias of a coin or the object that was observed. However, if the conditions themselves change in an unknown manner from event to event, the statements weaken considerably.

We mention this because the imprecisely known motion of an orbital object between photographs represents a change in 'conditions" which limits the strength of conclusions that can be deduced from the resulting sequence of photographs. This situation can be ameliorated to some degree by compensating for the image motion in the observation system or in the resulting processor. However, since the object itself is presumed to be unknown, its motion cannot be completely compensated for and a residual error will remain in the form of image degradation due to compensation. This clearly suggests that the restoration and motion compensation problems should be considered jointly. However, in the interests of simplicity, we focus here upon restoration in the absence of motion.

Aside from its contribution to the problem of processing data collected with existing systems, statistical theory can also be used to suggest and evaluate modifications that might improve the performance of the over-all system. This is particularly useful when one considers major innovations such as measurement of the instantaneous atmospheric point-spread function. Thus, as we do later, one can inquire into the value and best use of knowledge pertaining to the atmospheric point-spread function and into the relative advantages of short- and long-term photographic exposures. Such applications of statistical theory are particularly valuable because they sometimes serve to guide one's intuition beyond its initial boundaries.

The preceding discussion has been somewhat obvious at times, but it seemed best to set forth the perspective that underlies the remainder of the

paper. With that accomplished, we devote ourselves to a more quantitative treatment of statistical techniques in image restoration. As a first step in that direction, we establish an appropriate mathematical model for the degradation of the objects and choose some performance measures.

THE MODEL

We suppose that the viewed object is incoherently illuminated by white light and is completely characterized by an intensity function $o(\vec{r})$, where $\vec{r}$ is a position vector over the surface of the object. In the absence of any additive noise, this object gives rise to an intensity distribution, $i(\vec{r})$, at the image plane of an optical instrument. These two intensities are linearly related through the expression

$$i(\vec{r}) = \int h(\vec{r} - \vec{\rho})\, o(\vec{\rho})\, d\vec{\rho} \quad , \tag{1}$$

where $h(\cdot)$ is the combined intensity point-spread function of the atmospheric-optical system.

In writing the above expression, we have supposed that the object is contained within an isoplanatic patch of the atmospheric-optical system. If this is not true, $h(\vec{r} - \vec{\rho})$ must be replaced by a spatially variant point-spread function, $h(\vec{r}, \vec{\rho})$. The supposition is not a crucial one and is made only for expediency.

The simplification afforded by the isoplanatic assumption resides in the Fourier transform of $o(\cdot)$ and $i(\cdot)$. Specifically,

$$I(\vec{k}) = H(\vec{k})\, O(\vec{k}) \quad , \tag{2a}$$

where $\vec{k}$ denotes the spatial frequency vector and where $O(\cdot)$, $H(\cdot)$, and $I(\cdot)$ are, respectively, the Fourier transforms of $o(\cdot)$, $h(\cdot)$, and $i(\cdot)$, *e.g.*,

$$H(\vec{k}) = \int d\vec{r}\, h(\vec{r})\, \exp(-j2\pi\, \vec{k} \cdot \vec{r}) \, . \tag{2b}$$

Although it has not been explicitly indicated, the quantites of Eqs. (1) and (2) are, in general, functions of time because of the time variations of the atmosphere and, perhaps, of the object. We now consider the character of this time dependence for the frequency domain representation of Eq. (2); entirely analogous remarks apply to Eq. (1).

Suppose that $O(\vec{k})$ does not change with time so that the time variation of $I(\vec{k})$ is entirely due to the time variation of the atmospheric transfer function $H(\vec{k})$. If the data collector integrates the received intensity over the basic observation interval, as it almost invariably will, the recorded intensity and its transform

$I(\vec{k})$ will be related to the object transform $O(\vec{k})$ as indicated by Eq. (2) with $H(\vec{k})$ being the integrated transfer function of the atmosphere over the basic observation interval. Consequently, if $O(\vec{k})$ is fixed during the observation interval, time enters the problem only through the statistical description of $H(\vec{k})$. On the other hand, if the atmospheric transfer function does not change during the basic observation interval whereas the object does, the integrated value of $I(\vec{k})$ will be given by Eq. (2) with $O(\vec{k})$ being replaced by its time integral over the observation interval, i.e., the recorded image will be blurred due to object motion. More generally, if both object and atmospheric changes occur during the observation interval, each side of Eq. (2) must be integrated with respect to time over that interval and the resulting $I(\vec{k})$ will not be a product of integrals of $O(\vec{k})$ and $H(\vec{k})$.

In the sequel we suppose that there is no object motion, so that the image is related to the object through Eq. (2) with $H(\vec{k})$ being the time integral of the instantaneous atmospheric transfer function over the observation interval. This supposition is a restrictive one since object motion is important for orbital objects; indeed the supposition will lead us to conclude that long observation periods should be employed — periods so long that object motion will most assuredly occur. Unfortunately, the restriction has been dictated by the limited time available for the study.

Thus far, the presence of recording and processing noise has been ignored. In many existing systems such noise is of secondary importance, but there are statistical as well as physical arguments which suggest that it should be as important as atmospheric effects in a well-designed system. Stated alternately, one can ameliorate atmospheric degradation by resorting to techniques which tend to increase "processing noise," and these techniques are profitable until the noise becomes as bothersome as the atmospheric degradation.

The detailed character of the processing noise will depend upon the particular recording mechanism employed; but to a first approximation it may be supposed to be a zero-mean white Gaussian process which is added to the recorded intensity $i(\vec{r})$. In the frequency domain, this noise will also appear as a noise, $N(\vec{k})$, added to $I(\vec{k})$, i.e.,

$$I(\vec{k}) = O(\vec{k})\ H(\vec{k}) + N(\vec{k}). \tag{3}$$

Moreover, the real and imaginary parts of this noise will, in a half space (say $k_1 \geq 0$), comprise zero-mean white Gaussian random processes which are independent of each other. The process in the remaining half space is the complex conjugate of the given half-space process.

It is worth noting that the transformation of the white noise into the frequency domain constitutes an abuse of the mathematics, as indeed does the use of white noise. However, the resultant model is a plausible one and can be placed upon a firm mathematical basis by appropriately limiting the bandwidth of the original noise spectrum. This is the first of many instances in which we will use sloppy mathematics to obtain results which are generally agreed to be valid.

It remains to generalize the model to allow for sequences of observations. That is, we suppose that one can collect a sequence of images or, equivalently, a sequence of the transforms, $I_i(\vec{k})$, $i = 1, \ldots n$, where each such transform is given by the expression

$$I_i(\vec{k}) = O(\vec{k})\, H_i(\vec{k}) + N_i(\vec{k}), \qquad i = 1, \ldots, n. \tag{4}$$

We employ the very reasonable assumption that the noises added to the various transforms are statistically independent of each other. We also suppose that the atmospheric transfer functions, $H_i(\vec{k})$, associated with the various observations are statistically independent of each other and are identically distributed. This assumption is not a crucial one but it does lead to a substantial simplification of the ensuing analysis. Moreover it is often satisfied in practice.

We have supposed in the writing of Eq. (4) that $O(\vec{k})$ does not change from observation to observation, i.e., the target is presumed to be stationary during the complete observation sequence. This condition will invariably be violated for an orbital object as the total observation interval increases, so there is again the conflict between atmospheric degradation on the one hand and object motion on the other. It is possible to allow for some such motion, e.g., translation, in the subsequent analysis, but the complication is unwarranted in this preliminary investigation.

To complete the model, it is necessary to specify the statistical properties of the $H_i(\vec{k})$. There is some reason to believe that they comprise either Gaussian random processes or log normal processes, but neither assumption can yet be considered valid. For the most part, only first and second moments are required in the sequel so the model will be left in its present uncompleted form. The development of a complete statistical model is, however, a prerequisite to more comprehensive analysis.

PERFORMANCE MEASURES

The choice of quantitative performance measures, difficult in any problem, is very nearly a black art when the human senses are involved. The "best" possible restored images are desired, yet "best" cannot be defined until the question to be answered from the image has been formulated, and this formulation often takes place after the image has been studied; hence the circle is closed. There are, in addition, performance principles which provide reasonable guidelines for system development without establishing a measure of performance; one such principle is that of maximum likelihood.

Integrated Squared Error

If the restored image of the object, $o(\vec{r})$, is denoted by $\hat{o}(\vec{r})$, the integrated squared error, ϵ, between $o(\vec{r})$ and $\hat{o}(\vec{r})$ is defined to be

$$\epsilon = \int d\vec{r} \left| o(\vec{r}) - \hat{o}(\vec{r}) \right|^2 . \tag{5a}$$

Alternately, by Parseval's theorem,

$$\epsilon = \int d\vec{k} \left| O(\vec{k}) - \hat{O}(\vec{k}) \right|^2 , \tag{5b}$$

where $O(\vec{k})$ and $\hat{O}(\vec{k})$ are, respectively, the Fourier transforms of $o(\vec{r})$ and $\hat{o}(\vec{r})$.

The integrated squared error is one of the most familiar, maligned, and abused performance measures in existence. Its merits are simplicity, analytical tractability, and the assurance that, if ϵ is zero, $o(\vec{r})$ and $\hat{o}(\vec{r})$ are identical. Its great limitation is its tenuous relationship to subjective picture quality, e.g., a reduction in ϵ does not always result in a better restored image. Some such limitations are inherent in any attempt to measure quality by a single number, but some are attributable to this particular measure. For example, a restored image with a few large errors may possess a larger value of ϵ than an image with many small errors, whereas subjectively the picture with a few large errors may be preferred. Or, for another example, the difference in contrast between the restored and original image may result in a large value of ϵ, whereas in reality such differences are rarely important.

On the other hand, there is little reason to believe that any single parameter measure of quality is really satisfactory for all applications and the integrated squared error is probably, by and large, no worse than any other. Moreover, there is some comfort in the observation that any system which minimizes the

integrated squared error between an object and a restored image will also minimize the integrated squared error between any derivative of the object and the derivative of the restored image.* Such derivatives are of interest, for example, if the object is known to be black and white.

Given that we choose to employ the integrated squared error as a measure of restored image quality, it remains to reduce this to a useful measure of system performance. Because of turbulence and noise, the restored image will in general be a random function for any given object; thus it is appropriate to regard ϵ as a random variable characterized by its conditional probability distribution given the object. This distribution function, given as a functional of $O(\vec{k})$, provides a very complete and useful description of ϵ for any given system. It is, however, often difficult to compute and even more difficult to optimize; therefore, we will employ a simpler description in spite of our belief that the distribution function of ϵ provides a much more meaningful description.

Specifically we shall employ $\bar{\epsilon}$, the conditional expected value of ϵ with respect to the atmospheric and noise fluctuations, given the object. By virtue of Eq. (5), this mean integrated squared error is equal to the integrated mean-square error, i.e.,

$$\bar{\epsilon} = \int d\vec{r} \left\{ \left| o(\vec{r}) - \hat{o}(\vec{r}) \right|^2 \right\}$$

and

$$\bar{\epsilon} = \int d\vec{k} \left\{ \left| O(\vec{k}) - \hat{O}(\vec{k}) \right|^2 \right\}, \tag{6}$$

where the super bar is employed to denote the expected value given the object.

It will also be convenient to speak of the conditional average error of the restored image given the object. If this conditional average error is zero for all values of $\vec{r}$ or $\vec{k}$ and for all objects of interest, the system is said to provide an unbiased estimate of the object. That is, a system is unbiased if, for every object of interest, the conditional expected value of $\hat{o}(\vec{r})$ equals $o(\vec{r})$ for all $\vec{r}$, or, equivalently, if the conditional expected value of $\hat{O}(\vec{k})$ equals $O(\vec{k})$ for all $\vec{k}$.

Although the conditional integrated mean-squared error may provide an adequate means of comparing specific systems, it is of less value as a guide for

*This is easily seen in the frequency domain, since differentiation appears there as multiplication by frequency.

system development and optimization. The difficulty is that any attempt to minimize $\bar{\epsilon}$, say, is likely to result in the conclusion "given that $O(\vec{k})$ was the object say that $O(\vec{k})$ was the object." This occurs because it is difficult to prevent the minimization procedure from exploiting the conditioning which underlines $\bar{\epsilon}$. This degeneracy can, in principle, be remedied by a variety of techniques which suppress the conditioning of the performance measure upon the object, e.g., by averaging $\bar{\epsilon}$ over an ensemble of objects. That is, one can consider

$$\bar{\bar{\epsilon}} = \int d\vec{k} \, \overline{\overline{\left| O(\vec{k}) - \hat{O}(\vec{k}) \right|^2}}, \tag{7}$$

where the double super bar has been employed to denote the average over the triple ensemble which comprises the object, the atmosphere, and the additive noise.

In the sequel, we will have occasion to consider the structure of systems that minimize $\bar{\bar{\epsilon}}$ subject to some restrictions. We do this, primarily, because of the comparative and supporting results obtained rather than because of any belief that $\bar{\bar{\epsilon}}$ is inherently an appropriate performance measure. One might, however, argue that it is appropriate to the low-priority bulk processing discussed above under The Role of Statistical Techniques.

One escape from the conceptual and analytical difficulties associated with an ensemble of objects and with $\bar{\bar{\epsilon}}$ is provided by the maximum likelihood principle. According to this principle, one presumes that the object was that one, among all possible objects, for which the conditional probability of the recorded data given the object is greatest, i.e., the estimate $\hat{O}(\vec{k})$ given $I(\vec{k})$ is chosen so that

$$p(I/\hat{O}) \geq p(I/O) \quad \text{for all } O \neq \hat{O}. \tag{8}$$

In this expression, $p(I/O)$ has been used to denote formally the conditional probability of $I(\vec{k})$ given $O(\vec{k})$. More precisely, it denotes the conditional probability density of the set of observables that are used to represent the recorded random process. We note in passing that it makes no difference whether one applies the maximum likelihood principle in the frequency domain or in the intensity domain; the resulting processors are identical.

There is no fundamental reason why one should employ a maximum likelihood estimator in perference to an estimator which, say, minimizes the integrated mean-squared error; both are rather arbitrary. However, the maximum likelihood estimator at least eliminates the need to specify an object ensemble and is also frequently feasible to implement when the minimum integrated mean-squared error estimator is not. Moreover, it is intimately related to the minimum

integrated mean-squared error estimator in an important asymptotic situation. Specifically, if in accord with Eq. (4), a sequence of n degraded images is used to obtain a maximum likelihood estimate of the object, the estimate will, under rather general conditions, converge to the minimum integrated mean-squared error estimate with increasing n. Precisely stated, the estimate will become unbiased, the conditional integrated mean-squared error will vanish, and for every object, the ratio of the conditional integrated mean-squared error to the minimum attainable integrated mean-squared error will approach unity.* This result, which is usually referred to as the asymptotic efficiency of the maximum likelihood estimator, provides a powerful argument for its use when a large number of independent pictures are available.

There is another reason for the use of a processor which bases its decisions upon the likelihoods, if not upon the maximum of those likelihoods. Specifically, if only a finite number of objects can occur and if the objective is to minimize the unconditional probability that an error is made in identifying which object is present, the optimum processor maximizes the a posteriori probability. That is, the processor presumes that the object is that one, say $\hat{O}(\vec{k})$, for which

$$p(\hat{O}/I) \geq p(O/I) \text{ for all } O \neq \hat{O} , \tag{9}$$

where we have again used a purely formal notation for the probability in question. Equivalently, the processor chooses $O(\vec{k})$ so that

$$p(I/\hat{O})\ p(\hat{O}) \geq p(I/O)\ p(O) \text{ for all } O \neq \hat{O} ,$$

where p(O) denotes, formally, the a priori probability of the object $O(\vec{k})$.

It is worth noting that this a posteriori probability rule differs from the maximum likelihood estimate only through the presence of the a priori probabilities p(O). Stated alternately, if all the possible objects are equiprobable, the maximum a posteriori probability estimator reduces to the maximum likelihood estimator. Although the maximum a posteriori probability estimator has the disadvantage that the a priori probabilities of the different objects must be specified, it is significant that these probabilities enter only as biases which can often be left as adjustable parameters in the system. This is very desirable since it permits the operator to resolve the very subjective question of how the a priori probabilities should be chosen.

*A complete discussion of these topics will be found in Cramer[1]; the transition from Cramer to the statement given here follows upon expressing $\bar{\epsilon}$ in terms of the real and imaginary parts of $O(\vec{k})$ and $\hat{O}(\vec{k})$ and then applying Cramer's result to each part separately.

Our discussion of performance measures and principles is now complete, and we proceed to apply them to the model of the previous section. This will first be done under the assumption that the atmospheric transfer function is known albeit random, and then that it is unknown.

KNOWN ATMOSPHERIC TRANSFER FUNCTION

We suppose that the atmospheric transfer function has been measured precisely during the recording of the data, so that the only uncertainty in the collected data is due to the additive noise. This assumption is a useful one in that it provides a measure of the best possible performance that could ever be obtained; thus one has a standard against which to compare systems that do not presume knowledge of the atmosphere. Of course it may be that the standard is not attainable. It will also lead us to conclude that linear processing is all that is required when the atmospheric transfer function is known; that is, nonlinear processing should only be employed in an attempt to measure the atmospheric transfer function.

It is well known that the (unconditional) integrated mean-squared error is minimized by taking the estimate of the object to be the conditional expected value of the object given the recorded data. However, there are few problems for which this conditional expected value can be expressed in any useful form. Two almost trivial but instructive extremes occur when the additive noise is either extremely large or extremely small. In both these extremes the resultant processor is linear and this suggests that linear processors might be effective, albeit nonoptimum, no matter what the noise level. Therefore we next inquire as to the minimum integrated mean-squared error that can be realized with linear processors. Specifically we take the estimate $\hat{O}(\vec{k})$ to be of the form

$$\hat{O}(\vec{k}) = \sum_{i=1}^{n} G_i(\vec{k})\, I_i(\vec{k}) \quad , \tag{10}$$

where $G_i(\vec{k})$ is the transfer function of the linear filter through which the ith recorded image is passed. We seek to determine the $G_i(\vec{k})$ that minimize the conditional, and also the unconditional, integrated mean-squared error with and without the constraint that the estimate be unbiased. The use of the conditional mean-squared error is appropriate in this instance because the linearity (and spatial invariance) of the processor prevents the degeneracy discussed in the preceding section. The results we seek are as follows; their derivation is contained in Exhibit A.

For any given object, the conditional integrated mean-squared error is minimized by choosing the linear estimate to be

$$\hat{O}(\vec{k}) = \left[\sum_{i=1}^{n} I_i(\vec{k})\, H_i^*(\vec{k})\right]\left[\frac{\sigma^2}{|O(\vec{k})|^2} + \sum_{j=1}^{n} |H_j(\vec{k})|^2\right]^{-1}, \qquad (11)$$

where the $H_i(\vec{k})$ are the known atmospheric transfer functions, σ^2 is the power density of the additive noise, and * denotes conjugation.

The linear estimate with the minimum unconditional integrated mean-squared error differs from Eq. (11) only in that $|O(\vec{k})|^2$ is replaced by its average value. In either event the processor passes each image through a filter which is matched to the corresponding atmospheric transfer function, adds the outputs together, and then passes the result through a filter whose transfer function depends upon either the particular object of interest or the average object according to whether the unconditional or conditional integrated mean-squared error is being considered.

The linear processor of Eq. (11) requires a knowledge of the object or its statistics. This knowledge is not utilized in the linear processor which minimizes the integrated mean-squared error subject to the constraint that the estimate be unbiased. In fact, this unbiased linear estimate is

$$\hat{O}(\vec{k}) = \left[\sum_{i=1}^{n} I_i(\vec{k})\, H_i^*(\vec{k})\right]\left[\sum_{j=1}^{n} |H_j(\vec{k})|^2\right]^{-1}, \qquad (12)$$

which is just what the general linear minimum mean-squared error estimator becomes as the noise density, σ^2, approaches zero. It is quite significant that the estimator of Eq. (12) provides as small a conditional, and unconditional, integrated mean-squared error as can any unbiased estimator be it linear or not. This statement follows easily from the Cramer-Rao inequality for unbiased estimates.[1]

Processors of the form described by Eqs. (11) and (12) are also suggested by maximum likelihood arguments. In fact the maximum likelihood estimator over the set of all possible functions $O(\vec{k})$ is just the unbiased estimator defined by Eq. (12). Moreover, the maximum a posteriori probability processor discussed in the preceding section need only extract the quantity

$$\sum_{i=1}^{n} I_i(\vec{k})\, \frac{H_i^*(\vec{k})}{\sum_{j=1}^{n} |H_j(\vec{k})|^2}$$

from the recorded data, i.e., this quantity is a sufficient statistic.

It is significant that all the processors considered above compute a quantity of the form

$$\hat{O}(\vec{k}) = Q(\vec{k}) \sum_{i=1}^{n} I_i(\vec{k})\, H_i^*(\vec{k}) \quad , \tag{13a}$$

where

$$Q(\vec{k}) = \left[W(\vec{k}) + \sum_{i=1}^{n} \left| H_i(\vec{k}) \right|^2 \right]^{-1} \quad , \tag{13b}$$

and where $W(\vec{k})$ is determined by the particular criterion employed. This certainly suggests that Eq. (13) provides a very reasonable estimate of $O(\vec{k})$ for some choice of $W(\vec{k})$ and that the choice might well be left to an operator, or at least determined by experiment. Alternately in the interests of simplicity one might merely set $W(\vec{k}) = 0$.

UNKNOWN ATMOSPHERIC TRANSFER FUNCTION

The repeated appearance of linear estimators when the atmospheric transfer function is known raises the question of their utility when it is unknown. Therefore, we next determine the linear estimator which minimizes the integrated mean-squared error under these more general conditions. Finally, we determine the maximum likelihood estimator that results when the noise dominates each of the recorded transforms but does not necessarily dominate the restored image. Although the maximum likelihood processor becomes linear in some circumstances, it is usually nonlinear. Thus, we conclude that the almost universal optimality possessed by the linear processor when the atmospheric transfer function is known is lost when it is unknown.

Linear Estimators

We again consider an estimate of the form given by Eq. (10) and seek to minimize the resulting integrated mean-squared error. For brevity we only consider the unbiased estimator; an expression for the biased estimator is presented in Exhibit A, where the unbiased estimator is also derived. To facilitate comparisons with the results of the preceding section, we do not invoke the supposition that successive observations are identically distributed.

The minimum integrated mean-squared error linear unbiased estimate of the object, given a sequence of image transforms $I_i(\vec{k})$, i = 1, ..., n, and given only a statistical description of the atmosphere is

$$\hat{O}(\vec{k}) = \left[\sum_{i=1}^{n} I_i(\vec{k}) \, \overline{H}_i^*(\vec{k}) \right] \left[\sum_{i=1}^{n} \left| \overline{H}_i(\vec{k}) \right|^2 \right]^{-1} , \tag{14}$$

where $\overline{H}_i(\vec{k})$ denotes the average value of the atmospheric transfer function associated with $I_i(\vec{k})$. As was true for a known atmospheric transfer function, the requirement that the estimate be unbiased causes it to be independent of the statistics of the object ensemble.

Comparison of Eqs. (12) and (14) shows that, lacking knowledge of the atmospheric transfer function, one supposes that it is equal to its average value. Of course, if the atmospheric transfer function is almost always equal to its average, the error introduced by this supposition will be quite small. To gain some further insight into the effects of these uncertainties upon the system performance, we now consider the integrated mean-squared error of the unbiased linear estimator when the atmospheric transfer function is known and when it is unknown.

It is easy to show that the integrated mean-squared error, averaged over the atmospheric and the noise fluctuations but not over the object, is given by the expression

$$\bar{\epsilon} = \frac{\sigma^2}{n} \int d\vec{k} \, \frac{1}{\left| \overline{H}(\vec{k}) \right|^2} + \frac{1}{n} \int d\vec{k} \left| O(\vec{k}) \right|^2 \frac{V(\vec{k})}{\left| \overline{H}(\vec{k}) \right|^2} \tag{15a}$$

when the atmospheric transfer functions are unknown, and by the expression

$$\bar{\epsilon} = \sigma^2 \int d\vec{k} \, \overline{\left[\sum_{i=1}^{n} H_i(\vec{k})^2 \right]^{-1}} \geq \frac{\sigma^2}{n} \int d\vec{k} \, \frac{1}{\left| H(\vec{k}) \right|^2} \tag{15b}$$

if they are known. In writing Eq. (15a) and in the subsequent discussion, we reinstate the condition that successive atmospheric transfer functions are identically distributed. The super bar again denotes average and

$$V(\vec{k}) = \overline{\left| H(\vec{k}) \right|^2} - \left| \overline{H}(\vec{k}) \right|^2 .$$

The right-most member of Eq. (15b) follows from a convexity argument.

The degraded performance that results when the linear estimator is employed with an unknown atmospheric transfer function, as compared with a known atmospheric transfer function, is immediately apparent from Eq. (15). Indeed, even if there is no additive noise, the error in the latter instance will be zero, whereas in the former instance

$$\bar{\epsilon} = \frac{1}{n} \int d\vec{k} \left|O(\vec{k})\right|^2 \frac{V(\vec{k})}{\left|\bar{H}(\vec{k})\right|^2} . \tag{16}$$

Of course this residual error can be decreased as much as desired by increasing n. Alternately, if the time statistics of the atmosphere are stationary, it can also be decreased by increasing the basic observation interval, or exposure time, T.

These two approaches are almost equivalent for stationary processes, since $\bar{H}(\vec{k})$ can be shown to increase linearly with T whereas $V(\vec{k})$ increases no more rapidly than linearly. Consequently, the error is upper bounded by a quantity that is inversely proportional to nT. Thus the simple linear estimator can be made to yield an arbitrarily small error, in the absence of noise, by resorting to sufficiently large values of n or T. However, for any fixed value of n and T, the system which lacks knowledge of the atmosphere will be inferior to that which possesses such knowledge.

It is worth noting that, for an unknown atmosphere, the integrated mean-squared error of the linear processor will approach zero as 1/n even when noise is present. Moreover, if the time statistics of the atmosphere are stationary, the error will also vanish as 1/T. In fact, as T increases,

$$\sigma^2 = T\,\sigma_o^2 ,$$

$$H(\vec{k}) = T\,\bar{H}_o(\vec{k}) ,$$

and

$$V(\vec{k}) \leq T\,V_o(\vec{k}) ,$$

where the subscript o denotes that the quantity is independent of T. Consequently, for an unknown atmospheric transfer function

$$\epsilon \leq \frac{1}{nT} \int d\vec{k} \left[\frac{\sigma_o^2 + V_o(\vec{k}) \left|O(\vec{k})\right|^2}{\left|\bar{H}_o(\vec{k})\right|^2} \right] . \tag{17}$$

Although the error can be made to vanish as $1/T$, this often is not an efficient means of operation, because as T increases, the high-frequency response of the atmospheric transfer function will diminish and hence the integrated mean-squared error may actually increase. That is, Eq. (17) does not insure that the error decreases monotonically with T and, in fact, it usually does not. Rather there is some optimum, finite value of T for any given values of σ, $V(\vec{k})$, and $\overline{H}(\vec{k})$.

Finally, it is important to note from Eq. (15) that knowledge of the atmosphere can improve the performance of a linear processor but it cannot alter the rate at which the integrated mean-squared error vanishes with n; that rate is proportional to $1/n$ in the presence or absence of such knowledge. It is significant that no processor, linear or nonlinear, can yield an integrated mean-squared error which vanishes at a faster rate. This result follows easily from the Cramer-Rao inequality.[1] Thus the only improvement that can be realized from more sophisticated data processing, insofar as it is measured by the integrated mean-squared error, is a reduction in the constant of proportionality. Of course, such a reduction may be of considerable value when a limited number of images are available. To gain some preliminary appreciation of the possibilities for improvement and the means of their realization, we consider maximum likelihood estimators.

Maximum Likelihood Estimators

The structure of the maximum likelihood estimator is in general a complicated function of the atmospheric statistics. There are, however, a few simple and important situations for which it can be determined.

One of these occurs when the additive noise is negligible; then the problem can be transformed into the classical signal plus noise problem by taking the logarithm of the recorded Fourier transforms, $I_i(\vec{k})$. Whether or not the resultant maximum likelihood processor is linear will depend upon the statistics of the logarithm of the atmospheric transfer function. If the transfer function itself is log normal, its logarithm will be normal and the resulting estimator will be linear in the logarithms of the $I_i(\vec{k})$, but it will not be linear in the $I_i(\vec{k})$ themselves. On the other hand, if the atmospheric transfer function comprises a Gaussian random process, the maximum likelihood estimator will be a nonlinear function of both the $I_i(\vec{k})$ and also of their logarithms.

It turns out that the processors which result from these two analyses are not very similar, so the question of which statistical description is appropriate becomes important. In reality, both are probably inadequate, so lacking further statistical knowledge of the atmosphere at this time, we will not pursue the

subject further; rather we will examine another situation for which a complete determination is possible.

We noted earlier that a long basic observation interval, or exposure time, is sometimes undesirable; better high-frequency response can be obtained by decreasing the exposure time. This decrease causes a relative increase in the noise level of the resulting image, but the object may still be recoverable if a sufficient number of images are available. These comments suggest that the problem wherein the noise dominates each of the recorded images may be an important one. It is also an appealing one in that the maximum likelihood estimate can be described fairly simply if n, the number of images, is not too large. Stated precisely, it can be determined in the limit of increasing noise density, σ^2, for any given n. The derivation is presented in Exhibit B.

The resulting estimate, $\hat{O}(\vec{q})$, is given by the solution of a Fredholm integral equation of the second kind. Specifically

$$\sigma \overline{H^*(\vec{q})}\, Y(\vec{q}) = \hat{O}(\vec{q})\, \overline{|H(\vec{q})|^2} - \frac{1}{2} \int d\vec{k}\, \hat{O}(\vec{k})\, G(\vec{q}, \vec{k}) \qquad \text{all } \vec{q},$$

where

$$Y(\vec{q}) = \frac{1}{n\sigma} \sum_i I_i(\vec{q})$$

and

$$G(\vec{q}, \vec{k}) = -\frac{1}{n\sigma^2} \left[\sum_i I_i(\vec{q})\, I_i^*(\vec{k}) \right] \left[\overline{H^*(\vec{q})\, H(\vec{k})} - \overline{H^*(\vec{q})}\ \overline{H(\vec{k})} \right] .$$

This equation can be solved by Fredholm techniques. In particular, if the variance of the atmospheric transfer function is zero, $G(\vec{q}, \vec{k})$ vanishes, and the estimate reduces to that of Eq. (14), i.e.,

$$\hat{O}(\vec{q}) = \frac{\sigma \overline{H^*(\vec{q})}\ Y(\vec{q})}{|H(\vec{q})|^2} .$$

More generally, the resulting estimate will be a nonlinear function of $G(\vec{q}, \vec{k})$ but, even in the general case, the estimate depends on the recorded data only through the quantities

$$\sum_i I_i(\vec{k})$$

and

$$\sum_i I_i(\vec{q})\ I_i^*(\vec{k})\ .$$

Time has not been available to evaluate the improvement in performance that the maximum likelihood estimator provides over a simple linear estimator; however, it should be substantial. This is not to say that it is necessarily the best processor to use; rather it suggests that substantial improvements over simple linear processing may be possible. Further study is required to determine the appropriate atmospheric models, to establish the resulting optimum processors, and to evaluate their performance.

ACKNOWLEDGMENT

In the course of this work, the author benefited from many discussions with his colleagues at the study. The lengthy discussions with David Slepian were particularly helpful.

EXHIBIT A

Subject to the assumptions of the section beginning on page 72, estimates of the form

$$\hat{O}(\vec{k}) = \sum_i I_i(\vec{k})\, G_i(\vec{k}) \tag{A1}$$

yield conditional integrated mean-squared errors, $\bar{\epsilon}$, of the form

$$\bar{\epsilon} = \int d\vec{k} \left[\sigma^2 \sum_i \left| G_i(\vec{k}) \right|^2 + \left| O(\vec{k}) \right|^2 \overline{\left| \sum_i \left[\frac{1}{n} - H_i(\vec{k})\, G_i(\vec{k}) \right] \right|^2} \right] . \tag{A2}$$

To minimize this error with respect to the $G_i(\vec{k})$, we minimize the integrand for each $\vec{k}$ by finding its stationary points. The resulting stationary point equation is, by the independence of successive observations

$$\left[V_i(\vec{k}) + \frac{\sigma^2}{\left| O(\vec{k}) \right|^2} \right] G_i(\vec{k}) - \overline{H}_i^{\,*}(\vec{k}) + \overline{H}_i^{\,*}(\vec{k}) \sum_j G_j(\vec{k})\, \overline{H}_j(\vec{k}) = 0 \quad \text{all } i, \vec{k}, \tag{A3}$$

where

$$V_i(\vec{k}) = \overline{\left| H_i(\vec{k}) \right|^2} - \left| \overline{H_i(\vec{k})} \right|^2 .$$

The solution of this equation is clearly of the form

$$G_i(\vec{k}) = C(\vec{k}) \frac{\overline{H}_i^{\,*}(\vec{k})}{\frac{\sigma^2}{\left| O(\vec{k}) \right|^2} + V_i(\vec{k})} , \tag{A4}$$

where $C(\vec{k})$ is a function of $\vec{k}$ but not of i. Substituting Eq. (A4) into Eq. (A3) and solving for $C(\vec{k})$ yields

$$C(\vec{k}) = \frac{1}{1 + \sum_j \left[\frac{\left| \overline{H}_j(\vec{k}) \right|^2}{\frac{\sigma^2}{\left| O(\vec{k}) \right|^2} + V_j(\vec{k})} \right]} .$$

We next minimize the conditional error subject to the constraint that the estimate be unbiased. It is easy to show that the constraint requires that

$$\sum H_i(\vec{k})\, G_i(\vec{k}) = 1 \qquad \text{all } \vec{k} \ , \tag{A5}$$

Upon adding this to the integrand of Eq. (A2) as a Lagrangian and again seeking stationary points, one obtains

$$G_i(\vec{k}) = \lambda(\vec{k})\, H_i^*(\vec{k}) \ , \tag{A6}$$

where $\lambda(\vec{k})$, the Lagrangian multiplier, is independent of i. Finally, substitution of Eq. (A6) into Eq. (A5) yields

$$G_i(\vec{k}) = \left[\sum_j \left| H_j(\vec{k}) \right|^2 \right]^{-1} H_i^*(\vec{k}) \ . \tag{A7}$$

These results can be applied to known atmospheric transfer functions by replacing the $V_i(\vec{k})$ with zero. The result of Eq. (A4) is easily extended to ensembles of objects (unconditional integrated mean-squared error); the result is Eq. (A4) with $\left| O(\vec{k}) \right|^2$ replaced by $\overline{\left| O(\vec{k}) \right|^2}$.

EXHIBIT B

Maximum Likelihood Estimator

To determine the maximum likelihood estimator, we first represent the transforms of the image, object, and noise by samples. Since these transforms are completely determined by their behavior in a half space (say $k_1 \geq 0$), it suffices to employ only samples in that half space. The conditional probability density of the image samples given the object samples is then expressed as the average, over the atmospheric transfer function, of the sample density conditioned both upon the object and upon the atmospheric transfer function. The result is next differentiated with respect to the real and imaginary parts of the object samples to obtain maximum likelihood equations. These operations yield, after some manipulation,

$$\overline{\left[\sum_i \left(I_g^i - \hat{O}_g H_g^i\right)^* H_g^i \exp\left\{-\sum_{n,k} \frac{1}{2\sigma^2}\left|I_k^n - \hat{O}_k H_k^n\right|^2\right\}\right]} = 0 \quad \text{all } g, \tag{B1}$$

where I_g^i and H_g^i denote the gth frequency samples of the ith recorded transform and its associated atmospheric transfer function, $\hat{O}_g$ denotes the gth sample of the object transform, and the super bar denotes average with respect to atmospheric fluctuations. The form of this expression follows from the facts that the noise is white and Gaussian and that the real and imaginary parts of the noise are uncorrelated over a half space.

The supposition that the atmospheric transfer function is statistically independent from observation to observation permits one to express Eq. (B1) as

$$\sum_i \frac{\left[\left(I_g^i - \hat{O}_g H_g^i\right)^* H_g^i \exp\left\{-\frac{1}{2\sigma^2}\sum_k \left|I_k^i - \hat{O}_k H_k^i\right|^2\right\}\right]}{\left[\exp\left\{\frac{1}{2\sigma^2}\sum_k \left|I_k^i - \hat{O}_k H_k^i\right|^2\right\}\right]} = 0 . \tag{B2}$$

We next suppose that the noise density, σ^2, is quite large. The noise term then dominates I_k^i and, as σ increases,

$$F_k^i \triangleq \frac{I_k^i}{\sigma}$$

has an average value that approaches zero and a variance that approaches unity. Upon introducing F_k^i into Eq. (B2) and expanding to first order in the $1/\sigma$, one obtains

$$\sigma \overline{H_g}^* Y_g - \hat{O}_g \overline{|H_g|^2} + \frac{1}{2}\left[\sum_k \hat{O}_k \tilde{V}_{gk} \tilde{Z}_{gk}\right]^* + \frac{1}{2}\left[\sum_k \hat{O}_k V_{gk} Z_{gk}\right] = 0, \tag{B3}$$

where

$$Y_g = \frac{1}{n} \sum_i F_g^i ,$$

$$Z_{gk} = \frac{1}{n} \sum_i F_g^i \left(F_k^i\right)^* ,$$

$$V_{gk} = \overline{H_g^* H_k} - \overline{H_g^*}\ \overline{H_k} ,$$

$$\tilde{Z}_{gk} = \frac{1}{n} \sum_i \left[F_g^i F_k^i\right]^* ,$$

$$\tilde{V}_{gk} = \overline{H_g H_k} - \overline{H_g}\ \overline{H_k} .$$

Upon letting the samples become dense in the positive half space and passing, formally, to an integral, Eq. (B3) becomes

$$\sigma \overline{H^*(\vec{g})}\, Y(\vec{g}) - \hat{O}(\vec{g}) \overline{|H(\vec{g})|^2} + \frac{1}{2} \int d\vec{k}\, \hat{O}(\vec{k})\, V(\vec{g},\vec{k})\, Z(\vec{g},\vec{k})$$

$$+ \frac{1}{2} \int d\vec{k} \left[\hat{O}(\vec{k})\, \tilde{V}(\vec{g},\vec{k})\, \tilde{Z}(\vec{g},\vec{k})\right]^* = 0 , \tag{B4}$$

where the integral is over a half space and where $\vec{g}$ and $\vec{k}$ now represent vectors of spatial frequencies, and

$$Y(\vec{g}) = \frac{1}{n\sigma} \sum_i I_i^i(\vec{k}) ,$$

$$Z(\vec{g},\vec{k}) = \frac{1}{n\sigma^2} \sum_i I_i^i(\vec{g})\, I_i^{i*}(\vec{k}) ,$$

$$V(\vec{g},\vec{k}) = \overline{H^*(\vec{g})\,H(\vec{k})} - \overline{H^*(\vec{g})}\;\overline{H(\vec{k})} ,$$

$$\tilde{Z}(\vec{g},\vec{k}) = \frac{1}{n\sigma^2} \sum_i \left[I_i(\vec{g})\, I_i(\vec{k}) \right]^* ,$$

$$\tilde{V}(\vec{g},\vec{k}) = \overline{H(\vec{g})\,H(\vec{k})} - \overline{H(\vec{g})}\;\overline{H(\vec{k})} .$$

Finally, we employ the conjugate symmetry possessed by transforms of real functions to eliminate the second integral in Eq. (B4) and to extend the domain of validity to the entire frequency space. The result is

$$\sigma \overline{H^*(\vec{g})}\, Y(\vec{g}) = \hat{O}(\vec{g}) \left|\overline{H(\vec{g})}\right|^2 - \frac{1}{2} \int_{\infty}^{0} d\vec{k}\, \hat{O}(\vec{k})\, G(\vec{g},\vec{k}) = 0 \qquad \text{all } \vec{g}, \text{ (B5)}$$

where

$$G(\vec{g},\vec{k}) = Z(\vec{g},\vec{k})\, V(\vec{g},\vec{k}).$$

REFERENCE

1. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, Princeton, N. J. (1946), Chapter 32.

APPENDIX 13

ON LINEAR LEAST-SQUARES FILTERING OF DISTORTED PHOTOGRAPHS

David Slepian

INTRODUCTION AND SUMMARY

Consider a rectangular distorted photograph that occupies the region

$$|x_1| \leq L_1, \qquad |x_2| \leq L_2 \tag{1}$$

in the $x_1 - x_2$ plane. We describe its density in some suitable units by

$$f(x_1, x_2), \qquad |x_1| \leq L_1, \qquad |x_2| \leq L_2 \ , \tag{2}$$

and we extend $f(x_1, x_2)$ to the entire plane by defining

$$f(x_1, x_2) \equiv 0, \qquad |x_1| > L_1, \qquad |x_2| > L_2 . \tag{3}$$

We suppose that this distorted photograph is the result of noise added to a filtered version of a desired undistorted photograph with density $g(x_1, x_2)$. Specifically, we write

$$f(x_1, x_2) = \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2\, k(x_1 - y_1,\ x_2 - y_2)\, g(y_1, y_2)$$

$$+ \ w(x_1, x_2)\, n(x_1, x_2), \qquad -\infty < x_1, x_2 < \infty \ . \tag{4}$$

Here

$$w(x_1, x_2) = \begin{cases} 1, & |x_1| \leq L_1, |x_2| \leq L_2 \\ 0, & |x_1| > L_1, |x_2| > L_2 \end{cases} \tag{5}$$

is a window function and $n(x_1, x_2)$ is a sample of a spatially stationary two-dimensional noise process. We suppose the undistorted photograph to be of limited extent and extend the definition of its density by

$$g(x_1, x_2) = 0, \quad |x_1| > a_1, \ |x_2| > a_2. \tag{6}$$

Similarly, we suppose k to be of limited support so that

$$k(x_1, x_2) \equiv 0, \quad |x_1| > b_1, \quad x_2 > b_2 \ . \tag{7}$$

Finally, to avoid cumbersome edge effects we suppose that

$$a_1 + b_1 < L_1, \quad a_2 + b_2 < L_2 \tag{8}$$

so that the first term on the right of Eq. (4) vanishes for $|x_1| > L_1$, $|x_2| > L_2$, and hence that equation holds for all points in the plane as already indicated.

The foregoing assumptions form a model of the process of photographing a small object against a uniform background. The background is taken to yield density zero on the photograph. The imaged object must be smaller than the size of the photograph. The function $k(x_1, x_2)$ is the point-spread function of the optical system. We complete our hypotheses by assuming that k is stochastic in nature, i.e., for each different photograph made of the object, a point-spread function $k(x_1, x_2)$ is drawn independently from an ensemble of point-spread functions. We further suppose that for all x_1, x_2, x'_1, x'_2 the random variables $k(x_1, x_2)$ and $n(x'_1, x'_2)$ are independent.

Consider now processing the distorted picture by a linear filtering operation to obtain a restored picture with density

$$\hat{g}(x_1, x_2) = \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2\, q(x_1 - y_1,\ x_2 - y_2)\ f\ (y_1, y_2)\ . \tag{9}$$

The mean-squared error between the restored picture and the undistorted picture is

$$\epsilon = E \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, [\,\hat{g}(x_1, x_2) - g(x_1, x_2)\,]^2\ , \tag{10}$$

where the expectation is taken over the point-spread function ensemble and the noise ensemble. In the following section we determine the processing filter $q(x_1, x_2)$ that minimizes ϵ. We conclude by describing some possible shortcomings of this minimal mean-squared error criterion for choosing a processing filter.

The least-mean-square filter that we derive in the following section is most easily specified by its transfer function

$$Q(\vec{\lambda}) \equiv Q(\lambda_1, \lambda_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 e^{-i(\lambda_1 x_1 + \lambda_2 x_2)}\, q(x_1, x_2). \tag{11}$$

We show that

$$Q(\vec{\lambda}) = \frac{|G(\vec{\lambda})|^2\ E\overline{K}(\vec{\lambda})}{\eta(\vec{\lambda}) + |G(\vec{\lambda})|^2\ E|K(\vec{\lambda})|^2} \tag{12}$$

Here

$$G(\vec{\lambda}) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 e^{-i\vec{\lambda}\cdot\vec{x}} g(x_1, x_2) \tag{13}$$

is the complex amplitude spectrum of the undistorted photograph and

$$K(\vec{\lambda}) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 e^{-i\vec{\lambda}\cdot\vec{x}} k(x_1, x_2) \tag{14}$$

is the transfer function of the optical system. The over-bar denotes complex conjugate. The quantity $\eta(\vec{\lambda})$ is a filtered noise spectrum related to the power density spectrum $N(\vec{\lambda})$ of the noise $n(x_1, x_2)$ by

$$\eta(\vec{\lambda}) = 16 \int_{-\infty}^{\infty} d\lambda'_1 \int_{-\infty}^{\infty} d\lambda'_2 \left[\frac{\sin(\lambda_1 - \lambda'_1) L_1}{\lambda_1 - \lambda'_1} \cdot \frac{\sin(\lambda_2 - \lambda'_2) L_2}{\lambda_2 - \lambda'_2} \right]^2 N(\lambda'_1, \lambda'_2). \tag{15}$$

The noise power density spectrum is of course related to the noise covariance by

$$\rho(x_1, x_2) = En(y_1, y_2)\, n(y_1 + x_1, y_2 + x_2)$$

$$= \int_{-\infty}^{\infty} d\lambda_1 \int_{-\infty}^{\infty} d\lambda_2 e^{i\vec{\lambda} \cdot \vec{x}} N(\vec{\lambda}) . \tag{16}$$

We have assumed that $En(x_1, x_2) = 0$. [If this is not the case, e.g., if $En(x_1, x_2) = m(x_1, x_2)$, one need only replace $f(x_1, x_2)$ by $f(x_1, x_2) - w(x_1, x_2)$ in the analysis as is seen from Eq. (4).]

As is to be expected, the optimal filter depends only on the second-order statistical properties of the optical transfer function and these enter only through the quantities $E\overline{K}(\vec{\lambda})$ and $E|K(\vec{\lambda})|^2$. It is noteworthy that if no noise is present, i.e., $\eta(\vec{\lambda}) \equiv 0$, the optimal restoring filter is independent of the undistorted picture. In this case

$$Q(\vec{\lambda}) = \frac{E\overline{K}(\vec{\lambda})}{E|K(\vec{\lambda})|^2} . \tag{17}$$

In many practical situations, the film and processing noise $n(x_1, x_2)$ is very small compared with turbulence or optical noise represented by the variations in K. The universal filter given by Eq. (17) can then be expected to give nearly optimal results for a wide class of undistorted photographs.

The mean-squared error obtained using the optimal filter given by Eq. (12) will be shown to be

$$\epsilon_{min} = \int_{-\infty}^{\infty} d\lambda_1 \int_{-\infty}^{\infty} d\lambda_2 \left|G(\vec{\lambda})\right|^2 \left[1 - \frac{|G(\vec{\lambda})|^2 \, |EK(\lambda)|^2}{\eta(\vec{\lambda}) + |G(\vec{\lambda})|^2 \, E|K(\vec{\lambda})|^2} \right]. \tag{18}$$

It will be shown that the quantity in square brackets here can never be negative and cannot exceed unity. The expression vanishes only if the additive noise $n(x_1, x_2)$ vanishes and if the optical transfer function $K(\vec{\lambda})$ is deterministic (i.e., not stochastic).

If one wishes to minimize Eq. (10) averaged over an independent ensemble of undistorted pictures as well as over the other ensembles already considered, the optimal filter is given by Eq. (12) with $|G(\lambda)|^2$ replaced by $E|G(\lambda)|^2$. The same substitution in Eq. (18) yields the minimum mean-squared error in this case.

DERIVATION OF RESULTS

To avoid unduly complicated formulas we adopt vector notation and write

$$\int dx \quad \text{for} \quad \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2$$

with a similar convention for integration with respect to the components of other 2-vectors.

In the present notation, Eqs. (4) and (9) are, respectively,

$$f(\vec{x}) = \int dy\, k(\vec{x}-\vec{y})g(\vec{y}) + w(\vec{x})n(\vec{x}), \tag{19}$$

and

$$\hat{g}(\vec{x}) = \int dy\, q(\vec{x}-\vec{y})f(\vec{y}) . \tag{20}$$

Combining these statements, we write

$$\hat{g}(\vec{x}) = \hat{g}_1(\vec{x}) + n_1(\vec{x}) , \tag{21}$$

where

$$\hat{g}_1(\vec{x}) = \int dv\, q(\vec{x}-\vec{y}) \int du\, k(\vec{y}-\vec{u})g(\vec{u}) . \tag{22}$$

and

$$n_1(\vec{x}) = \int dy\, q(\vec{x}-\vec{y})w(\vec{y})n(\vec{y}) . \tag{23}$$

For the mean-squared error, Eq. (10), we now find

$$\epsilon = E\int dx\,[\hat{g}(\vec{x}) - g(\vec{x})]^2 = E\int dx\,[\{\hat{g}_1(\vec{x}) - g(\vec{x})\} + n_1(\vec{x})]^2$$

$$= E\int dx\,[\hat{g}_1(\vec{x}) - g(\vec{x})]^2 + E\int dx\,[n_1(\vec{x})]^2 , \tag{24}$$

since for the cross product term we have

$$E \int dx\, n_1(\vec{x})\,[\hat{g}_1(\vec{x}) - g(\vec{x})] = \int dx\, En_1(\vec{x})\, E[\hat{g}_1(\vec{x}) - g(\vec{x})] = 0$$

from the assumed independence of $n(\vec{x})$ and $k(\vec{x})$ and the zero mean assumption for n. We proceed to treat the two terms of Eq. (24) separately.

From Parseval's theorem, we have

$$E \int dx\, [\hat{g}_1(\vec{x}) - g(\vec{x})]^2 = E \int d\lambda \left|\hat{G}_1(\vec{\lambda}) - G(\vec{\lambda})\right|^2 , \tag{25}$$

where upper-case letters are used to denote the two-dimensional Fourier transforms of quantities denoted by corresponding lower-case letters. From Eq. (22) and the convolution theorem,

$$\hat{G}_1(\vec{\lambda}) = Q(\vec{\lambda})\, K(\vec{\lambda})\, G(\vec{\lambda}) ,$$

so that Eq. (25) becomes

$$E \int dx\, [\hat{g}_1(\vec{x}) - g(\vec{x})]^2 = \int d\lambda \left|G(\vec{\lambda})\right|^2 E \left|Q(\vec{\lambda})\, K(\vec{\lambda}) - 1\right|^2 . \tag{26}$$

For the second term of Eq. (24) we find from Eq. (23),

$$E \int dx\, [n_1(\vec{x})]^2 = \int dx \int dy \int dy'\; q(\vec{x}-\vec{y})\, q(\vec{x}-\vec{y}')\, w(\vec{y})\, w(\vec{y}')\, \rho(\vec{y}-\vec{y}') , \tag{27}$$

where ρ, the covariance function of the noise, is defined in Eq. (16). Now for the x integration in Eq. (27) we have

$$\int dx\, q(\vec{x}-\vec{y})\, q(\vec{x}-\vec{y}') = \int dx\, \ell_{\vec{y}}(\vec{x})\, \ell_{\vec{y}'}(\vec{x}) = \int d\lambda\, L_{\vec{y}}(\vec{\lambda})\, \overline{L}_{\vec{y}'}(\vec{\lambda}) , \tag{28}$$

where we have again used Parseval's theorem and made the definition

$$\ell_{\vec{y}}(\vec{x}) = q(\vec{x}-\vec{y}) = \int d\lambda\, e^{i\vec{\lambda}\cdot(\vec{x}-\vec{y})} Q(\vec{\lambda}) .$$

It follows from the transform of this equation that

$$L_{\vec{y}}(\vec{\lambda}) = e^{-i\vec{\lambda}\cdot\vec{y}} Q(\vec{\lambda}) ,$$

so that Eq. (28) becomes

$$\int dx\, q(\vec{x}-\vec{y})\, q(\vec{x}-\vec{y}') = \int d\lambda \left|Q(\vec{\lambda})\right|^2 e^{i\vec{\lambda}\cdot(\vec{y}'-\vec{y})}.$$

Equation (27) now becomes

$$\int dx\, [n_1(\vec{x})]^2 = \int d\lambda \left|Q(\vec{\lambda})\right|^2 \int dy \int dy'\, e^{i\vec{\lambda}\cdot(\vec{y}'-\vec{y})} w(\vec{y})\, w(\vec{y}')\, \rho(\vec{y}-\vec{y}')$$

$$= \int d\lambda \left|Q(\vec{\lambda})\right|^2 \eta(\vec{\lambda}) . \qquad (29)$$

Here we have set

$$\eta(\vec{\lambda}) = \int_{-L_1}^{L_1} dy_1 \int_{-L_2}^{L_2} dy_2 \int_{-L_1}^{L_1} dy'_1 \int_{-L_2}^{L_2} dy'_2\, e^{i\vec{\lambda}\cdot(\vec{y}'-\vec{y})}\, \rho(\vec{y}-\vec{y}')$$

$$= \int d\lambda' N(\vec{\lambda}') \int_{-L_1}^{L_1} dy_1 \int_{-L_2}^{L_2} dy_2 \int_{-L_1}^{L_1} dy'_1 \int_{-L_2}^{L_2} dy'_2\, e^{i\vec{\lambda}\cdot(\vec{y}'-\vec{y})}$$

$$\times\ e^{-i\vec{\lambda}'\cdot(\vec{y}'-\vec{y})}$$

$$= 16 \int_{-\infty}^{\infty} d\lambda'_1 \int_{-\infty}^{\infty} d\lambda'_2 \left[\frac{\sin(\lambda_1-\lambda'_1)L_1}{\lambda_1-\lambda'_1} \cdot \frac{\sin(\lambda_2-\lambda'_2)L_2}{\lambda_2-\lambda'_2}\right]^2 N(\vec{\lambda}')$$

in agreement with Eq. (15). We have also used Eqs. (5) and (16).

Inserting Eqs. (26) and (29) in Eq. (24). we find for the mean-squared error

$$\epsilon = \int d\lambda \left[\left|G(\vec{\lambda})\right|^2 E\left|Q(\vec{\lambda})\, K(\vec{\lambda}) - 1\right|^2 + \left|Q(\vec{\lambda})\right|^2 \eta(\vec{\lambda})\right] . \qquad (30)$$

We seek to minimize this expression by proper choice of $Q(\vec{\lambda})$. Since each term in the integrand is nonnegative, the minimum value of ϵ is obtained by minimizing the integrand at each point in the $\lambda_1 - \lambda_2$ plane. To carry out the minimization, we adopt the notation

$$Q(\vec{\lambda}) = x + iy,$$

$$EK(\vec{\lambda}) = \alpha + i\beta, \tag{31}$$

$$E\left|K(\vec{\lambda})\right|^2 = A,$$

where on the right side we have suppressed the dependence on $\vec{\lambda}$. In this notation the integrand of Eq. (30) is

$$p \equiv \left[\left|G\right|^2 E\left|QK-1\right|^2 + \left|Q\right|^2 \eta\right]$$

$$= \left|G\right|^2 \left[A(x^2+y^2) - 2\alpha x + 2\beta y + 1\right] + (x^2+y^2)\eta \ . \tag{32}$$

The minimum is found by setting $\partial p/\partial x = \partial p/\partial y = 0$. These conditions give

$$x = \frac{\alpha|G|^2}{A|G|^2+\eta} \ , \qquad y = -\frac{\beta|G|^2}{A|G|^2+\eta} \ . \tag{33}$$

Equations (31) and (33) yield our main result, Eq. (12). Substitution of Eq. (33) in Eq. (32) gives

$$p_{min} = |G|^2 \left[1 - \frac{(\alpha^2+\beta^2)\,|G|^2}{A|G|^2+\eta}\right]. \tag{34}$$

Inserting this in Eq. (30) and using the definitions Eqs. (31) yields Eq. (18).

The magnitude of the quantity in brackets in Eq. (34) can be readily bounded by using the complex Schwartz inequality

$$\left|\int dx\, f(\vec{x})\, \bar{g}(\vec{x})\right|^2 \leq \int dx \left|f(\vec{x})\right|^2 \int dy \left|g(\vec{y})\right|^2 \tag{35}$$

with equality only if f is proportional to g. To apply this inequality set $f(\vec{x}) = (x_1 + ix_2)\sqrt{p(x_1,x_2)}$ and $g(\vec{x}) = \sqrt{p(x_1,x_2)}$. Here we have set $K(\vec{\lambda}) = x_1 + ix_2$ and have written $p(x_1,x_2)$ for the joint density of the random variables x_1 and x_2. The inequality (35) now reads

$$\left| EK(\vec{\lambda}) \right|^2 = \left| \int dx\, (x_1 + ix_2) \sqrt{p(\vec{x})} \sqrt{p(\vec{x})} \right|^2$$

$$\leq \int dx\, (x_1^2 + x_2^2)\, p(\vec{x}) \int dy\, p(\vec{y}) = E\left| K(\vec{\lambda}) \right|^2 . \qquad (36)$$

Equality holds only if p degenerates to a delta function.

From the inequality (36),

$$0 \leq \left| G(\lambda) \right|^2 \left| EK(\vec{\lambda}) \right|^2 \leq E\left| K(\lambda) \right|^2 \left| G(\lambda) \right|^2 + \eta(\lambda) ,$$

so that

$$0 \leq \frac{\left| G(\lambda) \right|^2 \left| EK(\vec{\lambda}) \right|^2}{\eta(\vec{\lambda}) + \left| G(\vec{\lambda}) \right|^2 E\left| K(\vec{\lambda}) \right|^2} \leq 1$$

with the right equality holding only if $\eta = 0$ and K is deterministic. The last two sentences below Eq. (18) then follow.

Returning to Eq. (30), we see that if $g(\vec{x})$ is also regarded as stochastic and independent of $k(\vec{x})$ and $n(\vec{x})$, then further averaging ϵ over the undistorted pictures results in replacing $\left| G(\vec{\lambda}) \right|^2$ in Eq. (30) by $E\left| G(\vec{\lambda}) \right|^2$. The full paragraph following Eq. (18) follows from this observation.

DISCUSSION

The analysis presented here is open to criticism on a number of points. In the first place, a not very serious shortcoming is the infinite range of integration which we have adopted for mathematical convenience in the definition (Eq. 10) of the mean-square error. It will be recalled that the original distorted photograph occupies the region $\left| x_1 \right| \leq L_1$, $\left| x_2 \right| \leq L_2$, so that a comparison of the processed picture with the undistorted picture only over this region would form a more reasonable error criterion. Such an analysis can readily be carried out. The results, which are much more complicated than Eq. (12), are presented in Exhibit A. They do not permit an explicit solution for the filter characteristic. By using an infinite range in Eq. (10) for the sake of mathematical simplicity, we have allowed part of the restoration potential of the filter to be expended in matching artificial white backgrounds that extend to infinity. In cases of practical interest, this loss is probably most small, especially in those cases where the inequalities (8) can be replaced by $a_1 + b_1 \ll L_1$, $a_2 + b_2 \ll L_2$.

Open to more serious criticism is our adoption of

$$\epsilon = \int_{-L_1}^{L_1} dx_1 \int_{-L_2}^{L_2} dx_2 \, [\hat{g}(\vec{x}) - g(\vec{x})]^2 \tag{37}$$

as a measure of the quality of a single restored photograph. This integral can have a given value because $\hat{g}$ and g have a certain small mismatch everywhere throughout the rectangle or the same value of ϵ can be achieved with $\hat{g}$ identically g except in a minuscule area where the discrepancy is large. In general, in practice the latter restored photograph would be much preferred to the former, though our analysis treats them as equally desirable.

Even if one adopts the smallness of Eq. (37) as a reasonable criterion for judging the merit of a single restoration, one can still with reason challenge the philosophy of choosing a filter to minimize ϵ averaged over an ensemble of many pictures of an object. A given mean for ϵ averaged over many photos can be obtained either when all of the photos have a given moderate individual ϵ value or when many of the photos are perfect ($\epsilon = 0$) and a few have large errors. In practice the latter situation is generally preferred by far, although our analysis treats them as of equal worth.

We have treated the case of a minimum mean-squared error filter because of its mathematical tractability and because of our inability to define a more suitable tractable measure of restoration quality. We conclude with the obvious support for the criterion adopted: if the mean-squared error is small enough, at least one of the restored photos is bound to be good.

EXHIBIT A

When the error criterion (Eq. 10) is replaced by

$$\epsilon = \int dx\, w(\vec{x})\, [\hat{g}(\vec{x}) - g(\vec{x})]^2 \tag{A1}$$

the least-mean-squared filter has a transfer function $Q(\vec{\lambda})$ given as the solution of the integral equation

$$\int d\lambda'\, H(\vec{\lambda}, \vec{\lambda}')\, Q(\vec{\lambda}') = R(\vec{\lambda}) \qquad -\infty \leq \lambda_1, \lambda_2 \leq \infty \,. \tag{A2}$$

Here

$$H(\vec{\lambda}, \vec{\lambda}') = W(\vec{\lambda} - \vec{\lambda}')\, [G(\vec{\lambda})\, \overline{G}(\vec{\lambda}')\, \rho_K(\vec{\lambda}, \vec{\lambda}') + \eta(\vec{\lambda}, \vec{\lambda}')] \,, \tag{A3}$$

$$R(\vec{\lambda}) = \int d\lambda'\, W(\vec{\lambda} - \vec{\lambda}')\, G(\vec{\lambda})\, \overline{G}(\vec{\lambda}')\, E\overline{K}(\vec{\lambda}) \,, \tag{A4}$$

where G and K are as in Eqs. (13) and (14) and $W(\vec{\lambda})$ is the transform of the window w of Eq. (5). The other quantities are given by

$$\rho_K(\vec{\lambda}, \vec{\lambda}') = EK(\vec{\lambda})\, \overline{K}(\vec{\lambda}') \,,$$

$$\eta(\vec{\lambda}', \vec{\lambda}'') = 16 \int d\lambda\, N(\vec{\lambda})\, \frac{\sin(\lambda_1 - \lambda'_1) L_1}{(\lambda_1 - \lambda'_1)}\, \frac{\sin(\lambda_2 - \lambda'_2) L_2}{(\lambda_2 - \lambda'_2)}$$

$$\times\, \frac{\sin(\lambda_1 - \lambda''_2) L_1}{(\lambda_1 - \lambda''_1)}\, \frac{\sin(\lambda_2 - \lambda''_2) L_2}{(\lambda_2 - \lambda''_2)} \,.$$

These results follow in a rather straightforward manner. We list only a few key steps in the derivation. Equation (25) becomes

$$E \int dx\, [w(\vec{x}) \{\hat{g}_1(\vec{x}) - g(\vec{x})\}^2] = E \int d\lambda \left| \int d\lambda' W(\vec{\lambda}-\vec{\lambda}')[\hat{G}_1(\vec{\lambda}') - G(\vec{\lambda}')] \right|^2$$

$$= \int d\lambda' \int d\lambda'' W(\vec{\lambda}' - \vec{\lambda}'') G(\vec{\lambda}') \overline{G}(\vec{\lambda}'')$$

$$\times \left[Q(\vec{\lambda}') \overline{Q}(\vec{\lambda}'') \rho_K(\vec{\lambda}', \vec{\lambda}'') - \overline{Q}(\vec{\lambda}'') E\overline{K}(\vec{\lambda}'') - Q(\vec{\lambda}') EK(\vec{\lambda}') + 1 \right]. \quad (A5)$$

Equation (28) becomes

$$\int dx\, w(\vec{x})\, q(\vec{x}-\vec{y})\, w(\vec{x})\, q(\vec{x}-\vec{y}')$$

$$= \int d\lambda' \int d\lambda'' W(\vec{\lambda}' - \vec{\lambda}'')\, Q(\vec{\lambda}')\, \overline{Q}(\vec{\lambda}'')\, e^{-i\vec{\lambda}'\cdot\vec{y}} e^{i\vec{\lambda}''\cdot\vec{y}'} ,$$

so that

$$\int dx\, w(\vec{x})\, [n_1(\vec{x})]^2 = \int dy \int dy' \int d\lambda' \int d\lambda''\, W(\vec{\lambda}'-\vec{\lambda}'')\, Q(\vec{\lambda}')\, \overline{Q}(\vec{\lambda}'')$$

$$\times\, e^{-i\vec{\lambda}'\cdot\vec{y} + i\vec{\lambda}''\cdot\vec{y}'} w(\vec{y})\, w(\vec{y}')\, \rho(\vec{y}-\vec{y}')$$

$$= \int d\lambda' \int d\lambda''\, W(\vec{\lambda}'-\vec{\lambda}'')\, Q(\vec{\lambda}')\, \overline{Q}(\vec{\lambda}'')\, \eta(\vec{\lambda}', \vec{\lambda}'') . \quad (A6)$$

The mean-squared error is the sum of Eqs. (A5) and (A6). Straightforward calculus of variations arguments then yield the results (A2), (A3), and (A4). If W in Eqs. (A3) and (A4) is replaced by the delta function, Eq. (A2) yields Eq. (12) once more and Eq. (A1) becomes Eq. (10).

APPENDIX 14

SOME NOTES ON FILM GRAIN NOISE

T. S. Huang

INTRODUCTION

We show that for the pictures we are interested in, such as those taken by E. T. Tyson at Cloudcroft, the grain noise is approximately multiplicative. Therefore, in analyzing the limitations of various image-processing methods, the case of multiplicative noise should be considered.

A MODEL

A model for a film transparency is shown in Figure 1. This model is reasonable under the assumption that the reading aperture is much smaller than the sizes of the point-spread functions of all the systems prior to the aperture, but is much larger than the grain size.

THE H-D CURVE

By proper development, one can make the H-D curve such that over a range of densities (a typical range $\approx 0.3 - 2.7$) $\Delta T \propto \Delta E$, or $\Delta \sqrt{T} \propto \Delta E$. ($\Delta D = \gamma \Delta \log E$. Therefore by controlling γ, one can make $\Delta T \propto \Delta E$, or $\Delta \sqrt{T} \propto \Delta E$. We have used Δx to denote an increment in x.)

NOISE IN TERMS OF D

Extensive measurements have indicated that, in terms of density, the grain noise is approximately Gaussian and white, and that, for a fixed aperture, the standard

deviation of the noise satisfies approximately the relation

$$\sigma_D = kD^p ,$$

where k and p are constants, and $0 < p < 1$. According to the well-known Selwyn law, $p = 1/2$. However, more recent measurements[1] indicate that it is more accurate to take $p = 1/3$. Therefore, using the notation of Figure 1, we have

$$D_r = D_s + kD_s^{1/3} n , \tag{1}$$

where n is Gaussian and white over the bandwidth of the aperture, has mean zero and variance 1, and is independent of D_s; and k is a constant.

NOISE IN TERMS OF T

$$D_r = -\log T_r , \tag{2}$$

$$D_s = -\log T_s , \tag{3}$$

$$T_r = 10^{-D_r} , \tag{4}$$

$$T_s = 10^{-D_s} , \tag{5}$$

$$T_r = 10^{-D_s} 10^{-kD_s^{1/3} n} = T_s 10^{-kD_s^{1/3} n} = T_s 10^{-k(-\log T_s)^{1/3} n} . \tag{6}$$

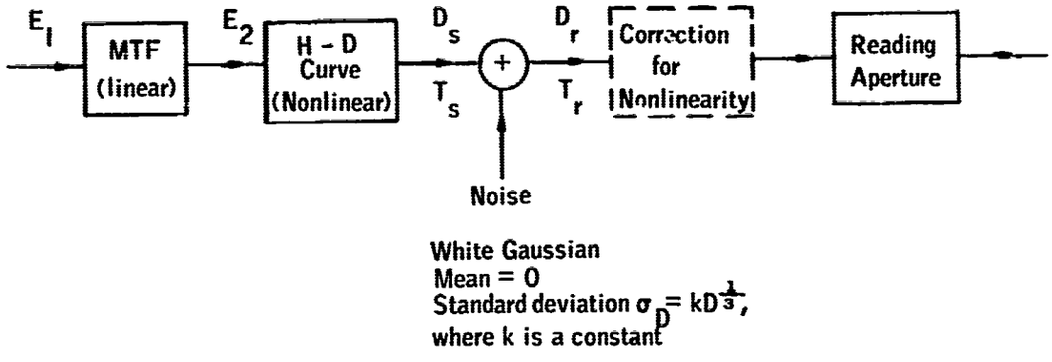


FIGURE 1 Model for a film transparency. E, exposure; T, intensity transmission; D, density.

THE CASE OF SMALL NOISE

Assume

$$k\,(-\log T_s)^{1/3} \ll 1\,. \tag{7}$$

Then, roughly speaking,

$$T_r \approx T_s\,[\,1 - k \log e\,(-\log T_s)^{1/3}\,n\,] = T_s - k \log e\, T_s(-\log T_s)^{1/3}\,n\,. \tag{8}$$

CORRECTION FOR NONLINEAR H-D CURVES

Suppose the H-D curve is nonlinear in T:

$$T = f(E)\,, \tag{9}$$

where f is a nonlinear function. Since T_s is inaccessible to us, we can only apply the correction $g(\cdot) = f^{-1}(\cdot)$ to T_r:

$$\begin{aligned} E &= g(T_r) \\ &\approx g[\,T_s - k \log e\, T_s\,(-\log T_s)^{1/3}\,n\,] \\ &\approx g(T_s) - g'(T_s)\,k \log e\, T_s\,(-\log T_s)^{1/3}\,n\,, \end{aligned} \tag{10}$$

assuming that the noise is small and that the function g is approximately linear over any range the size of the noise.

Let

$$E_s = g(T_s) \tag{11}$$

be the corrected (i.e., linearized) value of T_s. Then

$$E = E_s - k \log e\, f(E_s)\,[-\log f(E_s)\,]^{1/3}\,\frac{1}{f'(E_s)}\,n\,, \tag{12}$$

where f is the function representing the intensity transmission versus exposure curve as indicated in Eq. (9). An alternative expression for Eq. (12) is

$$E = E_s + kD_s^{1/3}\left(\frac{dE}{dD}\bigg|_{D=D_s}\right)n = E_s + k\,\frac{1}{\frac{d}{dE}(D^{2/3})\Big|_{E=E_s}}\,n\,. \tag{13}$$

Assume that over a range of densities

$$D = \gamma \log E + \delta \tag{14}$$

where γ and δ are constants. Then over this range, Eq. (13) becomes

$$E = E_s + k \frac{3}{2} \frac{1}{\log e} \gamma^{-1} E_s D_s^{1/3} n$$

$$= E_s + k \frac{3}{2} \frac{1}{\log e} \gamma^{-2/3} E_s (\log E_s + \frac{\delta}{\gamma})^{1/3} n \ . \tag{15}$$

Assume that in the range of interest, $D^{1/3}$ varies much more slowly than E. Then the noise will be approximately multiplicative. Let us consider an example. For 4X negative films using a normal development procedure, a typical usable density range is $0.3 < D < 3$. The range of exposure required to obtain this density range is 5 decades. Therefore, the noise is approximately multiplicative.

LINEAR T-E CURVE CASE

In the case where the T-E curve is linear, we can take T_s in Eq. (8) as the signal, and the noise is given by the second term of the same equation. The signal-dependent factor in the noise is

$$t = T_s (-\log T_s)^{1/3} \ . \tag{16}$$

This is sketched in Figure 2. It is seen that t is approximately linearly proportional to T_s for $0 < T_s < 0.5$, which is the usual range of intensity transmission. Therefore, the noise in Eq. (8) is approximately multiplicative for the range $0 < T_s < 0.5$.

CONCLUDING REMARKS

It is shown that the film grain noise is approximately multiplicative both for the case of a linear T-E curve and for the case of a nonlinear T-E curve (the nonlinearity is afterwards corrected, on a computer, say), under the following assumptions: (a) The reading aperture is much smaller than the sizes of the point-spread functions of all the systems prior to the aperture, but is much larger than the grain size. (b) The grain noise is small compared with the signal. The grain noise is approximately additive (independent of the signal), if the additional assumption is made that the ac signal is much smaller than the dc level, i.e., the picture is of low contrast.

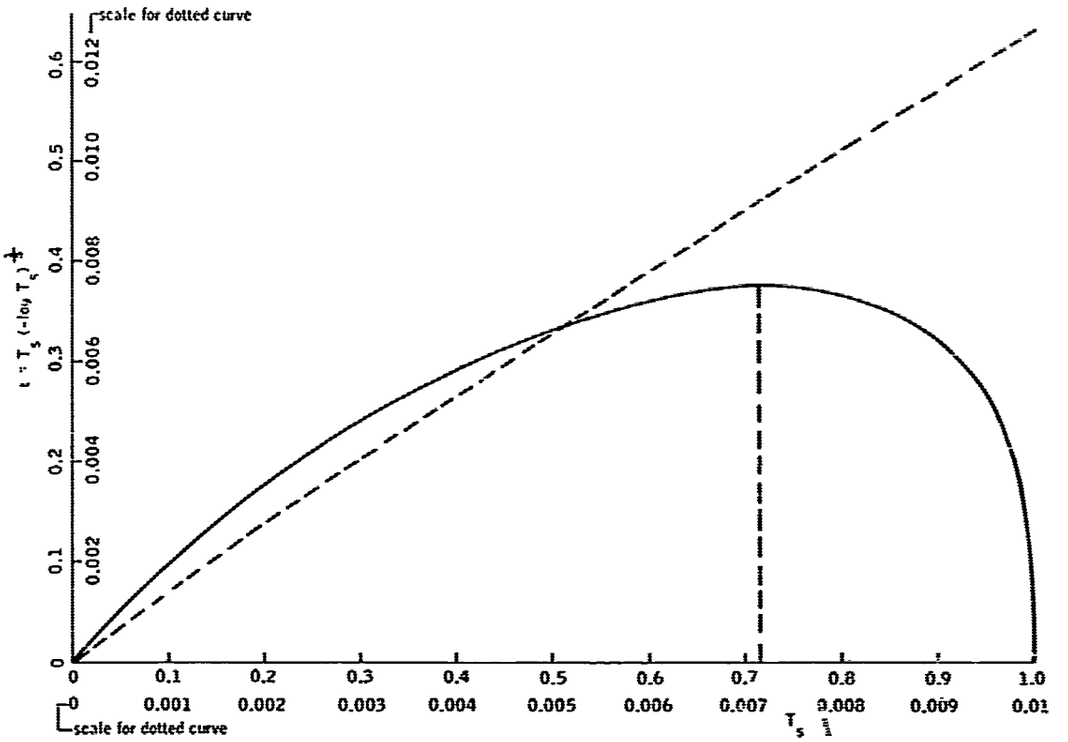


FIGURE 2 Approximate linearity of the function $t = T_s (-\log T_s)^{1/3}$.

For the pictures we are interested in, the contrast is by no means low (Tyson is hoping for a film dynamic range of 10^5): therefore the final assumption is not valid. However, in most cases assumptions (a) and (b) roughly hold. The grain noise can then be considered as approximately multiplicative.

We propose therefore that in analyzing the limitations of various processing methods, the case of multiplicative noise should be considered.

ACKNOWLEDGMENT

The professional help of J. H. Altman is greatly appreciated.

REFERENCE

1. E. F. Haugh, J. Phot. Sci. 11, 64 (1963).

APPENDIX 15

IMAGE ANALYSIS WITH THE ISODENSITRACER

Daniel H. Lufkin

In the great majority of cases, the point-spread function associated with turbulent perturbation of an image proves to be of a nearly Gaussian form. If we only require that the point-spread function be unimodal, and not even necessarily symmetrical, this criterion will be met by nearly all photographic systems. Of course, as exposure time decreases, as the angular subtense of the object decreases, or as the turbulence takes on certain special characteristics, the point-spread function loses its smoothed form and the simple criterion of unimodality is no longer met.

Until then, however, the image, even though degraded, has an important property: each point on the image receives the greatest part of its illumination from its conjugate point on the object. Thus a relatively bright point on the object is always imaged as a relatively bright point on the image; only the contrast between the point and its surround has been reduced. Thus, the image still carries some amount of information so long as the point-spread function still exhibits any peak at all.

This information is of no use, naturally, if it falls below the detection threshold of the receiver. In usual photographic practice, the receiver is the combination of the silver halide emulsion and the human eye. Nearly always the human eye is the member of the partnership which sets the limit to the ultimate performance of the system. The commonly accepted lower limit to visual contrast detection is $(B_1 - B_2)/(B_1 + B_2) \geq 0.04$, where B_1 and B_2 are the luminances of a point and its surround. Almost all usual photographic emulsions will record

contrast about an order of magnitude lower. In short, there is nearly always some information, even in a badly degraded image, which has been recorded but is not visible without some form of contrast amplification.

The simplest method of amplifying photographic contrast is to make a straightforward print on high-contrast emulsion. This method is very widely used in photointerpretation units and, given a skilled technician and a certain amount of cut and try, the results are often gratifying. Various edge-enhancement techniques involving the use of photographic masks are also used occasionally, but these are not easy to control and often introduce confusing artifacts into the image.

Direct measurement of the photographic density by instrumental means has the fundamental advantage that the whole process can be kept under close physical control. The distribution of density in the photographic image can be mapped at a convenient scale and presented to the eye in such a way that human ability to perceive weak contrast is no longer a limiting factor.

For practical use, a specialized instrument, the Isodensitracer,[1] has been developed and marketed by Technical Operations, Inc., Burlington, Massachusetts. The Isodensitracer (IDT) is an ingenious adaptation of the Joyce-Loebl double-beam microdensitometer, an instrument very widely used in densitometric work. (See Figure 1.)

The photographic transparency to be analyzed is fixed to a specimen table (1) which is drawn through a beam of light which originates at a lamp (2) and, passing through a conventional microscope optical train, is brought to a focus at an adjustable field aperture (4). Another beam, the reference beam, originates at the same lamp and passes through an attenuator wedge (3). The beams pass alternately

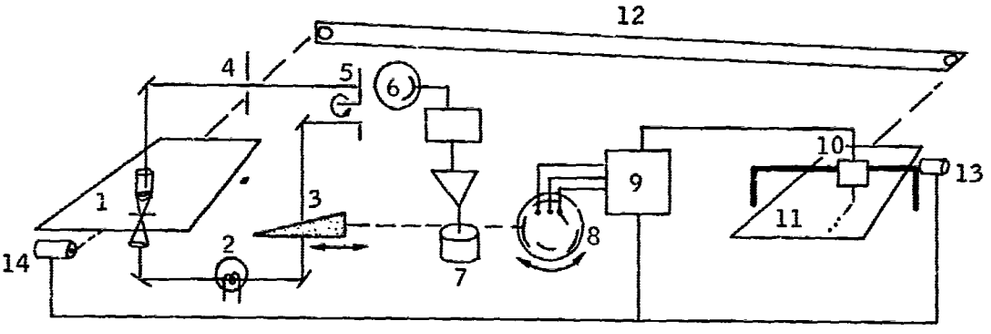


FIGURE 1 IDT schematic.

through a chopper (5) and fall on a photomultiplier (6). A circuit compares the intensities of the two beams and, by means of a servomotor (7), adjusts the position of the attenuator wedge until the reference beam and the specimen beam are equal. The position of the wedge is then a measure of the density of the transparency at some particular point.

The wedge servomotor also drives a commutator disk (8) which contacts three pick-ups sequentially. These pick-ups, in turn, control circuits in a power unit (9), which drives a solenoid (10) to move a ballpoint pen.

The pen is mounted on a lead-screw over the recording table (11). As the specimen table moves under the scanning beam, the recording table is driven by the lever (12) at a speed ratio which can be varied from 1:1 to 1:2,000. As the scanned density varies during the table stroke, the movements of the reference wedge are translated by the commutator into signals to the pen solenoid which cause the pen to stay up, to oscillate up and down, or to stay down. This sequence of signals will leave a record of a blank, a line of dots, and a solid dash on a sheet of recording paper attached to the table. The density value of any code mark may be determined by inserting a photometric step-wedge into the specimen beam, and the density increment required on the scanning table to change from one code mark to the next is determined by the density gradient of the reference wedge and by the pitch of the commutator. In practice, absolute density measurements are rarely required. The density increment for one code change can vary between 0.005 and 0.12 density units. By convention, the sequence blank-dot-dash indicates the direction of increasing density.

After the code marks for one scan-line have been recorded, the control unit operates two stepper motors (13 and 14) which move the two tables laterally for a new scan. The ratio between the lateral movements is adjustable; normally it is set to match the lever-arm ratio so that an undistorted enlargement results. The minimum scan separation on the specimen is 1.25 μ.

Figure 2 shows a 40X enlargement of a photograph of the satellite Pegasus A taken with a 900-in. focal length camera on 4X film. Slant range to the object was 432 miles. The IDT recording of the negative from which Figure 2 was printed is shown in Figure 3. The density increment represented by each full code cycle is 0.24 D, so that a change from one code element to another (dots to line, for example) represents a contrast of about twice the detection threshold of the eye. Even so, none of the detail is visible to the eye because the density gradients are confused with the grain noise.

Looking at the closed isodensity contours, the analyst can see that the ends of the "wings" are progressively more squared off on the inner contours, while the end of the "body" remains rounded. Choosing the innermost closed contour

FIGURE 2 Satellite Pegasus A at a slant range of 432 miles.

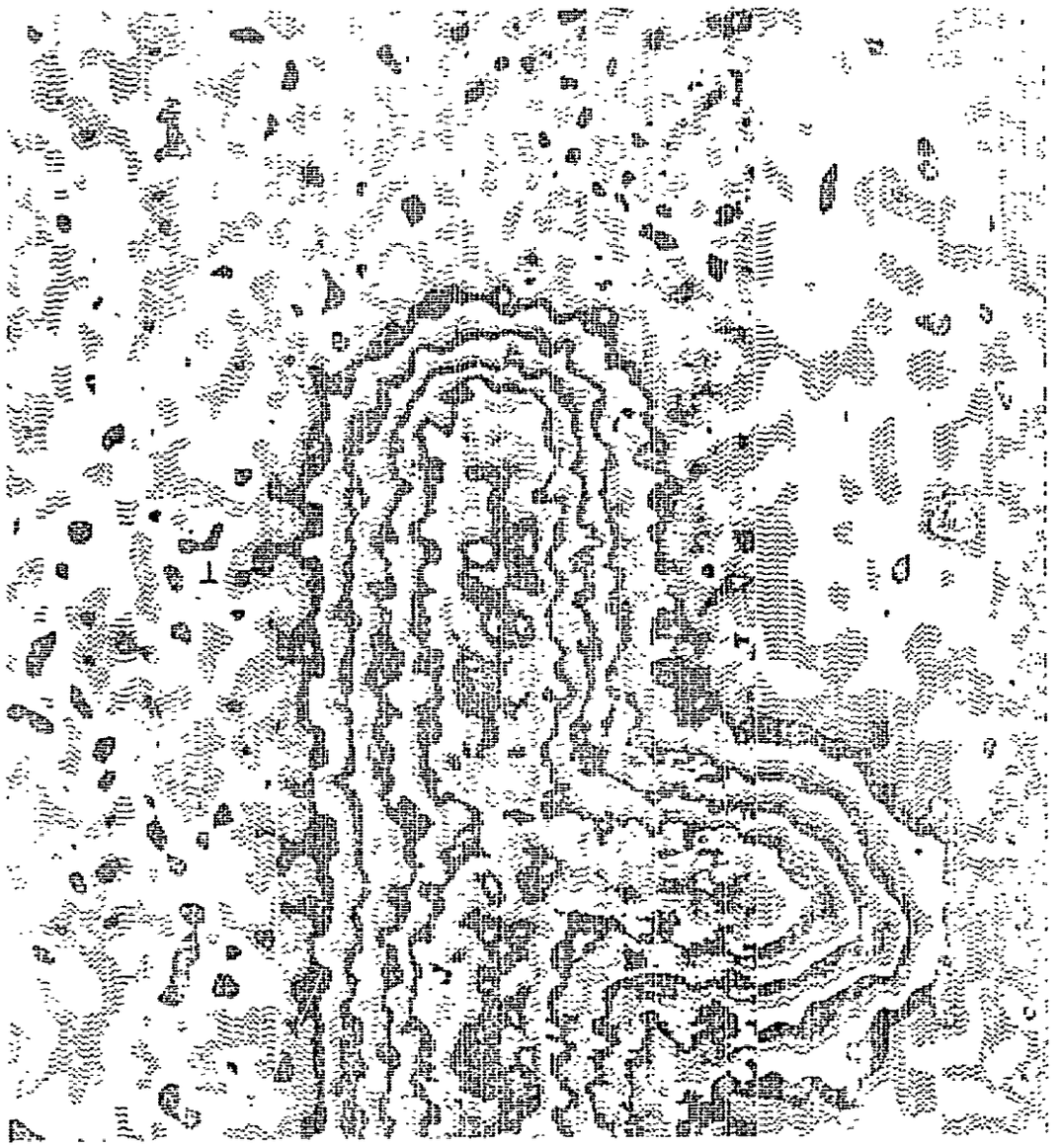
as a working boundary for the object, the analyst can sketch an approximate outline of the object. By comparing the IDT records made with different aperture sizes, the analyst can form a mental picture of the effects of grain noise, which decreases with increasing aperture size, and of the effects of the instrument transfer function, which effects decrease with decreasing aperture size. With these effects in mind, some of the smaller details of the isodensity distribution can be interpreted in terms of likely object structure sketched in Figure 4.

Figure 5 shows an artist's rendering of the actual satellite. The shape of the body and the "wings," the ribbed structure of the panels, and the general proportions of the object are seen to correspond well with the analysis.

This aspect of the IDT analysis is still largely subjective. Only by comparing details of the distribution in question with those of IDT analyses of known geometrical forms can the analyst interpret the patterns he sees.

In this connection, the role of a priori information in the analysis must be mentioned. In general, the analyst will know that the object is man-made, composed of an assembly of a limited number of simple geometric forms such as plane squares and rectangles, cylinders, spheres, thin straight lines, and cones. He knows, furthermore, that the image is composed of random photographic grains which are being integrated in density over the area of the scanning aperture. The analyst will normally have made more than one IDT analysis with varying apertures. In this way, he can mentally evaluate the effect of the scanning aperture's transfer function upon the analysis. He will often have a "library" of IDT analyses of the typical images of various objects, which he will use as a photointerpretation key to help him understand the meaning of the isodensity contours. In this way, the analyst has the advantage of information which helps him restrict the range of decisions affecting his interpretation. At the same time, this a priori knowledge does not usually force the analyst into an uncomfortably narrow range of interpretation. The number of different configurations which can be built up out of a modest store of basic shapes is very large indeed.

There is no question at all that the goal of an image reconstruction program should be to produce an end product so clear and unambiguous that anyone would be able to recognize and describe the object at first sight. It is important, nevertheless, to bear in mind that the problem of extracting information from



DATE 7 JUL 66 RECORDING NO. 1057.5 IDENTIFICATION PEGASUS

SPOT HEIGHT	25μ	WEDGE NUMBER	D	MAGNIFICATION:	200 =	OPERATOR	DHL
SPOT WIDTH	25μ	ΔD INCREMENT	0.08		200 '	COMMENTS:	
OBJECTIVE	10	DIFFERENTIAL CONTROL	4.5	SCALE: ONE INCH=			1085 #8
CONDENSER	10	PEN DAMPING		ONE CENTIMETER=			

FIGURE 3 IDT recording of negative of Figure 2. 1 sec of arc = 21 mm on the figure; 1 sec of arc = 11 ft at 432 miles.

RECORDING # 1057.5 PEGASUS (1085) Frame 8

200X Enlargement Scanning aperture 25x25 microns

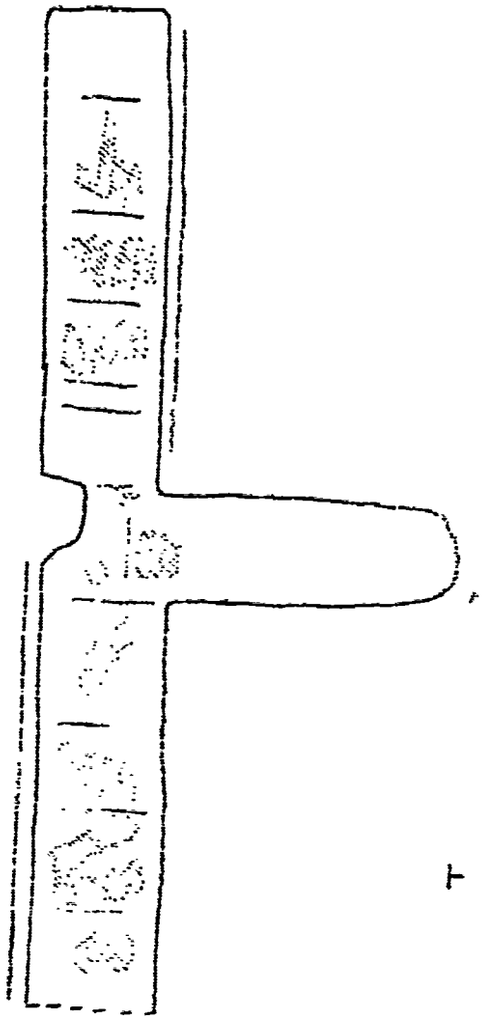


FIGURE 4 Preliminary interpretation of Figure 3.

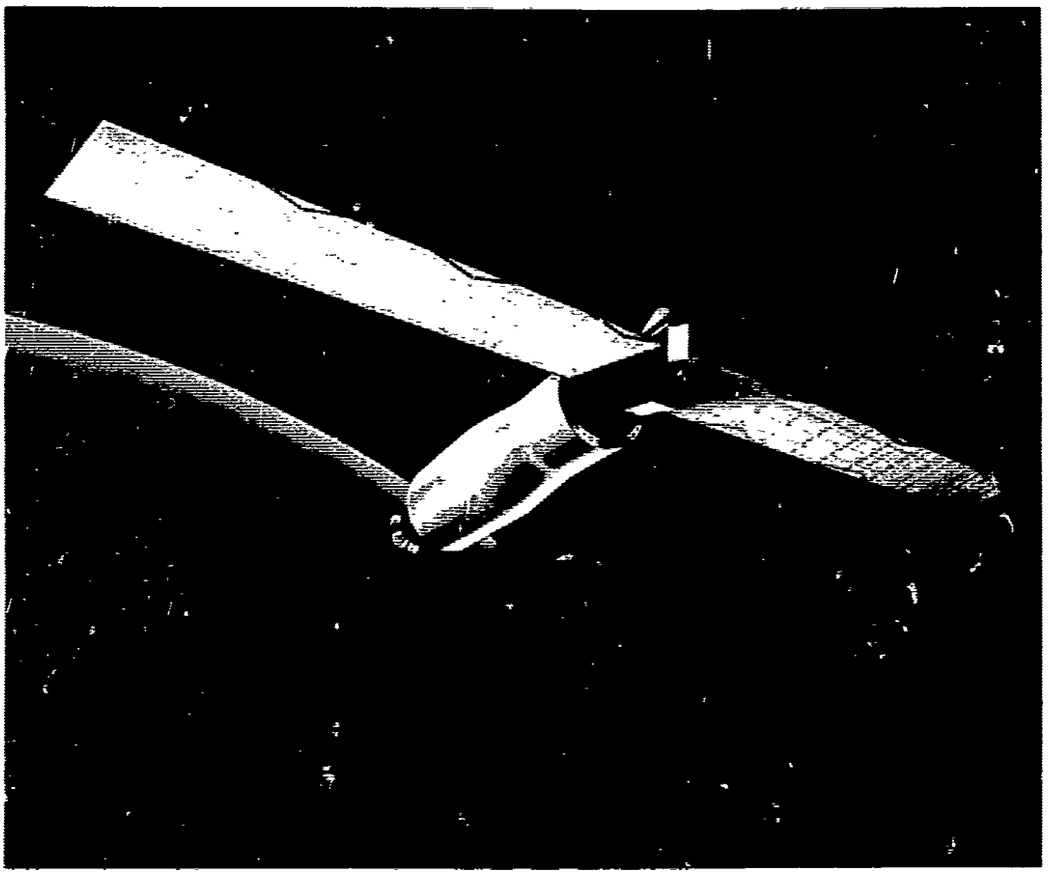


FIGURE 5 Artist's conception of the satellite Pegasus A in earth orbit.

imperfect images has had a history of reasonable successes for many years under the pressure of economic and military demands. Painstaking analysis of images, even when these are grossly imperfect, will often reveal surprising stores of information.

REFERENCE

1. C. S. Miller, F. G. Parsons, and I. L. Kofsky, Nature 202, 1196 (June 20, 1964).

APPENDIX 16

DYNAMIC RANGE OF BRIGHTNESS OF SATELLITES ILLUMINATED BY THE SUN

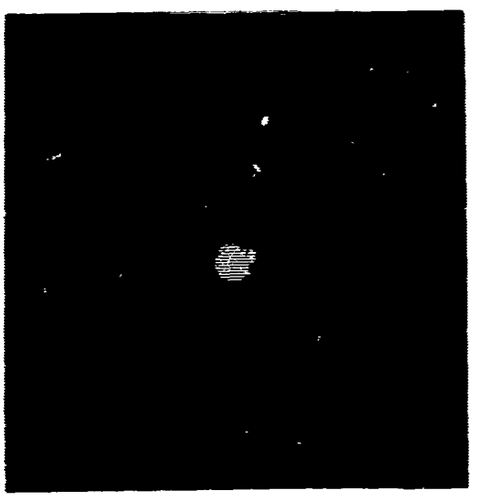
Edmund T. Tyson

One of the problems encountered at Cloudcroft during satellite image recording experiments has been the large dynamic range of surface brightness exhibited by satellites. Typical sensors that have been used for recording are image orthicons with a dynamic range of 2.5 decades and films with ranges from 2 to 6 decades. The purposes of this note is to define the range numerically. With this in view, several satellite shapes are used to describe the ranges encountered in practice.

The source of illumination is the sun, whose surface brightness is 225,000 cd/cm^2. The sun produces an illuminance of 15.3 lm/cm^2 at the distance of the earth. This illuminance will produce a brightness of $15.3/\pi = 4.9$ cd/cm^2 on a diffuse reflecting surface normal to the sun's rays.

The brightness of a diffusely reflecting surface element is proportional to the cosine of the angle between the normal to the surface and the direction of the sun's rays. The dynamic range of a diffusely reflecting, spherical satellite is a function of how close to the terminator one wishes to obtain an image. For example, at 5° from the terminator, the dynamic range is over 10:1; at 1°, it is nearly 60:1.

The observed intensity of a specular, spherical satellite is due to a mirror-like reflection of a solar image from a small region of the satellite's surface. The diameter of the solar image is the product of the spherical radius and half the angular diameter of the sun. The image is a few centimeters in diameter on even large satellites such as Echo I (r = 3000 cm).



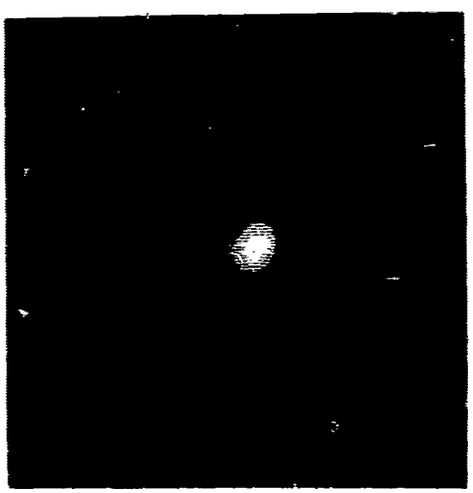
Slant range: 1,020 statute miles

Angular size: 8 sec of arc

Exposure time: 1/250 sec

Remarks: Visually, this object appears as a small crescent, pearly in texture, with a brilliant specular reflection halfway between the limb and the terminator.

FIGURE 1 Satellite Echo II, showing effects of specular reflection.



Slant range: 560 statute miles

Exposure time: 1/125 sec

Remarks: This picture illustrates the effect of a specular reflection from a normally diffuse reflecting surface. A line image of the sun, reflected from the cylindrical S4 rocket stage, caused diffusion in the emulsion and completely obscured the tank and one meteorite detection panel.

FIGURE 2 Satellite Pegasus C, showing effects of specular reflection.

Echo II exhibits both diffuse and specular characteristics. The measured intensity of the satellite can be accounted for by the specular reflection. However, visual observations and photographic records show the spherical shape of the satellite. The brightness ratio of the solar image to the diffuse background is of the order of 225,000:4.9 or 50,000:1. The ratio may well be larger, since the diffuse component probably has a reflection coefficient of less than 0.1. On the other hand, the apparent brightness ratio on the film is reduced, because the bright specular point has been diffused by "seeing" and by diffusion in the emulsion.

A cylindrical satellite with the normal paint surface applied for temperature control exhibits both specular and diffuse reflection characteristics. This surface, in the proper orientation, will reflect a line image of the sun while presenting a diffusely illuminated area. The brightness ratio will be comparable with that noted for Echo II.

More complex structures, such as a cylindrical satellite with solar paddles, can provide specular and diffuse reflections from separate components. A transit of an Agena with two solar paddles attached was observed during which one of the solar paddles was in a specularly reflecting attitude. Diffusion in the photographic emulsion was sufficient to completely obscure the image of the rocket body and the second paddle.

The brightness of the night sky is 10^{-8} cd/cm^2 and is so far down it does not enter into the problem. The brightness of the day sky is 10^{-1} cd/cm^2 and will set a lower limit to the brightness range that can be observed.

Figures 1 and 2 are photographs illustrating the dynamic range problem in satellite photography. Both were taken through the 48-in. telescope at Cloudcroft at a focal length of 900 in. and recorded on Eastman Kodak 4X 35-mm film using a Leica IIc camera.

APPENDIX 17

LINEARITY AND DYNAMIC RANGE OF FILMS

J. H. Altman

In this paper, the terms linearity and dynamic range will refer to the relation between density and log exposure in the common photographic sensitometric curve. The first requirement of wide-range photography is that this curve maintain a useful gradient – i. e., not "shoulder off" – over the necessary exposure range. Dynamic range is defined here as the exposure range over which such a useful gradient is available. Linearity then refers to the relation between density and log exposure on the straight-line portion of the curve; this is not, of course, linearity in the optics-electronics sense. Linearity is desirable but not absolutely necessary for recording, so that in many cases, the useful dynamic range will exceed the range over which the D-log E curve is straight.

As normally used, black and white photographic materials will cover a range of perhaps 10^3 or less before shouldering. One available emulsion which covers significantly more range is the XR (extended-range) film of Edgerton, Germeshausen and Greer. An early form of this film has been described by Wyckoff.[1] It is a three-layer material similar in construction to Kodacolor, except that the three emulsion layers are differentiated by speed rather than by spectral sensitivity. In use, a very low exposure will register only in the fastest layer. On the other hand, a very bright object will overexpose this fastest layer, but register at correct exposure in one of the slower layers. After exposure, the film is put through the Kodacolor process, producing a dye image in each of the three layers and removing all silver. The observer obtains the image from the correctly exposed layer by viewing with a filter whose color is complementary to the dye in

this proper layer. The fact that the faster layers are badly overexposed is no problem because with a subtractive color system, only the layer complementary to the viewing filter can be seen. The others are effectively transparent.

The film is reported[1] to cover an exposure range of 10^8. At the time of writing (July 1966), it is understood that a modified XR film has been coated experimentally, and that information and possibly samples are available from P. Emerson of Edgerton, Germeshausen and Greer.

In addition to this, it also turns out that some common black and white materials will respond over a range of approximately 10^6 (though this is not true of all materials). Two in particular which have been shown to work are Kodak Plus-X Roll Film and Eastman 4X Negative Film (type 5224). Others can no doubt be found. Processed to normal gammas over extended exposure ranges, such materials will give maximum densities of 4.5 or greater, which may make read-out difficult. This problem can be eased simply by decreasing the development and working at very low gamma, the implications of which will be discussed later. At this point it is worth noting that a phenidone developer recently described by M. Levy[2] for extended-range processing was tested with 5224. The resulting D-log E curves are shown in Figure 1; the plots are essentially linear over a range of 10^6 with slopes of the order of 0.15. The minor departures from linearity may well be real.

However, it is not enough to consider only the sensitometry of the film - developer combination; we should also consider the structure of the resulting image. This has been done by Courtney-Pratt in Appendix 11, and there is no point in repeating his analysis here. However, some additional notes on dynamic range may be useful, as well as a comment on the problem of image "blooming."

In Appendix 11, Courtney-Pratt has derived a quantity J which is to be maximized for most efficient detection of satellite images. For the present writer, it is convenient to rewrite J in terms of more familiar photographic quantities. From Courtney-Pratt's Eq. (12):

$$J = \frac{U}{\sigma_{10}} \frac{dD}{dU},$$

where U signifies exposure. Rewriting,

$$J = \frac{1}{\sigma_{10}} \frac{UdD}{dU}.$$

But $U/dU = 1/d \log_e U = 0.434/d \log_{10} U$, and by definition $dD/d \log_{10} U$ is photographic gamma. Therefore,

$$J = 0.434 \gamma/\sigma_{10},$$

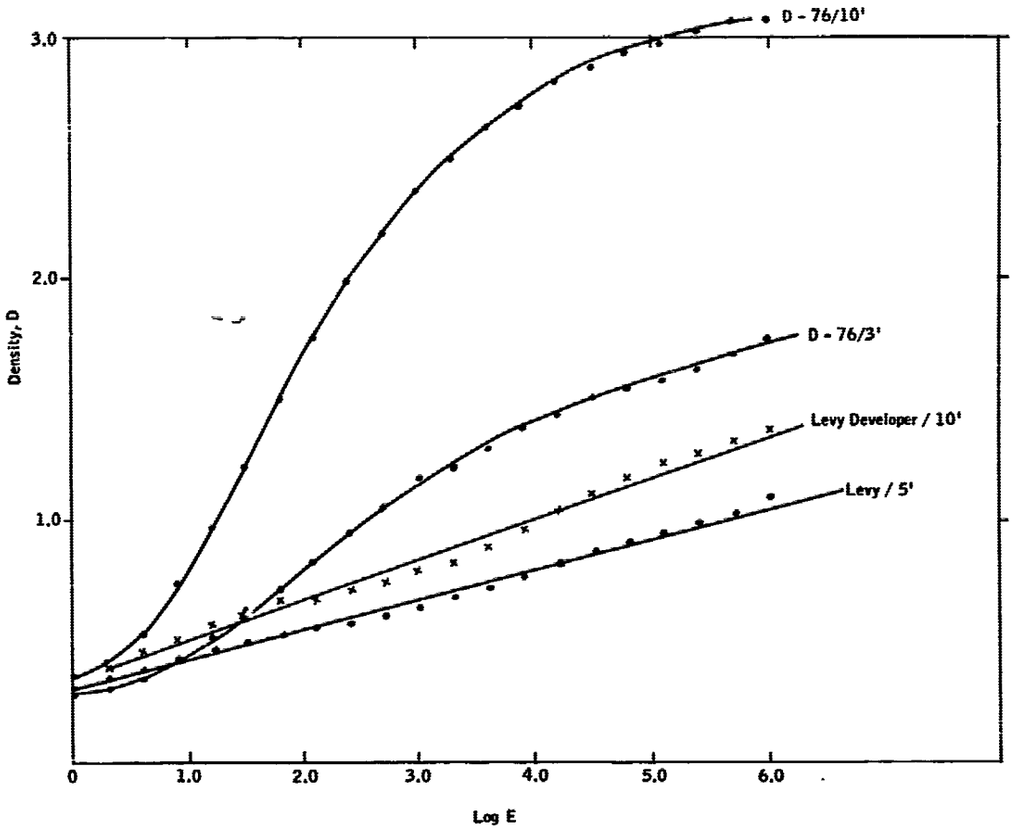


FIGURE 1 D-log E curves for 4X negative film type 5224 processed as shown at 68°.

where σ_{10} signifies the rms granularity of the image area as measured with a 10-μ scanning aperture. Since by Selwyn's law the rms granularity varies inversely as the scanning aperture, we have finally that $J \propto \hat{\gamma}/\sigma$, where σ is measured for any image size of interest, and the symbol $\hat{\gamma}$ refers to the slope of any point on the D-log E curve (not necessarily the straight-line portion). This is the signal-to-noise (S/N) ratio of the output image.[3]

Since, in general, σ for black and white films increases as $D^{1/3}$, where D is the density,[4] it follows immediately that optimum S/N ratios are achieved by recording at the lower end of the straight-line portion of the D-log E curve.

When a very wide range of exposures must be recorded, it should be useful to determine curves of $\hat{\gamma}/\sigma$ for the material. For some wide but finite range, the optimum exposure region is that which gives maximum area under the curve. Alternatively, the exposure range over which some selected S/N ratio could be maintained can also be determined from such a curve. Finally, it should be

instructive to compare very low gamma processes such as Levy's to more typical processes with the aid of this criterion. (Levy's developer gives lower granularity on 5224 than does D-76, but quantitative data are not available at this writing.)

Even this is not the whole story, however. Photographs provided by Tyson (see Appendix 16) indicate that the "blooming" of very heavily exposed images is at least as serious a problem as the shouldering of the emulsion. The blooming can arise from three factors: (a) image spread caused by the atmosphere and other system factors; (b) image spreading within the emulsion layer; and (c) halation from the support returning to the surface of the emulsion. Data are needed on the relative importance of these three effects. Effect (a), of course, cannot be controlled by emulsion factors. Effect (c), if it is important, can be reduced by using better protection than gray b. e — such as an "antihalation undercoat" (AHU). A measure of the magnitude of effect (b) is given by the point-spread function of the emulsion. Such spread functions can be calculated from published modulation transfer curves.[5,6]

At least in some applications, the emulsion spread function should be small compared to the system spread function, and in any case the two can be convolved using well-known techniques.[6] Note, however, that the combined spread function is the spatial distribution of exposure in the image, and that the resulting distribution of density in the record may be somewhat different after development.

At any rate, it appears that a figure of merit for wide-latitude photography of satellites might be of the form

$$[\log E_2 - \log E_1]/a_s ,$$

where a_s is the area of the system spread function, and where E_2 and E_1 are the exposures between which some specified value of S/N ratio can be maintained. At present, data are lacking on minimum useful values of S/N, as well as on the relative importance of the several factors. Note that it may be preferable to replace the spread function term $1/a_s$ with a factor $1/w$, the rate of widening of the developed image with exposure.

Tyson has suggested that as an operational matter, the time of exposure be limited to not longer than 0.01 sec. If exposure time is stated, this converts the abscissa from log E to log I, the illuminance in the image plane in meter candles. And for a camera of given F-number, this illuminance can be related to the brightness of the object by the well-known expression

$$I = \frac{\pi BT}{4(F/No)^2} ,$$

where T is the transmission factor of the system. Thus the effective exposure latitude for a given emulsion looking at a target of given brightness range can be derived. Courtney-Pratt has discussed this aspect of the problem also.

REFERENCES

1. C. W. Wyckoff, "An Experimental Extended Exposure Response Film," Soc. Phot. Instr. Engrs. Newsletter, 16 (June-July 1962).

2. M. Levy, "Wide Latitude Photography," Preprints of Paper Summaries, 1966 Annual Conference of the Society of Photographic Scientists and Engineers, San Francisco, May 9-13, 1966, p. 188.

3. J. H. Altman, "Image Structure for Data Reduction," paper delivered at the 100th Technical Conference, Society of Motion Picture and Television Engineers, Los Angeles, October 2-7, 1966.

4. E. F. Haugh, J. Phot. Sci. 11, 65 (1963).

5. "Modulation Transfer Data for Kodak Films," Kodak Pamphlet P-49, Eastman Kodak Company, Rochester, N. Y. (1962).

6. C. E. K. Mees and T. H. James, The Theory of the Photographic Process, Macmillan, New York (1966), pp. 501-505.

APPENDIX 18

SOME FUNDAMENTAL ASPECTS OF FILM SCANNING

James L. Harris, Sr.

The purpose of this paper is to discuss some fundamental aspects of the problem of scanning film to obtain data input for computerized image processing. Many of the requirements for such a scanner are similar to those associated with precision microdensitometry. There are, however, a few important distinctions. We must keep in mind that image processing amounts to an attempt to obtain approximate solutions to a convolution integral equation involving the irradiance map which constitutes the image. The purpose of scanning the film is to make measurements which will allow determination of the image irradiance map to the precision required by the subsequent processing operations.

"TRANSMISSION" OF THE FILM

A piece of film has rather complex transmission properties. If we picture a narrow collimated beam incident at some angle with respect to the normal on one side of the film, then on the other side of the film we find a rather broad angular distribution of transmitted flux, generally peaked at the angle of incidence. If we now repeat the operation for a piece of film of higher density, i.e., lower transmittance, we find that in addition to a reduction of total flux passing through the film, the shape of the angular distribution of flux coming out of the film has been altered. Similarly, if we change the angle of incidence of our narrow beam, we see changes in the angular distribution of the transmitted flux. These observations lead to an important conclusion; i.e., the "apparent transmission" (transmitted flux divided by incident flux) depends on the angular distribution of the

flux used to illuminate the film, and on the geometry of the angular acceptance cone associated with the measurement of transmitted flux.

It is for this reason, primarily, that two projection densitometers of different optical design can read the same film gray scale wedge and produce significantly different renderings of the characteristic curve of the film, including dynamic range and gamma, for example. It is therefore more proper to say that we have measured "a transmission" rather than "the transmission." These observations should also make it clear that the process of scanning a gray scale along with images is an operation which serves to provide calibration, not only for the film, but also for the properties of the scanning device itself.

GENERAL SCANNER LAYOUT

Figure shows a typical optical layout for a projection densitometer.

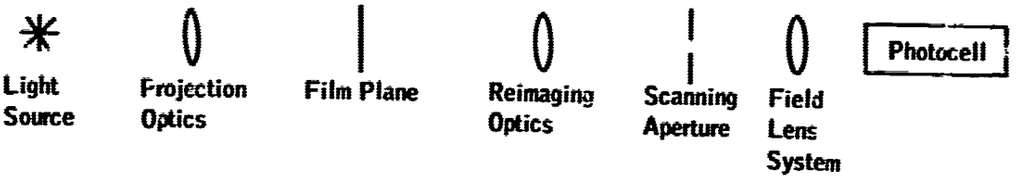


FIGURE 1 Projection densitometer optical layout.

The light source and projection optics furnish illumination of the film. The film is reimaged in the scanning aperture plane and the flux passing through the aperture is deposited on the photocathode of a multiplier phototube, usually by means of a field lens which images the exit pupil of the reimaging optics onto the photocathode.

The system is called an object space scanner if the scanning operation is accomplished by moving the film, and an image space scanner if the scanning operation is accomplished by moving the scanning aperture.

OPTICAL TRANSFER FUNCTION OF SCANNER

Consideration must be given to the image-forming properties of the projection system. As an example, suppose that we are recording images with a diffraction-limited telescope of f-number $(f/\ )_T$. Then the spatial frequency cutoff on the film will be

$$\xi_{max} = \frac{1}{\lambda(f/\ )_T} \text{ cycles/meter} , \qquad (1)$$

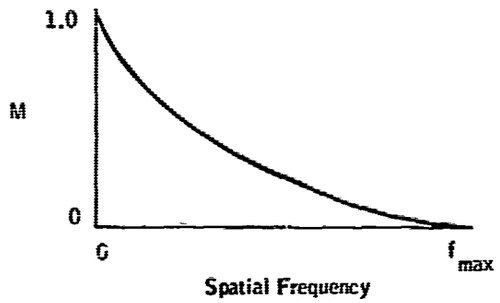


FIGURE 2 Schematic optical transfer function.

where λ is the wavelength in meters. A similar relationship holds for the optical transfer function of the scanner reimaging optics. If we assume that the magnification is large, i.e., film located in the focal plane of the lens, then

$$f_{max} = \frac{1}{\lambda (f/\ )_S} \text{ cycles/meter,} \quad (2)$$

where $(f/\ )_S$ is the f-number of the scanner reimaging optics, provided that the scanner optical system was diffraction limited. Since the optical transfer function is a function of the type shown in Figure 2, it follows that $(f/\ )_S \ll (f/\ )_T$ if we are to pass the image on the film without substantial degradation. But what about the assumption of diffraction-limited optics in the scanner?

An important factor in determining whether an optical system can be diffraction limited is the angular field of view which it is required to present. In the telescope the field of view associated with a film image of width w is

$$\theta_T = \frac{w}{F_T} \text{ radians,} \quad (3)$$

where F_T is the focal length of the telescope. In the scanner,

$$\theta_S = \frac{w}{F_S} \text{ radians,} \quad (4)$$

so that the ratio of the two fields of view is given by

$$\frac{\theta_S}{\theta_T} = \frac{F_T}{F_S} , \quad (5)$$

which can be a substantial factor. Whether or not this is any problem depends entirely on the specific values of w and F_S in Eq. (4). One reason why most good densitometers are film plane scanners is that since the scanner always operates on the axis of the optical system, high-quality imagery only needs to be maintained over a field of view which is determined by the aperture dimensions.

Suppose that we have designed reimaging optics which are diffraction limited over a satisfactory field of view and suppose that the optics are such that $(f/\ )_S \ll (f/\ )_T$. Have we satisfied our requirements for resolution in the scanner? Not necessarily, because the optical transfer function associated with our reimaging

operation depends not only on the optical properties of the lens system, but also on the manner in which the optics are illuminated. For example, in most inexpensive projection systems, visual observation of the exit pupil of the lens system reveals a dramatically nonuniform illumination, frequently in the form of a filament image. The optical transfer function for such a system would be found by performing the autocorrelation function over the exit pupil and would be much worse than that associated with the lens system. Therefore, if we are to maintain good resolution in the system, the light source and projection optics must be designed so that the exit pupil of the reimaging optics is uniformly illuminated.

In theory it can be argued that we can tolerate great loss of resolution in the scanner because the scanner optical transfer function acts on both the film signal and the film noise, and therefore does not lower the signal-to-noise (S/N) ratio providing the scanner noise is sufficiently low. This argument is not usually valid, because in most scanners there is significant partial coherence in the plane of the film and the coherent optical transfer function must be used. In this case we cannot fully recover from the loss of resolution by making corrections with incoherent transfer functions. A less sophisticated but important objection to the argument is that in the real world we seldom can make full recovery of losses which take place for a whole variety of frustrating nonfundamental reasons.

FLARE

Stray light of any origin is a matter which deserves serious consideration in any precision film reading system. With a good optical design, the stray light can usually be reduced to that associated with the reimaging lens system. It is important to keep this lens dust free. Even so, some residual dust, imperfections in the optics, and multiple reflections will generate stray light. Where the resolution and field of view requirements will allow, it is best to select a lens system having a small number of elements to reduce the effect. This can be accomplished with object space scanning, i.e., moving the film, since the field of view is very small. Another advantage of the object space scan is that a small aperture can be placed just behind the film, so that only that small region of the film to be examined is illuminated. This greatly reduces the total flux incident on the reimaging lens and therefore greatly reduces the flare. This cannot be generally accomplished with image plane scanners.

The effect of flare can be quite pronounced, and flare is not automatically compensated by gray scale calibration procedures. This fact can be visualized from the following example. Suppose we observe a piece of film on which we have a small gray scale with a dark surround. We read the gray scale with the scanner. Now we substitute an identical gray scale except that it has a bright surround. The

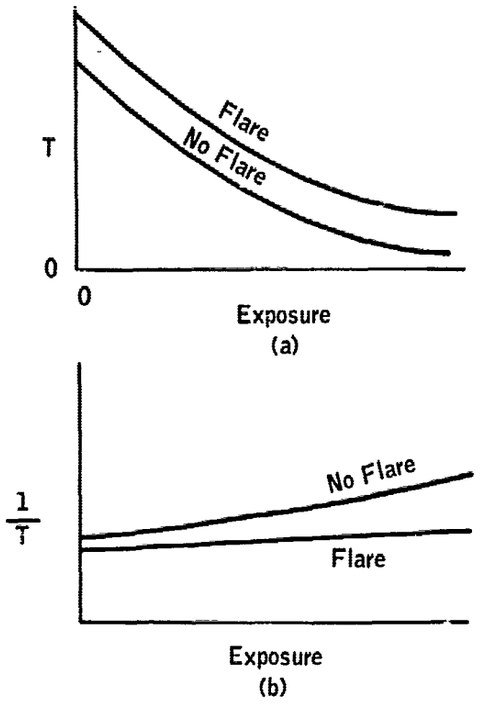


FIGURE 3 (a) Transmission readings; (b) reciprocal transmission readings.

flux incident on the reimaging lens will be increased and therefore the flare will be increased. The results of our transmission readings for the two cases might be as shown in Figure 3(a) or 3(b). Obviously, then, if the flare conditions for the images to be scanned are different from the flare conditions for the gray scales, the result will be a failure to calibrate properly for film nonlinearity.

For the scanning of small images, flare can be much reduced by masking the film to just the region to be scanned. Whether or not flare is an important consideration for a particular scanning operation must be determined by study of the specific conditions associated with the task.

FILM AND SCANNER NOISE LEVEL

We assume that noise on a photographic emulsion is described by Selwyn's equation,

$$\sigma_D = G(D/A)^{1/2} , \tag{6}$$

where σ_D is the standard deviation in density D, G is a constant, and A is the area of the scanning aperture. But

$$\sigma_D = K\sigma_T/T = G(D/A)^{1/2} , \tag{7}$$

so that

$$\sigma_T = (GT/K)\ (D/A)^{1/2} . \tag{8}$$

If we illuminate the film to be scanned with a flux density E, then the flux passing through the scanning aperture will be

$$F = EAT . \tag{9}$$

The change in F due to a film noise ΔT is

$$\Delta F = EA\Delta T , \tag{10}$$

so that

$$\frac{\Delta F}{F} = \frac{\Delta T}{T} = \frac{G}{K} (D/A)^{1/2} , \tag{11}$$

or in terms of photocell output

$$\frac{\Delta i}{i} = \frac{G}{K} (D/A)^{1/2} .$$

The photon shot noise associated with the reading would be

$$\sigma_N = \sqrt{2\epsilon EATS\Delta f} , \tag{12}$$

where S is the sensitivity in amperes per lumen, ϵ is the electronic charge, and Δf is the electrical bandwidth. The ratio of shot noise to the average current is

$$\sigma_N/i = \sqrt{2\epsilon \Delta f / EATS} . \tag{13}$$

We can rate the efficiency of the scanner by forming the figure of merit,

$$\gamma = \frac{\Delta i/i}{\sigma_N/i} = \frac{G}{K} \left[\frac{DETS}{2\epsilon \Delta f} \right]^{1/2} . \tag{14}$$

We make the observation from Eq. (14) that γ is independent of aperture size.

The total noise standard deviation will be

$$\sigma_T = \sigma_F \left[1 + (1/\gamma^2) \right]^{1/2} , \tag{15}$$

where σ_F is the rms film noise. It would be good to have $\gamma = 10$ in order to be sure that the film noise is truly limiting the system. From Eq. (14) we note that the lowest values of γ will result when T is small. We must therefore design for that condition. Equation (14) allows us to relate flux requirements to scanning speed. The G to be used in the equation should be selected for the finest-grain film which will be used in the scanner. There are no fundamental reasons why we cannot achieve scanning rates as high as Eq. (14) will allow.

HOW MANY GRAY SCALE STEPS?

Equation (7) indicates that σ_T/T is relatively insensitive ($D^{1/2}$) to film density. This suggests that a logarithmic gray scale is best suited to the task. We would like to have steps smaller than the film noise if we are not to lose information. If we specify a maximum and minimum transmittance and stipulate that the steps will be smaller than film noise by a factor K_1, then we can easily add up the number of steps since each step is an equal increment in density. The formula then becomes

$$n = \frac{K_1 (D_{max} - D_{min})}{\sigma_D} , \tag{16}$$

where n is the number of steps required. By substituting Selwyn's equation and approximating D by its average value, we obtain

$$n = \frac{K_1 (D_{max} - D_{min}) A^{1/2}}{G D_{av}^{1/2}} , \tag{17}$$

which makes it clear that the number of gray scale steps depends on the quality of the film through G, and also on the dynamic range of the film ($D_{max} - D_{min}$), the scanning aperture size A, and our precision factor K_1. For the films we have been using and for our 100-μ scanning aperture and a precision factor of $K_1 = 4$, a range of density of 2 would require approximately 2,000 steps.

APPENDIX 19

SOME CONSIDERATIONS WITH RESPECT TO THE DESIGN OF SCANNING MECHANISMS FOR RESTORATION OF PHOTOGRAPHIC IMAGES

R. V. Shack

The considerations to be discussed here are made under the assumption that the scanning process is intended to convey the pictorial information in a photographic negative or transparency with the minimum distortion of the pictorial information yet with the maximum discrimination against noise (grain structure), taking into account the fact that the image scanned is nonlinearly related to the exposure image which produced the picture.

First we note that the variable describing the scanned image is the transmission of the image, if we are to treat the scanning process as a linear operation, so that we can simply correct for the blur occurring in the scanning process, e.g., from diffraction by the reimaging lens and from the finite size of the scanning aperture. That this correction is necessary under optimal conditions will be demonstrated later. The proper calibration function relating the transmission of the developed image to the exposure which produced it is the transmission versus exposure curve. A typical curve is schematically illustrated in Figure 1. It is clear that high-contrast images will be transferred from exposure to transmission in a highly nonlinear fashion.

Now the exposure image is band-limited by the diffraction of the optical system which took the picture, and no pictorial information exists in the exposure image at spatial frequencies beyond the cutoff determined by diffraction. However,

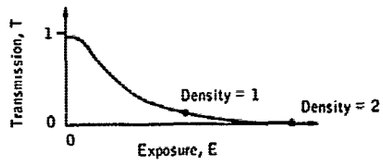


FIGURE 1 Transmission versus exposure curve.

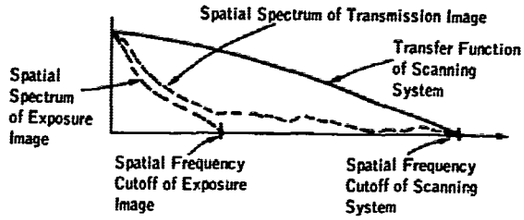


FIGURE 2 Spatial spectra of exposure and transmission images.

the transmission image, being derived from the exposure image through a nonlinear transfer, will contain spatial frequency components beyond this cutoff, and if they were to be removed in the scanning process, then the implied exposure image obtained by transferring back through the nonlinear conversion would be distorted. Thus the transfer function of the scanning system must extend further out in the spatial frequency range than the transfer function of the system which took the picture (Figure 2).

Even apart from the above consideration of generated high-frequency content, it is desirable to have the transfer function of the scanning system extend further than that of the taking system, not because of the extended frequency range, but in order to retain a reasonably high response over the frequency range of the taking system.

The scanning system also passes noise in the form of grain structure in the image, and the small scanning spot implied from the long transfer function means that the scanning output will of course be much noisier than it would be if its cutoff were made to match that of the taking system. This condition, however, is unimportant with appropriate treatment, as will be seen later.

Now let us consider the sampling interval for digitizing. First we note that if a sampling interval appropriate to the cutoff frequency of the exposure image were chosen, some of the frequency content generated by the nonlinear process would be folded back into the spatial frequency range of the exposure image; but what is more important, the noise spectrum would be folded back too, making the apparent noise level higher than is necessary. If instead we choose a sampling interval appropriate to the scanning system so the structural information passed through it is not aliased, then we can correct for the drooping transfer function without introducing additional noise in the system.

Note that this sampling interval implies considerable overlap in the spread function of the scanning system from one sample point to another.

After the latter correction is applied to the scanned image, we have an accurate representation of the transmission image out to the frequency limit of the scanning system. If we are careful to have this limit far enough out, no ringing effects will be observed in the transmission image structure, but the sharp cutoff will affect the appearance of the grain. In any case, the sharp cutoff can be ameliorated by rounding off the correction at the high-frequency end. This rounding-off corresponds to suppressing the wings of the corrective convolution function.

Once this accurate representation of the transmission image is obtained, the image in the spatial domain can be transferred back through the T versus E curve to obtain the exposure image. The latter will now be band-limited to the cutoff of the taking system, and our sample interval contains redundant information about the image. This redundancy can be used to reduce the apparent noise by increasing the data interval to that appropriate to the cutoff frequency of the exposure image by appropriate averaging (corrective convolution or equivalent operation in the spatial frequency domain), thus reducing the noise level to that which existed in the transmission image over the same frequency range. This residual noise is irreducible for a single image, and the implied exposure image we now have is in the best shape we can obtain for restoration processing.

APPENDIX 20

A CONCEPTUAL DESIGN FOR AN IMAGE SPACE SCANNER FOR IMAGE PROCESSING

H. L. Kasnitz

In this paper a conceptual design is given for an image space scanner of limited precision but very high speed, built around a standard vidicon. The design provides for production of up to 250,000 picture elements. The system is self-storing and no computer storage of these input data is required. In addition, the scanner may be used on-line in conjunction with a coherent optical processing facility and provides a mechanism to couple optical processing techniques with digital computer processing techniques, as suggested by Huang and Kasnitz in Appendix 28.

LINCOLN LABORATORY VIDICON SCANNER

A new vidicon scanner[1] is being developed at Lincoln Laboratory. The most important characteristic of this development is that by appropriate treatment of a standard vidicon, a linear dynamic range of three orders of film density has been achieved.

This system operates at standard TV rates of 30 frames/sec. It is presently being used to provide a high magnification analog viewer for the examination of film recorded spectroscopic data. The readout beam on the retina of the vidicon is approximately 50 μ in diameter, and with the 20 times optical magnification usually used, this corresponds to a 2.5-μ spot on the filmed data. The system has resolved a set of 50-μ bars on the retina with 15 to 20 percent modulation.

The system is regularly used to study images recorded on Kodak 4X negative film. With this film, grain noise is clearly discernible and electrical noise is much lower than grain noise.

The system will read density to an accuracy of 0.015 to 0.020 density unit, corresponding to an accuracy of about 3 to 5 percent in transmitted light. The system is linear over three orders of film density with a gamma of about 0.7.

The output of the present system is a conventional TV display. A logarithmic amplifier is used to convert the intensity readings to density. This also compresses the range of the output to allow for convenient display on the TV monitor. An adjustable and positionable synthetic electronic slit is available, and an A-scope arrangement displayed on a test oscilloscope presents selected data as a conventional trace of density versus distance on the film.

As in all computer-associated scanners, stability or repeatability is of prime importance. The computer can linearize the position coordinates of any stable scanning device. With image space scanners, stability in the response of the system to input light intensity is also required. Photosensitive surfaces will vary in their response as a function of position on the photosensitive surface. This variation on the vidicon is of the order of 15 to 20 percent. It can be corrected by mapping the intensity response of the system as a function of position and applying a correction function in the computer up to the uncertainty inherent in the measurement of density.

The investigation of the wide dynamic range vidicon is still in progress and all results are to be considered very preliminary.

DESIGN CONSIDERATIONS FOR AN IMPROVED SCANNER FOR IMAGE RESTORATION PROCESSING

The vidicon described above provides a good match to the films used for recording the images which must be restored. The design of a suitable scanning system will be detailed.

Geometrical Considerations

We use a vidicon of the general characteristics listed in the preceding section.

Resolution element size on the vidicon: 50-μ diameter
No. of resolution elements: 500 x 500 = 250,000
Field size at 1 to 1 magnification: 25 x 25 mm
Field size at 50 to 1 magnification: 500 x 500 μ

A set of lenses installed in a turret will give a selected set of resolution elements and corresponding frame sizes varying over this range. At the highest

magnifications, the lens design may limit the effective resolution. Optical distortion and vignetting effects due to each lens may be mapped along with the vidicon characteristics and the entire system corrected by the computer in one operation. The active lens is coded into the computer, and the computer may, if desired, be programmed to select the appropriate magnification.

Assume, for illustrative purposes, that it is desired to read, at any single resolution available with the lens turret, an image which extends over a 10 in. x 10 in. plate. For such a requirement, one may purchase as a stock item an air-bearing stage which can be motor driven at a rate of 2 in. per sec or faster, while the motion of the stage is continuously monitored to a precision of 1 μ. In operation, the stage would be stopped at a known coordinate position, and the data would be read providing for a matrix of 500 x 500 resolution elements. When these data are processed by the computer, the stage would be moved a precisely known distance in one or both axes, and the next frame read. Precision servos are not required. Precision encoding is necessary. In general, the stage would be moved a little less than the size of the scanning field that is used. The computer would have precise information as to the amount of movement and could reject the redundant data points.

The computer-scanner combination can thus bootstrap itself to read its way across a large plate. It could, for example, read data at the center of a plate and then find and read edge fiducial marks. The lens turret should not be moved during such an operation, as it is not possible to mechanically reindex a lens axis to better than a few thousandths of an inch accuracy with a quick-change turret.

Increased stage position accuracy could be obtained with laser interferometers if needed.

It is thus seen that the limitation of the scanner to a 500 x 500-element field need not limit the applicability of such a system. With suitable transports and stages, filmed data of any size and any format may be accommodated to the highest resolution of the scanner.

Data Rates

It is arbitrarily decided that the scanner will be operated at a 30-cycle framing rate to produce 30 noninterlaced fields, each containing the full 250,000 picture elements. In order to get satisfactory intensity readings from the vidicon, this rate cannot be changed and must be accurately clocked. To insure repeatability, digital circuits will be used to generate the line and point deflection system. The point driving system must operate at the rate of

$$500\ (500)\ 30 = 7.5\ \text{Mc/sec}.$$

The line stepping rate is

500 (30) = 15,000 cycles/sec.

Flyback time is ignored in this calculation. These rates may easily be achieved with modern solid-state digital circuits.

The digital position generator would drive analog sweep circuits with a stability of one part in 2,000 or better. Thus, for each sample point in the scanning matrix, there would exist a unique digital address that should be repeatable to the order of one quarter of the beam resolution element in each coordinate. Some CRT scanners do better than this. The actual system will include line synchronization marker codes, and may have an automatic settling time built in to allow for flyback and line increment transients.

There is an operational reason for insisting on the 30-cycle framing rate. We wish to provide on a TV monitor a continuously available picture of the film plane being scanned. In image-restoration processing, this means that the degraded image will always be visible as long as the film is being scanned. Operator decisions, based on the image being scanned, may be made when necessary and communicated to the computer by means of a light pen or other appropriate indicator directly from the picture. The TV-type display is driven from the same circuits that drive the scanner, although the precision of the visual display need be no better than one part in a few hundred if it is to be used for viewing only. If this display is also to be used for photocopying of the restored data, full precision must be preserved. The use of separate equipments for visual display and for photography is strongly recommended.

The Measurement of Incident Light Intensity

It is assumed that the scanner covers a range of densities of three orders of magnitude, corresponding to a range of input light intensity to the scanner of 10^3. One arbitrarily decides to digitize this range to 11 bits or one part in 2,000. It is true that the basic performance of the scanner may be good to only 3 to 5 percent; but the computer typically handles words of 16 or more bits and will carry the extra bits in any case. This gives a grand total of

500 x 500 x 30 x 11 = 82.5 megabits/sec

produced by the scanner.

The TV monitor display discussed above will not have a dynamic range of 10^3. It can most economically be driven, in parallel with the digitization process, directly by the analog signal. Adjustable clipping and threshold levels can be

introduced into the analog signals which drive the display (they would not affect the digital intensity values), and different portions of the intensity range of the original picture could be displayed at the operator's option. It may be desirable to compress this signal with a logarithmic amplifier so that a signal proportional to the film density would be produced. Digitization of the density and introduction of this value into the computer is possible. Digitization to 512 parts or 9 bits would give a density increment of about 0.006, which is of the order of a factor of 2 better than the performance claimed for the vidicon in the first section of this paper. In image-restoration processing, there seems to be no point to unnecessarily introduce the film density.

The Input Buffer

Let us now consider whether a buffer can be built to accept 11-bit words at a 7.5-mcps rate. This can definitely be done. One has a choice of using modern integrated flip-flop logic or a high-speed core buffer. Lincoln Laboratory is presently building a system which includes buffers capable of accepting words at a 10-mcps rate.

We must also ask whether it is possible to digitize a voltage to 11 bits at a 7.5-mcps word rate. The answer to this is a qualified yes. There exist A/D converters capable of providing 9 bits at a 10-mcps rate, and it is highly possible that 11 bits at 7.5 mcps has already been accomplished. We may, at worst, have to sacrifice one or two bits at this point, but a digital resolution of ±2 parts in 1,000 is probably better than the film.

Thus a suitable input buffer can be achieved. A buffer to handle the entire scanning field would require 250,000 words. This is manifestly impractical.

The Photograph is a Disk File

One should recall that the requirements for the scanner driver called for line and point counters so that a unique digital address was associated with every resolvable picture element. The picture being scanned in effect constitutes a disk file with an access time to any point (or any predetermined group of points) of 33 msec. This is an illustration of the principle that when a proper scanner is available, a picture is its own best storage.

One may now determine how large the special purpose high-speed input buffer must be. The buffer must contain the number of points that the computer and the program can effectively use in a 33-msec interval.

The optimum size of the buffer can only be determined when the computer configuration and the programming details are approximately known. One can, however, make some estimates. For a 1-μsec cycle computer, the buffer cannot

be larger than 33,000 words. Indeed, it cannot be anywhere near that large, since one must allow time for the incoming data to be processed. Assume, for example, that the buffer contains 1,000 words. Then 33 computer cycles are available to process each input element of data. This estimate is probably good to within a factor of 4.

Note that it is not meaningful to ask how long it would take to read the data into storage. The data being scanned <u>are</u> in storage. If it should require 33 computer cycles (of 1-μsec duration) to input process each data point for image restoration processing, then the entire field could be input processed in 500(500) x 33×10^{-6} = 8.25 sec. In general, it will not be necessary to process all of the field since there is no point in processing areas on the film that do not contain data of interest. The TV-type input display could be used by the operator to indicate the areas to be processed to the computer with a light pen. As the program becomes more sophisticated, it could search the field, find the areas of data, and read only those regions.

The Pseudo-Random Image Space Scanner

In order for this system to work, the operation of the buffer must be synchronized with the scanner, and an addressing system must be provided so that the computer may direct the buffer to read in the required block of data on the next scan. In the most general case, every digital address in the picture must be available to the computer as the starting point of a data block. It is probably sufficiently general to require that the computer be able to select rectangular blocks parallel to the coordinate axes of any extent in both coordinates as long as the buffer size is not exceeded. The computer must associate these coordinate read orders with each data block, so that the geometry of the picture may be reconstructed.

A CRT object space scanner is a powerful instrument because it is completely addressable. This image space scanner is also completely randomly accessible for a single point at a 33-msec rate. This is interesting but not very efficient. However, there is no built-in restriction in this system which requires that successive blocks be contiguous or that successively demanded blocks have the same shape. Thus a true random scanning capability is achieved in the selection of successive blocks of data. The necessary repetitive scanning of the complete frame by the vidicon is not disturbed. It neither knows nor cares what the buffer is doing.

Block operation is suitable for operation with a large time-sharing computer. A CRT scanner under complete computer control requires an order every 20 μsec or so. Time-sharing computers are not available at this frequency. A set of orders and a data transfer from the scanner to the main frame every 33 msec is quite in order with modern time-sharing concepts.

Tricolor operation for XR film is easily available by sequentially changing color filters under computer control. The system can read a Fourier transform, imaged on its retina, with the same facility that it can read a film negative, so that the inherent advantages of both coherent optics and computer processing may be at least conceptually realized.

REFERENCE

1. J. Rotstein and R. Billups, "An Imaging Microdensitometer," presented to the Society of Photo-Optical Instrumentation Engineers, 11th Technical Symposium, St. Louis, Mo., Aug. 22-26, 1966.

APPENDIX 21

PHOTOELECTRIC ARRAYS OR EQUIVALENTS

James L. Harris, Sr.

The success of image processing is greatly dependent on the signal-to-noise (S/N) ratio associated with the original image recording. Any step in the direction of improving this S/N ratio is therefore a substantial step in the direction of giving maximum yield from image processing techniques. Photon noise limited measurement is the desired goal.

The sensor elements which are convenient to use, such as film and TV camera systems, fall considerably short of attaining the precision of measurement set by the photon limit. This paper is directed toward a discussion of photoelectric sensor units which can achieve substantially higher S/N ratio.

Before we consider possible instrumentation, let us study the potential improvement which is possible. For this purpose it is convenient to use the numerical example which has been worked out for the case of film in Appendix 11. In that appendix, Courtney-Pratt shows that it is not unreasonable with Plus-X film to photograph a satellite of luminance B = 1 candle cm^{-2}, with a "resolution" on the order of 0.1 sec of arc, with a telescope entrance pupil diameter of 48 in., a focal length of 900 in., a transmission factor of 0.25, and an exposure time of 3.5×10^{-3} sec. My interpretation is that he feels Plus-X film is a reasonable choice for this task.

He finds that the film noise corresponding to 0.1-sec-of-arc resolution is equivalent to an rms density variation of 0.063. If

$$D = \gamma \log E, \tag{1}$$

where D is the density, γ is the slope of the D-log E curve and E is the exposure, then

$$D + \Delta D = \gamma \log (E + \Delta E) , \tag{2}$$

and subtracting Eq. (1) from Eq. (2),

$$\Delta D = \gamma [\log (E+\Delta E) - \log E] , \tag{3}$$

or

$$\Delta D = \gamma \log [1 + (\Delta E/E)] , \tag{4}$$

so that

$$\Delta E/E = \log^{-1} (\Delta D/\gamma) - 1 . \tag{5}$$

For a ΔD of 0.063 and $\gamma = 1$,

$$\Delta E/E = 0.16 . \tag{6}$$

Now let us examine the performance of a photoelectric device. The flux, in lumens, associated with the observation is

$$F = B \Omega A_L T_T \text{ lm} , \tag{7}$$

where B is the luminance of the object, Ω is the solid angle associated with the resolution element, T_T is the transmission of the telescope system, and A_L is the area of the entrance pupil of the telescope. If this flux were incident on the photocathode of a multiplier phototube having a photocathode sensitivity of S amperes per lumen, the current would be

$$i = B \Omega A_L T_T S \text{ A} . \tag{8}$$

The photon shot noise associated with the measurement would be

$$i_n = [2 \epsilon i \Delta f]^{1/2} \text{ A} , \tag{9}$$

where i_n is the rms noise current, ϵ is the electronic charge, and Δf is the electrical bandwidth of the receiver. For an equivalent to $\Delta E/E$ we form

$$i_n/i = \left[\frac{2 \epsilon \Delta f}{B \Omega A_L T_T S} \right]^{1/2} \tag{10}$$

For the numerical values of Courtney-Pratt's example, and a sensitivity S of 10^{-4} A/lm,

$$i_n/i = 0.0173,$$

which shows a precision of measurement improved by approximately a factor of 10 over that given in Eq. (6).

An S/N ratio improvement of a factor of 10 is very significant. It would mean that, for the observation described in the example, image processing could be achieved at spatial frequencies whose modulus is down by a factor of 10 from that achievable with a film image, i.e., significantly higher resolution. It can also mean any of the compromises implied by Eq. (10), such as operation on objects having 1/100th of the luminance, or exposure times of 3.5×10^{-5} sec (important for the case of very rapidly tumbling objects). Additional benefits of the photoelectric array might include increased dynamic range and maximum use of available observation time.

If we agree that there would be great merit in achieving the S/N ratio improvement just described, what type of instrumentation is required? The calculation was made as if we had an array of multiplier phototubes. For a 5-sec-of-arc field of view and 0.1-sec-of-arc resolution we would require 50 x 50 = 2500 elements in the array, or for a 10-sec-of-arc field of view, 100 x 100 = 10,000 elements.

If we picture conventional large multiplier phototubes, the prospect of arrays with 10,000 or more elements is awe inspiring. A more plausible picture is obtained by considering the developments currently under way in small channel multipliers. Arrays of such elements could provide a reasonably compact sensor package. Perhaps there are solid-state photosensitive elements in development which rival photoemissive sensors in terms of S/N ratio.

There is also the question of data recording from such an array; i.e., what do we do with the 10,000 outputs? We might visualize terminating each output in a simple low-pass filter and sequentially sampling the array. If we think ahead to the time of possible on-line operation with special purpose computers, then perhaps it is not unreasonable to assume that we might carry the image matrix into the computer in parallel, rather than filling an internal computer matrix sequentially.

Such a large array of detectors would undoubtedly have significant variations in individual sensitivities and calibration procedures would be required, as, for example, periodic uniform illumination of the array. Stability of the sensitivities would be required to the level of the precision of measurement (in the example, 1 percent) during periods between calibration.

A possible substitute for the concept of an array may be that of special use of television-type systems, perhaps involving image intensifiers and vidicons, specifically designed in terms of scan rates and spot size, for example, to meet the needs of this application. The critical question is simply: How close can such a system come to the S/N performance achievable with an array?

The number of if's, perhap's, and maybe's contained in this paper is sufficient indication that, within the limitation of knowledge of the author, an array or its performance equivalent in a TV system will not be placed in operation in the very near future. However, we would conclude that the final image quality achievable with such a sensor package warrants a reasonable research and development effort directed toward this goal.

APPENDIX 22

DIGITAL FOURIER ANALYSIS

T. S. Huang

THE PROBLEM

We wish to calculate the Fourier transform of a complex function g (x, y) of two spatial variables x and y. The function is zero outside a bounded region in the x-y plane. We wish to do the calculation on a digital computer.

For the sake of simplicity, we consider the one-dimensional case. The extension to two dimensions is straightforward.

Let g(x) be a complex function of one spatial variable x, and let g(x) = 0 for $x < -A$ and $x \geq A$. We wish to obtain the Fourier transform of g(x):

$$G(u) = \int_{-\infty}^{\infty} g(x)\, e^{-i 2\pi u x}\, dx \qquad (1)$$

where u is the spatial frequency in cycles per unit length. The function G(u) is in general complex, and has an infinite duration. See Figure 1(a).

To calculate G(u) on a digital computer, we have to sample and quantize g(x). The effect of amplitude quantization will be discussed later. Now, we only consider the effect of sampling. Assume that samples are taken from g(x) with a period Δx, to get

$$\hat{g}(x) = g(x) \sum_{k=-A/\Delta x}^{(A/\Delta x)-1} \delta(x - k\,\Delta x), \qquad (2)$$

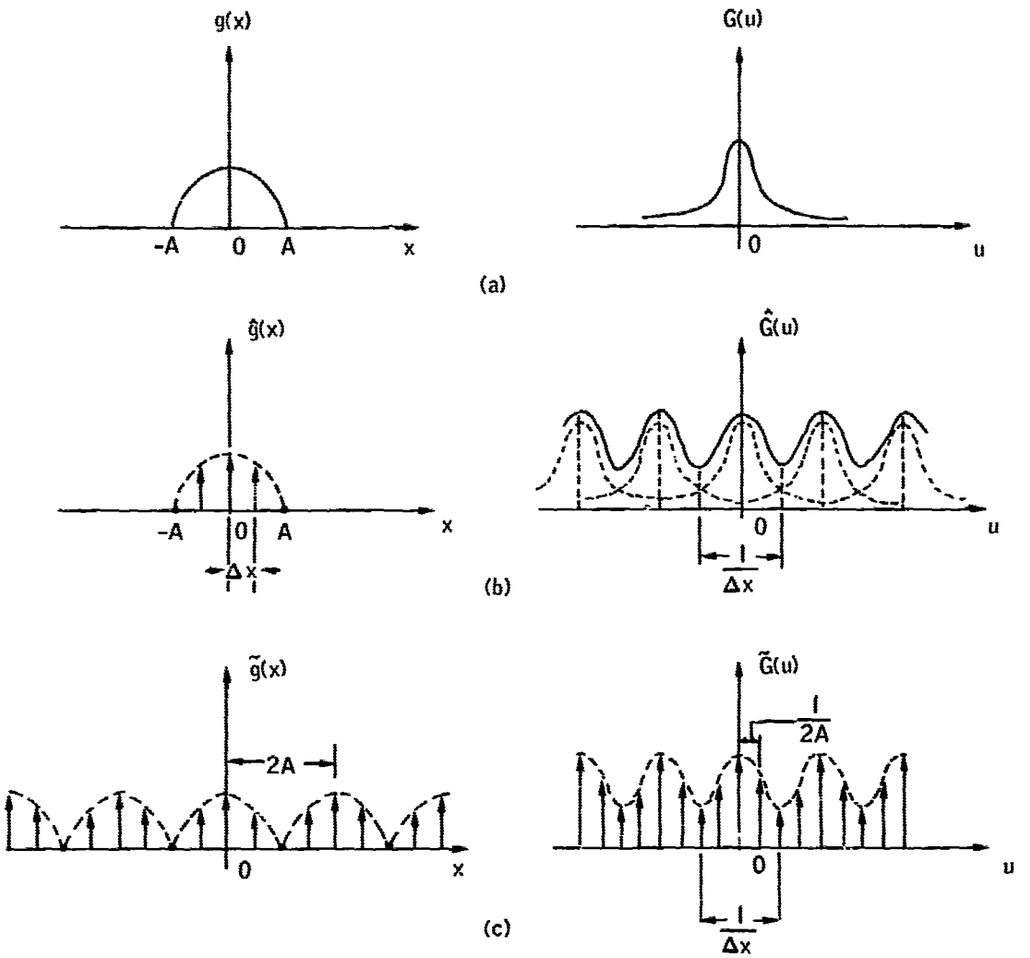


FIGURE 1 A function g(x) and its Fourier transform G(u). (a) Function and transform defined over an infinite interval; (b) effect of sampling g(x); (c) effect of sampling $\hat{G}(u)$.

where $\delta(x)$ is the Dirac delta function. The total number of samples is

$$N_i = 2A/\Delta x , \tag{3}$$

assuming, for simplicity, that $A/\Delta x$ is an integer. What we calculate on the computer is then actually the Fourier transform of $\hat{g}(x)$:

$$\hat{G}(u) = \sum_{k=-\infty}^{\infty} G\left[u - k\,(1/\Delta x)\right] , \tag{4}$$

which is periodic with a period $1/\Delta x$. See Figure 1(b). It is only necessary to calculage $\hat{G}(u)$ for one period, say $-1/2\Delta x \leq u < 1/2\Delta x$. Since G(u) has an infinite duration, aliasing occurs in $\hat{G}(u)$, and

$$\hat{G}(u) \neq G(u), \quad -1/2\Delta x \leq u < 1/2\Delta x \ . \tag{5}$$

However, one can always choose Δx small enough so that (5) becomes essentially an equality.

On a digital computer, the answer also comes in sampled form. Since $\hat{g}(x)$ is strictly limited in duration, it is sufficient to sample $\hat{G}(u)$ with a period $1/2A$ (or smaller). Assume that a sampling period $1/2A$ is used so that we are actually obtaining

$$\tilde{G}(u) = \hat{G}(u) \sum_{k=-\infty}^{\infty} \delta[u - k(1/2A)] \ , \tag{6}$$

which is the Fourier transform of

$$\tilde{g}(x) = \sum_{k=-\infty}^{\infty} \hat{g}(x - k2A) \ , \tag{7}$$

which is periodic with a period 2A. See Figure 1(c). We are therefore actually carrying out a Fourier series calculation, rather than a Fourier transformation, on the digital computer.

Since we only need to calculate $\tilde{G}(u)$ for $-1/2\Delta x \leq u < 1/2\Delta x$, the number of samples in the answer is

$$N_o = \frac{1/\Delta x}{1/2A} = \frac{2A}{\Delta x} \ . \tag{8}$$

We observe that

$$N_o = N_i \ , \tag{9}$$

i.e., the number of output samples is equal to the number of input samples.

Notice that one can get $\tilde{g}(x)$ back by calculating the inverse Fourier transform (or rather series) of $\tilde{G}(u)$. This $\tilde{g}(x)$ will agree with g(x) at the sampling positions, whether or not aliasing occurred in $\hat{G}(u)$.

Let the values of the samples of g(x) be g_0, g_1, g_2, ..., g_{N-1} and those of $\hat{G}(u)$ be G_0, G_1, G_2, ..., G_{N-1}. Then

$$G_k = \sum_{j=0}^{N-1} g_j w^{-jk} , \tag{10}$$

where

$$w = e^{i(2\pi/N)} . \tag{11}$$

Equation (10) is a discrete approximation of Eq. (1), and can be written in matrix form:

$$G] = W g] , \tag{12}$$

where

$$G] = \begin{bmatrix} G_0 \\ G_1 \\ . \\ . \\ . \\ G_{N-1} \end{bmatrix} \tag{13}$$

$$g] = \begin{bmatrix} g_0 \\ g_1 \\ . \\ . \\ . \\ g_{N-1} \end{bmatrix} \tag{14}$$

and

$$W = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w^{-1} & w^{-2} & \cdots & w^{-(N-1)} \\ 1 & w^{-2} & w^{-4} & \cdots & w^{-2(N-1)} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 1 & w^{-(N-1)} & w^{-2(N-1)} & \cdots & w^{-(N-1)^2} \end{bmatrix} . \qquad (15)$$

TIME ESTIMATES – DIRECT METHOD VERSUS COOLEY-TUKEY ALGORITHM

Assume that we take N samples from g(x). Then we need N samples for the Fourier transform, which can be obtained by carrying out the matrix multiplication of Eq. (12). To do this matrix multiplication directly, we need N^2 basic operations, a basic operation being defined as a complex multiplication plus a complex addition. To get each of the N output samples, we need N basic operations on the input samples. Therefore, the computer time required for the direct method is

$$T_d = k_d N^2 , \qquad (16)$$

where k_d is a constant, depending on the particular computer and the particular program.

An efficient method of digital Fourier analysis, which had been suggested by Good,[1] was recently developed by Cooley and Tukey.[2] This method is based on the following theorem.

Let $N = r_1 \times r_2 \times \cdots \times r_n$, where $r_1, r_2, \ldots, r_n$ are positive integers. Then the N x N matrix W of Eq. (15) can be factored into n matrices:

$$W = W_1 W_2 \cdots W_n , \qquad (17)$$

where the N x N matrix W_i has only $r_i N$ nonzero elements.

It follows from this theorem that Eq. (12) can be written as

$$G] = W_1 W_2 \cdots W_n g] . \qquad (18)$$

If we do the matrix multiplication step by step, multiplying W_n and g] first, then multiplying W_{n-1} and the product $W_n g]$, etc., then the computer time required will be

$$T_c = k_c(r_1 + r_2 + \ldots + r_n)\, N, \tag{19}$$

where k_c is a constant.

The case $r_1 = r_2 = \ldots = r_n = 2$ offers important advantages for computers with binary arithmetic, both in addressing and in multiplication economy. In particular, the entire calculation can be performed within the array of 2N storage locations used for the input. (The input samples are in general complex; hence, each sample takes 2 storage locations.) For this special case the computer time required is, according to Eq. (19),

$$T_c = k_c\, 2N \log_2 N \ . \tag{20}$$

The constant k_c, of course, depends on the particular computer and the particular program.

We note that while in the direct method the computer time, T_d, is proportional to N^2, in the Cooley-Tukey method the computer time, T_c, is proportional to $N \log_2 N$. Therefore, for large values of N, considerable savings in computer time can be achieved by using the Cooley-Tukey algorithm.

Wolf[3] has written a program for the Cooley-Tukey method in the MAD language, and has run it on an IBM 7094 computer, which has a cycle time of 1 μsec. It was found that

$$k_c \approx 30\ \mu\text{sec} \ . \tag{21}$$

For the direct method on the same computer, a reasonable estimate for k_d is

$$k_d \approx 30\ \mu\text{sec} \ . \tag{22}$$

Using these figures, we have

For N = 64 x 64,	T_d = 8 min,	T_c = 3 sec	(23)
For N = 256 x 256,	T_d = 30 hr,	T_c = 1 min	
For N = 512 x 512,	T_d = 20 days,	T_c = 5 min	
For N = 1024 x 1024,	T_d = 1 yr,	T_c = 20 min	

The above estimates were made based on the assumption that the computer memory is large enough to store all the input samples. The problem of insufficient storage will be discussed later. Also, the estimates do not include the input/output time. If a CRT is used to scan a real image on the film transparency, a reasonable estimate for input time is 30 μsec per sample. The total input time is then

$$T_{in} = 30\ N\ \mu sec\ , \tag{24}$$

which is small compared to T_d and T_c.

PRINCIPLE OF THE COOLEY-TUKEY ALGORITHM

We shall now sketch an indirect proof of the theorem of the preceding section, for the special case of $N = 2^n$. The proof can readily be extended to the general case.

For the sake of simplicity, let us take the case of $N = 2^3 = 8$. The input samples are $g_0, g_1, \ldots, g_7$ (at $x = 0, \Delta x, 2\Delta x, \ldots$), and the output samples $G_0, G_1, \ldots, G_7$ (at $u = 0, 1/2A, 2/2A, \ldots$). Instead of calculating the G_i's directly, we do it in several steps:

1. Let

$$\hat{g}(x) = \sum_{j=0}^{7} g_i\, \delta(x - j\,\Delta x) = \hat{g}_a(x) + \hat{g}_b(x - \Delta x)\ , \tag{25}$$

where

$$\hat{g}_a(x) = \sum_{j=0,2,4,6} g_i\, \delta(x - j\,\Delta x) \tag{26}$$

and

$$\hat{g}_b(x) = \sum_{j=1,3,5,7} g_i\, \delta[x - (j-1)\,\Delta x]\ . \tag{27}$$

Let the Fourier transforms of $\hat{g}_a(x)$ and $\hat{g}_b(x)$ be $\hat{G}_a(u)$ and $\hat{G}_b(u)$, respectively. Then

$$\hat{G}(u) = \hat{G}_a(u) + \hat{G}_b(u)\, e^{-\Delta x 2\pi u}\ . \tag{28}$$

Instead of calculating $\hat{G}(u)$ directly, we can calculate $\hat{G}_a(u)$ and $\hat{G}_b(u)$ and then use Eq. (28) to obtain $\hat{G}(u)$. Only 4 samples are required of each of $\hat{G}_a(u)$ and $\hat{G}_b(u)$, which have a period half that of $\hat{G}(u)$.

Let the values of the samples of $\hat{G}_a(u)$ and $\hat{G}_b(u)$ be (G_0', G_2', G_4', G_6') and $G_1', G_3', G_5', G_7')$, respectively; and assume these samples are taken at $u = 0, 1/2A, 2/2A, \ldots$. Then G_i and G_i' are related as shown in Step 3 of Figure 2 in flow graph notation.

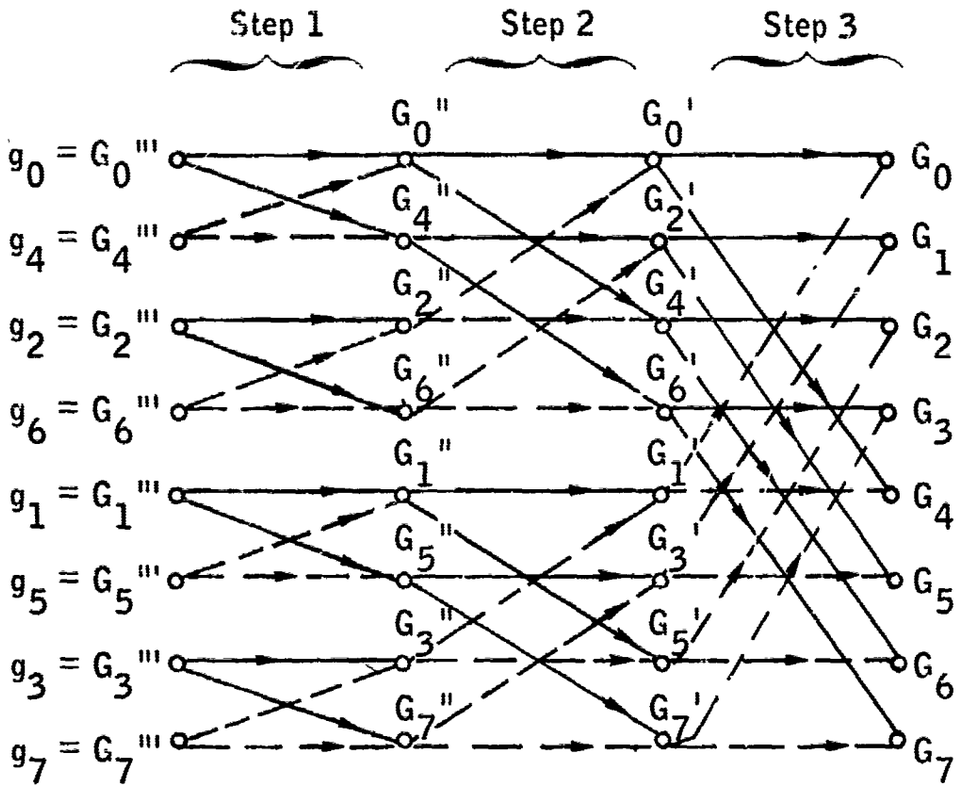


FIGURE 2 Flow graph. Solid line indicates unity factor; dashed line indicates nonunity factor.

2. Let

$$\hat{g}_a(x) = \sum_{j=0,2,4,6} g_i \, \delta(x - \Delta x) = \hat{g}_{aa}(x) + \hat{g}_{ab}(x - 2\Delta x) \quad , \tag{29}$$

where

$$\hat{g}_{aa}(x) = \sum_{j=0,4} g_i \, \delta(x - j\Delta x) \tag{30}$$

and

$$\hat{g}_{ab}(x) = \sum_{j=2,6} g_i \, \delta[x - (j-2)\,\Delta x] \,. \tag{31}$$

Let the Fourier transforms of $\hat{g}_{aa}(x)$ and $\hat{g}_{ab}(x)$ be $\hat{G}_{aa}(u)$ and $\hat{G}_{ab}(u)$, respectively. Then

$$\hat{G}_a(u) = \hat{G}_{aa}(u) + \hat{G}_{ab}(u)\, e^{-i2\Delta x 2\pi u} \,. \tag{32}$$

Instead of calculating $\hat{G}_a(u)$ directly, we can first calculate $\hat{G}_{aa}(u)$ and $\hat{G}_{ab}(u)$ and then use Eq. (32) to obtain $\hat{G}_a(u)$. Only two samples are required of each of $\hat{G}_{aa}(u)$ and $\hat{G}_{ab}(u)$, which have a period half that of $\hat{G}_a(u)$. Let the values of the samples of $\hat{G}_{aa}(u)$ and $\hat{G}_{ab}(u)$ be (G_0'', G_4'') and (G_2'', G_6''), respectively; and assume the samples are taken at u = 0, 1/2A, 2/2A, ... Then G_i'' and G_i' are related as shown in Step 2 of Figure 2.

A similar discussion applies to $\hat{G}_b(u)$.

3. Let

$$\hat{g}_{aa}(x) = \sum_{j=0,4} g_i \, \delta(x - j\Delta x) = \hat{g}_{aaa}(x) + \hat{g}_{aab}(x-4\Delta x) \,, \tag{33}$$

where

$$\hat{g}_{aaa}(x) = g_0 \, \delta(x) \tag{34}$$

and

$$\hat{g}_{aab}(x) = g_4 \, \delta(x) \,. \tag{35}$$

Let the Fourier transforms of $\hat{g}_{aaa}(x)$ and $\hat{g}_{aab}(x)$ be $\hat{G}_{aaa}(u)$ and $\hat{G}_{aab}(u)$, respectively. Then

$$\hat{G}_{aa}(u) = \hat{G}_{aaa}(u) + \hat{G}_{aab}(u)\, e^{-i4\Delta x 2\pi u} \,. \tag{36}$$

Instead of calculating $\hat{G}_{aa}(u)$ directly, we can first calculate $\hat{G}_{aaa}(u)$ and $\hat{G}_{aab}(u)$ and then use Eq. (36) to obtain $\hat{G}_{aa}(u)$. Only one sample is required of each of $\hat{G}_{aaa}(u)$ and $\hat{G}_{aab}(u)$, which have a period half that of $\hat{G}_{aa}(u)$. Let the values of the samples of $\hat{G}_{aaa}(u)$ and $\hat{G}_{aab}(u)$ be G_0''' and G_4''', respectively; and assume the samples are taken at u = 0, 1/2A, 2/2A, ... Then, the G_i''' and G_i'' are related as shown in Step 1 of Figure 2.

Similar discussions apply to $\hat{G}_{ab}(u)$, $\hat{G}_{ba}(u)$, and $\hat{G}_{bb}(u)$.

4. The calculation of the Fourier transform of a single sample is trivial. In fact,

$$G_j''' = g_j . \tag{37}$$

In summary, in this case of $N = 2^3 = 8$, we can calculate G_j ($j = 0, 1, \ldots, 7$) in $3 = \log_2 8$ steps:

$$g_j = G_j''' \rightarrow G_j'' \rightarrow G_j' \rightarrow G_j \qquad \text{(Figure 2)}$$

At each step, one does $16 = 2 \times 8$ basic operations. The total number of basic operations is $2 \times 8 \times \log_2 8$.

It is obvious that for $N = 2^n$, the total number of basic operations is $2N \log_2 N$. This proves the theorem of the preceding section. From a flow graph such as the one shown in Figure 2, one can figure out the factor matrices W_1, W_2, ...

COMPUTER STORAGE REQUIREMENTS

As mentioned earlier, the storage requirement of the Cooley-Tukey method is 2N, where N is the number of input samples. After the calculation, the output samples will appear in these same 2N storage locations. No separate storage is required for the output. Of course, some additional storage is required for the program and for generating the exponential factors.

It is clear from Figure 2 that each output sample depends on all the input samples. Therefore, in order to achieve the time estimates stated earlier, it is necessary to store all the input samples in the computer before the calculation. If the computer memory is smaller than 2N, what can we do? One possible answer is to use some auxiliary storage and do the calculation in several stages.

We see from Figure 2 that:

1. At each step, each output sample of the step depends on only two of the input samples of that same step.

2. Each output sample of the kth step depends on 2^k original input samples (i.e., input of Step 1).

Let the number of input samples be $N = 2^n$. Assume the computer can store at any one time only 2^k samples, where $k < n$. Then we can first calculate the output samples of the kth step (cf. Figure 2) in $2^n / 2^k = 2^{n-k}$ chunks. For each chunk of 2^k output samples (of the kth step) only 2^k samples of the original input are required. Therefore, we read in the original input samples 2^k at a time,

calculate 2^k samples for the output of step k, and put them in auxiliary storage (disk file, say). After we have all the output samples of step k in auxiliary storage, we start to calculate the output samples of the (k + 1)th step. Each output sample of the (k + 1)th step depends only on two output samples of the kth step. However, to save time in communication between the computer and the auxiliary storage, we probably should bring in from the auxiliary storage 2^k output samples of the kth step at a time (2^{k-1} from the first chunk and 2^{k-1} from the second chunk, to start with) to calculate 2^k output samples of the (k + 1)th step. These 2^k newly calculated samples can be put in the auxiliary storage in the same locations where the 2^k output samples of the kth step were. After all the output samples of the (k + 1)th step are calculated and put into the auxiliary storage, we start to calculate the output samples of the (k + 2)th step, and so forth, until we get the final output; i.e., the output samples of the nth step.

It is seen that in this procedure, the required capacity of the auxiliary storage is 2N locations (for N complex samples). The input samples have to be read in only once, but not sequentially. Therefore, unless the read-in device (e.g., a CRT) has random access to the input data, the entire input probably has to be read in (but not stored) 2^{n-k} times. The total traffic between the main computer and the intermediate storage is (2n - 2k - 1)N samples. The total time required for this procedure, excluding output time, is therefore

$$T_c' = T_c + T_{in}' + T_s , \tag{38}$$

where

$$T_c = k_c\, 2\, N \log_2 N , \tag{39}$$

$$T_{in}' = k_{in}\, 2^{n-k}\, N = k_{in}\, 2^{2n-k} , \tag{40}$$

and

$$T_s = k_s\, (2n - 2k - 1)\, N . \tag{41}$$

Since the traffic between the main computer and the intermediate storage is essentially sequential, a reasonable estimate for k_s in the case of disk files is

$$k_s \approx 100\ \mu\text{sec} \tag{42}$$

As discussed before, $k_c \approx 30\ \mu$sec, and $k_{in} \approx 30\ \mu$sec.

Let us consider a numerical example. Suppose $N = 256 \times 256 = 2^{16}$ ($n = 16$), and suppose the main computer can handle only $2^{12} = 64 \times 64$ ($k = 12$) samples (as in the case of the IBM 7094 computer which has a capacity of 32K 36-bit words). Then

T'_{in} = 30 sec,

T_s = 1 min,

T_c = 1 min,

T_c' = 2-1/2 min.

QUANTIZATION AND ROUND-OFF ERRORS

Aside from the effect of sampling, the errors in the output of a digital computation are due to two factors: initial quantization of the input samples and round-off during calculation. The number of quantization levels to use depends on the noise in the input data. Usually, there is no advantage in using a quantum step much smaller than the rms noise.

For the type of input we are interested in, 1,000 levels (~ 10 bits) probably are sufficient. The existing Cooley-Tukey programs use floating-point arithmetic. A floating-point number in the IBM 7094 has 28 significant bits. Therefore, the errors due to round-off are negligible compared to those due to the initial quantization.

Incidentally, the round-off errors in the Cooley-Tukey algorithm are smaller than those in the direct method, because the former involves a smaller number of basic operations.

SPECIAL-PURPOSE DIGITAL COMPUTERS

If Fourier analysis is to be done routinely, then it is probably worthwhile to build a special-purpose digital computer. From Figure 2 it is clear that the operations during any given step could be done in parallel.

In the extreme, if all of the operations at each step are done in parallel, then the computer time can be cut down by a factor of N, the number of input samples. In general, if P operations are done in parallel, the computer time is cut down by a factor of P.

CONVOLUTION BY FOURIER ANALYSIS

Consider the convolution of two functions, f and g, of finite durations. Assume N samples are taken uniformly over the duration of f, and m samples over the

duration of g. The sampling periods in f and g are the same. The number of output samples is (N + m). Let us assume that m << N, as is usually the case.

To carry out the convolution directly according to the equation

$$h_k = \sum_{j=0}^{m-1} f_{k-j}\, g_j; \quad k = 0, 1, \ldots, (N + m - 1) \tag{43}$$

where $(f_j; j = 0, 1, \ldots, N-1)$, $(g_j; j = 0, 1, \ldots, m-1)$, and $(h_j; j = 0, 1, \ldots, N + m - 1)$ are the samples of f, g, and the output h, respectively, would require roughly mN basic operations. The required computer time is therefore

$$T_1 = k_d mN , \tag{44}$$

where k_d is a constant. For the IBM 7094 computer, as mentioned earlier,

$$k_d \approx 30\ \mu\text{sec} .$$

Then

$$T_1 = 30\ mN\ \mu\text{sec} . \tag{45}$$

We can also[4] calculate h indirectly by first calculating the Fourier transforms of f and g, multiplying them, and taking the inverse Fourier transform. The computer time required would roughly be

$$T_2 = k_c\, 4\, N \log_2 N. \tag{46}$$

As mentioned earlier, $k_c \approx 30\ \mu\text{sec}$. Therefore,

$$T_2 = 120\ N \log_2 N\ \mu\text{sec} . \tag{47}$$

From Eqs. (45) and (47), we see that the Fourier analysis method would require less computer time, if

$$m > 4 \log_2 N . \tag{48}$$

For N = 64 x 64, Eq. (48) becomes $m > 48 \approx 7 \times 7$.
For N = 256 x 256, Eq. (48) becomes $m > 64 = 8 \times 8$.

The indirect method of doing convolution via Fourier analysis is straightforward except for one minor point. As is evident from Figure 1, on the computer we are actually calculating the Fourier transforms of $\tilde{f}$ and $\tilde{g}$ (periodically extended versions of $\hat{f}$ and $\hat{g}$). Therefore, we would get $\tilde{f} \circledast \tilde{g}$ instead of $\hat{f} \circledast \hat{g}$.

Errors would be introduced in $h_0, h_1, \ldots, h_{m-1}$. This can be easily avoided by introducing m zero samples in $(f_0, f_1, \ldots, f_{N-1})$ to get

$$(\overbrace{0, 0, \ldots, 0}^{\text{m zeros}}, f'_m, f'_{m+1}, \ldots, f'_{m+N-1}),$$

where

$$f'_{m+j} = f_j; \quad j = 0, 1, \ldots, N-1,$$

and then discard the first m samples of the output.

CONCLUDING REMARKS

In the case of restoration of satellite images degraded by the atmosphere, the maximum number of input samples per image will probably be $N = 256 \times 256 = 2^{16}$. To calculate the Fourier transform of such an image on an IBM 7094 computer (cycle time = 1 μsec, memory = 32,786 36-bit words) with a disk file of $2^{17} = 131,072$ 36-bit words, would take about 3 min. For a computer which is 10 times faster than the IBM 7094 and has 4 times as many words in the memory (one such computer is the IBM 360), the same Fourier transform can be performed in about 1/2 minute. Special-purpose parallel processing digital computers can be built to cut down the time further. In the extreme case, where the computer has a 2^{17}-word memory and can perform 2^{17} basic operations in parallel, the time requirement would be $K \log_2 N = 16K$, where K is the time for one basic operation. For K = 30 μsec, say, the time required is 1/2 msec, excluding input and output.

REFERENCES

1. I. J. Good, J. Roy. Stat. Soc. B20, 361 (1958); B22, 372 (1960).
2. J. H. Cooley and J. W. Tukey, Math. Comp. 19, 297 (1965).
3. J. Wolf, Fourier Series Subroutine, Internal Memorandum, Research Lab. of Electronics, Mass. Inst. of Tech.
4. T. G. Stockham, Jr., "High-Speed Convolution and Correlation," presented at 1966 Joint Spring Computer Conf., AFIPS Conf. Proc. 28, 229 (1966).

APPENDIX 23

TWO-DIMENSIONAL CONVOLUTION WITH A SEPARABLE KERNEL

T. S. Huang

INTRODUCTION

When the kernel in a two-dimensional convolution is separable, the computation time can be greatly reduced. This is probably the case in restoring long-exposure atmospherically degraded images.

THE CONTINUOUS CASE

Consider the convolution of a two-dimensional function f(x, y) with a kernel h(x, y):

$$g(x, y) = \int_{-\infty}^{\infty}\int f(x-\xi,\ y-\eta)\, h\,(\xi, \eta)\, d\xi\, d\eta \quad . \tag{1}$$

Assume h(x, y) is separable, i.e.,

$$h(x, y) = h_1\,(x)\, h_2\,(y) \quad . \tag{2}$$

Then Eq. (1) can be written as

$$g(x, y) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f\,(x-\xi,\ y-\eta)\, h_1\,(\xi)\, d\xi \right] h_2\,(\eta)\, d\eta \,. \tag{3}$$

Therefore, the problem of evaluating a two-dimensional integral is reduced to that of evaluating sequentially two one-dimensional integrals.

THE DISCRETE CASE

The discrete version of Eq. (1) is

$$g_{jk} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} f_{j-a,\ k-b}\ h_{ab} \tag{4}$$

$$(j,\ k = 0,\ 1,\ \ldots,\ N + m - 1)\ ,$$

assuming that f has N^2 samples, and g has m^2 samples.

Assume that $m << N$, as is usually the case. Then the number of basic operations required in calculating Eq. (4) is approximately N^2m^2. One basic operation is defined as one multiplication plus one addition. If the calculation is done on a digital computer, the required time is

$$T = k\,N^2m^2\ , \tag{5}$$

where k is a constant. For an IBM 7094 computer, a reasonable estimate for k is

$$k \approx 30\ \mu\text{sec}\ . \tag{6}$$

If N = 200, and m = 20, for example, then T = 8 min.

If the kernel h is separable,

$$h_{ab} = h'_a\,h''_b\ , \tag{7}$$

then Eq. (4) can be written as

$$g_{jk} = \sum_{a=0}^{m-1} \left[\sum_{b=0}^{m-1} f_{j-a,\ k-b}\ h''_b \right] h'_a \tag{8}$$

$$(j,\ k = 0,\ 1,\ \ldots,\ N + m - 1)\ .$$

We can calculate the inner summation first, then the outer summation. For each g_{jk}, we need only 2m basic operations, instead of m^2 as in Eq. (4). The total number of basic operations for all g_{jk} is therefore approximately

$$T_1 = \tfrac{1}{2} N^2 \, 2m \ . \tag{9}$$

Again, if N = 200, and m = 20, then T_1 = 48 sec. In general, if h is separable then the computation time can be cut down by a factor m/2.

LONG-EXPOSURE IMAGES

In the long-exposure case, the restoring kernel can probably be approximated by a two-dimensional Gaussian function

$$h(x, y) = A \, e^{ax^2 + by^2} \ , \tag{10}$$

where A, a, and b are constants. This function h(x, y) is clearly separable, and the discussions of the preceding sections apply.

In general, in any case where the exact shape of the kernel of the convolution is not important, it pays to look for an approximation which is separable or is a sum of a small number (<m/2) of separable functions.

APPENDIX 24

COMPLEX SPATIAL FILTERS FOR OPTICAL SYSTEMS

G. O. Reynolds and C. E. Thomas

GENERAL DESCRIPTION OF OPTICAL FILTER GENERATION TECHNIQUES

In general, the filter in a coherent optical processing system must be complex; it must vary the phase as well as the amplitude of the Fourier transform light distribution. The light amplitude distribution can be arbitrarily modified by density modulation of fine-grain silver halide film. Real-time amplitude variation will soon be operational via electron-beam recording on photochromic film.

To modify the transform phase distribution, two general techniques are applicable. The thickness of a transparent substance can be varied as a function of distance. This relief image phase-modulates the light distribution. Manufacturing techniques are difficult and in general not sufficiently reproducible for routine phase function generation.

INTERFEROMETRIC FILTER TECHNIQUES

A more generally applicable technique has been developed by A. B. Vander Lugt[1] for producing arbitrary complex spatial filters. The general approach is to form an interference pattern which contains both the phase and amplitude information. This pattern can be recorded on film.

Consider a complex light distribution:

$$F(u, v) = A(u, v)e^{i\phi(u, v)} , \qquad (1)$$

where

$F(u, v)$ = Fourier transform of processing point-spread function,

$A(u, v)$ = Amplitude of transform,

$\phi(u, v)$ = Phase of transform.

A normal photograph records only the intensity,

$$I(u, v) = F(u, v)\, F^*(u, v) = A^2(u, v) \quad , \tag{2}$$

which loses phase information.

If we add a coherent reference beam represented by $Re^{i\omega v}$, the interferometric intensity pattern becomes

$$I(u, v) = [\, A(u, v)e^{i\phi(u, v)} + Re^{i\omega v} \,]\, [\, A(u, v)e^{-i\phi(u, v)} + Re^{-i\omega v} \,]$$

$$= [\, A(u, v) \,]^2 + R^2 + 2RA(u, v) \cos [\, \omega v - \phi(u, v) \,] \quad , \tag{3}$$

or

$$I(u, v) = \left|F(u, v)\right|^2 + R^2 + RF(u, v)e^{-i\omega v} + RF^*(u, v)e^{i\omega v}. \tag{4}$$

Equation (3) indicates that the resulting intensity pattern contains both amplitude and phase information. When this filter is reinserted into the optical system, three images are generated. Equation (4) defines these three images. The "dc" image proceeds along the optical axis and contains the first two terms of Eq. (4). The power spectrum of the filter, $\left|F(u, v)\right|^2$, in general extends over much of the (u, v) plane. The third term of Eq. (4) propagates at an angle determined by ω. It is modulated in amplitude and phase by the desired transfer function, $F(u, v)$. In this application, this is the desired term and would be separated from the other two images by the direction of its propagation. Finally, the last term is propagated in the mirror-image direction and is modulated by the complex conjugate of the transform, $F^*(u, v)$.

TYPICAL OPTICAL COMPLEX SPATIAL FILTERING SYSTEM

The system suggested by Vander Lugt to generate these complex filters is more stable than most interferometers. A laser light source provides sufficient temporal coherence to allow slightly unequal path lengths in the signal and reference beams, as shown in Figure 1. The small off-axis lens generates a reference beam which is collimated at an angle by the transform lens. This same lens Fourier transforms the "signal film." To make the complex filter, the processing point-spread function is recorded on the "signal film." Finally, the interference pattern previously described is recorded on the "complex filter plate."

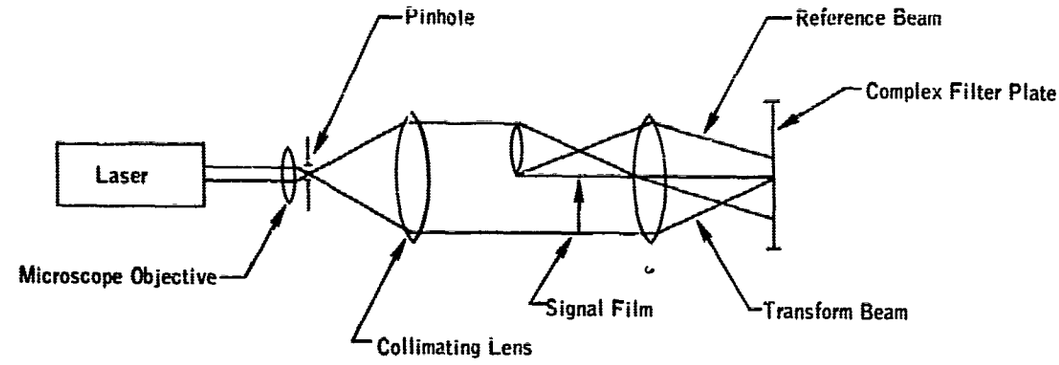


FIGURE 1 Complex filter generator.

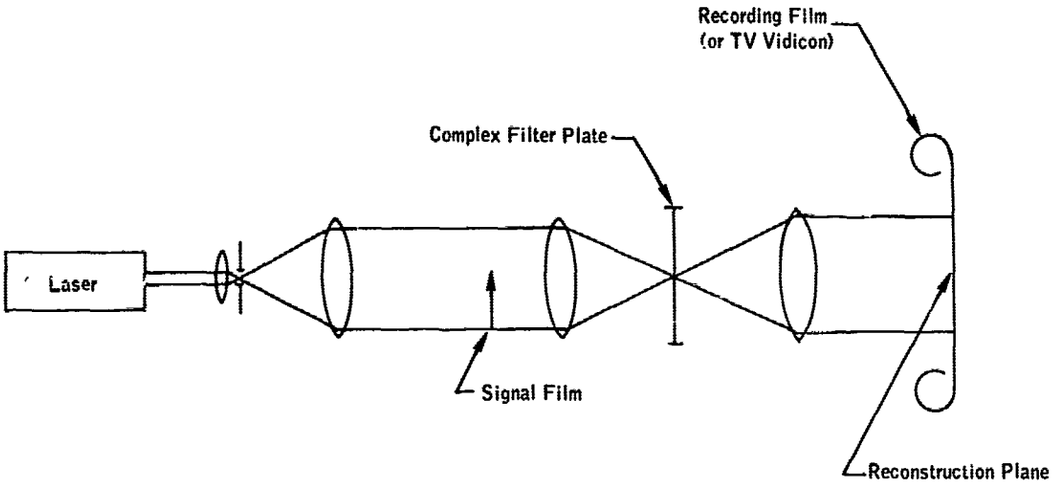


FIGURE 2 Optical restoration equipment schematic.

The same system is used for image processing as for filter generation (see Figure 2). The developed filter is reinserted into the system and the reference lens is covered to eliminate the reference beam. The degraded image replaces the impulse function. Another lens is added to reconstruct the filtered image in Figure 2. The restored image can be observed through a microscope, recorded on film, or viewed via a closed-circuit TV system.

REFERENCE

1. A. B. Vander Lugt, IEEE Trans. on Information Theory IT-10, 139 (1964).

APPENDIX 25

OPTICAL FREQUENCY PLANE SCALE SEARCH

G. O. Reynolds and C. E. Thomas

In many restoration problems, the form of the processing transfer function is known but the scale varies from one picture to the next. A digital processing system would have to sequentially process such a photograph with various sizes of transfer functions. The optical system of Figure 1 allows routine scale search of the transfer function while observing the processed image in real time.

In Figure 1, the degraded image transparency f(x, y) is illuminated by a plane wave, and the far-field diffraction pattern is proportional to

$$F(u, v) = \iint f(x, y) \exp\left\{\frac{2\pi i}{\lambda R}(ux + vy)\right\} dx\, dy \ .$$

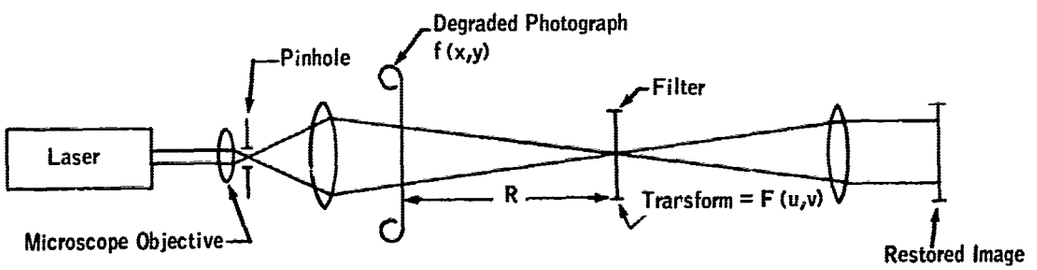


FIGURE 1 Coherent optical image restoration system with frequency plane scale search.

Hence the scale of the transformation varies linearly with R, the separation between the transform plane and the degraded negative. As the transform plane is moved, the last lens and the output sensing device must be moved in synchronism to maintain the restored image in focus. A closed-circuit television monitor can be used in this system to observe in real time the effect of changing transform scale.

APPENDIX 26

PHOTOGRAPHIC LINEARITY IN COHERENT OPTICAL SYSTEMS

G. O. Reynolds and C. E. Thomas

A linear relationship between the amplitude transmission of the degraded image film and the original exposure (product of light intensity and time, neglecting reciprocity failure) is essential to a coherent optical analog system. The following technique has been successfully used in the past to linearize the photographic process. Further experiments are needed to determine the utility of this technique for the atmospheric turbulence restoration problem.

Consider the Hurter-Driffield (H-D) curve which represents input-output characteristics of a given film-developer combination, as shown in Figure 1. Mathematically, the straight-line portion of this curve can be represented by

$$D_1 = -\log T_1 = \gamma_1 \log E_1 - \log k_1 \tag{1}$$

or

$$T_1 = k_1/E_1^{\gamma_1} \tag{2}$$

or

$$t_1 = \sqrt{T_1} = K_1/E_1^{\gamma_1/2}, \tag{3}$$

where T_1 is the intensity transmission and t_1 is the amplitude transmission. Now suppose that the original negative is contact printed. The amplitude transmission of the positive transparency will be

$$t_2 = K_2/E_2^{\gamma_2/2} \quad . \tag{4}$$

But

$$E_2 = K'T_1 = K''/E_1 \quad , \tag{5}$$

so

$$t_2 = KE^{\gamma_1\gamma_2/2} \quad . \tag{6}$$

From Eq. (6), linearity is achieved provided that

$$\gamma_1\gamma_2 = 2 \quad . \tag{7}$$

This derivation assumes that all exposures on the degraded photograph lie on a linear part of both films' H-D curves. Thus, if any part of the degraded image receives no light, then the initial film must be pre-exposed to a uniform light level, to bias the film above the toe of the H-D curve.

Figure 2 indicates the improvement in linearity resulting from this contact-printing procedure. A sinusoidal intensity pattern was used to expose a test film. This film was developed to a gamma of 1. A spatial spectral analysis of the film showed considerable second harmonic distortion resulting from film non-linearity. The top graph of Figure 2 shows the output from the spectrum analyzer.

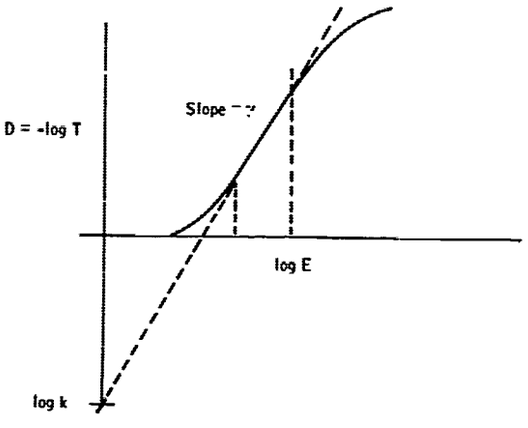


FIGURE 1 H-D curve, where γ = slope of curve, D = density of developed negative, E = exposure of negative (input), and T = intensity transmission of developed negative.

Next a contact print of this film was made. The contact print was developed to a gamma of 2, thereby satisfying the linearity criterion of Eq. (7). The second spectral analysis of Figure 2 shows no detectable second harmonic distortion. The gain of the system was increased in the last analysis to show no second harmonic above the noise level.

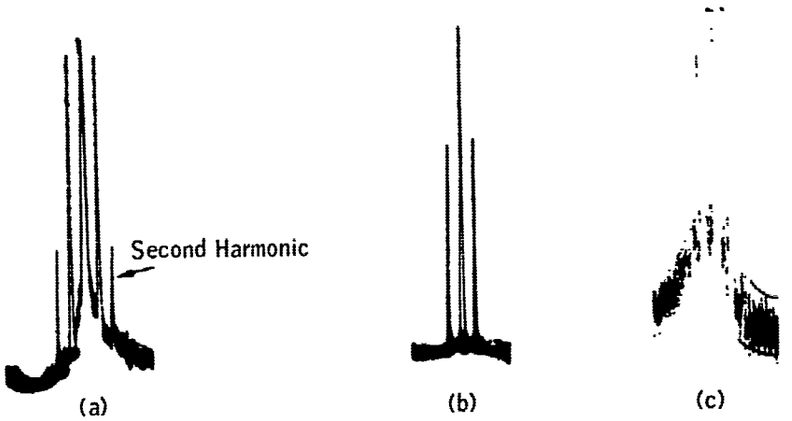


FIGURE 2 Second harmonic distortion test. (a) Spectrum of original negative, $\gamma = 1$; (b) spectrum of contact print, $\gamma = 2$; (c) spectrum of contact print, gain increased by 20.

Harmonic distortions of less than 1 percent in amplitude have been obtained routinely by this technique with manual film processing. The actual distortion may be much less. It is possible that by controlled machine processing, one may obtain less than 0.1 percent distortion.

APPENDIX 27

LINEARITY AND DYNAMIC RANGE IN COHERENT PROCESSING

J. H. Altman

For the purpose of this paper, linearity is defined somewhat differently from conventional photographic usage. Frequently in photography, "linearity" signifies a linear relation between the developed density and the logarithm of the exposure received by the emulsion. Such a system is nonlinear in the optics-electronics sense. In the present context, "linearity" will mean either that

(a) $T = KE$, where T is the transmittance of the developed sample and E is the exposure; or

(b) $\sqrt{T} = KE$.

Which of these is required depends upon the particular optical processing system involved. Coherent systems will require (b).

By definition, we write $T = KE^{-\gamma}$, γ being the slope of the straight-line portion of the normal photographic D versus log E curve. It follows directly that when $\gamma = -1$, then $T = KE$, which is case (a) above; and when $\gamma = -2$, then $\sqrt{T} = KE$, which is case (b). Physically, these cases mean positive images the slopes of whose D versus log E curves are 1 or 2 in the straight-line portion, as the case may be. Dynamic range can then be defined as the exposure range over which the curves do not depart significantly from the required gradient.

Sensitometric response of the $\sqrt{T} = KE$ type has been obtained by a positive-negative technique in which a low-gamma negative is printed onto a high-gamma positive so that the resulting gamma product = 2. A simpler method has been

demonstrated in which Kodak Minicard Film is reversed by a procedure which gives $\gamma = -2$ with a dynamic range of at least 16:1 (1.2 log units).[1] This particular emulsion is quite slow (< MicroFile) but is very sharp and fine grained. Maximum high-contrast resolution is of the order of 1,500 lines/mm. Spectral sensitivity is either ortho or pan. The process can be adjusted to give $\gamma = -1$ by changing some of the immersion times; in this case, the indicated dynamic range is about 100:1 (2.0 log units). The resulting silver deposits are not quite neutral, but slightly greenish. For monochromatic or narrow-band illumination, however, this merely means a suitable adjustment in the process times so that the densities measured at the wavelengths of interest give the desired gamma. Details of this process are available from the Eastman Kodak Company.

Finally, it might be noted for completeness that a new type of emulsion which gives a direct positive image on normal single-stage "negative" development has recently been introduced by Kodak. The presently available version is identified as SO-136; it is a fine-grain slow material giving $\gamma = -1$ with a dynamic range of 25:1 (1.4 log units).[2] The gamma change which can be induced by processing variations is limited.

REFERENCES

1. B.K. Boller, "Reversal Duplicating of Aerial Negatives," Phot. Sci. Eng. 9, 318 (1965).
2. H.J. Fromm, "A New Direct Duplicating Silver Film," to be published in the Proceedings of the 1966 Convention of the National Microfilm Association, Washington, D.C., May 17-19, 1966.

APPENDIX 28

COMPARISON AND POSSIBLE COUPLING OF COHERENT OPTICAL AND DIGITAL COMPUTER PROCESSING TECHNIQUES FOR RESTORING DEGRADED IMAGERY

T. S. Huang and H. L. Kasnitz

INTRODUCTION

The restoration of degraded images calls for both linear and nonlinear mathematical operations. The linear operations (Fourier transformation and filtering) can be done either by a coherent optical analog system or by a digital computer. Each of these two methods possesses certain advantages. There are some differences of viewpoint between advocates of analog and digital methods. This paper will compare these two methods and explore the possibility of coupling them together.

COMPARISON OF COHERENT OPTICAL AND DIGITAL COMPUTER PROCESSING

Flexibility

Coherent optical systems are essentially limited to linear operations on the amplitude transmission variations of a film transparency. On the other hand, digital computers can be used to do linear operations on either the amplitude transmission, or the intensity transmission, or the density. More importantly, digital computers can also be used to do nonlinear operations. In fact, they can be used to do essentially any operations one can specify.

Capacity and Speed

In a coherent optical system, the film is used as the storage, resulting in an enormous capacity. More importantly, the data on the film can be operated on in parallel, so that the speed is limited, in principle, only by the speed of light. Although a digital computer usually has a limited memory, any amount of auxiliary memory can be attached to it. If the computer has a film scanner, then films can also serve as its storage. However, present-day digital computers operate essentially sequentially on the data. Therefore, if a large number of data points need to be operated on, it takes a long time to bring the data into the central processor, and still a longer time to process them.

Let us consider an example. A giant coherent optical system[1] at the Institute of Science and Technology, University of Michigan, is capable of processing 70-mm films with a resolution of 100 cycles/mm. It can therefore do a Fourier transformation on approximately 2×10^8 data points essentially instantaneously.

Now suppose we do the same thing on a digital computer, e.g., an IBM 7094, which has a cycle time of 1 μsec. It would take more than an hour just to read in the data points, if the computer film reader reads at 30 μsec/point. Assume that the efficient Fourier analysis algorithm of Cooley and Tukey[2] is used, and assume that the computer had a core memory of more than 4×10^8 words. (The actual memory size is 32 K.) It still would take about 100 hr to perform the Fourier transform.

We note in passing that in a coherent optical system, the actual operation time is negligible compared with the time spent in processing the film, aligning the optical system, and making the spatial filters. These chores could be quite time consuming, but in most cases, they would certainly take much less than 100 hr.

Accuracy

In digital processing, there are inherent errors due to sampling and amplitude quantization. These errors, however, can be made arbitrarily small by increasing the sampling rate and the number of quantization levels. In practice, the accuracy of digital computer image processing is limited by the film scanner. It is probably difficult at present to build a film scanner with an accuracy better than 0.01 percent.

In a coherent optical system, there are various sources of errors:

(a) Imperfect optical components

(b) Dust and scratches on optical components

(c) Film grain noise and nonlinearity

(d) Spurious thickness variations of film emulsions

(e) Errors in spatial filters

(f) Nonuniformity of light beam across the input aperture

(g) Imperfect alignment of the optical system

These errors are not easy to control. One can get an accuracy of only 3 to 5 percent in a coherent optical system.

Cost

Coherent optical systems are usually cheaper than digital computers. However, large-aperture diffraction-limited coherent optical systems can be rather expensive. The giant system at the University of Michigan which we mentioned earlier cost about $500,000.

In summary, the main advantages of a coherent system are its information storage capacity and processing speed, and the main advantages of a digital computer are its flexibility and accuracy.

It strikes one that the two methods should be coupled together, preferably in real time, so that the advantages of both can be utilized. For example, a high-resolution image on film may first be corrected for nonlinearity on a computer, then its Fourier transform may be taken on a coherent optical system, then the computer may be used to do some nonlinear operation on the Fourier transform, and finally the coherent optical system may be used to take the inverse Fourier transform to reconstruct the image.

AVAILABILITY OF HARDWARE FOR COUPLING COHERENT OPTICAL SYSTEMS AND DIGITAL COMPUTERS

Suitable hardware is required to allow communications between a coherent optical system and a digital computer. A conceptual design for a high-speed film scanner built around a standard vidicon has been described earlier.[3] By a technique described in the following section, it is possible to read in both the amplitude and the phase of the light distribution representing the Fourier transform of an image.

A coherent optical system operates on hard copies (or their equivalents) of images. To transfer data from a digital computer to a hard copy, a CRT display can be used to expose some medium. Possible media for this purpose are silver halide film, photochromic glass or film, and oil film (Eidophor system). The latter two media have the advantage of being erasable. A real-time link from the computer to the coherent optical system can, for example, be provided by a loop of photochromic film whose motion is controlled by the computer and which passes through the computer CRT, the coherent optical system, and a suitable

eraser. The same setup can be used to make spatial filters by the computer. Complex data can be handled as suggested in the section on page 187.

DATA TRANSFER FROM COHERENT OPTICAL SYSTEMS TO DIGITAL COMPUTERS

In restoring degraded imagery, the coherent optical system will probably be called upon to do two operations: convolving a positive real image by a real kernel (linear filtering) and Fourier transformation.

For the linear filtering, the output image is usually positive and real. Therefore, the output image can be focused directly on the computer scanner for read-in. If for some reason, the output image is partially negative, a dc bias can be introduced to make it positive.

In the case of Fourier transformation, the output image is in general complex, and we are interested in both its amplitude and phase. If the Fourier transform is focused on the computer scanner, only the square of the amplitude will be read in and the phase information will be lost. However, by the following scheme, one can preserve the phase information.

Let $g(x, y)$, where x and y are spatial coordinates, be a positive real image whose Fourier transform $G(u, v)$, where u and v are spatial frequencies, we wish to read into the computer. Then Re $G(u, v)$, the real part of $G(u, v)$, is the Fourier transform of

$$(1/2)\,[\,g(x, y) + g(-x, -y)\,] \quad ,$$

and Im $G(u, v)$, the imaginary part of $G(u, v)$, is the Fourier transform of

$$(1/2i)\,[\,g(x, y) - g(-x, -y)\,] \quad .$$

If we choose the origin so that

$$g(x, y) = 0, \quad \text{for} \quad x < 0, \quad x > A \quad \text{and} \quad y < 0, \quad y > A$$

where A is the width of the image, then the nonzero parts of $g(x, y)$ and $g(-x, -y)$ will not overlap, and $g(x, y) + g(-x, -y)$ can be synthesized by combining two identical transparencies, both with amplitude transmission $g(x, y)$, one of these transparencies being rotated 180° around the origin. We can put this combined transparency in the input plane of the optical system, get Re $G(u, v)$ in the frequency plane, and read $\left|\text{Re } G(u, v) + K\right|^2$ into the computer, where K is a properly selected dc bias to make $[\,\text{Re } G(u, v) + K\,]$ nonnegative. Similarly, using the same input transparency and appropriate phase plates, we can record $\left|\text{Im } G(u, v) + K_1\right|^2$, where K_1 is a constant.

In reading in $\left|\text{Re } G(u, v) + K\right|^2$ and $\left|\text{Im } G(u, v) + K_1\right|^2$, the components at u = 0 = v may have to be suppressed to reduce the dynamic range requirement of the computer scanner.

DATA TRANSFER FROM DIGITAL COMPUTERS TO COHERENT OPTICAL SYSTEMS

In restoring degraded imagery, the computer will probably be called upon to transfer a two-dimensional function f(x, y) to the coherent optical system to get its Fourier transform.

If f(x, y) is positive and real, then we can display it on a CRT, which in turn exposes a piece of film. A photochromic material may be very useful here. The computer can be programmed to compensate for the nonlinearities of the CRT and the photochromic film so that the amplitude transmission of the exposed film will be f(x, y).

If f(x, y) is partially negative, a dc bias can be added to make it completely positive.

If f(x, y) is complex, we can generate a transparency with amplitude transmission

$$\left|f(x, y) + Be^{jbx}\right|^2,$$

where B and b are constants. If this transparency is placed in the input plane of the coherent system, the Fourier transform of f(x, y) will appear in the frequency plane off axis, if the constants B and b are properly chosen.

MAKING SPATIAL FILTERS BY DIGITAL COMPUTER

In some cases, although the filtering is best done by a coherent optical system, the spatial filter is most conveniently made on a digital computer.

The technique of making spatial filters on a digital computer is the same as the technique described in the preceding section for transferring data from a digital computer to a coherent optical system.

A silver halide or photochromic transparency is made from the computer CRT, with amplitude transmission H(u, v), the filter transfer function, if H is real and positive, or

$$\left|H(u, v) + Be^{jbu}\right|^2,$$

where B and b are appropriate constants, if H(u, v) is partially negative or complex. Again, the computer can be programmed to compensate for the nonlinearities of

the CRT and the film. This piece of film transparency is put in the frequency plane as the filter. The desired filtered output will appear in the output plane, either on axis or off axis, depending on whether H(u, v) or

$$\left|H(u, v) + Be^{jbu}\right|^2$$

was recorded as the filter.

CONCLUDING REMARKS

It is obvious that digital computers and coherent optical systems have their respective advantages and disadvantages. We propose that they should be coupled together so that the advantages of both can be utilized. It is seen from our discussions that data transfer between a digital computer and a coherent optical system is at least conceptually possible. Much theoretical analysis and experimental study need to be done to determine the accuracy and speed of the specific methods described in this paper as well as other possible methods.

REFERENCES

1. A. Kozma and N. Massay, "A Precision Optical Processor for High-Resolution Radar," presented at the 1964 Annual Radar Symp., Institute of Science and Technology, University of Michigan.

2. T.S. Huang, this volume, Appendix 22, page 153.

3. H.L. Kasnitz, this volume, Appendix 20, page 141.

APPENDIX 29

ANALYSIS OF WAVEFRONT-RECONSTRUCTION IMAGING THROUGH RANDOM MEDIA

J. W. Goodman

In this paper we study a number of analytical problems associated with wavefront-reconstruction imaging through random media. The reader is referred to the first and second sections of Chapter IV of Volume I of this report to obtain perspective as to the circumstances in which these problems arise.

RIGOROUS ANALYSIS BASED ON DIFFRACTION INTEGRALS

A rigorous formulation of the problem of wavefront-reconstruction imaging through random media can be constructed from the Kirchhoff diffraction integral as a starting point. Let the object and reference point sources be located at coordinates (ξ, η) and $(0, 0)$ respectively, in a plane parallel with, and at distance R from, the collecting aperture. The object-reference separation is thus $\delta = [\xi^2 + \eta^2]^{1/2}$. The random screen is again assumed to be located at distance ρ from the collecting aperture. Using the usual quadratic approximations to spherical waves, the fields incident at point (x, y) on the random screen from the reference and object can be written, respectively, as

$$^{(1)}E_r(x,y) = A_r \exp\left\{j\frac{\pi}{\lambda(R-\rho)}(x^2 + y^2)\right\}, \quad (1)$$

$$^{(1)}E_o(x,y) = A_o \exp\left\{j\frac{\pi}{\lambda(R-\rho)}\left[(x-\xi)^2 + (y-\eta)^2\right]\right\}, \quad (2)$$

where A_o and A_r are simply complex constants, and λ is the optical wavelength. Representing the complex amplitude transmittance of the random screen by $t(x, y)$, the fields incident on the collecting aperture may be written (using the Fresnel approximation) as

$$^{(2)}E_r(\alpha,\beta) = A_r \iint_{-\infty}^{\infty} t(x,y) \exp\left\{j\frac{\pi}{\lambda(R-\rho)}(x^2+y^2)\right\} \cdot \exp\left\{j\frac{\pi}{\lambda\rho}\left[(\alpha-x)^2+(\beta-y)^2\right]\right\} dxdy, \quad (3)$$

$$^{(2)}E_o(\alpha,\beta) = A_o \iint_{-\infty}^{\infty} t(x,y) \exp\left\{j\frac{\pi}{\lambda(R-\rho)}\left[(x-\xi)^2+(y-\eta)^2\right]\right\} \cdot \exp\left\{j\frac{\pi}{\lambda\rho}\left[(\alpha-x)^2+(\beta-u)^2\right]\right\} dxdy . \quad (4)$$

With considerable algebraic manipulation the incident fields can be rewritten as

$$^{(2)}E_r(\alpha,\beta) = A_r \exp\left\{j\frac{\pi}{\lambda R}(\alpha^2+\beta^2)\right\} \cdot \iint_{-\infty}^{\infty} t(x,y) \cdot \exp\left\{j\frac{\pi R}{\lambda\rho(R-\rho)}\left[\left(x-\frac{R-\rho}{R}\alpha\right)^2+\left(y-\frac{R-\rho}{R}\beta\right)^2\right]\right\} dxdy , \quad (5)$$

$$^{(2)}E_o(\alpha,\beta) = A_o \exp\left\{j\frac{\pi}{\lambda R}(\alpha^2+\beta^2)\right\} \exp\left\{j\frac{\pi}{\lambda R}(\xi^2+\eta^2)\right\} \cdot \exp\left\{-j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)\right\} \cdot \iint_{-\infty}^{\infty} t(x,y) \cdot \exp\left\{j\frac{\pi R}{\lambda\rho(R-\rho)}\left[\left(x-\frac{R-\rho}{R}\alpha+\frac{\rho}{R}\xi\right)^2+\left(y-\frac{R-\rho}{R}\beta+\frac{\rho}{R}\eta\right)^2\right]\right\} dxdy . \quad (6)$$

The method of stationary phase may now be employed to evaluate the two integrals. The results of this approximation are

$$ {}^{(2)}E_r(\alpha,\beta) = A'_r \exp\left\{ j \frac{\pi}{\lambda R} (\alpha^2 + \beta^2) \right\} \quad t \left(\frac{R-\rho}{R} \alpha, \frac{R-\rho}{R} \beta \right) , \tag{7} $$

$$ {}^{(2)}E_o(\alpha,\beta) = A'_o \exp\left\{ j \frac{\pi}{\lambda R} (\alpha^2 + \beta^2) \right\} $$

$$ \cdot \exp\left\{ -j \frac{2\pi}{\lambda R} (\alpha\xi + \beta\eta) \right\} t \left(\frac{R-\rho}{R} \alpha - \frac{\rho}{R} \xi, \frac{R-\rho}{R} \beta - \frac{\rho}{R} \eta \right) , \tag{8} $$

where the approximations are valid only when

$$ \sqrt{\frac{\lambda \rho (R-\rho)}{R}} << \Delta , \tag{9} $$

where Δ is the correlation width of the wavefront emerging from the random screen. The intensity distribution across the interference pattern is

$$ I(\alpha,\beta) = \left| {}^{(2)}E_r(\alpha,\beta) \right|^2 + \left| {}^{(2)}E_o(\alpha,\beta) \right|^2 $$

$$ + {}^{(2)}E_r^*(\alpha,\beta) \; {}^{(2)}E_o(\alpha,\beta) + {}^{(2)}E_r(\alpha,\beta) \; {}^{(2)}E_o^*(\alpha,\beta) . \tag{10} $$

For a phase screen,

$$ t(x,y) = \exp\left\{ j \phi (x,y) \right\} \tag{11} $$

and

$$I(\alpha,\beta) = \left|A_r'\right|^2 + \left|A_o'\right|^2$$

$$+ A_r'A_o'^* e^{-j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)} e^{j\left[\phi(\frac{R-\rho}{R}\alpha, \frac{R-\rho}{R}\beta) - \phi(\frac{R-\rho}{R}\alpha - \frac{\rho}{R}\xi, \frac{R-\rho}{R}\beta - \frac{\rho}{R}\eta)\right]}$$

$$+ A_r'^* A_o' e^{+j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)} e^{-j\left[\phi(\frac{R-\rho}{R}\alpha, \frac{R-\rho}{R}\beta) - \phi(\frac{R-\rho}{R}\alpha - \frac{\rho}{R}\xi, \frac{R-\rho}{R}\beta - \frac{\rho}{R}\eta)\right]} \tag{12}$$

The condition required for an uncorrupted interference pattern is thus

$$\phi(\frac{R-\rho}{R}\alpha, \frac{R-\rho}{R}\beta) - \phi(\frac{R-\rho}{R}\alpha - \frac{\rho}{R}\xi, \frac{R-\rho}{R}\beta - \frac{\rho}{R}\eta) << 1 , \tag{13}$$

or equivalently,

$$\delta = \sqrt{\xi^2+\eta^2} << \frac{R}{\rho}\Delta , \tag{14}$$

in which case the uncorrupted interference pattern,

$$I(\alpha,\beta) = \left|A_r'\right|^2 + \left|A_o'\right|^2$$

$$+ A_r'A_o'^* e^{-j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)}$$

$$+ A_r'^* A_o' e^{j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)} , \tag{15}$$

is obtained.

Thus sufficient conditions for the formation of an unperturbed interference pattern in the presence of the screen are the inequalities (9) and (14). Condition (9) requires that the coherence width of the wavefront emerging from the screen must be larger than the first Fresnel zone generated at the screen by a point-source located at distance $\rho\,[\,1 - (\rho/R)\,]$ from the screen. Equivalently, diffraction effects must be negligible. Condition (14) requires that the object and reference be within a single "isoplanatic patch."

EFFECTS OF VIOLATING THE FRESNEL-ZONE CONDITION

While sufficient conditions for the formation of an unperturbed interference pattern have been found, it is extremely important to consider the effects of violating one of those conditions, namely the Fresnel-zone condition (9). In effect, we are simply investigating the effects of diffraction by the turbulent medium. Since it is certain that diffraction effects are introduced by high-altitude turbulence, the investigation is an important one.

To find the effect of violating (9), let the result of the integration in Eq. (5) be represented by some unspecified complex function as follows:

$$\int_{-\infty}^{\infty}\int t(x,y)\exp\left\{j\,\frac{\pi R}{\lambda\rho\,(R-\rho)}\left[\left(x-\frac{R-\rho}{R}\alpha\right)^2+\left(y-\frac{R-\rho}{R}\beta\right)^2\right]\right\}dx\,dy$$

$$\triangleq\; a_1(\alpha,\beta)\,\exp\left\{j\theta_1(\alpha,\beta)\right\}. \tag{5a}$$

Similarly we represent the integral of Eq. (6) by another complex function

$$\int_{-\infty}^{\infty}\int t(x,y)\exp\left\{j\,\frac{\pi R}{\lambda\rho\,(R-\rho)}\left[\left(x-\frac{R-\rho}{R}\alpha+\frac{\rho}{R}\xi\right)^2+\left(y-\frac{R-\rho}{R}\beta+\frac{\rho}{R}\eta\right)^2\right]\right\}dx\,dy$$

$$\triangleq\; a_2(\alpha,\beta)\exp\left\{j\theta_2(\alpha,\beta)\right\}. \tag{6a}$$

Now note that when $\theta_1 = \theta_2$ and $a_1 = a_2$ ($\triangleq a$), the intensity distribution across the (α,β) plane becomes a purely amplitude-modulated fringe pattern,

$$I(\alpha,\beta) = \left|A'_r\right|^2 a^2(\alpha,\beta) + \left|A'_o\right|^2 a^2(\alpha,\beta)$$

$$+ A'_r A'^{*}_o a^2(\alpha,\beta)\, e^{-j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)}$$

$$+ A'^{*}_r A_o a^2(\alpha,\beta)\, e^{j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)} \qquad (15a)$$

Such a result is obtained, of course, only when the two integrals above are equal. To find the condition required for equality, make the change of variables

$$x' = x + (\rho/R)\,\xi\,,$$

$$y' = y + (\rho/R)\,\eta\,,$$

in (6a), yielding

$$\int_{-\infty}^{\infty}\int t\left(x' - \frac{\rho}{R}\xi,\ y' - \frac{\rho}{R}\eta\right)$$

$$\cdot \exp\left\{j\frac{\pi R}{\lambda\rho(R-\rho)}\left[\left(x' - \frac{R-\rho}{R}\alpha\right)^2 + \left(y' - \frac{R-\rho}{R}\beta\right)^2\right]\right\} dx'\,dy'$$

$$= a_2(\alpha,\beta)\ \exp\left\{j\theta_2(\alpha,\beta)\right\}.$$

Now note that if $(\rho/R)\xi \ll \Delta$ and $(\rho/R)\eta \ll \Delta$, we have

$$t[x' - (\rho/R)\,\xi,\ y' - (\rho/R)\,\eta] = t(x', y')$$

and the two integrals are therefore identical. Hence the amplitude-modulated fringe pattern of Eq. (15a) is obtained, with the only requirement being that the isoplanatic condition is satisfied.

To summarize, when the isoplanatic condition (14) and the Fresnel-zone condition (9) are both satisfied, a perfect interference pattern is obtained and clean diffraction-limited images can be formed. When the Fresnel-zone

condition is not satisfied, the result is a random amplitude modulation of the fringe pattern. The effects of this random modulation on the quality of the reconstructed images is considered in the analysis that follows. The effects are extremely important; as we shall now demonstrate, they introduce a random noise in the reconstructed image, thus providing a fundamental limit to the quality of the image data.

Under the isoplanatic condition, the intensity modulations of each of the incident waves are identical, and the intensity distribution (Eq. 15) across the interference pattern can be again written

$$I(\alpha,\beta) = \left|A'_r\right|^2 a^2(\alpha,\beta) + \left|A'_o\right|^2 a^2(\alpha,\beta)$$

$$+ A'_r A'^*_o a^2(\alpha,\beta)\, e^{-j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)}$$

$$+ A'^*_r A'_o a^2(\alpha,\beta)\, e^{j\frac{2\pi}{\lambda R}(\alpha\xi+\beta\eta)}, \qquad (16)$$

where the function $a(\alpha,\beta)$ represents the common amplitude modulation of the two incident fields.

The intensity distribution $\mathcal{I}(u,v)$ in the plane of the reconstructed images is proportional to the squared modulus of the Fourier transform of I. The resulting expression contains sixteen terms, of which only a smaller number are of interest in the weak-object case (i.e., when $\left|A'_o\right| \ll \left|A'_R\right|$). As a consequence of the random modulation of the hologram transmittance, two major sources of noise arise: light is deflected out of the strong zero-order component and may tend to obscure parts of the image that lie close to the reference; and a portion of the light in the images themselves is spread out, yielding a diffraction-limited image superimposed on a blurred background. In the analysis that follows, the signal-to-noise (S/N) ratios due to these two effects are found. In each case, only the intensity of the deflected light is calculated. Thus interference effects are neglected, an approximation that will yield accurate predictions only for high S/N ratios. Again attention is restricted to a point-source object, but the results also apply directly to an arbitrary object.

Consider first the light deflected out of the strong zero-order reconstruction component. The noise arises from the variations of a^2 about its mean in the term

$$\mathcal{J}_o(u,v) = \left|A'_R\right|^4 \left| \int_{-L/2}^{L/2}\int a^2(\alpha,\beta)\, e^{-j2\pi(u\alpha+v\beta)}\, d\alpha d\beta \right|^2 . \tag{17}$$

On the other hand, the desired twin images arise from the mean value of a^2 in the terms

$$\mathcal{J}_1(u,v) = \left|A'_R\right|^2 \left|A'_o\right|^2 \cdot \left| \int_{-L/2}^{L/2}\int a^2(\alpha,\beta)\, e^{-j2\pi\left[(u\pm\frac{\xi}{\lambda R})\alpha + (v\pm\frac{\eta}{\lambda R})\beta\right]} d\alpha d\beta \right|^2 . \tag{18}$$

Let the variations of a^2 about its mean be represented by

$$\Delta a^2(\alpha,\beta) = a^2(\alpha,\beta) - \overline{a^2}(\alpha,\beta) . \tag{19}$$

Then the average noise intensity across the image plane can be reduced to

$$\mathcal{J}_N(u,v) = \left|A'_R\right|^4 \int_{-\infty}^{\infty}\int \Lambda\left(\frac{\tilde{\alpha}}{L}\right)\Lambda\left(\frac{\tilde{\beta}}{L}\right) R_{\Delta a^2}(\tilde{\alpha},\tilde{\beta}) \cdot \exp\left\{-j2\pi\left[(u-\frac{\xi}{\lambda R})\tilde{\alpha} + (v-\frac{\eta}{\lambda R})\tilde{\beta}\right]\right\} d\tilde{\alpha} d\tilde{\beta} , \tag{20}$$

where

$$\Lambda(x) \triangleq \begin{cases} 1-|x| & |x| \le 1 , \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

In a similar fashion, the intensity in either of the twin images reduces to

$$\mathcal{J}_s(u,v) = \left|A'_R\right|^2 \left|A'_o\right|^2 (\overline{a^2})^2 \iint_{-\infty}^{\infty} \Lambda\left(\frac{\tilde{\alpha}}{L}\right) \Lambda\left(\frac{\tilde{\beta}}{L}\right)$$

$$\cdot \exp\left\{-j2\pi\left[\left(u - \frac{\xi}{\lambda R}\right)\tilde{\alpha} + \left(v - \frac{\eta}{\lambda R}\right)\tilde{\beta}\right]\right\} d\tilde{\alpha}\, d\tilde{\beta} \,. \tag{22}$$

Assuming a Gaussian autocorrelation function for the intensity fluctuations,

$$R_{\Delta a^2}(\tilde{\alpha}, \tilde{\beta}) = \sigma_{a^2}^2 \exp\left\{-\left[\left(\frac{\tilde{\alpha}}{l}\right)^2 + \left(\frac{\tilde{\beta}}{l}\right)^2\right]\right\} , \tag{23}$$

then under the condition that the intensity-correlation interval is much less than the aperture width L, the S/N ratio reduces to

$$\frac{\mathcal{J}_s\left(\frac{\xi}{\lambda R}, \frac{\eta}{\lambda R}\right)}{\mathcal{J}_N\left(\frac{\xi}{\lambda R}, \frac{\eta}{\lambda R}\right)} \approx \frac{\left|A'_o\right|^2}{\left|A'_R\right|^2} \left[\frac{\overline{a^2}}{\sigma_{a^2}}\right]^2 \left[\frac{L}{l}\right]^2 \frac{\exp\left[\pi^2 \frac{l^2\theta^2}{\lambda^2}\right]}{\pi} , \tag{24}$$

where θ is the angular separation of the object and reference.

Thus for narrow-angle scenes (small θ), a major source of noise arises from light deflected out of the "zero-order" reconstruction component, a consequence of the random intensity modulation of the interference pattern. The average noise intensity is strongest at object points nearest the reference and grows smaller as the object-reference separation increases. Note that to achieve a specified S/N ratio in the reconstructed image, the reference intensity should not exceed the object intensity by more than some calculable factor. By way of illustration, suppose that an S/N ratio of 10 is required for all points further than 1 sec of arc (5×10^{-6} rad) from the reference. Then, from Eq. (24), the ratio of reference to object intensities must satisfy

$$\frac{\left|A'_R\right|^2}{\left|A'_o\right|^2} < \frac{1}{10} \left[\frac{\overline{a^2}}{\sigma_{a^2}}\right]^2 \left(\frac{L}{l}\right)^2 \frac{\exp\left[(5\times 10^{-6})^2 \frac{\pi^2 l^2}{\lambda^2}\right]}{\pi} \,.$$

As typical values we take

$$\lambda = 7 \times 10^{-7} \text{m (ruby laser)},$$

$$L = 1.2 \text{ m (Cloudcroft)},$$

$$\ell \cong 0.12 \text{ m},$$

$$a^2/\sigma_{a^2} \cong 3,$$

yielding

$$\frac{\left|A'_R\right|^2}{\left|A'_o\right|^2} < 5 \times 10^3 \quad .$$

Thus the reference intensity should not be more than 5,000 times the object intensity. The S/N ratio will, of course, substantially exceed 10 at points on the object that are further than 1 sec of arc from the reference. Furthermore, the S/N ratio at all points is improved by increasing the aperture size.

Consider next the second source of noise, namely light deflected out of the reconstructed images themselves. Such noise results from the variations of a^2 about its mean in expression (18) for $\mathcal{J}_1(u, v)$. Calculation of the S/N ratio due to this effect yields

$$\frac{\mathcal{J}_s}{\mathcal{J}_N} = \left[\frac{\overline{a^2}}{\sigma_{a^2}}\right]^2 \frac{L^2}{\pi \ell^2} \quad . \tag{25}$$

For the numbers of the previous example this yields an S/N ratio of about 300. However, this result is overly optimistic, for a number of bright object points will generally lie close enough to the one point source assumed here to contribute additional noise at the image point of interest. The analysis becomes rather complicated in this more general case, and we must refer to the experimental results shown in Figure 21 of Volume 1 of this report to verify that usable S/N ratios can be achieved in practice. This particular aspect of image quality requires further exploration. In particular, the crucial question of what S/N ratios can be achieved in the real atmosphere must be answered.

EFFECTS OF TARGET MOTION

We consider now the effects of target motion during the illumination period. Of particular interest are the limitations to exposure time that are implied by linear translation of the satellite vehicle, and by rotation of the vehicle. To

analyze the effects of such motions we first develop a temporal filtering property of holograms.

Let the time interval over which the reference wave illuminates the detector be represented by T. This interval is determined by the duration of the transmitted optical illumination, or alternatively by some shuttering mechanism. The total energy density $\mathcal{E}$ incident on the detector at point (α, β) is thus

$$\mathcal{E}(\alpha,\beta) = \int_{-T/2}^{T/2} I(\alpha,\beta;\, t)\, dt \,, \tag{26}$$

where I is the point-by-point intensity at the detector, and for convenience we have defined the time origin at the center of the integration period.

Even though the transmitted light is approximately monochromatic, the light incident on the detector will no longer be monochromatic. Rather, each point on the object and reference will introduce a Doppler shift determined by its particular radial component of velocity at the time of illumination. We assume that the reference is a single point source with a well-defined Doppler shift, and we again write the complex amplitude of the reference at the detector as ${}^{(2)}E_r(\alpha,\beta)$. The complex amplitude of the incident object wave can then be written as ${}^{(2)}E_o(\alpha,\beta;t)$, where all phasors are taken to rotate at the angular frequency of the reference. The total energy incident at (α,β) can then be written

$$\mathcal{E}(\alpha,\beta) = \int_{-T/2}^{T/2} \Big\{ \Big|{}^{(2)}E_r(\alpha,\beta)\Big|^2 + \Big|{}^{(2)}E_o(\alpha,\beta;t)\Big|^2 + {}^{(2)}E^*_r(\alpha,\beta)\,{}^{(2)}E_o(\alpha,\beta;t) + {}^{(2)}E_r(\alpha,\beta)\,{}^{(2)}E^*_o(\alpha,\beta;t) \Big\}\, dt. \tag{27}$$

The last two terms are responsible for generating the twin images. Restricting attention to one of the twin images, consider only the term

$$\mathcal{E}'(\alpha,\beta) = {}^{(2)}E^*_r(\alpha,\beta) \int_{-T/2}^{T/2} {}^{(2)}E_o(\alpha,\beta;t)\, dt. \tag{28}$$

To find a convenient interpretation of the effects of the object time variations, we manipulate the expression for $\mathcal{E}'(\alpha,\beta)$ to yield a frequency-domain relation.

Let Eq. (28) be rewritten as

$$\dot{e}(\alpha,\beta) = {}^{(2)}E^{*}_{r}(\alpha,\beta) \int_{-\infty}^{\infty} \text{rect}\,\frac{t}{T}\ {}^{(2)}E_{o}(\alpha,\beta;t)\,dt\,, \tag{29}$$

where

$$\text{rect}\,\frac{t}{T} \triangleq \begin{cases} 1 & -\frac{T}{2} \le t \le \frac{T}{2}\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{30}$$

We now use Parseval's theorem to write

$$\dot{e}(\alpha,\beta) = {}^{(1)}E^{*}_{r}(\alpha,\beta)\,T \int_{-\infty}^{\infty} \text{sinc}\,Tf\ {}^{(2)}F_{o}(\alpha,\beta;f)\,df\,, \tag{31}$$

where

$$\text{sinc}\,T\,f \triangleq (\sin \pi Tf)/\pi Tf \tag{32}$$

and ${}^{(2)}F_{o}(\alpha,\beta;f)$ is the Fourier transform of ${}^{(2)}E_{o}(\alpha,\beta;t)$.

To interpret this result, note that the uniform time exposure of duration T is equivalent in the frequency domain to a linear filtering operation with (temporal) transfer function

$$H(f) = \text{sinc}\,Tf\,. \tag{33}$$

Thus the intensity of an object point with a Doppler shift that differs from that of the reference by f cps is suppressed in the reconstructed image by a factor $\text{sinc}^2 Tf$. If the object point happens to be moving such that its Doppler shift is n/T cps (n is an integer) larger or smaller than that of the reference, that object point will be entirely suppressed in the reconstructed image. Thus for high-quality imaging of the object, the maximum allowable Doppler difference between the object and reference points is about 1/2T cps.

To estimate how short the exposure time must be for typical satellite problems, consider first the lateral translation problem illustrated in Figure 1. A rigidly attached point object and point reference are traveling parallel to the ground with velocity v. The angle between object and reference when the vehicle is directly overhead is represented by ϕ. When the vehicle is not directly overhead, the angle of departure from the normal is represented by θ. The radial

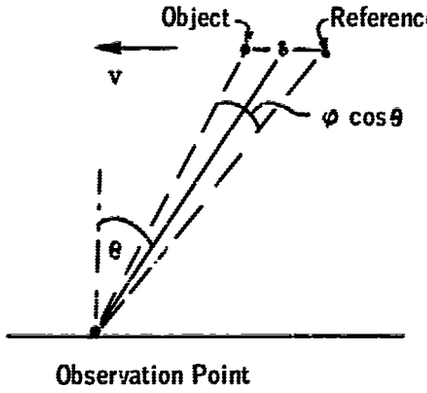


FIGURE 1 Linear translation of the satellite.

velocity components of the reference and object can be written, respectively, as

$$v_R = v \sin [\theta + (\phi \cos \theta)/2)] \ , \qquad (34)$$

$$v_o = v \sin [\theta - (\phi \cos \theta)/2] \ ,$$

or, for small ϕ,

$$v_R = v [\sin \theta + \tfrac{\phi}{2} \cos^2 \theta] \ , \qquad (35)$$

$$v_o = v [\sin \theta - \tfrac{\phi}{2} \cos^2 \theta] \ ,$$

The radial velocity difference is thus

$$\Delta v = v \phi \cos^2 \theta \ . \qquad (36)$$

As might be expected, the maximum radial velocity difference occurs directly overhead ($\theta = 0$), yielding a maximum Doppler difference of

$$\Delta v = 2(v/c) f_o \phi = 2(v/\lambda) \phi \ , \qquad (37)$$

where f_o is the frequency of the transmitted light and c is the velocity of light. Since the maximum Doppler difference must be less than $1/2T$ for high-quality imaging, the exposure interval must satisfy

$$T < \lambda/4v\phi \ . \qquad (38)$$

For the following typical numbers,

$\lambda = 7 \times 10^{-7}$ m,

$v = 8000$ m/sec,

$\phi = 20$ sec of arc (10^{-4} rad),

we find that the maximum exposure period is about 0.2 μsec (200 nsec). Longer exposure periods can be used at larger satellite angles and for smaller satellites.

Consider next the problem of angular rotation of the object. To take the worst possible case, suppose that the rotation axis is normal to the line of sight. We also suppose that the reference is on one side of the vehicle, rotating toward

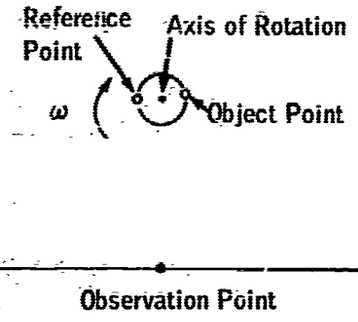


FIGURE 2 Rotation of the satellite.

(or away from) the receiving point. To obtain a satisfactory image of the portion of the vehicle on the opposite side of the rotation axis from the reference (see Figure 2), the maximum Doppler difference must be less than $1/2T$ cps. Let the angular rotation rate be ω and represent the maximum distance between object and reference (normal to the rotation axis) by δ. The radial velocities of the reference and object (assuming they lie symmetrically on opposite sides of the rotation axis) are

$$v_R = +\omega(\delta/2) ,$$
$$v_o = -\omega(\delta/2) , \tag{39}$$

or a velocity difference of $\omega\delta$. The Doppler difference is therefore

$$\Delta f = 2\omega\delta/\lambda . \tag{40}$$

For a high-quality image the exposure time must satisfy

$$T < \lambda/(4\omega\delta) . \tag{41}$$

Alternatively, if we specify the shortest pulse duration that we can conveniently generate (T_{min}), the maximum angular rotation rate ω_{max} can be written

$$\omega_{max} = \lambda/(4\,T_{min}\,\delta) . \tag{42}$$

To illustrate with a typical example, suppose

$\delta = 10$ m,

$T_{min} = 20$ nsec,

$\lambda = 7 \times 10^{-7}$ m.

Then

$$\omega_{max} = \frac{7 \times 10^{-7}}{8 \times 10^{-7}} \cong 1 \text{ rad/sec} .$$

Thus angular rotation rates less than 1 rad per sec can be tolerated in this worst geometry. When the rotation axis is not normal to the line of sight, or when the satellite is smaller than 10 m, higher rotation rates can be tolerated.

In summary, angular rotation places far more severe restrictions on exposure duration than does linear translation. To obtain high-quality pictures of rotating satellites, it would appear imperative that short-pulse Q-switched lasers be used as the optical source. Even when such a source is used, there will occur some extreme geometries for which successful imaging cannot be performed over the full extent of the satellite. The occurrence of such unfavorable situations would appear rather rare, however.

APPENDIX 30

DIFFRACTION EFFECTS IN WAVEFRONT-RECONSTRUCTION IMAGING

J. L. Walsh

INTRODUCTION

Goodman has proposed a method of imaging through a random medium by what is essentially a holographic technique.

In the laboratory experiments conducted thus far, one finds as expected that good images can be obtained on reconstruction when the phase perturbing screen is located either close to the object or close to the film on which the hologram is made. As one moves the phase-perturbing screen to a position midway between object and the hologram being produced, the quality of the reconstructed image is degraded. As we shall see in the following discussion, this difficulty arises because diffraction effects are most important when the phase perturbation is midway between the object and the hologram being produced. In this discussion we wish to consider specifically the following question: What can be said about the effects diffraction phenomena occurring in the real atmosphere are likely to have on the reconstructed images?

Let us begin by considering a simple model of the atmosphere. It seems to be fairly well established[1] that under good seeing conditions at least, the scintillation of starlight is primarily a diffraction phenomena which can be considered to be caused by a layer at an altitude $\approx$ 10 km which behaves as a random phase grating. It can be shown that since the intensity fluctuation which is observed on the ground is a sizable fraction of the average intensity, the rms phase variation

must be ≈ 1 rad. In addition, Hufnagel[2] finds that the mean-square variation in optical path between two rays for a plane wave entering the top of the atmosphere and penetrating to an altitude ≈ 10 km is given by

$$\overline{(\ell_1 - \ell_2)^2} = 6 \times 10^{-12}\, r^{5/3} , \tag{1}$$

where ℓ_1 and ℓ_2 are the optical paths of the two rays in centimeters and r is their separation in centimeters.

As a simple model, then, we assume a phase grating of zero thickness parallel to the earth's surface for which the phase delay is given by

$$\phi = \phi_o \sin (2\pi x/\ell) . \tag{2}$$

In our model, based on the previous considerations, we are concerned with a layer at an altitude of 10^6 cm for which $\phi_o \approx 1$ rad and $\ell \approx 20$ cm.

Immediately above the phase grating we consider two plane waves incident as shown in Figure 1. We can then write for the disturbance at y = 0 immediately below the phase grating

$$U(x, 0) = A \exp\left(i\phi_o \sin \frac{2\pi x}{\ell}\right) + a \exp\left(ik\gamma x + i\phi_o \sin \frac{2\pi x}{\ell}\right) , \tag{3}$$

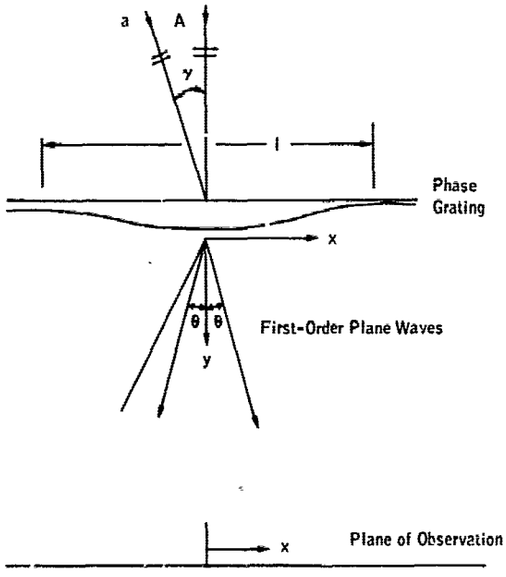


FIGURE 1 Reference wave A and signal wave a incident on a phase grating.

where A is the amplitude of the reference wave, a that of the signal wave separated from the reference by the angle γ, $k = 2\pi/\lambda$, and the time dependence exp (-iωt) is suppressed. We will also assume that a << A. In order to write the solution for y > 0, we require the expansion

$$\exp\left[i\phi_o \sin \frac{2\pi x}{\ell}\right] = \sum_{m=-\infty}^{\infty} J_m(\phi_o) \exp\left[im \frac{2\pi x}{\ell}\right] . \tag{4}$$

In the region y > 0, we see from the form of Eq. (4) that we can express

$U(x, y)$ as an infinite series of plane waves of the form

$$\exp(i\alpha x + i\beta y) ,$$

where α and β can be complex subject to the requirement imposed by the wave equation, namely, $\alpha^2 + \beta^2 = k^2$. For $y > 0$ we obtain, if $\gamma, \theta \ll 1$,

$$U(x, y) = \sum_{m=-\infty}^{\infty} J_m(\phi_o) \exp\left[\mathrm{Im}\ \frac{2\pi x}{\ell}\right] \times \left\{A \exp ik\left[1 - \frac{m^2\theta^2}{2}\right] y + a \exp ik\left[\gamma x + y\left(1 - \frac{m^2\theta^2}{2} - \gamma m\theta - \frac{\gamma^2}{2}\right)\right]\right\} , \quad (5)$$

where θ is given by

$$\theta = \lambda/\ell . \quad (6)$$

We note that Eq. (5) corresponds to a spectrum of plane waves making angles with the y axis given by $m\theta$ for the A waves and $\gamma + m\theta$ for the a waves, where m takes on positive and negative integral values. The amplitudes of the various components are given by $J_m(\phi_o)$, which is significant only for m values not much greater than ϕ_o.

In terms of our model, θ has a value $\theta \approx 2 \times 10^{-6}$ rad for visible light and we find for $\phi_o \approx 1$ rad significant energy only for the first few terms in the Bessel function expansion. In the actual case, of course, the phase screen has a spectrum associated with it, but the spreading will still be of the order of a few times θ and the sharp angular spectrum will be washed out. We might note further that θ given above is consistent with the value of seeing produced by the layer responsible for scintillation, namely several tenths of a second of arc.

FRINGE PATTERN

Let us turn now to a consideration of the fringe pattern. Squaring the real part and taking the time average of Eq. (5) we obtain for the intensity

$$\left\langle \left|U(x,y)\right|^2 \right\rangle = \sum_{m,n=-\infty}^{\infty} J_m(\phi_o)\, J_n(\phi_o) \left\{ \frac{A^2}{2} \cos\left[(m-n)\frac{2\pi x}{\ell} + \frac{k\theta^2}{2}(n^2-m^2)\, y \right] \right.$$

$$\left. + \; aA \cos\left[\left(m-n-\frac{\gamma}{\theta}\right)\frac{2\pi x}{\ell} + \frac{k\theta^2}{2}\left(n^2-m^2+\frac{2n\gamma}{\theta}+\frac{\gamma^2}{\theta^2}\right) y \right] + O(a^2) \right\} . \tag{7}$$

Examining the interference term aA in Eq. (7), we note that there are several fringe patterns superposed. The (0, 0) term in the sum gives fringes with the correct spacing in x, namely,

$$\cos\left(2\pi\gamma x/\theta\ell\right) = \cos k_\gamma x .$$

We note, however, that when $\phi_o \approx 1$ the (0, 1) and (1, 0) terms in the sum give rise to a fringe spacing corresponding to object points located at $\gamma + \theta$ and $\gamma - \theta$. Thus it appears that we begin to have a problem with resolution of points whose angular separation is smaller than θ. In the real atmosphere, of course, we have a continuous spectrum of θ values and the result will be an enhanced noise contribution making resolution increasingly difficult for angular separations less than $\theta \sim 0.5$ sec of arc. This point requires further consideration.

In the foregoing discussion we have considered Eq. (7) on a term by term basis. One must be rather careful in this approach, however, because it neglects important interference effects which can occur between different terms in the sum. There are two cases of particular interest in which we can sum the series. They are:

1. y small, i.e., $y \ll \ell^2/2\lambda$;

and

2. $y = \ell^2/2\lambda$, the first value of y for which the A wave has only intensity variation and no phase variation with x.

EXPLICIT EVALUATION FOR SMALL VALUES OF y

For small y we write Eq. (7) in complex form, and expand the exponentials keeping only the first order terms in y. The necessary summation formulas are given in Exhibit A. After some manipulation, we obtain

$$\langle |U(x,y)|^2 \rangle = \frac{A^2}{2}\left[1 + k\theta^2 y \sin\frac{2\pi x}{\ell}\right] + aA\left\{\cos(k\gamma x)\left[1 + k\theta^2\phi_o y \sin\frac{2\pi x}{\ell}\right] + ky\left[\frac{\gamma^2}{2} + \gamma\theta\phi_o \cos\frac{2\pi x}{\ell}\right]\sin k\gamma x\right\}, \quad (8)$$

subject to the condition $y \ll \ell^2/2\lambda$, which in our example is

$$y \ll \frac{\ell^2}{2\lambda} = \frac{(20)^2}{2(7 \times 10^{-5})} \approx 30 \text{ km} . \quad (9)$$

We note here, as has been pointed out by Goodman, that as $y \to 0$ the interference pattern is unaffected by the presence of the phase screen. As y increases, the hologram begins to acquire spatial frequencies characteristic of the period of the phase grating, which is equivalent to an additional object at an angle $\gamma = \theta$; and again we see that in the general case resolution between objects separated by angles $\approx \theta$ or less will become increasingly difficult.

In terms of our model, the separation between the phase screen and the hologram at which this effect begins to set in depends on the relative magnitudes of the signal and reference. Thus from Eq. (8) we see that the additional fringe pattern begins to appear when $Ak\theta^2 y \approx a$ or in terms of our model when

$$y \approx \frac{a}{A}\,\frac{\ell^2}{2\lambda} \approx \frac{30a}{A} \text{ km} . \quad (10)$$

Depending of course on the ratio a/A, we see that phase screens having $\phi_o \approx 1$ rad and located many meters from the aperture can have negligible effects on the hologram which is produced.

EXPLICIT EVALUATION FOR $y = \ell^2/2\lambda$

In this case it is easier to carry out the sum in Eq. (5) and then square and average. Using the results given in Exhibit A, we find after some manipulation

$$< \left|U^2(x, \ell^2/2\lambda)\right| > = \frac{A^2}{2}\left[1 + \sin\left(2\phi_o \sin\frac{2\pi x}{\ell}\right)\right] + aA\cos\left(k\gamma x - \frac{\pi\gamma^2}{2\theta^2}\right)$$

$$\times \left\{\cos\left[2\phi_o \sin\frac{\pi\gamma}{2\theta}\cos\left(\frac{2\pi x}{\ell} - \frac{\pi\gamma}{2\theta}\right)\right]\right.$$

$$\left. + \sin\left[2\phi_o \cos\frac{\pi\gamma}{\theta}\sin\left(\frac{2\pi x}{\ell} - \frac{\pi\gamma}{2\theta}\right)\right]\right\} , \qquad (11)$$

and in this case substantial sidebands are produced, corresponding to additional objects at angles $\gamma+\theta$ and $\gamma-\theta$, so that resolution to angles smaller than θ would seem to be impossible. In terms of our model we saw that $y = \ell^2/2\lambda \approx 30$ km, which is of the same order of magnitude as the height of the layer responsible for scintillation. Thus the effects to be expected in practice would probably correspond more closely to those given in Eq. (11).

SUMMARY

In the simple model we have chosen to represent the seeing layer, it appears that resolution of points separated by angles $\lesssim 0.5$ sec of arc becomes increasingly difficult. At interesting altitudes of the order of 100 to 1,000 km, the corresponding distances on the body are 20 to 200 cm. Admittedly the calculation does not fully incorporate the correct statistical description of the atmosphere down to the 10-km level.

When this refinement is included, the result will be to introduce noise into the system, making angular resolution better than θ increasingly difficult.

On the other hand, it appears that one can render much of the close-in turbulence relatively ineffective. Thus we may find that what we have accomplished with Goodman's approach is to achieve the good seeing conditions associated with the presence of only the scintillation layer, a much larger fraction of the time than would be possible with incoherent imaging.

EXHIBIT A

We begin with the expansion

$$\exp(iz \sin\theta) = \sum_{m=-\infty}^{\infty} J_m(z) \exp(im\theta) \quad . \tag{A1}$$

Differentiating twice with respect to θ and multiplying by the complex conjugate of Eq. (A1), we obtain

$$iz \sin\theta + z^2 \cos^2\theta = \sum_{m,n=-\infty}^{\infty} m^2 J_m(z) J_n(z) \exp i\theta(m-n) \tag{A2}$$

and its complex conjugate

$$-iz \sin\theta + z^2 \cos^2\theta = \sum_{m,n=-\infty}^{\infty} n^2 J_m(z) J_n(z) \exp i\theta(m-n) \quad . \tag{A3}$$

Similarly, we can easily see that

$$z \cos\theta = \sum_{m,n=-\infty}^{\infty} nJ_m(z) J_n(z) \exp i\theta(m-n) \quad . \tag{A4}$$

Multiplying Eq. (A1) by its complex conjugate we obtain

$$1 = \sum_{m,n=-\infty}^{\infty} J_m(z) J_n(z) \exp\left[i(m-n)\theta\right] \quad . \tag{A5}$$

In carrying out the summations we also make use of the properties of the squares of integers

$$\exp\left(-im^2 \frac{\pi}{2}\right) = \begin{cases} -i, & m \text{ odd}, \\ 1, & m \text{ even}. \end{cases} \tag{A6}$$

REFERENCES

1. C. G. Little, Mon. Notices Roy. Astron. Soc. 111, 289 (1951).

2. R. E. Hufnagel, private communication.

APPENDIX 31

AIRBORNE RETROREFLECTORS FOR HOLOGRAPHIC TECHNIQUES

S. A. Collins, Jr.

This paper concerns the possibility of creating an artificial reference beam for the holographic technique by placing a corner-cube reflector above the tropopause, but within the same isoplanatic region as the target. The thought is to use either a high-flying balloon or airplane, with appropriate retroreflector affixed.

The principal difficulty associated with this idea is getting the plane sufficiently close to the field of view. To take a very favorable case, let the isoplanatic region be 40 sec of arc across (twice the size of a representative satellite). For an airplane at an altitude of 80,000 ft, this means the airplane has to be positioned to within ±8 ft. In other words, the airplane has to find a point moving with variable velocity (because it is between the satellite and telescope) along a path 16 ft wide. It appears to be extremely difficult[1] to bring an airplane consistently closer than one-half mile of a given point, let alone a moving one.

There is a chance that the majority of atmospheric effects are caused by low-level turbulence, i.e., below the tropopause. If that were the case, isoplanatic regions for the reference corner cube might be larger, and techniques for maneuvering airplanes might be better developed. It would be desirable to perform an experiment to check the effect of the various atmospheric levels. Nevertheless, it still seems very improbable that airborne corner-cube references will be used for the holographic scheme in the near future.

REFERENCE

1. Col. T.S. Byington, personal communication.

APPENDIX 32

REMOVAL OF MULTIPLE RECONSTRUCTED IMAGES DUE TO MULTIPLE REFERENCE SOURCES IN WAVEFRONT RECONSTRUCTION

Raoul F. vanLigten

The reference beam in Goodman's system is derived from a specular reflection on the object. There can, however, occur more than one such reflection with the consequence that in the reconstruction a multiplicity of overlapping images will appear. A method for isolating a single image is described here.

In Figure 1(a) the case is shown for two reference sources. Figure 1(b) shows the reconstruction with one source. It should be noted that the schemes of Figures 1(a) and 1(b) are to illustrate how multiple images can be separated. Later the connection is drawn with the Goodman system. Figure 1(b) shows the two overlapping images and the corresponding two bundles which appear to originate at the illumination source in Figure 1(a). If only one reference source were present the total bundle, as seen from the hologram plane, would be completely reconstructed. This follows from the fact that the hologram has arrested a propagating field in both amplitude and phase. Upon reconstruction this field continues to propagate in the same fashion as it would without the hologram and from the same two-dimensional plane in which the hologram was placed. To the observer it still seems that in the reconstruction the light originates from the illumination source, passes through the object (where it is diffracted), and subsequently arrives at the observer. When now two reference beams are present each at a different offset angle from the image bundle, there will be two reconstructed

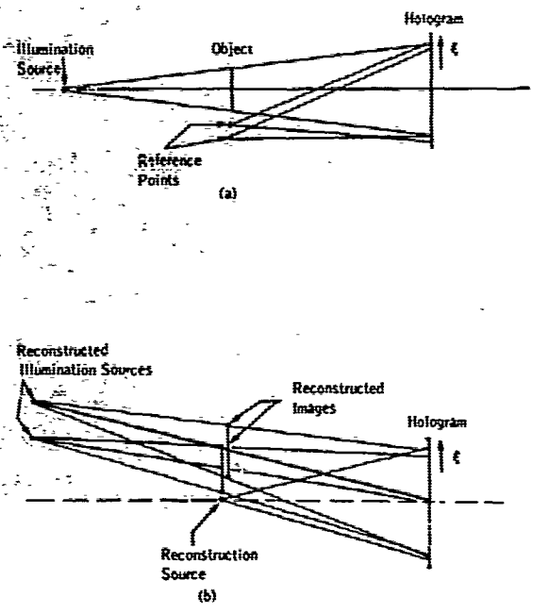


FIGURE 1 (a) A hologram made with two reference sources. (b) The reconstruction of the hologram made in (a) obtained with a single reconstructing beam. The twin images are not shown.

images and two corresponding twin images. [In Figure 1(b) only the first two are shown.] This can be demonstrated by the following argument. It is assumed that the object wave at the hologram plane is given by the complex function $F(\xi)$ and that the two reference waves are given, respectively, by the complex functions $A_1(\xi)$ and $A_2(\xi)$. The hologram contains these functions in the form

$$H(\xi) = [A_1 + A_2 + F] \cdot [A_1 + A_2 + F]^* \tag{1}$$

[see for instance Eq. (5) in reference 1]; that is,

$$H(\xi) = |A_1|^2 + |A_2|^2 + |F|^2 + A_1A_2^* + A_2A_1^* + A_1F^* + A_2F^* + A_1^*F + A_2^*F \tag{2}$$

As follows, for instance, from Eqs. (22), (23), and (24) of reference 1, the first three terms in Eq. (2) will in the reconstruction, with one source, yield a wave that propagates unperturbed in the direction of the reconstruction wave and will carry information with little or no resemblance to the original object. The term $A_1A_2^*$ will be a uniform wave propagating in the direction given by the difference of the offset angles of A_1 and A_2; the term $A_1^*A_2$ will yield a uniform wave again propagating in a direction given by the difference of the offset angles of A_1 and A_2, but the difference angle has opposite sign to that of $A_1A_2^*$. The reconstructed waves $A_1A_2^*$ and $A_1^*A_2$ are not shown in Figure 1(b). Neither are the two waves A_1F^* and A_2F^*. These bundles each carry an image in a direction given by the offset angle determined respectively by A_1 and A_2.

Finally, the two terms A_1^*F and A_2^*F yield two images, with an offset angle from the reconstruction source given by A_1^* (= - angle of A_1) and A_2^* (= - angle A_2). The bundles produced by the latter two terms are shown in Figure 1(b). From what was shown earlier, each of the bundles is reconstructed in its entire form as they propagate from the hologram plane.

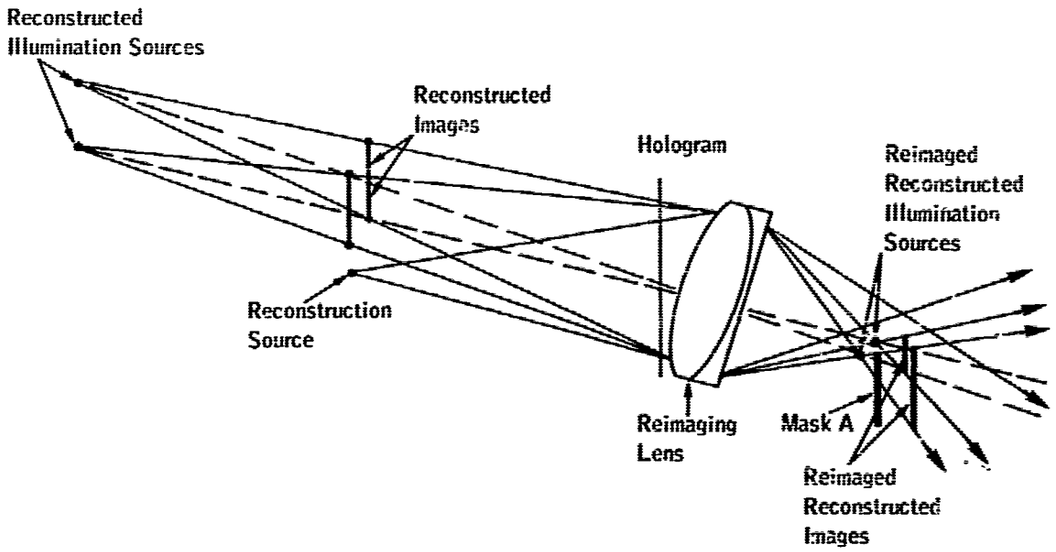


FIGURE 2 A scheme to eliminate one of the overlapping images by reimaging and insertion of an occluding mask A.

By placing an appropriate lens behind the hologram as shown in Figure 2, the two images and their bundles are reimaged. Subsequently, an occluding mask is placed such that only one of the two reimaged illumination sources is permitted to pass, causing the associated reimaged reconstructed image only to remain.

The other terms of Eq. (2) have all their own converging points in Figure 2 if they were drawn in, but they will be all masked out when their convergence centers are different from the one permitted to pass.

The reimaged illumination sources do not have the same form and extent in amplitude and phase as the original sources. They are influenced by diffraction and scatter at the object. When the hologram is made of an object from which the light reflects, the description is still the same as given above, but with due account taken of the scattering properties of the surface and of the diffraction from the object.

It is clear that when the scattering and diffraction at the object are sizable and the reconstructed sources become larger in extent than the image, this technique has no advantage. In this case, to obtain nonoverlapping images with multiple reference-point sources, the separation between the sources must exceed the size of the object.

The question now arises how this technique can be used in the case of the Goodman system. In Figure 3 a sketch of this system is given. The object is to provide a reference source from specular reflection of the illumination source on the ground. After the laser light has reflected from the object, this object can

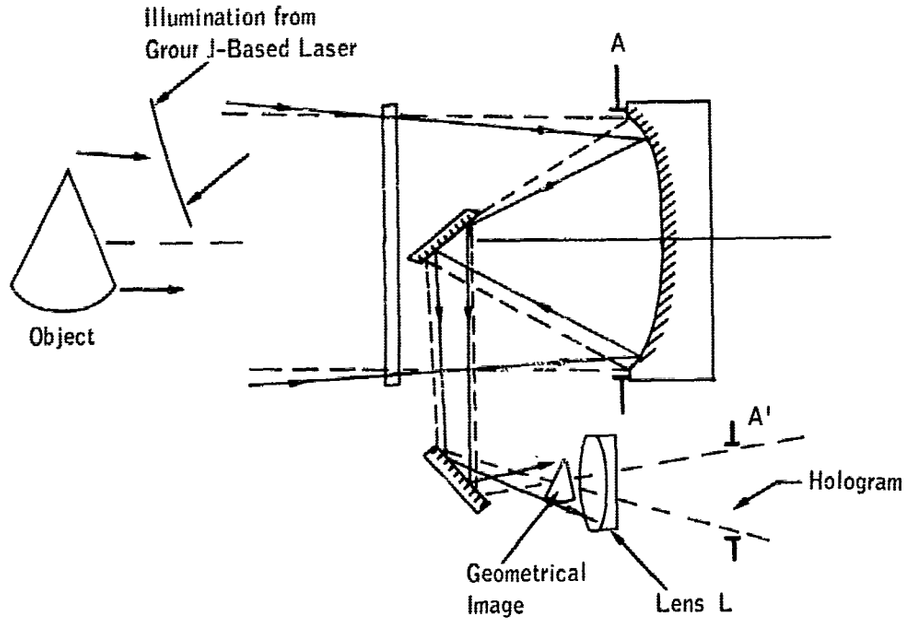


FIGURE 3 The Goodman system with a lens L which images the telescope aperture to a smaller size at A', in which plane a hologram-producing detector is placed. A = (48-in.) telescope pupil; A' = image of A produced by lens L. Dashed line = rays coming from axial point of the object. The detector array is positioned at A', to produce a hologram.

be considered as a transparency with phase-retardation, which is related to the three-dimensional shape of the object, and an amplitude attenuation which depends on the intensity of reflected light. The transparency appears to be illuminated by a spherical wave through a scattering medium, where the scattering medium is determined by the surface structure of the object.

Thus the case sketched in Figure 1(a) is obtained. Namely, the geometrical image in Figure 3 is the object, the reference source is provided by the object itself and the illumination source for the object lies somewhere between the 48-in. telescope aperture and the geometrical image. Although the original object is very far away, it cannot be considered at infinity, otherwise the geometrical image would be a point. Because it is assumed that under conditions free of atmospheric turbulence the geometrical image is finite in extent, the ground-based illumination source should be considered as a finite distant light source. Let this distance be R_o from the object. The equivalent distance at the geometrical image will be approximately M^2R_o, where M is the linear magnification from the original object to the geometrical image of the object.

When the object distance is 200 km and the focal length of the telescope is 300 in., the product M^2R_o is 2/7 mm. This is, of course, very small; however, when the reconstructed image is magnified with a linear magnification of 10X with a well-corrected lens, the aerial image of the illumination source will be $(10)^2 \times 2/7 \approx 28$ mm separated from the aerial image of the original object.

REFERENCES

1. R. F. vanLigten, J. Opt. Soc. Am. 56, 1 (1966).
2. R. F. vanLigten, J. Opt. Soc. Am. 56, 1009 (1966).

APPENDIX 33

ABERRATIONS IN WAVEFRONT-RECONSTRUCTION IMAGING CAUSED BY THE OPTICS AND THE DETECTOR

Raoul F. vanLigten

In the first part of this paper, we shall treat the magnification and third-order aberrations from object to telescope in the Goodman system as illustrated in Figure 3 of Appendix 32 (see page 218). It is assumed that the hologram was taken in the telescope aperture. It is found that the imagery will be aberration-free when: (1) the wavelength in taking the hologram is the same as that used for the reconstruction, (2) the reconstruction source is located on the line which is defined by the reference source and the origin of the coordinate system in the hologram, and (3) the radius of curvature of the reconstruction wavefront equals that of the reference wavefront but is opposite in sign.

Because in Goodman's system an additional imaging step is required by lens L, it is necessary to investigate how this lens will influence the image quality. In the second part of this paper, the imaging step from the 48-in. aperture A to the demagnified pupil A' in the figure on page 218 is investigated, keeping in mind that this is a holographic imaging process. It is concluded that the only requirement on the lens L is freedom from coma. A specific example is given and it is concluded that a good commercial photographic objective is probably well suited for this function.

In the final part of this paper, a review is given of the influence of the detector in the exit pupil. A more detailed numerical analysis cannot be given without additional information about the detector. In this final part it is concluded that

(1) nonaxial distortions in the detector array between making the hologram and reconstructing the image from it, will appear as astigmatism in the reconstructed image, (2) the spatial modulation transfer function of the detector array (or image orthicon) only limits the field of view of the reconstructed image, and (3) the dynamic range of the detector will impose a limit on the resolution in the reconstructed image when the hologram signal exceeds the dynamic range.

MAGNIFICATION AND THIRD-ORDER ABERRATION IN THE HOLOGRAPHIC SYSTEM OF J. W. GOODMAN

Magnification and third-order aberration of holographic images have been treated by R. W. Meier.[1] The starting conditions for Goodman's system, such as the angle between image wavefront and reference wavefront, are such that Meier's work can be applied. For the magnification in the image, Meier's expressions are used in object space. For the third-order aberrations, the image space is chosen. This can be done provided there are no aberrations in the imagery from entrance pupil to exit pupil.

Magnification

In Goodman's case, where for photometric reasons, it would be better to demagnify the pupil to obtain a reasonable exposure time, the magnification m in Meier's paper will have to take a magnitude < 1.

With the following definitions from Meier's paper,

Z_o = distance from film to object,

Z_r = radius of curvature of reference wavefront,

Z_c = radius of curvature of reconstructed wavefront,

Goodman's case corresponds to $Z_o = Z_r$.

The magnification $M_{v\,lat}$ will take the form

$$M_{v\,lat} = \frac{m\mu Z_c}{m^2 Z_o} = \left(\frac{1}{m}\ \mu\ \frac{Z_c}{Z_o}\right) ,$$

where

$$\mu = \lambda_c / \lambda_o .$$

The longitudinal magnification is

$$M_{long} = -\frac{1}{\mu}\frac{1}{m^2}\mu^2\left(\frac{Z_c}{Z_o}\right)^2 = -\frac{1}{m^2}\ \mu\ \left(\frac{Z_c}{Z_o}\right)^2 .$$

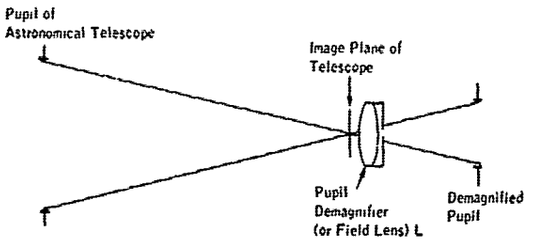


FIGURE 1 The lens L images the entrance pupil of the telescope into a demagnified exit pupil. The lens L is the same as lens L in Figure 3, page 218.

For $Z_c \approx Z_o$, the above reduces to

$$M_{v\,lat} = \mu/m$$

and

$$M_{long} = -\mu/m^2 .$$

The optical system we consider here to demagnify the pupil is shown in Figure 1. The demagnifier may, however, also be chosen as an eyepiece, which will image the telescope pupil as a real image.

Third-Order Aberrations

Spherical Aberration S With $Z_r = Z_o$, the spherical aberration S is equal to zero, for

$$S = \frac{\mu}{m^4}\left[0 - \frac{6\mu}{Z_c Z_o^2} + 0 + \frac{6\mu}{Z_c Z_o^2}\right] = 0.$$

Coma With $Z_r = Z_o$, the coma is

$$C_x = \frac{X_c}{Z_c^3} - \frac{\mu(X_o - X_r)}{m^2 Z_o^3} - \frac{a_R}{Z_R^3}$$

$$= \frac{X_c}{Z_c}\left[\frac{1}{Z_c^2} - \left(\frac{1}{Z_c} - 0\right)^2\right] - \frac{X_o}{Z_o}\frac{\mu}{m}\left[\frac{1}{m^2 Z_o^2} - \left(\frac{1}{Z_c} - 0\right)^2\right]$$

$$+ \frac{X_r}{Z_r}\frac{\mu}{m}\left[\frac{1}{m^2 Z_r^2} - \left(\frac{1}{Z_r} - 0\right)^2\right]$$

$$= 0 - \frac{X_o}{Z_o}\frac{\mu}{m}\left(\frac{1}{m^2 Z_o^2} - \frac{1}{Z_c^2}\right) + \frac{X_r}{Z_o}\frac{\mu}{m}\left[\frac{1}{m^2 Z_o^2} - \frac{1}{Z_c^2}\right]$$

$$= \frac{\mu}{m} \frac{1}{Z_o} \left(\frac{1}{m^2 Z_o^2} - \frac{1}{Z_c^2} \right) (X_r - X_o).$$

The coma would reduce to zero if $Z_c = \pm m Z_o$.

Astigmatism With $Z_o = Z_r$, the astigmatism is

$$A_x = \frac{X_c^2}{Z_c^2} - \frac{\mu}{m^2 Z_r^3} \left(X_o^2 - X_r^2 \right) - \left[\frac{X_c}{Z_c} - \frac{\mu}{m Z_r} (X_o - X_r) \right]^2 \left[\frac{1}{Z_c} \right]$$

$$= \frac{X_c^2}{Z_c^2} - \frac{\mu}{m^2 Z_r^3} \left(X_o^2 - X_r^2 \right) - \left[\frac{X_c^2}{Z_c^2} - 2 \frac{\mu X_c}{m Z_r Z_c} (X_o - X_r) + \frac{\mu^2}{m^2 Z_r^2} (X_o - X_r)^2 \right] \frac{1}{Z_c}$$

$$= - \frac{\mu}{m^2 Z_r^3} \left(X_o^2 - X_r^2 \right) - \frac{\mu^2}{m^2 Z_r^2 Z_c} (X_o - X_r)^2 + 2 \frac{\mu X_c}{m Z_r Z_c^2} (X_o - X_r).$$

In the case we deal with, $Z_r \to \infty$ and so

$$A_x = 0.$$

Note that we can still use a reconstruction wavefront with a finite radius of curvature.

If now the hologram is taken in the exit pupil of the telescope, we have the case where Z_r and Z_o are both finite and negative; then

$$A_x = \frac{\mu}{m^2 Z_r^2} \left[- \frac{1}{Z_r} \left(X_o^2 - X_r^2 \right) - \frac{\mu}{Z_c} \left(X_o^2 - 2 X_o X_r + X_r^2 \right) \right]$$

$$+ 2 \frac{\mu X_c}{m Z_r Z_c^2} (X_o - X_r)$$

$$= \frac{\mu}{m}\left[-\frac{X_o^2}{Z_r^2}\left(\frac{1}{mZ_r}+\frac{1}{mZ_c}\right)+\frac{2X_o}{Z_cZ_r}\left(\frac{X_c}{Z_c}+\frac{\mu X_r}{mZ_r}\right)\right.$$

$$\left.+\frac{X_r^2}{Z_r^2}\left(\frac{1}{mZ_r}-\frac{\mu}{mZ_c}\right)-\frac{2X_cX_r}{Z_c^2Z_r}\right].$$

In the reconstruction

$$m = 1, \quad \mu = 1,$$

and

$$A_x = \left[-\frac{X_o^2}{Z_r^2}\left(\frac{1}{Z_r}+\frac{1}{Z_c}\right)+\frac{2X_o}{Z_cZ_r}\left(\frac{X_c}{Z_c}+\frac{X_r}{Z_r}\right)+\frac{X_r^2}{Z_r^2}\left(\frac{1}{Z_r}-\frac{1}{Z_c}\right)\right.$$

$$\left.-\frac{2X_cX_r}{Z_c^2Z_r}\right].$$

If we make

$$\frac{X_c}{Z_c} = -\frac{X_r}{Z_r},$$

then

$$A_x = \left[-\frac{X_o^2}{Z_r^2}\left(\frac{1}{Z_r}+\frac{1}{Z_c}\right)+\frac{X_r^2}{Z_r^2}\left(\frac{1}{Z_r}-\frac{1}{Z_c}\right)+\frac{2}{Z_c}\frac{X_c^2}{Z_c^2}\right]$$

$$= -\frac{X_o^2}{Z_r^2}\left(\frac{1}{Z_r}+\frac{1}{Z_c}\right)+\frac{X_r^2}{Z_r^2}\left(\frac{1}{Z_r}+\frac{1}{Z_c}\right) = \left(\frac{X_o^2}{Z_r^2}-\frac{X_o^2}{Z_o^2}\right)\left(\frac{1}{Z_r}+\frac{1}{Z_c}\right).$$

If, furthermore, $|Z_r| = |Z_c|$ and the reconstruction point source is chosen to be real, then $A_x = 0$ because $[1/Z_r + 1/Z_c] = 0$.

Field Curvature

$$F = \frac{X_c^2+Y_c^2}{Z_c^3}-\frac{\mu(X_o^2+Y_o)^2}{m^2Z_o^3}+\frac{\mu(X_r^2+Y_r^2)}{m^2Z_r^3}-\frac{a_R^2+b_R^2}{Z_R^3}.$$

This form is the same as the form for astigmatism and therefore the same conditions which remove astigmatism will also remove field curvature.

Distortion

$$D_x = \frac{X_c^3 + X_c Y_c^2}{Z_c^3} - \frac{\mu(X_o^3 + X_o Y_o^2)}{m Z_o^3} + \frac{\mu(X_r^3 + X_r Y_r^2)}{m Z_r^3} - \frac{a_R^3 + a_R b_R^2}{Z_R^3} .$$

Where Y_c and Y_r are chosen to be zero, distortion becomes

$$D_x = \frac{\mu}{m}\left[\frac{X_o^3}{Z_o^3}\left(\frac{\mu^2}{m^2} - 1\right) - 3\frac{\mu}{m}\frac{X_o^2}{Z_o^2}\left(\frac{X_c}{Z_c} + \frac{\mu}{m}\frac{X_r}{Z_r}\right)\right.$$

$$+ 3\frac{X_o}{Z_o}\left(\frac{X_c}{Z_c} + \frac{\mu}{m}\frac{X_r}{Z_r}\right)^2 - \frac{X_r^3}{Z_r^3}\left(\frac{\mu^2}{m^2} - 1\right)$$

$$\left. - 3\frac{X_c X_r}{Z_c Z_r}\left(\frac{X_c}{Z_c} + \frac{\mu}{m}\frac{X_r}{Z_r}\right) + \frac{X_o Y_o^2}{Z_o^3}\left(\frac{\mu^2}{m^2} - 1\right)\right] .$$

In our case $\mu = m$ and $X_c/Z_c = -X_r/Z_r$ for the removal of coma and astigmatism. This yields

$$D_x = 0.$$

Conclusion If in Goodman's case, (1) the wavelength in taking the hologram is the same as the wavelength during reconstruction, (2) the reconstruction source is located on the line which is defined by the reference source and the origin of the coordinate system in the hologram, [1] and (3) the radius of curvature of the reconstruction wavefront equals that of the reference wavefront but is opposite in sign, then the reconstructed image will be free of third-order aberration.

PUPIL ABERRATIONS

Thus far it has been assumed that the aberrations from entrance pupil to exit pupil were zero. In general this is not the case and in the following this matter will be considered.

Let the diffracted wave coming from the object be given by

$$A_o e^{i\phi_o}$$

and that from the reference point by

$$F e^{i\phi_r} ;$$

then the field arriving at the entrance pupil is

$$E = A_o e^{i\phi_o} + F e^{i\phi_r} .$$

During the imagery from entrance pupil to exit pupil, aberrations will occur that can be classified as amplitude aberrations ΔA and phase aberrations $\Delta\phi$. These errors have the same effect on both waves and after transforming the wavefronts back to the entrance pupil, the two wavefronts interfering together can be expressed in the following form:

$$E + \Delta E = (A_o + \Delta A)\, e^{i[\phi_o + \Delta\phi]} + (F + \Delta A)\, e^{i[\phi_r + \Delta\phi]} .$$

The hologram now contains the energy of this combined wavefront:

$$[E+\Delta E]\cdot[E+\Delta E]^* = (A_o+\Delta A)^2 + (F+\Delta A)^2 + (A_o+\Delta A)(F+\Delta A)e^{i(\phi_o-\phi_r)} + (A_o+\Delta A)(F+\Delta A)e^{i(-\phi_o+\phi_r)} . \quad (1)$$

It is immediately obvious that the phase aberrations have disappeared completely.

The only aberrations left to be studied in the transfer of the wavefronts from entrance pupil to exit pupil are the ones affecting the magnitude of the waves. The latter aberrations are relatable to the geometrical aberrations occurring during the imagery from entrance pupil to exit pupil. A combination of geometrical aberrations in the pupil planes will cause an aberration in the reconstructed image. The diffraction limit of the pupil is given by whatever aperture limits the field of view of the image. To see the relation between the aberrations in the image plane and those in the object plane, the third-order aberration coefficients for a system of thin lenses can be determined.

The third-order coefficients are fixed by the following formula:

$$n'X' - Gn\overline{X} = \frac{1}{2}\left\{m_1\overline{X}^3 + m_2\overline{X}^2\overline{X}_1 + m_3\overline{X}\overline{X}_1^2 + m_4\overline{X}_1^3\right\}$$

in which

$$\overline{X} = GJnX/(S-G) \quad \text{and} \quad \overline{X}_1 = SJnX_1/(S-G) ,$$

where

X is the coordinate in the object plane,

X_1 is the coordinate in the entrance pupil plane,

G = object magnification,

S = pupil magnification,

J = power of the system of thin lenses,

n = refractive index in object space,

m_1 = distortion coefficient,

m_2 = astigmatism coefficient,

m_3 = coma coefficient,

m_4 = spherical aberration coefficient,

$R_m = n' [J/(S-G)]^2 m_2$ = meridional field curvature,

$R_s = \frac{n'}{3} [J/(S-G)]^2 m_2 - \frac{2}{3} n' \Sigma \frac{R}{P}$ = saggital field curvature,

R = 1/r, $P = nn'/(n'-n)$, where r and n' are quantities defined by the ray-tracing scheme shown in Figure 2,

e_o = spherical aberration in the pupil.

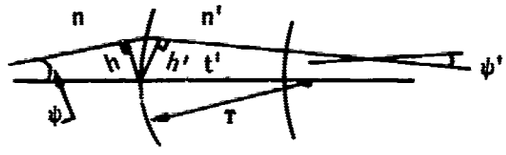


FIGURE 2 Notation for ray-tracing scheme.

One can now find different sets of object and image conjugate planes and relate the m coefficients of one set of conjugate planes to the $\overline{m}$ coefficients of another set of conjugate planes in the same system.

Let one set of conjugate planes be the object and image plane for which the coefficients m_1, m_2, m_3, and m_4 are in effect. Let the other set of conjugate planes be the entrance pupil and the exit pupil where the aberrations are given by $\overline{m}_1$, $\overline{m}_2$, $\overline{m}_3$, and $\overline{m}_4$.

One can now derive

$$\overline{m}_1 = \frac{1}{3} m_3 + \frac{S-G}{J} (G^2-1),$$

$$\overline{m}_2 = m_2 - 3 \frac{S-G}{J} (SG-1),$$

$$\overline{m}_3 = 3m_1 + 3 \frac{S-G}{J} (S^2-1),$$

$$\overline{m}_4 = -e_o.$$

For the case where the object is at infinity, the lateral object magnification G takes the magnitude 0. This is the case in an astronomical telescope, where the object plane is at infinity and the image plane at a finite distance behind the exit pupil.

For this instance, the transformation formulas will yield

$$\overline{m}_1 = \tfrac{1}{3} m_3 - \frac{S}{J},$$

$$\overline{m}_2 = m_2 + 3\frac{S}{J},$$

$$\overline{m}_3 = 3m_1 + 3\frac{S}{J}(S^2-1),$$

$$\overline{m}_4 = -e_o;$$

or

$$e_o = -\overline{m}_4,$$

distortion $$m_1 = \tfrac{1}{3}\overline{m}_3 - \frac{S}{J}(S^2-1),$$

astigmatism $$m_2 = \overline{m}_2 - 3\frac{S}{J},$$

coma $$m_3 = 3\overline{m}_1 + 3\frac{S}{J},$$

spherical aberration m_4 = independent of the third-order aberrations in the pupil.

The m coefficients can be computed once the geometrical data of the field lens involved are known. It is not in the scope of this study to design such a lens and to ray trace available lenses, so this can be saved for a study to be performed hopefully in the near future.

Thus it can be said that (a) spherical aberration generated during the imagery from entrance pupil into exit pupil does not influence any of the aberrations in the image and (b) aberrations from entrance pupil to exit pupil do not affect spherical aberration in the image.

When it is assumed that the telescope has been corrected, m_1, m_2, and m_3 are very small for the telescope alone. In the case that a "field" lens is placed near the image to demagnify the pupil of the astronomical telescope into an exit pupil, it is important to consider the aberrations in the field lens in order not to jeopardize the m_1, m_2, m_3, and m_4 of the telescope alone.

The field lens should be chosen so that certainly coma ($\overline{m}_3$) is balanced, because this will influence the distortion in the image. Distortion is not immediately a wavefront aberration and is not eliminated by the use of the

reference point. However, the field of view that is considered is so small that it would be reasonable to say that distortion is negligible here. The $\overline{m}_2$ and $\overline{m}_1$ influence the wave aberrations astigmatism and coma only, and are not important in view of the fact that these aberrations are compensated by the use of a reference point.

The magnitude of the cross terms in Eq. (1) is given by

$$(A_o + \Delta A)\,(F + \Delta A) \cong A_o F\left(1 + \frac{\Delta A}{A_o} + \frac{\Delta A}{F}\right) ,$$

in which the term $(\Delta A/A)^2$ has been neglected.

When the geometrical aberrations are not present or are very small, ΔA is zero or very near zero, and the magnitude of the cross terms can be written as $A_o F$. A note of caution should be sounded here. If it is assumed that $F >> A_o$, the aberration in the absolute magnitude would be $F(A_o + \Delta A)$. The reconstructed image thus obtained is the same as if there were no pupil aberration for the reference wavefront and only an aberration for the image-carrying wavefront. The treatment of the third-order aberrations now holds, strictly speaking, only for this latter situation. The value ΔA relative to A has here an accuracy which depends on the geometrical aberration. In the case that F is not many times larger than A_o, the term $A_o \Delta A$ enters again; this means that the aberrations in the image are in general larger than for the case of $F >> A_o$, and consequently larger than for the case represented by the third-order m coefficients.

The note of caution now is that by choosing the intensities of the reference and diffraction wavefronts relatively much different, cancellation of aberrations cannot be well achieved without correction for the geometrical aberrations in the imagery from entrance pupil to exit pupil.

Conclusions

The main conclusion is that the field lens need be corrected only for coma in order to minimize third-order aberrations from object to image.

For the astronomical telescope at Cloudcroft, the entrance pupil is 48 in. in diameter, the focal length is about 300 in., and the F-number is approximately F/6. This means that the total field of view for the field lens should be 1/60 rad, which is about 1°.

The field lens could have a diameter which is large enough to contain the image, namely about 1.5 mm in the Cloudcroft case. However, for finding reasons the field lens should be larger. Assume a demagnification of the 48-in. aperture to 1 in.; then the field lens should have a focal length of 6 in. If the field lens has

a diameter of about 20 image sizes, then the diameter will be 1.2 in. and the F-number is then F/5. For a field of view of 1° and an F-number of 5 and assuming that the resolution in the entrance pupil to exit pupil imagery due to astigmatism has to be good up to 16 lines/mm, the field lens has to be of rather good quality. It is considered that a lens of such image quality with a focal length of 6 in. with an F-number 5 and a field of view of 1° is a good commercial photographic objective. The other aberrations, $\overline{m}_2$ (astigmatism) and $\overline{m}_1$ (distortion), of the field lens can be expressed as wavefront aberration in the pupil of the astronomical telescope and thus as a phase error, which would cancel in this wavefront-reconstruction scheme.

INFLUENCE OF THE DETECTOR IN THE EXIT PUPIL ON THE RECONSTRUCTED IMAGE QUALITY

In the preceding sections the third-order aberrations were discussed in the absence of a detector, such as photographic film or an image orthicon. For this study, an image orthicon type of detector is chosen. In order to describe this phenomenon it is necessary to classify what the optical aberrations of an image orthicon are. Just as in the geometrical aberrations, the categorization can be mainly into (a) distortion and (b) lack of "sharpness" expressed by a spread function or a modulation transfer function of the screen.

Distortion, labeled by $\overline{m}_1$ in the previous section, causes a comatic type of aberration in the reconstructed image. Distortions in the detector that are not axially symmetrical will not appear as coma in the reconstructed image but as another aberration. If two orthogonal axes of symmetry in the face surface of the detector can be found, the aberration in the reconstructed image can generally be compared with astigmatism. Once such distortions are known, however, they can be corrected afterwards, either by lens systems during reconstruction or by digital or analog processing of the hologram before reconstruction.

The image degradation due to the "sharpness" degradation can be treated as described by vanLigten.[2,3] Reference 3 treats the case where the reference point is located in the same plane as the object, which is the same as in the Goodman system. The result is that the MTF of the image orthicon limits only the field of view in the reconstructed image. In the reconstructed image, this function will appear as an envelope on the intensity in the image. The zero frequency point will appear in the image plane where the line defined by the reference point and the center of the hologram plane intersects the image plane. The envelope in the image plane extends symmetrically on either side of this point and terminates at a point determined by the resolution limit of the detector. This limit sets the maximum field of view.

When the image orthicon has MTF's which differ in different directions on the faceplate, the amplitude envelope will accordingly be different in different directions in the image plane.

Dynamic Range of the Detector

Provided that the modulation depth of the object bundle on the reference bundle is sufficiently small, the dynamic range is given by the straight portion of the transmittance - exposure[4] curve of the detector and is of no importance to the aberrations in the reconstructed image. If the modulation depth is such that the record occurs over a nonlinear region of the T-E curve, defects in the image will manifest themselves in a manner described by Kozma.[4] Mainly the aberration will not occur in the phase distribution of the reconstructed image, but it will occur in the amplitude distribution. In general a degradation of the resolution will result. The dynamic range of the detector alone does not have any effect on the dynamic range in the image. The dynamic range of the detector together with the modulation depth mentioned above will, however, cause a noise level to occur in the reconstructed image.

REFERENCES

1. R. W. Meier, J. Opt. Soc. Am. 55, 987 (1965).
2. R. F. vanLigten, J. Opt. Soc. Am. 56, 1 (1966).
3. R. F. vanLigten, J. Opt. Soc. Am. 56, 1009 (1966).
4. A. Kozma, J. Opt. Soc. Am. 56, 428 (1966).

APPENDIX 34

THE GENERATION OF CONTOURS ON IMAGES

R. O. Harger

Since coherent sources are about to be used in systems forming images of satellites, there is the possibility of using novel methods to gather information about the images. One such method is the generation of contours of constant depth, or range, on the reconstructed image. The method essentially consists of using two holograms,[1] or interferograms, of the object which interfere to provide the contours. Since the image itself is present one can expect to infer the direction of change. The method has been investigated analytically and demonstrated experimentally in the laboratory.[2,3]

THE PRINCIPLE

There are two methods of generating images with contours, i.e., generating two holograms that will interfere.

1. One can use two illuminating sources at the same frequency but spatially separated. These two waves produce an interference pattern on the object, which is then used to form a hologram in the usual way (Figure 1).

2. One can use two illuminating sources at the same spatial position but separated in frequency. Two images are reconstructed which produce contours useful over a limited solid angle (Figure 2).

The first scheme is simpler and in some ways better, because the contours obviously appear on the image whereas in the second scheme they may not. The first scheme is less attractive for the present application in that the two illuminating sources need to be far to the side of the object, relative to the viewing

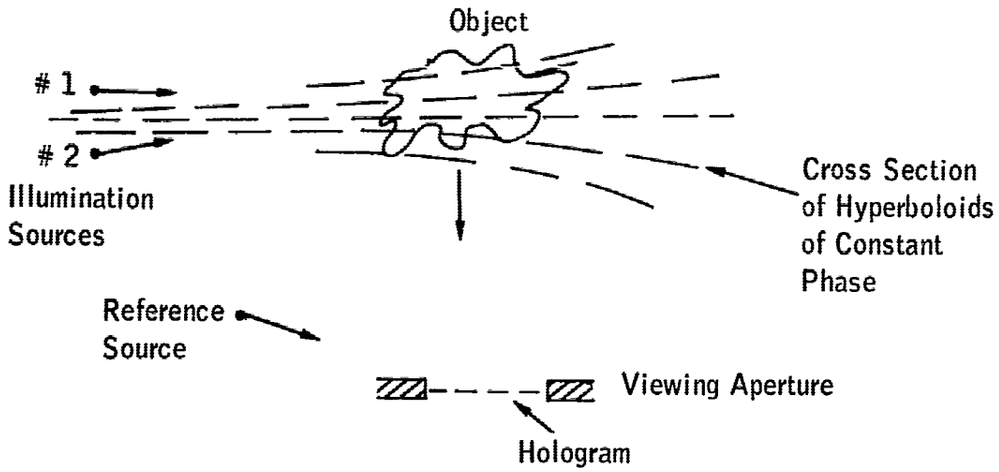


FIGURE 1 Geometry of contour generation using two spatially separated sources.

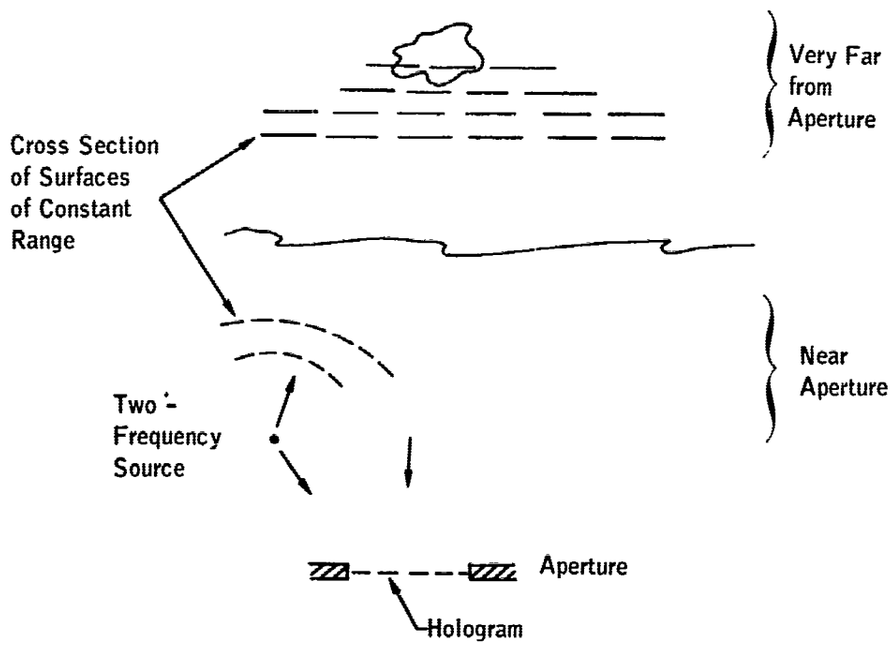


FIGURE 2 Geometry of contour generation using two-frequency source.

direction. Under such a geometrical arrangement the family of hyperboloids of constant phase intersects the object to give contours of constant phase approximately orthogonal to the viewing direction. They are then contours of constant depth, or range.

If the sources are at a great distance, the contours are separated by a range interval

$$\Delta R = \lambda / 2\Delta\gamma ,$$

where $\Delta\gamma$ is the angular separation of the two illuminating sources as viewed from the object. Since the geometry required does not appear to be favorable for this application, we do not consider it further.

The two-frequency method is more complicated in that the position and magnification of a reconstructed holographic image are a function of frequency. Two sources of wavelength λ_1 and λ_2 interfere constructively at distances r_n where

$$\phi_1 = (2\pi/\lambda_1)\, r_n = (2\pi/\lambda_2)\, r_n + n\pi = \phi_2$$

or

$$r_n = n \frac{\lambda_1 \lambda_2}{2(\lambda_2 - \lambda_1)} , \quad \lambda_2 > \lambda_1 .$$

The range or depth separation of contours is then

$$\Delta R = \frac{\lambda_1 \lambda_2}{2(\lambda_2 - \lambda_1)} \tag{1}$$

if the sources are placed at (near) the viewing aperture and the range is relatively large.

Since the positions of the two images are slightly different, only two corresponding points can be aligned from one viewing angle, and then only if the ray passes through the hologram aperture. As the viewing solid angle increases there will be a gradual loss of contouring; finding the largest usable solid angle is a complicated calculation[3] probably requiring computer assistance in any specific case.

As this second method appears to be more practical in viewing satellites, we will discuss it only below.

APPLICATION

The most obvious consideration is the fringe separation, or range/depth resolution, of the contours. From Eq. (1) we see that $\Delta\lambda$ will have to be very small, λ^2 is extremely small relative to the depth resolution that would be of interest here — say 1 m for order of magnitude.

Consider a resonant cavity of length L. The wavelengths λ_m and λ_n of the mth and nth modes of oscillation must satisfy the relation

$$\lambda_m - \lambda_n = \frac{\lambda_m \lambda_n}{2L/(n-m)} , \quad m < n . \tag{2}$$

Combining Eqs. (1) and (2) we find

$$\Delta R = L/(n-m), \quad m < n .$$

Thus the depth resolution cannot be coarser than the laser cavity length L and equals L if two adjacent modes are selected.

Thus in the present application a laser cavity of the order of 1 m is needed, operating in two adjacent (or nearly adjacent) modes. This appears to be within the present capability of laser technology. A 48-in. aperture has a theoretical resolution of about 0.1 sec of arc; at 200 nautical miles this gives a resolution of roughly 1 ft. A most optimistic estimate of instrument degradation and seeing conditions would be a real resolution of 1 m. It would seem that generally one would want depth resolution of about the same order.

The energy requirement for the contouring scheme is the same as that for making an ordinary coherent image of the object. This is discussed in detail elsewhere.

The coherence question requires serious consideration. There should be no problem in getting the required spatial coherence across the aperture. The amount of temporal coherence that is required is determined by the exposure time if this is very short; e.g., in this application with a pulsed laser the exposure time, with proper time gating, is the pulse length, which may be of the order of 10^{-6} sec. The recorded interferogram is of the form

$$A(x,y) \int_T \cos [\phi_t + \theta(x,y)] \, dt ,$$

where T is the exposure time. To enable reconstruction we must have

$$A(x,y) \int_T \cos [\phi_t + \theta(x,y)] \, dt \approx cA(x,y) \cos [\phi_o + \theta(x,y)] ,$$

where c and ϕ_o are constants. If the fluctuations of ϕ_t are a small part of a radian during T, this approximation will hold; e.g., if $\phi_t = at$, corresponding to a change in frequency during the transit time of the pulse, then we need

$$a \ll 1/T .$$

If $T = 10^{-6}$ sec, we need $a \ll 1$ Mcps, corresponding to a laser rate of change of frequency of 1 Mcps/sec. Generally, of course, one would expect the behavior of the laser to give a much more complicated ϕ_t.

SUMMARY

The two-frequency source method of constructing an image of a satellite with contours of constant depth, or range, is a very interesting technique. It appears to be within the laser technology of the present or near future. To generate contour information by high-power, very short-pulsed lasers appears to be much farther in the technological future. Preliminary analytical and experimental investigation has been encouraging and further work should be pursued.

REFERENCES

1. E. N. Leith and J. Upatnieks, J. Opt. Soc. Am. 52, 1123 (1962); 53, 1377 (1963); 54, 1295 (1964).

2. K. Haines and B. P. Hildebrand, "Contour Generation by Wavefront Reconstruction," Phys. Letters 19 10 (1965).

3. K. Haines and B. P. Hildebrand, "Contour Generation by Wavefront Reconstruction," to be published; partial contents presented at Optical Society of America meeting, March 1966, paper WB-17.

APPENDIX 35

A DOPPLER SPREAD IMAGING SYSTEM

R. E. Hufnagel

INTRODUCTION

The purpose of this paper is to describe a type of imaging system which, like the wavefront-reconstruction system described in Chapter IV of Volume I of this report, uses the coherence properties of laser light to overcome the degrading effects of atmospheric turbulence.*

The basic principle of operation is as follows.

If the whole satellite is illuminated from the ground with single-frequency laser light, then the return signal will have a large Doppler shift owing to the motion of the center of gravity of the satellite relative to the receiver. Suppose that, in addition, the satellite is rotating about an axis normal to the line of sight. Then reflections from some parts of the satellite will be Doppler shifted up and some will be shifted down in frequency with respect to the mean frequency. Figures 1 and 2 show these frequency shifts and how they originate

If the angular velocity of rotation of the satellite is approximately known, then the frequency shifts caused by its rotation can be converted into resolution of distance from its axis of rotation. For example, if the satellite is attitude stabilized to the local nadir and orbits at 200 km altitude, its angular velocity of rotation will be about 0.03 rad/sec. A frequency resolution of 10 kc for visible

*R. E. Hufnagel, in Perkin-Elmer Technical Report No. 8025A (June 1965), has given a more complete description of this system, along with technical back-up material.

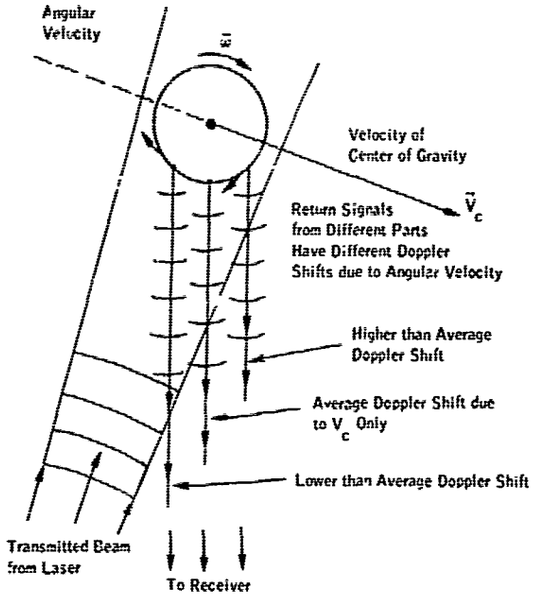


FIGURE 1 Doppler shifts from different parts of a rotating satellite.

wavelength light would convert into a resolution of distance from the axis of rotation of about 10 cm.

Since, as we will show, this order of frequency resolution is possible even in the presence of normal atmospheric turbulence, we have in principle the possibility of obtaining considerably better resolution perpendicular to the axis of rotation through the use of a Doppler spread imaging device, than by conventional means.

A picture of the satellite can be formed by signal processing as shown in Figure 3. Here, optical heterodyne signals from an array of detectors giving 0.5-sec angular resolution along the y axis go to multichannel optical frequency analyzers. The "image" of the satellite is formed in an array display of the spectral analyzer outputs.

Not shown in Figure 3 is a frequency shifter in the local oscillator path to compensate for gross Doppler shifts caused by $\vec{V}_c$.

POSSIBILITY OF IMAGE ENHANCEMENT IN THE SECOND DIMENSION

Various proposals have been made to correct turbulence-induced image defects through the use of real time deformable optics or their electronic equivalents. The successful development of these particular techniques has never come about because of the complex two-dimensional nature of the problem and the optical phase ambiguities brought about by the use of incoherent image-detection techniques.

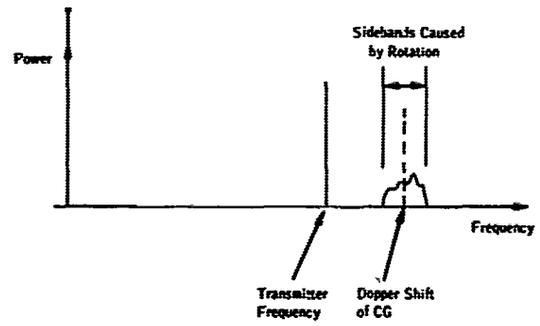


FIGURE 2 Spectrum of return signal from a rotating satellite.

In this system, however, it may be seen from Figure 3 that imagery in the normal sense is done in only one direction (along the axis of rotation) rather than in two and that the detection process is coherent.

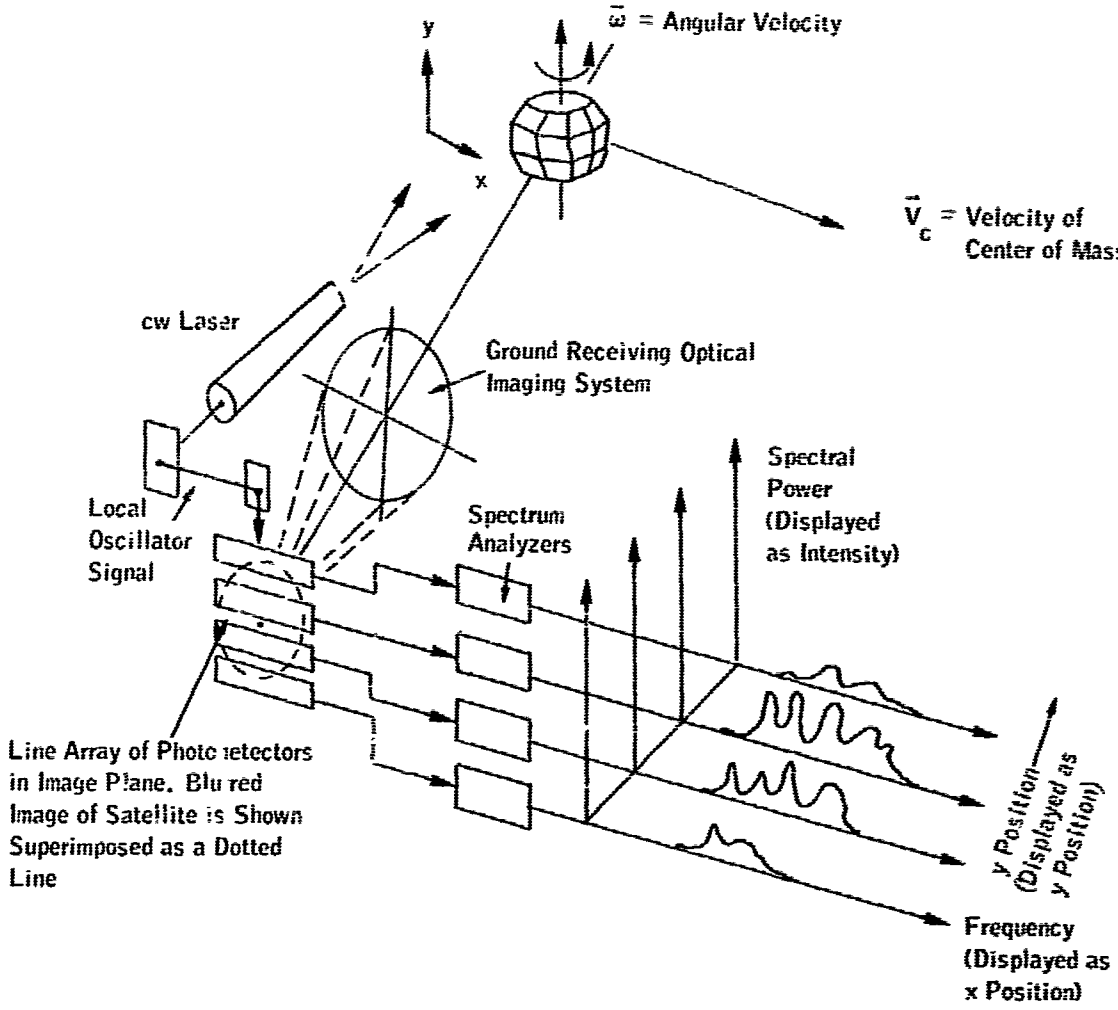


FIGURE 3 Schematic (simplified) of signal-processing system.

This raises the possibility that because of the simplified nature of the problem, it may be practical to perform image enhancement in the vertical direction as well as in the horizontal direction.

EFFECT OF ATMOSPHERIC TURBULENCE ON THE MEASUREMENT OF OPTICAL FREQUENCY

If one recalls that frequency is the time derivative of phase angle, then it is apparent that atmospheric turbulence which disturbs optical signal phase angles (wavefronts) will also perturb the light's apparent optical frequency.

Wavefront perturbations manifest themselves as blurring, image motion, and (in the far field) intensity scintillation. Wiener spectra of the measured temporal variations in these effects (for visible starlight) show that there exist no spectral components for any of these effects for frequencies beyond 2,000 cps.

We may, therefore, conclude that the wavefronts themselves do not vary by more than 2,000 visible wavelengths λ per second.

The most rapid of these wavefront fluctuations are associated with atmospheric turbulence arising from high winds in the vicinity of the tropopause (altitude of tropopause $\cong$ 12 km). For wind speeds of 50 m/sec, we then have an equivalent wind velocity V_w to altitude H ratio of

$$\frac{V_w}{H} = \frac{50 \text{ m/sec}}{12 \times 10^3 \text{ m}} = 4 \times 10^{-3} \text{ rad/sec.}$$

The V/H ratio for the satellite is 0.03 rad/sec. Hence we expect the visible light "seeing spectrum" to have frequency components up to a maximum of 2,000 cps x 0.03/0.004 = 15 kc/sec. The maximum frequency component at 10 μ can be computed by using an approximate inverse-wavelength scaling factor. 0.5 μ to 10 μ is a factor of 20 in wavelength. Therefore, the maximum frequency component at 10 μ is approximately 1/20th of the maximum frequency component at 0.5 μ. This gives a value of 750 cps for the maximum frequency component at 10 μ. We will conservatively use a frequency uncertainty Δf of 1 kc/sec, although most of the time it will be less than that value.

RESOLUTION LIMITS

Figure 3 illustrates the geometry for the simplified case where both $\vec{V}_c$ and $\vec{\omega}$ are perpendicular to the line of sight. In this case the y axis coincides with the rotation axis $\vec{\omega}$.

The resolution limit, Δy, is set by the receiver aperture diffraction and/or atmospheric turbulence. Typically the resolution limit at $\lambda = 10$ μ will be about 1 sec of arc, which at a range of 200 km corresponds to $\Delta y = 1$ m.

The Δx resolution limit may be computed from the relationship

$$\Delta x = \lambda \Delta f / 2\omega ,$$

where Δf is the optical signal frequency uncertainty caused by atmospheric turbulence. Since Δf is (at least roughly) proportional to λ^{-1} it follows that Δx is essentially independent of wavelength. If $\Delta f = 1{,}000$ cps at $\lambda = 10$ $\mu = 10^{-3}$ cm and $\omega = 0.03$ rad/sec, then $\Delta x = 16.7$ cm $\cong$ 7 in.

DESCRIPTION OF A FEASIBLE SATELLITE INSPECTION SYSTEM

In this section it is our purpose to describe in detail a potentially workable system made up of components which could soon be made available. It is not a

mathematically optimized system, but it has the attractive feature that it could be built with today's technology.

Choice of Laser

At the moment, the most practical and feasible laser capable of generating kilowatt levels of cw light is the CO_2 molecular laser operating at 10.6078 μ. The atmospheric absorption line at 10.6078 μ absorbs only a fraction (~40 percent) of the laser light and thus does not prohibit its use for this application.* As will be shown, adequate detectors and receiving telescopes are available for these wavelengths. The required frequency stability has been achieved[1] for He-Ne lasers and this technology is directly applicable to CO_2 lasers. The transmitter aperture need be only 20 cm in diameter to achieve the desired beam spread at the target.

The primary mirror in the receiver would typically be 2 m high by 1 m wide. The 2-m height is required to achieve a diffraction-limited vertical resolution of 1 sec of arc for 10-μ radiation. The mirror figure would be held accurate to about 4 wavelengths of visible light (= λ/5 at 10 μ).

The cryogenically cooled photodetector could be a mercury- or copper-doped germanium type. It would have an array of about 20 side-by-side elements each 1/2 mm thick by 5 or 10 mm long. With this dimension and an f/100 focus, the system would still be diffraction-limited in the y direction. The field of view would be about 10 sec of arc square. The tracking accuracy would have to be about 3 sec of arc with short time precision of 1 sec of arc per msec. These requirements are severe but still within today's technology.

Real-time signal processing could be done by either electronic or electro-optical devices.

Signal-to-Noise (S/N) Considerations

Let the transmitted signal be a 1-kW carrier wave. For 10 μ wavelength light, there are 5×10^{22} photons per second. We will assume that the light beam is spread over an area of 100 sq m, or enough to cover the size of a medium-sized satellite. From a resolution element area $\Delta x \times \Delta y = 16.7 \text{ cm} \times 1 \text{ m} = 0.17 \text{ m}^2$ on the satellite with a diffuse reflectivity r = 0.001, the reflected light per resolution area will be $(0.17/100) \times 0.001 = 1.7 \times 10^{-6}$ times the total incident flux of 5×10^{22} photons per second. For a ground receiver with an effective area of 2 m^2, the captured return signal will be

*The use of CO_2 isotopes could completely eliminate the absorption problem.

$5 \times 10^{22} \times (2/\pi R^2) \times 1.7 \times 10^{-6}$ photons/second.

For a satellite range R of 200 km, the captured return signal is equal to $1.4 \times 10^{+6}$ photons/second/ resolution area.

For an 18 percent (one-way air path transmission = 0.6, optics efficiency = 0.5) over-all optical system - air path transmission, we obtain a net average return of 2.5×10^5 photons per second per resolution area per kilowatt of laser power for a 0.1 percent target reflectivity. This corresponds to an average signal power $\overline{P}_s$ of 4.8×10^{-15} W at the receiver per target resolution element.

The sky background spectral irradiance at 10 μ for 60° zenith angle viewing is 3.4×10^{-4} W/μ-sr-cm^2. We will assume a 0.3-μ predetection optical bandpass filter at 10 μ and a net optical system transmittance of 50 percent. The one-way atmospheric transmission at λ = 10 μ is approximately 60 percent. The collector area is 2 m^2 = 2×10^4 cm^2 and the field of view of each detector element is 10 m^2/R^2 = 2.5×10^{-10} sr. Therefore, the sky background average power passed by the 0.3 μ predetection optical filter is equal to 1.5×10^{-11} W.

The self-emission from the satellite impinging on a small detector element with the same optical system is computed to be equal to 6.4×10^{-11} W, assuming that the satellite is a 320° K blackbody radiator.

The additive detector noise power in the absence of background radiation is estimated by using as an upper bound the noise equivalent power (NEP) determined from knowledge of spectral detectivity measured with a small field of view, and the sensitive area of the cell. Under these conditions, a detectivity of 10^{11} cm(cps)$^{1/2}$/W for Ge:Hg can be achieved. With a sensitive area of 1/2 mm x 10 mm, and a resolution element bandwidth Δf of 1 kc/sec, we obtain a noise equivalent power of 7×10^{-11} W.

The S/N ratio per frequency resolution element Δf for heterodyne detection can now be computed from

$$\frac{S}{N} = \frac{2\eta\,(P_S/2N)\,P_o}{2h\nu\Delta f\,[P_o + P_B + P_S] + 2\eta\,(\tilde{P}_B/2)\,P_o + \eta P_D{}^2} \quad , \qquad (1)$$

where

P_S = average received signal power,

P_o = average local oscillator power,

P_D = NEP of detector in resolution bandwidth Δf,

P_B = combined sky background and target self-emission average power in the predetection optical bandpass filter,

$\tilde{P}_B$ = amount of combined sky background and target self-emission in Δf bandwidth

η = detector quantum efficiency,

Δf = bandwidth associated with resolution element,

N = constant allowing for degree of nonoverlap of signal beam and local oscillator beam; typically N = 5.

Note that signal and background power are divided by a factor of 2 to account for depolarization effects. Since heterodyne beats can only arise from like polarizations between signal and local oscillator beam, any change of polarization state of the signal beam relative to the local oscillator results in a reduction of heterodyne signal.

The first term in the denominator of Eq. (1) constitutes the shot noise component arising from the signal, local oscillator, and background. The second term represents the thermal noise contribution of components of the background within the resolution bandwidth Δf. The third term is the additive noise of the detector.

In the case where the local oscillator power P_o predominates over P_S and P_B, Eq. (1) simplifies to

$$\frac{S}{N} = \frac{2\eta P_S P_o/10}{2h\nu\Delta f P_o + 2\eta (\tilde{P}_B/2) P_o + \eta P_D^2} . \quad (2)$$

For low background radiation and additive noise of the detector, we may neglect $2\eta (\tilde{P}_B/2)$ and $\eta P_D^2/P_o$ in comparison with $2h\nu\Delta f$, and we obtain:

$$\frac{S}{N} = \frac{\eta P_S}{10h\nu\Delta f} . \quad (3)$$

In order to achieve the S/N performance given by Eq. (3), the local oscillator power P_o must be such that the local oscillator shot noise predominates over the other components of shot noise. However, P_o must be less than what is necessary to saturate the detector. A local oscillator power of 20 μW probably meets both conditions.

Now for the case where the quantum efficiency is 20 percent, $P_S = 4.8 \times 10^{-15}$ W, and $h\nu\Delta f = 2 \times 10^{-17}$ W, we obtain an S/N ratio per resolution element Δf of approximately 5. This S/N ratio is for a resolution element per millisecond. If we integrate the signal for 0.1 sec, we then have an equivalent S/N ratio of 480, which integrates out the noise and also the speckle pattern.

TABLE 1 Summary of System Parameters for Heterodyne Earth - Satellite Link

Range	200 km
Transmitter power	1,000 W
Wavelength	10.6078 μ
Satellite diffuse reflectance coefficient	0.001
Receiver area	2m x 1m = $2m^2$
Receiver optical predetection filter bandpass	0.3 μ
Net optical system transmission efficiency	0.5
Bandwidth corresponding to one resolution element	1,000 cps
One-way atmospheric transmission at 10.6078 μ	0.6
Sky background spectral irradiance at 10 μ for 60° zenith angle viewing	3.4×10^{-4} W/μ-sr-cm^2
Satellite self-emission spectral irradiance at 10 μ	1.4×10^{-3} W/μ-sr-cm^2
Detector quantum efficiency	0.2
Area of each detector element in array	1/2 mm x 10 mm = $5mm^2$

Table 1 summarizes the parameter values for the heterodyne earth - satellite link.

REFERENCE

1. M.S. Lipsett and P.H. Lee, Paper TF-16 presented at the 1965 Annual Meeting of the Optical Society of America, Oct. 5-8, 1965.

APPENDIX 36

ALTERNATIVE SATELLITE OBSERVATION SCHEME

S. A. Collins, Jr.

The following is a rather different scheme for observing satellites, using laser illumination. The experiment makes the point that a large telescope might be used not as a high-grade receiver but as a high-grade transmitter. The advantage to be gained is improved signal-to-noise (S/N) ratio.

The scheme is shown schematically in Figure 1, where we see a laser beam fed through appropriate optics into a large telescope. The telescope is adjusted so that the beam is focused to a diffration-limited spot (broadened and reoriented somewhat by the atmosphere) onto the satellite.

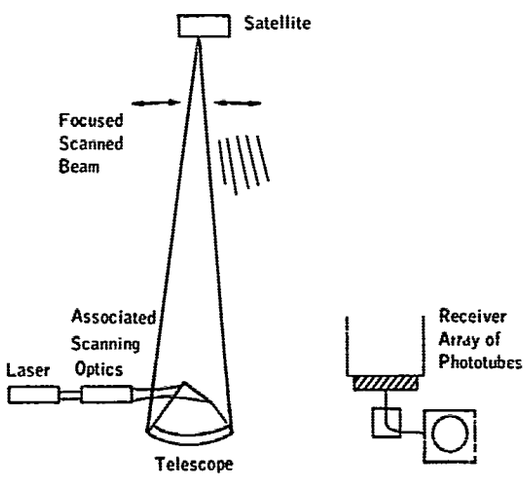


FIGURE 1 Schematic of a laser beam fed through appropriate optics into a large telescope.

The beam is scanned back and forth across the satellite by means of the auxiliary optics, and the reflected light is monitored as a function of time at the receiver. The received signal is then composed into an image, as a television image is recomposed.

This scheme has the following advantages.

1. All the transmitted power impinges on the target.

2. The receiver is of the "photon bucket" type, not limited in size. An

array of photomultipliers or a large low-grade telescope could be used.

3. The point-spread function can be obtained each time the beam is swept past an edge.

4. By monitoring laser frequency Doppler shifts, the contours of the satellite could be obtained.

5. One might hope to use the scheme 24 hr a day.

One main component, the beam scanner, is available. Scanners have been constructed that sweep a laser beam 100 beamwidths at rates well over 1 kc/sec.

Indeed this scheme is in a preliminary state of development. However, it is presented here as a contribution to the "for what it is worth" department.

APPENDIX 37

NOTES ON THE GREGORY REAL-TIME SAMPLING TECHNIQUE FOR MINIMIZING IMAGE DISTURBANCE

R. L. Gregory and J. S. Courtney-Pratt

OPERATING STRETEGY

It is important to design and set the device to extract the maximum number of occasions when the image most closely matches the master. This involves an operating strategy which is controlled by a simple analog computer which we have designed and built for the purpose.

Before using the device to build up a final picture, we must supply it with a master negative. This is given sufficient exposure time to make it reflect the statistical properties of the disturbance reliably, and it is processed as rapidly as possible. Since the quality of the master does not have to be high, we can use forced drying, and our total processing time is down to under 2 min. No doubt this could be improved. The use of a phosphor or other nonphotographic technique could cut this time waste dramatically.

There are a number of variables to consider when matching the technique to the prevailing conditions and requirements. The essential point is to adopt the most stringent mismatch criterion possible, consonant with the available total sampling time and prevailing seeing conditions. If the seeing conditions generally remained constant over periods of an hour or more, there would be little need for a computer, but we included it principally to deal with the rapidly changing conditions of the English climate. It also serves to give an objective estimate of the prevailing conditions for the initial setting of the device. It has outputs which, we hope, will serve to give useful information on the physics of the turbulence.

The first parameter to feed into the computer is the total available sampling time. This is determined by considerations such as rate of changes in the object or the lighting, and, in the case of satellites, by the transit time of the object. For the moon, we estimate that sampling time should be limited to between 15 and 30 min for the terminator, where shadows change relatively rapidly and occupy a large proportion of the viewed object. For Mars, with a rotation period of 24 hr, 37 min, we should have an hour or two available for the central regions of the disk. But Jupiter (9 hr, 55 min), Saturn (10 hr, 14 min), Uranus (10 hr, 40 min), and Neptune (15 hr, 40 min) all have rotation periods shorter than Mars, which will restrict our sampling time to 1 hr or less.

Sampling time available for satellites will be less than 5 min, and in the case of a rotating or tumbling object may be further limited, if we select for sampling only those moments when it presents a particular orientation.

For the case of very dim objects such as distant nebulae, where maximum effective resolution is most desirable, the available sampling time will be limited by the duration of darkness at the observatory (about 8 hr at Mt. Palomar).

Telescopic tracking errors do not set a limit to total sampling time, for (a) manual or electronic corrections can be made as required, and (b) the device will reject images displaced by tracking errors just as it will reject atmospherically disturbed images. The only effect of tracking errors is to reduce the effective sampling time, since they give systematic mismatch with the master, preventing exposure of the final picture plate during a tracking error. This loss in useful time will be taken care of by the computer and need not be estimated by the operator. It is, however, important to know the relative contributions of tracking errors and turbulence if the system is to be used for research into the physics of turbulence, since it does not distinguish between these two kinds of disturbance.

Having arrived at a maximum allowable sampling time, the computer is set to examine the mismatch signals from the photocell. First, it notes the lowest voltage (minimum mismatch signal) which occurs. Second, it estimates the rate of occurrence and the average duration of these signals. Third, it raises the gate level until there should be just sufficient total exposure time to expose the plate in the available total sample time. At this point, the shutter is switched on, starting the sampling procedure. Now the computer will follow changes in the disturbance during the building up of the final picture. (Operational amplifiers with adjustable droop characteristics slowly lower the gate level in the absence of accepted signals.) This ensures that it will always produce a fully exposed plate. By setting the droop rate to a low value, the plate is unlikely to be exposed significantly before the end of the allotted sampling time.

The present electromechanical shutter allows a minimum exposure of about 10^{-3} sec, so any "accept" signals of shorter duration must not affect the computer's estimate. It is clearly important to increase the shutter speed for such work as photographing satellites, where it is essential to use only the best matches with the master and in the shortest possible total sampling time, which means that the shortest occurring matches must be included in the total sample and not lost through limited shutter speed.

Disturbance errors may be simple lateral translation of the entire image or shifts of one or more comparatively small regions. In general, a given displacement of a large region will give a larger reject signal (greater output from the photomultiplier) than the same shift from a small region. But this is not the whole story, for the mismatch is not of areas but rather of contours. This has an implication of some importance. Since it is contours which convey information in pictures — regions of constant intensity are informationally redundant — we derive our accept/reject signal primarily from those regions of the image which convey most information, and fortunately it is generally these which most need to be preserved. Thus although a "cell" of disturbance occurring on a structureless region may not give a reject signal consonant with the area or extent of the displacement, this will not in general matter. In the case of satellite photography this is indeed a useful feature, for cells of disturbance in other regions of the field (except for bright stars) will not prevent sampling of the wanted object, with resulting waste of total sampling time. The only kind of case where this feature of the system is not an advantage is for lunar photography of large homogeneous maria, where there are but few contours to give reject signals. If the effective reject signals came from pronounced features widely separated in the image plane from the wanted regions of low contrast, we should expect poor rejection of disturbance in these mare regions. But this is a special case, and it can be countered by selecting fields largely free of distant pronounced contours, by suitable choice of magnification and telescope positioning. It would even be possible to mask out such features from the master, for a really difficult case, but this is unlikely to arise.

The shutter may be operated in various ways. In particular, it may be set to give only one exposure time which will be repeated while the computer delivers an "accept" signal, or it may be held continuously open during the "accept" signal, whatever its length. We have adopted the latter procedure, this being the more efficient.

Our preliminary tests show that the device is not difficult to operate under typical conditions, though no doubt there is still a lot to learn to make it maximally efficient.

EXPOSURE LEVELS

We will note some of the points to be considered, though it may not be possible to determine what are the best conditions until further experiments have been performed.

Master Negative

To get the best cancellation of signal when the incoming decision beam falls precisely on the master negative, the master should be exposed and processed to a gamma of unity. The over-all exposure should be adjusted so that shadow detail in the image should produce faintly detectable density changes. For an object like the full moon that has a relatively small brightness range, the peak density of the master will thus be small. For a satellite with strong highlights, the peak density of the master will be high.

The exposure time for making the master must obviously be long enough to satisfy the density requirements above, but, in addition, it is central to the whole scheme that the exposure time be long compared with turbulence periodicities. A second is probably the lower safe limit (until at least we have other experimental evidence). For rolling satellites this will mean selection of suitable short times during each revolution.

The exposure energy can be adjusted as necessary by the use of attenuating filters or beam splitters.

Integrated Negative

The optimum size of image and the total integrated energy for this negative can be computed as described in Appendix 11 on optimum photographic exposures by Courtney-Pratt. We do not know what will be the optimum values of the individual exposure times to be summed to make up the integrated negative. We have measured the frequency spectra of scintillation of stars and planets and have found no significant amplitudes at frequencies above 1,000 cps. Minimum exposure times for the elemental exposures might thus appropriately be in the 10^{-4} to 10^{-5} sec range. Perhaps we could even reasonably limit them to $\geq 10^{-3}$ sec, which would allow electromechanical shutters as in the present experimental apparatus.

Beam Splitters

Suppose the mean density of the master negative were 2.0, and that we wished to detect a 1 percent increase in light from perfect match. The number of photons per second in the image of a given satellite can easily be calculated for any given telescope. We then find that for operation of the photoelectric detector in say 10^{-5}

sec, we would need to reflect at least half the incoming light to the photocell for an 8-in. aperture telescope. On the other hand, we would only need about 2 percent of the incoming beam for a 48-in. aperture telescope.

It is desirable to use no more than necessary in the decision beam so that as much as possible of the light can be utilized in building up the integrated negative.

VARIANTS OF THE SCHEME

Frozen Real-Time Operation

So far we have assumed that the device is being used in real time. But just as a series of existing photographs can be compiled to give a composite, so we might use this method to give automatically a composite of many existing pictures. We will call this working in "frozen real time." In place of the telescopic image, we would use a series of pictures taken with a cine camera and projected in sequence into this device.

There would be a number of advantages:

1. There would be as much time as one wished to make and process the master negative.
2. Several different masters could be tried.
3. Several different decision levels could be tried.
4. One could arrange to change the master as the lighting on the object changed.
5. One could select every nth picture, so that one was working always at some given stage of rotation of a rotating satellite.

To achieve these advantages, however, one needs to take long sequences of pictures; and the exposure time of each frame ought ideally to be as short as the minimum elemental exposures used for integration in real time. In Appendix 11, it was shown that one could take pictures of optimum density with exposure times down to about 3 msec, using a 48-in. telescope. It would be well worthwhile trying some sequences of pictures taken at say 200 pictures/sec, and at 3-msec exposures to see how the scheme would work. Probably, however, the minimum elemental exposures should be shorter than 1 msec. We could arrange therefore to take sequences at, say, 5,000/sec with 10^{-4} sec exposures, if we incorporated any one of the currently available electronic image intensifiers at the recording end of the telescope system.

Such an electronic device might also sensibly incorporate the ideas of image stabilization or centration in the frame, presented on page 85 of Chapter V of Volume 1. This latter idea is not essential but might well be an advantage, as

image stability is sure to be a problem, particularly at the long focal lengths and high tracking rates required for satellite work. (Focal length will have to be increased by a small factor when using electronic image intensifiers as compared with direct photography, as the resolution in cycles/millimeter of most electronic imaging devices is lower than the resolution of typical photographic film.)

One might have said that one should examine all these pictures individually, or process them all digitally, but the huge numbers involved make this impracticable. The automatic features of the Gregory system make it most attractive, and in any case the output picture from this system (whether from real time or frozen real time) can, of course, be processed by one of the passive image-processing systems described in Chapter III of Volume 1.

Bootstrapping

There are cases where it might be advisable to "bootstrap" the operation by using a final picture as the statistical master for a second final picture. In principle, this could be continued any number of times. Experiment shows that a surprisingly poor master can give marked improvement; it seems that but little information is required to give the sampling criterion, at any rate for complex structures such as the moon. We have not tried it on star fields or satellites, when it may be important to have the best possible master. Bootstrapping could be worthwhile, at any rate for satellites.

We would, however, prefer to develop the system fully before considering "bootstrapping" as a potentially important technique for extracting maximum information. It is most likely to be useful (a) for extended objects such as the moon viewed through large-aperture telescopes when there will be a large number of cells of disturbance, and(b) when it is desired to extract the maximum possible information from satellites, on a series of transits. The final picture of each transit might be used for the succeeding transit, rather than sticking to the same master throughout a series.

Match-Mismatch with an Interferometer or Schlieren System

One might use a fairly similar scheme to those described above, but instead of using the match-mismatch of the incoming light with respect to a master negative, one might measure some other feature of the incoming wavefront. For example, one could envisage the use of a wavefront shearing interferometer or some other interferometer which produced interference between the right and left (and/or top and bottom) halves of the telescope aperture. Again, one might use a schlieren system for examination of the uniformity of illumination of the aperture.

No experiments have as yet been performed on such variants but they do seem to warrant consideration.

Multiple-Level Operation

One might build generally similar, but somewhat more complicated, apparatus with several decision levels. One would arrange to "pigeonhole" the differently selected images in different parts of the f ne, using, for example, a deflecting image intensifier. This scheme might have some advantage with transient events, for if one had not collected enough elemental exposures at the most critical decision levels, then at least there would be a larger summation of images at the less critical levels, so that (approximately) the best picture available would have been recorded.

USE OF THE TECHNIQUE FOR INVESTIGATING THE PHYSICS OF TURBULENCE

We started by confessing that this method is not based on a mathematical analysis of turbulence. It is based, rather, on the assumption that there will be reasonably frequent short moments when turbulence is minimal. It turns out that there are but few available data on turbulence as it disturbs images, and here we may consider this technique as a means for obtaining objective data on image fluctuation. The device may be set to track any convenient target, such as part of the moon, and if the photocell current is then monitored by an oscillograph (an ultraviolet paper recorder is convenient) we obtain a measure of the disturbance as integrated over the area of the target. By selecting a smaller target (and for the simplest case a bright star, as a point source), we may learn a good deal about the physics of the turbulent atmosphere. The technique could be used not only for celestial targets at various declinations, but also for shorter air paths, using target balloons. (Here the image size could be held constant, as the target balloon rises, with a zoom optical system.)

SUPPLEMENTARY

INFORMATION

NOTICE OF CHANGES IN CLASSIFICATION, DISTRIBUTION AND AVAILABLITY

69-18 15 SEPTEMBER 1969

AD-806 879 Advisory Committee to the Air Force Systems Command (NAS-NRC), Washington, D. C. RESTORATION OF ATMOSPHERICALLY DEGRADED IMAGES, VOLUME 2. 1966 Contract AF 18(600)-2891	No Foreign without approval of Research and Technology Div., Rolling AFB, D. C.	No limitation	AFSC ltr, 5 Feb 69