2mif

132919

University of California, San Diego
Grant NGL-05-009-079

UNIVERSITY OF CALIFORNIA

San Diego

Optical Signal Processing: Poisson Image Restoration and Shearing Interferometry

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Information and Computer Science

by

Yie-Ming Hong

Committee in charge:

Professor Carl W. Helstrom, Chairman
Professor Adolf W. Lohmann
Professor Elias Masry
Professor Harold W. Sorenson
Dr. Charles C. Wang

1973

The dissertation of Yie-Ming Hong is approved, and it is acceptable in quality and form for publication on microfilm:

Elias Masry

Harold W. Sorenson

Charles P. Wang

O. Tufunaga

Carl W. Helstrom

Committee Chairman

University of California, San Diego

1973

Page intentionally left blank

Table of Contents

Page

List of Figures and Tables

Acknowledgments

I would like to express my sincere appreciation to Professor Carl W. Helstrom who has led me to do independent research and provided stimulating suggestions and encouragement during the course of this work. I also wish to thank Professor Adolf W. Lohmann who introduced me to the field of coherent optics and carefuly read and constructively criticized the manuscript of the coherent optics portion of this dissertation.

Much of this research was supported by the National Aeronautics & Space Administration under grant NGL-05-009-079, whose assistance is gratefully acknowledged.

My warmest thanks go to my wife, Yen-Yen, for her encouragement and consideration during my extended studies at the University of California, San Diego.

Vita

[redacted]

1965- B.S., Electrical Engineering
National Taiwan University, Taipei, Taiwan

1969- M.S., Information and Computer Science
University of California, San Diego

1973- C.Phil., Applied Physics
University of California, San Diego

1967-1973 Research Assistant, Department of Applied Physics and Information Science, University of California, San Diego

Publications

"Restoration of Images on Photoelectrically Emissive Surfaces," C. W. Helstrom and Y. M. Hong, J. Opt. Soc. Am., Vol. 63, 480 (1973).

"Circular Carrier-Frequency Photography for Observing Phase Objects," D. Silva and Y. M. Hong, Opt. Comm., Vol. 7, 283 (1973).

Fields of Study

Major Field: Signal Processing
Detection and Estimation Theory
Communication Theory
Coherent Optics and Holography
Electronic Engineering

ABSTRACT OF THE DISSERTATION

Optical Signal Processing: Poisson Image Restoration

and Shearing Interferometry

by

Yie-Ming Hong

Doctor of Philosophy in Information and Computer Science

University of California, San Diego, 1973

Professor Carl W. Helstrom, Chairman

Optical signal processing can be performed in either digital or analog systems. This dissertation is concerned with the use of digital computers and coherent optical systems in optical signal processing, such as image restoration, phase-object visualization, image contrast reversal, optical computation, image multiplexing, and fabrication of spatial filters.

Part I of the dissertation is concerned with digital optical data processing. It deals with restoration of images degraded by

signal-dependent noise. When the input data of an image restoration system are the numbers of photoelectrons received from various areas of a photosensitive surface, the data are Poisson distributed with mean values proportional to the illuminance of the incoherently radiating object and background light. In this model, the shot noise is predominant in causing random variation of the observed data. The noise is signal-dependent. A nonlinear restoration scheme based on the principle of maximum likelihood is derived, which is optimum for a broad class of objects with preassumed average contrast and size of typical details. The nonlinear estimation scheme can be solved by linear iteration, or methods of nonlinear programming.

One-dimensional objects distorted by diffraction or motion blur are estimated in computer simulations. The performance of the estimator is measured by percentage squared error. When the background illuminance is low, the maximum-likelihood estimate is much superior to the linear least-squares estimate.

Part II of the dissertation is concerned with optical signal processing using coherent optical systems. Following a brief review of the pertinent details of Ronchi's diffraction grating interferometer, moiré effect, carrier-frequency photography, and achromatic holography, two new shearing interferometers based on them are presented. Both interferometers can produce variable shear.

New methods for optical testing and phase-object visualization, spatial differentiation, complex amplitude addition and subtraction, image contrast reversal, two-dimensional photometry, image multiplexing, and fabrication of spatial filters can be achieved easily by using these two new interferometers. All the new methods are compared with the existing methods. They all have some definite advantages in either flexibility, simplicity, light efficiency, or real-time operation. Finally, a new method for focusing and alignment is also developed.

Introduction

Because of the two-dimensional nature of pictures, an optical communication system can transmit much more information than electronic systems per unit time interval. An optical system can be thought of as a bank of parallel electronic communication systems. Due to the problem of the already over-crowded radio communication channels, much effort has been put into the study of optical communication systems, which afford wide temporal bandwidth and multiple spatial channel capacity. Optical signal processing is as important in optical communications as its counterpart in electronic communications. In this thesis we will discuss methods of optical signal processing using both digital computer and coherent optical systems.

In part I of this thesis we will present methods for the restoration of images degraded by signal-dependent noise. Processing of radio signals (deterministic or stochastic) degraded by a radio communication system is based on a linear model of signal formation, in which the noise is independent of the desired signal and combined with it additively; the sum of these constitutes the observed data. In signal estimation, the data are passed through a linear spatial filter whose transfer function is designed to reproduce the original signal within the smallest possible mean-square error. This kind of filter is known as the Wiener filter or matched filter.[1-3] This additive-noise model has been used extensively with success.

The signal processing techniques used at radio frequencies have been extended to the optical frequency region. Methods for restoring images degraded by an optical system are often based on a linear model of the optical process: a two-dimensional object function has passed through a spatial filter to whose output random noise has been added to produce the observed data.[4-9] On the basis of these data and the known filter transfer function, the original object function is to be estimated with minimum mean-square error.

When imaging incoherent light, it is the radiance of the object plane that is to be estimated, not the value of the light field itself. The light field is a complex spatio-temporal Gaussian random process, to whose mutual coherence function the radiance is related. This radiance like its counterpart in the image plane, is a quadratic functional of the field. Furthermore, linear filtering theory does not explicitly distinguish the two basic types of noise, that due to the inherent fluctuation of the light fields of object and background, and that associated with the process of recording the light. Instead, it makes certain assumptions about the statistical properties of the noise, which is usually assumed to be white and Gaussian and to leave such parameters as its mean-square amplitude and its bandwidth to be measured separately. Such a treatment is incapable of determining the fundamental limitations on restoring degraded images.

Helstrom[10] has shown how the statistical properties of the information-bearing quadratic functional of the light field modify the usual

Shannon formula[11] for the information transfer from an incoherently radiating object. We will take into account the physical nature of the light and the statistical properties of the recording process, and analyze an imaging system in which the light from the object plane is focused onto a photosensitive surface, from which it ejects photoelectrons. The surface is divided like a mosaic into a large number of small, insulated spots, from each of which the photocurrent can be measured. The measured values of the photocurrent constitute the data on which is to be based an estimate of the radiance distribution function of the object plane, which corresponds to the 'true image'.

In chapter 1, we first describe the statistical properties of the light field emitted by the incoherently radiating object, and then calculate the statistics of the instantaneous illuminance in the image plane, which is assumed to be a measurable quantity that provides the data on which a linear estimate of the object radiance function will be based. The object radiance function is considered as the sum of a uniform background B_0 and one of an ensemble of spatial random processes, which form the class of object to be estimated. The linear filter has the same form as for the estimation of the signal degraded by signal-independent additive noise. The equivalent noise spectral density contains a contribution not only from the background white noise, but also from the object-light fluctuation and the uniform object background, B_0. This is very different from the model treated before.[5]

In chapter 2, we will discuss the interaction of the light field with the recording medium. When the product WT of the bandwidth W of the incident light and the observation time T is much larger than 1, the numbers of photoelectrons emitted from the spots of the photoelectrically emissive recording surface are statistically independent with Poisson distributions whose mean values are proportional to the illuminance at the spots. The linear least-squares filter for the estimation of the object radiance function has the same form as those described by the Wiener filtering theory, but with an equivalent noise spectrum equal to the sum of a constant term arising from the interaction of the uniform object background B_0 and the background white noise with the recording medium, the shot noise, and a spatial variable term due to the fluctuation of the light field produced by object and background noise.[12] In particular, it is shown that under normal circumstances the shot noise far exceeds the noise associated with random fluctuations of the object and background noise. The shot noise is signal-dependent. Linear least-squares estimation, which is optimum for the situation when the signal is corrupted by signal-independent additive noise, is inadequate for the restoration of images degraded by signal-dependent noise. The principle of maximum likelihood[13] leads to nonlinear equations for the estimates, which can be solved by linear iteration or by methods of nonlinear programming.[14] The nonlinear estimator is signal-dependent and is asymptotically convergent to the linear estimator, when the received light is strong.

To provide the insights about the practical implications of the analytic result for the maximum-likelihood estimator, computer simulated experiments are performed to demonstrate the significance of the nonlinear estimation scheme in chapter 3. One-dimensional objects distorted by motion blur and diffraction have been estimated in computer simulations. When the background illuminance is weak and the contrast of the object is high, the maximum-likelihood estimate is much superior to the linear estimate.

In part II of this thesis we will present new methods of optical signal processing using only coherent optical systems.

Optics and electronics are means of information processing which have many common features. Communication theory is well developed in electronic signal processing. Recently it has been successfully adapted by optical physicists, especially in the development of holography[15] and spatial filtering.[16, 17] Many optical phenomena can be easily understood from the point of view of communication theory. Modulation, demodulation, heterodyning, harmonics generation, and filtering have been used successfully in electronic signal processing; we can find their counterparts in optics. In optics, modulation can be achieved by carrier-frequency photography[18, 19]; generation of harmonics by grating diffraction; demodulation and heterodyning by the moiré effect; and filtering by spatial filters. Carrier-frequency photography, diffraction by gratings, and the moiré effect

all have their unique functions in optical signal processing. Two new types of interferometers have been developed from the combined use of those concepts.

Diffraction by gratings and the moiré effect have been used extensively in optics. Gratings have been used as beam splitters.[21] Ronchi invented a shearing interferometer using a grating as a beam divider, which has been used extensively for testing optical components such as lenses. This technique is relatively simple and inexpensive. But the use of just one grating results in many drawbacks, such as (a) the fixed amount of shear, (b) the dependence of the fringe contrast on the degree of equality in light flux between different diffraction orders, and (c) the overlap of many diffraction orders.[21]

The drawbacks of Ronchi's interferometer can be overcome by the use of two superimposed gratings instead of just one grating as the beam divider. The amount of shear in Ronchi's interferometer is proportional to the spatial frequency of the diffraction grating used. If we can vary the spatial frequency, we can have variable shear. This can be done by the moiré effect of the two superimposed gratings. When two gratings are superimposed, owing to the moiré effect a variable spatial frequency is produced by rotating the angle between their axes. An interferometer, called the "grating interferometer," has been developed from two superimposed Ronchi rulings used as beam splitters. All the drawbacks in Ronchi's interferometer are overcome.

Carrier-frequency photography has some similarity to holography. In particular, both are two-step processes, and diffraction methods are generally used in the retrieval stages. Holography has been used very successfully in interferometry. Holographic interferometry can be operated either in real time or in two stages.[22]

A hologram's ability to record both the amplitude and phase of a wavefront is the reason for its success in interferometry. Carrier-frequency photography can register the amplitude as well as the phase variation of objects. When a grating is imaged through a complex object, the resultant image is an amplitude-modulated version of the grating with the slits shifted according to the phase variation of the object. This kind of encoding scheme, involving both amplitude-modulation and pulse-position modulation, called the "detour phase" by A. Lohmann, is the essence of computer-generated holography.[23,24] It is very difficult to retrieve the information about the phase variation of the object directly from the distorted grating image. But if we "beat" the distorted grating image with a regular grating (or master grating), moiré fringes appear which will reveal the phase variation. A new type of interferometer, based on carrier-frequency photography and the moiré effect, has been conceived. It is called the carrier-frequency interferometer. It can be operated in real time or in two stages. The system is achromatic and white light can be used.

In chapter 4, we briefly review diffraction by gratings, the moiré effect, carrier-frequency photography, and achromatic holography. This review will serve as the foundation of the following chapters.

The grating interferometer is the subject of chapter 5, and the carrier-frequency interferometer is described in chapter 6. Many kinds of masks can be used as modulation carriers in the carrier-frequency interferometer. We use linear gratings, checkerboard gratings, and circular gratings.

The applications of these two types of interferometers in optical information processing are presented in both chapters 5 and 6 for each interferometer. New methods for optical testing, phase object visualization, image polarity reversal and color complementarity, two-dimensional photometry, spatial differentiation, complex amplitude addition and subtraction, image multiplexing, pure spatial sinusoidal wave generation, fabrication of filters, focusing, and alignment are presented. Some of the applications can be used in optical computing, general image processing, and two-dimensional photometry.

Footnotes to Introduction

1. Helstrom [1], p. 112.
2. Van Trees [2], p. 467.
3. Wiener [3], p. 84.
4. Harris [4].
5. Helstrom [5].
6. Slepian [6].
7. Freiden [7].
8. Rushforth and Harris [8].
9. Falk [9].
10. Helstrom [10].
11. Wozencraft and Jacobs [11].
12. Helstrom [12].
13. Helstrom [1], p. 256.
14. Himmelblau [13], p. 64.
15. Leith and Upatnieks [14].
16. Elias [15].
17. O'Neill [16].
18. Lohmann and Werlich [17].
19. Armitage and Lohmann [18].
20. Nishijma and Oster [19].
21. Ronchi [20].

22. Collier, Burckhardt, and Lin [21], chapter 15.

23. Brown and Lohmann [22].

24. Lohmann and Paris [23].

Part I. Restoration of Images Degraded by Signal-Dependent Noise

In this part of the thesis we will discuss the restoration of images degraded by noise that is neither Gaussian nor additive.

When imaging incoherent light, it is the radiance distribution of the object plane that is to be estimated, not the value of the light field itself. The object radiance function, like the illuminance in the image plane, is a quadratic functional of the field. We are going to discuss the effect of the quadratic rectification of the field and the interaction of the light field with the recording medium in our imaging system. The optical system used is an imaging system in which the light from the object plane is focused onto a photosensitive surface, from which it ejects photoelectrons.

We begin by describing the statistical properties of the object light field and the image-plane light field. A linear estimator is derived for the restoration of the object radiance function based on the measured data which is assumed to be the illuminance in the image plane. The interaction of the light field with the recording medium is then taken into account. Linear and maximum-likelihood estimators of the object radiance function based on the measured numbers of photoelectrons emitted from the recording medium have been derived. Computer simulations are performed to demonstrate the significance of the maximum-likelihood estimate, which is superior to the linear estimate when the light level is low.

Chapter 1

Statistical Description of Object Field and Linear Estimation of Object Radiance Function Based on Measured Image Illuminance

In this chapter, we will first discuss the statistical representation of the object field and derive the linear least-squares estimation of the object radiance function based on the illuminance distribution in the image plane, which is assumed measurable. The interaction of the light field with the recording medium will be neglected here and will be discussed in the next chapter.

The electromagnetic field emitted by the object is assumed to be linearly polarized, quasimonochromatic, and represented by its "analytic signal."[1] The analytic signal of the object light field will be considered as a circular complex spatio-temporal Gaussian random process which is stationary, ergodic, and spectrally pure. Linear scalar diffraction theory is used for wave propagation. The relations among the covariance function of the object light field, that in the image plane, and the point spread function of optical system will be discussed. The radiance function of the incoherently radiating object is considered as one of an ensemble of spatial random processes, which form the class of the objects to be examined. The background noise is assumed to be a circular complex spatio-temporal, white Gaussian random process with spectral density N. The instantaneous intensity of the light field in the image plane is assumed to be

measurable and to constitute the data. A linear spatial filter based on these data is derived for the restoration of the object radiance function.

1.1. The Statistical Properties of the Object Light Field

The electromagnetic field emitted by the object is assumed to be linearly polarized and quasimonochromatic. The light field radiated by the object at point $\underset{\sim}{u} = (u_x, u_y)$ of the object plane can be specified by a scalar function

$$U_0(\underset{\sim}{u}, t) = R\ell\ \psi_0(\underset{\sim}{u}, t)\, e^{i\Omega t} \quad , \tag{1.1}$$

where $\Omega = 2\pi c/\bar{\lambda}$ is the central angular frequency, $\bar{\lambda}$ the wavelength at the central angular frequency, and $\psi_0(\underset{\sim}{u}, t)$ the complex envelope of the field. "$R\ell$" indicates the real part of a complex number. $\psi_0(\underset{\sim}{u}, t)\, e^{i\Omega t}$ is called the analytic signal. The object light field can be completely represented by the complex envelope of the object light field, $\psi_0(\underset{\sim}{u}, t)$.[2]

The complex envelope of the object light field is a circular complex spatio-temporal random process with the complex autocovariance function

$$\begin{aligned} \phi_0(\underset{\sim}{u}_1, \underset{\sim}{u}_2; t_1, t_2) &= \frac{1}{2} \langle \psi_0(\underset{\sim}{u}_1, t_1)\, \psi_0^*(\underset{\sim}{u}_2, t_2) \rangle \\ &= \phi_0^*(\underset{\sim}{u}_2, \underset{\sim}{u}_1; t_1, t_2) \quad , \end{aligned} \tag{1.2}$$

where $\langle \cdot \rangle$ denotes the ensemble average, and $*$ means the complex conjugate. The complex autocovariance function is proportional

to the mutual coherence function with a factor $\exp[i\Omega (t_1 - t_2)]$. Since the object is a circular complex Gaussian random process, the following relations[3] hold:

$$\langle \psi_0(\underset{\sim}{u}_1, t_1)\ \psi_0(\underset{\sim}{u}_2, t_2)\rangle = 0 \ , \tag{1.3}$$

and

$$\begin{aligned}
&\langle \psi_0(\underset{\sim}{u}_1, t_1)\ \psi_0^*(\underset{\sim}{u}_2, t_2)\ \psi_0(\underset{\sim}{u}_3, t_3)\ \psi_0^*(\underset{\sim}{u}_4, t_4)\rangle \\
&\quad = \langle \psi_0(\underset{\sim}{u}_1, t_1)\ \psi_0^*(\underset{\sim}{u}_2, t_2)\rangle\ \langle \psi_0(\underset{\sim}{u}_3, t_3)\ \psi^*(\underset{\sim}{u}_4, t_4)\rangle \\
&\quad + \langle \psi_0(\underset{\sim}{u}_1, t_1)\ \psi_0^*(\underset{\sim}{u}_4, t_4)\rangle\ \langle \psi_0^*(\underset{\sim}{u}_2, t_2)\ \psi_0(\underset{\sim}{u}_3, t_3)\rangle \\
&\quad = 4\, \phi_0(\underset{\sim}{u}_1, \underset{\sim}{u}_2;\ t_1, t_2)\ \phi_0(\underset{\sim}{u}_3, \underset{\sim}{u}_4;\ t_3, t_4) \\
&\quad + 4\, \phi_0(\underset{\sim}{u}_1, \underset{\sim}{u}_4;\ t_1, t_4)\ \phi_0(\underset{\sim}{u}_3, \underset{\sim}{u}_2;\ t_3, t_2) \ .
\end{aligned} \tag{1.4}$$

It can be shown that in general a quasimonochromatic wavefield is spectrally pure.[4] The spectrally pure light beam is characterized in such a way that the superposition of light beams will not affect the spectral distribution; the covariance function is then reducible to the product of a spatial covariance function $\phi_0(\underset{\sim}{u}_1, \underset{\sim}{u}_2)$ and a temporal covariance function $\chi(t_1 - t_2)$, i.e.,

$$\phi_0(\underset{\sim}{u}_1, \underset{\sim}{u}_2;\ t_1, t_2) = \phi_0(\underset{\sim}{u}_1, \underset{\sim}{u}_2)\, \chi(t_1 - t_2) \ . \tag{1.5}$$

As a temporal random process, the light field is considered as stationary and ergodic. The temporal covariance function is usually normalized so that $\chi(0) = 1$.

The bandwidth W of the object light field is defined as

$$W = |\chi(0)|^2 \Big/ \int_{-\infty}^{\infty} |\chi(\tau)|^2 \, d\tau$$

$$= \left[\int_{-\infty}^{\infty} X(\omega) \, d\omega/2\pi \right]^2 \Big/ \int_{-\infty}^{\infty} |X(\omega)|^2 \, d\omega/2\pi \quad , \tag{1.6}$$

where

$$X(\omega) = \int_{-\infty}^{\infty} \chi(\tau) \, e^{i\omega\tau} \, d\tau \quad , \tag{1.7}$$

is the Fourier transform of the temporal covariance function, or the temporal spectral density. The bandwidth of the object light field is so defined that if $X(\omega)$ is constant over frequency range $-\frac{1}{2} W < \omega < \frac{1}{2} W$ and zero otherwise, then the bandwidth defined in Eq. (1.6) equals to W. $\psi_0(\underset{\sim}{u}, t)$ and $\psi_0(\underset{\sim}{u}, t + W^{-1})$ can be considered statistically independent, and W^{-1} is usually called the correlation time. The correlation time of the signal is inversely proportional to its bandwidth.

For a spatially incoherently radiating object, the mutual coherence function takes the form

$$\phi_0(\underset{\sim}{u}_1, \underset{\sim}{u}_2; t_1, t_2) = (\lambda^2/4\pi) \, B(\underset{\sim}{u}_1) \, \delta(\underset{\sim}{u}_1 - \underset{\sim}{u}_2) \, \chi(t_1 - t_2) \, , \tag{1.8}$$

where $\delta(\underset{\sim}{u})$ is the two-dimensional delta function. $B(\underset{\sim}{u})$ is the radiance function of the object and can be decomposed as

$$B(\underset{\sim}{u}) = B_0 + b(\underset{\sim}{u}) \tag{1.9}$$

into a known uniform background B_0 and a deviation $b(\underset{\sim}{u})$. The deviation $b(\underset{\sim}{u})$ represents the scene of interest and is what we want to estimate. The radiance is so defined that $B(\underset{\sim}{u})/4\pi$ is the power emitted per unit area per steradian in a direction normal to the object plane.

Although the object light field is not spatially stationary, the object radiance function is very similar to the spectral density of a time series.[5, 6] To estimate $b(\underset{\sim}{u})$ is equivalent to estimating the spectral density in time series analysis. The variability and unbiasedness always conflict with each other.[6]

1.2. The Light Field in the Image Plane

The system we want to study is shown in Fig. 1.1. An optical observing aperture A is used to collect the light emitted by the object. The coordinate variables $\underset{\sim}{u}$, $\underset{\sim}{r}$ and $\underset{\sim}{x}$ represent the object, aperture and image plane respectively. The aperture contains a narrow-band filter used to block the background light outside the spectral region of the object light, and a lens to focus the light onto the image plane I. The distances from the aperture to the object plane and image plane are Z_o and Z_i respectively, and satisfy the image-forming relationship

$$Z_o^{-1} + Z_i^{-1} = f^{-1} ,$$

where f is the focal length of the lens used.

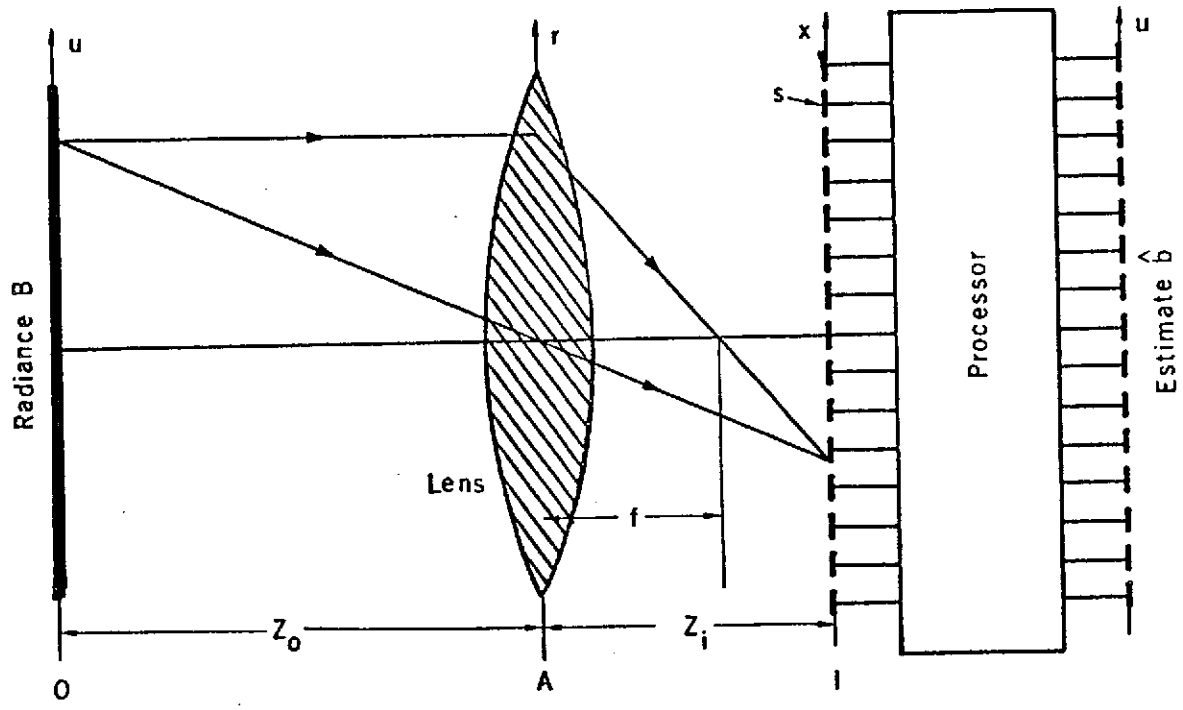


Figure 1.1. The object and image-processing system. O = object plane, A = aperture, I = image plane. A narrowband spectral filter for object and background light is not shown.

If the object, aperture, and image planes are assumed to be far apart, and the aperture is not too big, then all light rays can be considered paraxial, and it suffices to use scalar diffraction theory to describe the wave propagation from the object plane to the aperture plane and then to the image plane.

At point $\underset{\sim}{r}$ of the aperture plane A, the total field is

$$\psi_a(\underset{\sim}{r}, t) = \psi_s(\underset{\sim}{r}, t) + \psi_n(\underset{\sim}{r}, t) \ , \tag{1.10}$$

where $\psi_n(\underset{\sim}{r}, t)$ represents the background noise, and

$$\psi_s(\underset{\sim}{r}, t) = \int_0 \psi_0(\underset{\sim}{u}, t)\, S_0(\underset{\sim}{u}, \underset{\sim}{r}; t)\, d^2\underset{\sim}{u} \tag{1.11}$$

represents the component due to the object, or signal. Here $S(\underset{\sim}{u}, \underset{\sim}{r}; t)$ is the instantaneous amplitude point spread function describing the wave propagation from the object plane to the aperture plane, and 0 indicates an integration over the object plane. The point spread function $S(\underset{\sim}{u}, \underset{\sim}{r}; t)$ can be a time varying function, such as in the situation that the object light field propagated through a turbulent medium, or when the object is moving relatively to the aperture plane in the period of observation. For a diffraction limited system, in the paraxial approximation, it is

$$\begin{aligned} S_0(u, r; t) &= S_0(\underset{\sim}{u}, \underset{\sim}{r}) \\ &= i(\lambda Z_0)^{-1} \exp(iKZ_0 + iK|\underset{\sim}{u} - \underset{\sim}{r}|^2/2Z_0) \ , \end{aligned} \tag{1.12}$$

where $K = 2\pi/\lambda$ is the propagation constant.

The background noise $\psi_n(\underset{\sim}{r}, t)$ is a spatial and temporal white, circular-complex Gaussian random process whose mutual covariance function, after filtering by the temporal filter, is

$$\phi_n(\underset{\sim}{r}_1, \underset{\sim}{r}_2; t_1, t_2) = \frac{1}{2} \langle \psi_n(\underset{\sim}{r}_1, t_1) \psi_n^*(\underset{\sim}{r}_2, t_2) \rangle$$

$$= N \delta(\underset{\sim}{r}_1 - \underset{\sim}{r}_2) \chi(t_1 - t_2) , \qquad (1.13)$$

where N is the spectral density of the noise.

The field at the image plane is

$$\psi_i(x, t) = \int_A \psi_a(\underset{\sim}{r}, t) S_a(\underset{\sim}{r}, \underset{\sim}{x}) d^2\underset{\sim}{r}$$

$$= \psi_{si}(\underset{\sim}{x}, t) + \psi_{ni}(\underset{\sim}{x}, t) , \qquad (1.14)$$

where $S_a(\underset{\sim}{r}, \underset{\sim}{x})$ is the amplitude point spread function between the aperture and the image plane, and in the paraxial approximation

$$S_a(\underset{\sim}{r}, \underset{\sim}{x}) = i(\lambda Z_i)^{-1} \exp\left[iKZ_i + iK|\underset{\sim}{r} - \underset{\sim}{x}|^2/2Z_i - iK\underset{\sim}{r}^2/2F\right] . \qquad (1.15)$$

The last term in the exponent is the quadratic phase factor introduced by the lens.

The field due to the background noise is

$$\psi_{ni}(\underset{\sim}{x}, t) = \int_A \psi_n(\underset{\sim}{r}, t) S_a(\underset{\sim}{x}, \underset{\sim}{r}) d^2\underset{\sim}{r} , \qquad (1.16)$$

where A indicates an integration over the aperture plane. And the field due to the object light is

$$\psi_{si}(\underset{\sim}{x}, t) = \int_A \psi_s(\underset{\sim}{r}, t)\, S_a(\underset{\sim}{r}, \underset{\sim}{x})\, d^2\underset{\sim}{r}$$

$$= \int_A \int_0 \psi_0(\underset{\sim}{u}, t)\, S_0(\underset{\sim}{u}, \underset{\sim}{r}; t)\, S_a(\underset{\sim}{r}, \underset{\sim}{x})\, d^2\underset{\sim}{r}\, d^2\underset{\sim}{u}$$

$$= \int_0 \psi_0(\underset{\sim}{u}, t)\, S(\underset{\sim}{u}, \underset{\sim}{x}; t)\, d^2\underset{\sim}{u} \qquad (1.17)$$

where

$$S(\underset{\sim}{u}, \underset{\sim}{x}; t) = \int_A S_0(\underset{\sim}{u}, \underset{\sim}{r}; t)\, S_a(\underset{\sim}{r}, \underset{\sim}{x})\, d^2\underset{\sim}{r} \quad , \qquad (1.18)$$

is the amplitude point spread function between the object and image planes. For a diffraction limited system[13]

$$S(\underset{\sim}{u}, \underset{\sim}{x}; t) = S(\underset{\sim}{u}, \underset{\sim}{x})$$

$$= (\lambda^2 Z_o Z_i)^{-1} \exp[iK(Z_o + Z_i)] \exp\left[\frac{i\pi}{\lambda}\left(\frac{\underset{\sim}{u}^2}{Z_o} + \frac{\underset{\sim}{x}^2}{Z_i}\right)\right]$$

$$\times \int_A \exp\left[iK\left(\frac{\underset{\sim}{u}}{Z_o} + \frac{\underset{\sim}{x}}{Z_i}\right) \cdot \underset{\sim}{r}\right] d^2\underset{\sim}{r} \quad . \qquad (1.19)$$

If the aperture is large enough, the integral in Eq. (1.19) will approach a delta function located at $\underset{\sim}{x} = -(Z_i/Z_o)\underset{\sim}{u}$, and the field at point u will form its "geometrical image" at

$$\underset{\sim}{x} = -(Z_i/Z_o)\, \underset{\sim}{u}$$

$$= -M\underset{\sim}{u} \qquad (1.20)$$

where

$$M = Z_i/Z_o \tag{1.21}$$

is the magnification factor.

It is more convenient to deal with image coordinates having the same scale as those of the object, and we will define a new coordinate

$$\underset{\sim}{V} = -(Z_o/Z_i)\,\underset{\sim}{x}$$

$$= -M^{-1}\,\underset{\sim}{x} \quad . \tag{1.22}$$

The diffraction-limited point spread function, from Eq. (1.24), in the new coordinates is

$$S(\underset{\sim}{u},\underset{\sim}{v}) = (\lambda^2 Z_o Z_i)^{-1} \exp[iK(Z_o+Z_i)] \exp\left[\frac{i\pi}{\lambda Z_o}\left(\underset{\sim}{u}^2 + \frac{Z_i}{Z_o}\,\underset{\sim}{v}^2\right)\right]$$

$$\int_A \exp[iK(\underset{\sim}{u}-\underset{\sim}{v})\cdot \underset{\sim}{r}/Z_o]\, d^2\underset{\sim}{r}$$

$$= (\lambda^2 Z_o Z_i)^{-1} \exp[iK(Z_o+Z_i)] \exp\left[\frac{i\pi}{\lambda Z_o}(\underset{\sim}{u}^2 + M\underset{\sim}{v}^2)\right] \cdot A \cdot \vartheta(\underset{\sim}{u}-\underset{\sim}{v}) \,, \tag{1.23}$$

where

$$\vartheta(\underset{\sim}{u}) = A^{-1} \int_A \exp(2\pi i\, \underset{\sim}{u}\cdot\underset{\sim}{r}/\lambda Z_o)\, I_A(\underset{\sim}{r})\, d^2\underset{\sim}{r} \tag{1.24}$$

is the Fourier transform of the aperture indicator function

$$I_A(\underset{\sim}{r}) = \begin{cases} 1 & r \in A \\ 0 & r \notin A \end{cases} .$$

Since the noise background field $\psi_n(r,t)$ and object light field are independent, the covariance function of the field in the image plane, from Eq. (1.14), is

$$\phi_i(\underset{\sim}{v}_1, \underset{\sim}{v}_2; t_1, t_2) = \frac{1}{2} \langle \psi_i(\underset{\sim}{v}_1, t_1)\, \psi_i^*(\underset{\sim}{v}_2, t_2) \rangle$$

$$= \frac{1}{2} \langle \psi_{si}(\underset{\sim}{v}_1, t_1)\, \psi_{si}^*(\underset{\sim}{v}_2, t_2) \rangle$$

$$+ \frac{1}{2} \langle \psi_{ni}(\underset{\sim}{v}_1, t_1)\, \psi_{ni}^*(\underset{\sim}{v}_2, t_2) \rangle \ . \qquad (1.25)$$

In the image plane the covariance function of the light field originating from the object, from Eq. (1.17), is

$$\phi_{si}(\underset{\sim}{v}_1, \underset{\sim}{v}_2; t_1, t_2) = \frac{1}{2} \langle \psi_{si}(\underset{\sim}{v}_1, t_1)\, \psi_{si}^*(\underset{\sim}{v}_2, t_2) \rangle$$

$$= (\lambda^2/4\pi) \int_0 B(\underset{\sim}{u})\, S(\underset{\sim}{u}, \underset{\sim}{v}_1; t_1)\, S^*(\underset{\sim}{u}, \underset{\sim}{v}_2; t_2)\, d^2\underset{\sim}{u} \cdot \chi(t_1 - t_2) \ , \qquad (1.26)$$

and for the diffraction-limited case

$$= (\lambda^2/4\pi)(\lambda^2 Z_o Z_i)^{-2}\, \chi(t_1 - t_2) \exp\left[\frac{i\pi Z_i}{\lambda Z_o^2} (v_1^2 - v_2^2)\right]$$

$$A^2 \int_0 B(\underset{\sim}{u})\, \vartheta(\underset{\sim}{u} - \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u} - \underset{\sim}{v}_2)\, d^2\underset{\sim}{u} \qquad (1.27)$$

$$= (4\pi Z_i^2)^{-1} A \exp\left[\frac{i\pi Z_i}{\lambda Z_o^2} (\underset{\sim}{v}_1^2 - \underset{\sim}{v}_2^2)\right] \cdot \chi(t_1 - t_2)$$

$$\left\{ B_0\, \vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2) + (\lambda Z_o)^{-2} A \int_0 b(\underset{\sim}{u})\, \vartheta(\underset{\sim}{u} - \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u} - \underset{\sim}{v}_2)\, d^2\underset{\sim}{u} \right\} .$$

The background noise field $\psi_{ni}(v,t)$ has covariance function

$$\phi_{ni}(\underset{\sim}{v}_1, \underset{\sim}{v}_2; t_1, t_2) = \frac{1}{2} \langle \psi_{ni}(\underset{\sim}{v}_1, t_1) \psi^*_{ni}(\underset{\sim}{v}_2, t_2) \rangle$$

$$= N(\lambda Z_i)^{-2} \exp\left[\frac{i\pi Z_i}{\lambda Z_o^2} (\underset{\sim}{v}_1^2 - \underset{\sim}{v}_2^2)\right] A\, \vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2) \quad . \tag{1.28}$$

After combining Eqs. (1.25), (1.27), and (1.28), the covariance function of the light field in the image plane is

$$\phi_i(\underset{\sim}{v}_1, \underset{\sim}{v}_2; t_1, t_2) = \phi_i(\underset{\sim}{v}_1, \underset{\sim}{v}_2)\, \chi(t_1 - t_2) \quad , \tag{1.29}$$

$$\phi_i(\underset{\sim}{v}_1, \underset{\sim}{v}_2) = (4\pi Z_i^2)^{-1} A \cdot \exp\left[\frac{i\pi Z_i}{\lambda Z_o^2} (\underset{\sim}{v}_1^2 - \underset{\sim}{v}_2^2)\right]$$

$$\left\{ \bar{B}\, \vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2) + (\lambda Z_o)^{-2} A \int_o b(\underset{\sim}{u})\, \vartheta(\underset{\sim}{u} - \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u} - \underset{\sim}{v}_2)\, d^2\underset{\sim}{u} \right\}, \tag{1.30}$$

where

$$\bar{B} = B_0 + (4\pi/\lambda^2) N \tag{1.31}$$

is the equivalent background noise spectral density. Here we can see that the uniform object background, B_o, contributes to the noise spectral density. This is very different from the linear model of signal formation, in which the noise and the object radiance function are independent and combined additively. In the linear model, the known uniform object background B_o does not contribute to the noise spectral density.[7]

1.3. The Instantaneous Image Plane Light Field Intensity

In this chapter, we will assume that the instantaneous light field intensity is the measurable datum. A linear least-squares estimation of the object radiance function based on this measured datum will be derived.

The light field in the image plane is not measurable and is not of interest. The measurable datum is its intensity[8]

$$I_i(\underset{\sim}{v}, t) = \frac{1}{2} \left| \psi_i(\underset{\sim}{v}, t) \right|^2 . \tag{1.32}$$

The mean value of the intensity is

$$\begin{aligned} \langle I_i(v, t) \rangle &= \frac{1}{2} \langle \left| \psi_i(\underset{\sim}{v}, t) \right|^2 \rangle \\ &= \phi_i(\underset{\sim}{v}, \underset{\sim}{v};\ t, t) \\ &= (4\pi Z_i^2)^{-1} A \left\{ \bar{B} + (\lambda Z_o)^{-2} A \int_o b(\underset{\sim}{u}) \left| \vartheta(\underset{\sim}{u}-\underset{\sim}{v}) \right|^2 d^2\underset{\sim}{u} \right\} . \end{aligned} \tag{1.33}$$

The mean value of the image plane intensity is linearly proportional to the object radiance function. For an incoherent imaging system, the illuminance in the image plane is linearly related to the object radiance with point spread function $K(\underset{\sim}{u}, \underset{\sim}{v}) = \left| S(\underset{\sim}{u}, \underset{\sim}{v}) \right|^2$.

The covariance function of $I_i(\underset{\sim}{v}, t)$ is

$$\operatorname{cov}\{I_i(\underset{\sim}{v}_1, t_1),\ I_i(\underset{\sim}{v}_2, t_2)\}$$

$$= \langle I_i(\underset{\sim}{v}_1, t_1)\, I_i(\underset{\sim}{v}_2, t_2)\rangle - \langle I_i(\underset{\sim}{v}_1, t_1)\rangle \langle I_i(\underset{\sim}{v}_2, t_2)\rangle$$

$$= \frac{1}{4}\langle |\psi_i(\underset{\sim}{v}_1, t_1)|^2 |\psi_i(\underset{\sim}{v}_2, t_2)|^2\rangle - \frac{1}{4}\langle |\psi_i(\underset{\sim}{v}_1, t_1)|^2\rangle \langle |\psi_i(\underset{\sim}{v}_2, t_2)|^2\rangle$$

$$= \frac{1}{4}\left|\langle \psi_i(\underset{\sim}{v}_1, t_1)\, \psi_i^*(\underset{\sim}{v}_2, t_2)\rangle\right|^2$$

$$= \left|\phi_i(\underset{\sim}{v}_1, \underset{\sim}{v}_2;\ t_1, t_2)\right|^2$$

$$= (4\pi Z_i^2)^{-2} A^2 \cdot \left|\chi(t_1 - t_2)\right|^2 \left\{ \bar{B}^2 \left|\vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2)\right|^2 \right.$$

$$+ 2\bar{B} A(\lambda Z_o)^{-2} \vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2) \int_0 b(\underset{\sim}{u})\, \vartheta(\underset{\sim}{u} - \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u} - \underset{\sim}{v}_2)\, d^2\underset{\sim}{u}$$

$$+ (\lambda Z_o)^{-4} A^2 \iint b(\underset{\sim}{u}_1)\, b(\underset{\sim}{u}_2)\, \vartheta(\underset{\sim}{u}_1 - \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u}_1 - \underset{\sim}{v}_2)\, \vartheta^*(\underset{\sim}{u}_2 - \underset{\sim}{v}_1)\, \vartheta(\underset{\sim}{u}_2 - \underset{\sim}{v}_2)$$

$$\left. d^2\underset{\sim}{u}_1\, d^2\underset{\sim}{u}_2 \right\}, \qquad (1.34)$$

and its variance is

$$\operatorname{Var}(I_i(\underset{\sim}{v}, t) = \left|\phi_i(\underset{\sim}{v}, \underset{\sim}{v};\ t, t)\right|^2$$

$$= \left[\langle I_i(\underset{\sim}{v}, t)\rangle\right]^2 . \qquad (1.35)$$

Based on the statistics of the measurable data, $I_i(\underset{\sim}{v}, t)$, a linear least-square estimation of the object radiance function will be derived in the next section.

1.4. Linear Least-Squares Estimation

In this section we assume that the illuminance in the image plane is available for linear spatial filtering,[9, 10] the result is

$$J(\underset{\sim}{u}) = \frac{1}{T} \int_{I} \int_{0}^{T} L(\underset{\sim}{v})\, I_i(\underset{\sim}{u} - \underset{\sim}{v},\ t)\, d^2\underset{\sim}{v}\, dt\ , \tag{1.36}$$

where I indicates an integration over the image plane, T is the observation interval, and $L(\underset{\sim}{v})$ is the point-spread function of the spatial filter. The output of the rectified and filtered image plane light field $J(\underset{\sim}{u})$ is

$$\begin{aligned}
\langle J(u) \rangle &= \frac{1}{T} \int_{I} \int_{0}^{T} L(\underset{\sim}{v})\, \langle I_i(\underset{\sim}{u} - \underset{\sim}{v},\ t) \rangle\ d^2\underset{\sim}{v}\, dt \\
&= \int_{I} L(\underset{\sim}{v}) (4\pi Z_i^2)^{-1} A\ \overline{B} + (\lambda Z_o)^{-2} A \int_{0} b(\underset{\sim}{u}') \left| \mathcal{J}(\underset{\sim}{u}' - \underset{\sim}{u} + \underset{\sim}{v}) \right|^2 d^2u'\ d^2\underset{\sim}{v} \\
&= J_B + (4\pi\lambda^2 Z_o^2 Z_i^2)^{-1} A^2 \iint L(\underset{\sim}{v})\, b(\underset{\sim}{u}') \left| \mathcal{J}(\underset{\sim}{u}' - \underset{\sim}{u} + \underset{\sim}{v}) \right|^2 d^2\underset{\sim}{u}'\, d^2\underset{\sim}{v}\ ,
\end{aligned} \tag{1.37}$$

where

$$J_B = (4\pi Z_i^2)^{-1} A \cdot \overline{B} \int_{I} L(\underset{\sim}{v})\, d^2\underset{\sim}{v}$$

is the contribution from the background. Equation (1.36) can be re-written as

$$\langle J(\underset{\sim}{u}) \rangle = J_B + C \int \Delta(\underset{\sim}{u} - u')\, b(\underset{\sim}{u}')\, d^2\underset{\sim}{u}'\ , \tag{1.38}$$

where

$$\Delta(u') = \int L(\underset{\sim}{v}) \left| \vartheta(\underset{\sim}{u}' - \underset{\sim}{v}) \right|^2 d^2\underset{\sim}{v} \tag{1.39}$$

and

$$C = (4\pi\lambda^2 Z_o^2 Z_i^2)^{-1} A^2 \quad .$$

If we could choose a linear filter $L(v)$ such that $\Delta(u)$ is a delta function, then the estimate[11]

$$\hat{b}(\underset{\sim}{u}) = C^{-1} [J(\underset{\sim}{u}) - J_B] \tag{1.40}$$

of the deviation of the object radiance function would have a mean value equal to $b(\underset{\sim}{u})$ itself. Then the estimate $\hat{b}(\underset{\sim}{u})$ of $b(\underset{\sim}{u})$ would be unbiased, and the mean-square bias

$$\mathcal{E}_b = \langle \left| \langle \hat{b}(\underset{\sim}{u}) \rangle - b(\underset{\sim}{u}) \right|^2 \rangle \tag{1.41}$$

would be zero. But if we go too far in that direction, the mean-square fluctuation

$$\mathcal{E}_v = \langle \left| \hat{b}(\underset{\sim}{u}) - \langle \hat{b}(\underset{\sim}{u}) \rangle \right|^2 \rangle \tag{1.42}$$

begins to increase. The mean-square bias and mean-square fluctuation of the estimate $\hat{b}(\underset{\sim}{u})$ conflict with each other as in temporal power spectral estimation.[6] We have to choose a linear filter $L(\underset{\sim}{v})$ which achieves the minimum of error $\mathcal{E}_a + \mathcal{E}_v$. The mean-square bias, calculated from Eqs. (1.37) and (1.40), is

$$\mathcal{E}_b = b^2(\underset{\sim}{u}) - 2\, b(\underset{\sim}{u}) \int_0 \Delta(\underset{\sim}{u} - \underset{\sim}{u}')\, b(\underset{\sim}{u}')\, d^2\underset{\sim}{u}'$$

$$+ \int_0 \int_0 b(u_1)\, b(u_2)\, \Delta(\underset{\sim}{u} - \underset{\sim}{u}'_1)\, \Delta(\underset{\sim}{u} - \underset{\sim}{u}'_2)\, d^2u'_1\, d^2u'_2 \ ,$$

and the mean-square fluctuation is

$$\mathcal{E}_v = \text{var}\ (\hat{b}(\underset{\sim}{u}))$$

$$= c^{-2}\ \text{var}\ (J(\underset{\sim}{u})) \ . \qquad (1.43)$$

The variance of $J(\underset{\sim}{u})$, from Eqs. (1.36) and (1.33), is

$$\text{Var}(J(\underset{\sim}{u})) = \frac{1}{T^2} \int_0^T \int_0^T \iint_I \iint_I L(\underset{\sim}{v}_1)\, L(\underset{\sim}{v}_2)\, \text{cov}\left\{ I_i(\underset{\sim}{u} - \underset{\sim}{v}_1, t_1),\ I_i(\underset{\sim}{u} - \underset{\sim}{v}_2, t_2)\right\}$$

$$d^2\underset{\sim}{v}_1\, d^2\underset{\sim}{v}_2\, dt_1 dt_2$$

$$= (WT)^{-1}(4\pi Z_i^2)^{-2} A^2 \iint_I \iint_I L(\underset{\sim}{v}_1)\, L(\underset{\sim}{v}_2) \left\{ B^2 \left|\vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2)\right|^2 \right.$$

$$+ 2(\lambda Z_o)^{-2} \bar{B} A\, \vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2) \int_0 b(\underset{\sim}{u}_1)\, \vartheta(\underset{\sim}{u}_1 - \underset{\sim}{u} + \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u}_1 - \underset{\sim}{u} + \underset{\sim}{v}_2)\, d^2\underset{\sim}{u}_1$$

$$+ (\lambda Z_o)^{-4} A^2 \iint_0 b(\underset{\sim}{u}_1)\, b(\underset{\sim}{u}_2)\, \vartheta(\underset{\sim}{u}_1 - \underset{\sim}{u} + \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u}_1 - \underset{\sim}{u} + \underset{\sim}{v}_2)$$

$$\left. \times\ \vartheta^*(\underset{\sim}{u}_2 - \underset{\sim}{u} + \underset{\sim}{v}_1)\, \vartheta(\underset{\sim}{u}_2 - \underset{\sim}{u} + \underset{\sim}{v}_2)\, d^2\underset{\sim}{u}_1\, d^2\underset{\sim}{u}_2 \right\} d^2\underset{\sim}{v}_1\, d^2\underset{\sim}{v}_2$$

(1.44)

Any signal processing system is designed for a particular class of signal. The statistical properties assigned to the class reflect the nature of the signal to be processed. The prior knowledge about the statistical properties of signals is essential in signal processing.[7, 12] Here we will consider the scene of interest $b(\underset{\sim}{u})$ in the object plane as a realization of a homogeneous stochastic process with mean zero, and covariance function

$$E_b\{ b(\underset{\sim}{u}_1)\, b(\underset{\sim}{u}_2)\} = \varphi_b(\underset{\sim}{u}_1 - \underset{\sim}{u}_2) , \tag{1.45}$$

and

$$\sigma_b^2 = \varphi_b(\underset{\sim}{0}) \quad . \tag{1.46}$$

E_b denotes the average over the ensemble of the spatial processes $b(\underset{\sim}{u})$. The structure of the picture, i.e., $b(\underset{\sim}{u})$, is to some extent embodied in the covariance function $\varphi_b(\underset{\sim}{u})$, whose width will be of the order of the average size of details in the picture. The ratio σ_b^2/B_o^2 specifies the mean-square contrast. By averaging over the ensemble of possible pictures, we can express the mean-square error as

$$\begin{aligned} \mathcal{E} &= E_b\{ \langle [\hat{b}(\underset{\sim}{u}) - b(\underset{\sim}{u})]^2 \rangle \} \\ &= E_b\{ \langle [\hat{b}(\underset{\sim}{u}) - \langle \hat{b}(\underset{\sim}{u}) \rangle]^2 \rangle + \langle [\langle \hat{b}(\underset{\sim}{u}) \rangle - b(\underset{\sim}{u})]^2 \rangle \} \\ &= E_b\{ \mathcal{E}_v + \mathcal{E}_b \} \\ &= \mathcal{E}_1 + \mathcal{E}_2 \quad , \end{aligned} \tag{1.47}$$

where

$$\mathcal{E}_1 = E_b \{ \langle [\hat{b}(u) - \langle \hat{b}(u) \rangle]^2 \rangle = E_b \{ \mathcal{E}_v \} \tag{1.48}$$

represents the mean-square error due to the fluctuations of the light field itself, and

$$\mathcal{E}_2 = E_b \{ \langle [\hat{b}(u) - \langle b(u) \rangle]^2 \rangle \} = E_b \{ \mathcal{E}_b \} \tag{1.49}$$

represents the mean-square bias arising from the wave propagation and the post-rectification filtering.

From Eqs. (1.48), (1.43), and (1.45), after some calculation,

$$\begin{aligned}
\mathcal{E}_1 &= C^{-2} \cdot E_b \{ \mathrm{Var}\,(J(\underset{\sim}{u})) \} \\
&= C^{-2}(WT)^{-1}(4\pi Z_i^2)^{-2} A^2 \iint_I \iint_I L(\underset{\sim}{v}_1)\, L(\underset{\sim}{v}_2) \left\{ \bar{B}^2 \left| \vartheta(\underset{\sim}{v}_1 - \underset{\sim}{v}_2) \right|^2 \right. \\
&\quad + A^2 (\lambda Z_o)^{-4} \int_0 \int_0 \varphi_b(\underset{\sim}{u}_1 - \underset{\sim}{u}_2)\, \vartheta(\underset{\sim}{u}_1 - \underset{\sim}{u} + \underset{\sim}{v}_1)\, \vartheta^*(\underset{\sim}{u}_1 - \underset{\sim}{u} + \underset{\sim}{v}_2) \\
&\quad \left. \times \vartheta^*(\underset{\sim}{u}_2 - \underset{\sim}{u} + \underset{\sim}{v}_1)\, \vartheta(\underset{\sim}{u}_2 - \underset{\sim}{u} + \underset{\sim}{v}_2)\, d^2\underset{\sim}{u}_1\, d^2\underset{\sim}{u}_2 \right\} d^2\underset{\sim}{v}_1\, d^2\underset{\sim}{v}_2 \\
&= (WT)^{-1}(\lambda Z_o)^4 A^{-3} \left[\bar{B}^2 \int |\Lambda(\underset{\sim}{r})|^2 I_A^{(2)}(\underset{\sim}{r})\, d^2\underset{\sim}{r} \right. \\
&\quad \left. + \iint |\Lambda(\underset{\sim}{r})|^2 \Phi_b(s)\, I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})\, d^2\underset{\sim}{r}\, d^2\underset{\sim}{s} \right] ,
\end{aligned} \tag{1.50}$$

where

$$\Lambda(\underset{\sim}{r}) = \int_I L(\underset{\sim}{u}) \exp(-iK\underset{\sim}{u} \cdot \underset{\sim}{r}/Z_o)\, d^2\underset{\sim}{u} \tag{1.51}$$

$$\Phi(\underset{\sim}{r}) = (\lambda Z_o)^{-2} \int_0 \varphi_b(\underset{\sim}{u}) \exp(-iK\underset{\sim}{u} \cdot \underset{\sim}{r}/Z_o)\, d^2\underset{\sim}{u} \quad , \tag{1.52}$$

$$I_A^{(2)}(\underset{\sim}{r}) = A^{-1} \int_A I_A(\underset{\sim}{u})\, I_A(\underset{\sim}{r}+\underset{\sim}{u})\, d^2\underset{\sim}{u} \quad , \tag{1.53}$$

$$I_A^{(4)}(\underset{\sim}{r},\underset{\sim}{s}) = A^{-1} \int_A I_A(\underset{\sim}{u})\, I_A(\underset{\sim}{u}+\underset{\sim}{r})\, I_A(\underset{\sim}{u}+\underset{\sim}{s})\, I_A(\underset{\sim}{u}+\underset{\sim}{r}+\underset{\sim}{s})\, d^2\underset{\sim}{u} \, . \tag{1.54}$$

$\Lambda(\underset{\sim}{r})$ is the transfer function of the linear filter $L(\underset{\sim}{u})$. $I^{(2)}(r)$ is the correlation function of the aperture function. Combining Eqs. (1.50), (1.37), (1.40), (1.51), (1.52) and (1.53), we find

$$\begin{aligned} \mathcal{e}_2 &= E_b \{ \langle [\hat{b}(\underset{\sim}{u})\rangle - b(\underset{\sim}{u})]^2 \rangle \} \\ &= \int \Phi_b(\underset{\sim}{r}) [1 - (\lambda Z_0)^2 A^{-1} I_A^{(2)}(\underset{\sim}{r}) \Lambda(\underset{\sim}{r})]^2 \, d^2\underset{\sim}{r} \, , \end{aligned} \tag{1.55}$$

and the mean square-error is

$$\begin{aligned} \mathcal{e} &= \mathcal{e}_1 + \mathcal{e}_2 \\ &= \int |\Lambda(r)|^2 \left\{ (\lambda Z_o)^4 A^2 \Phi_b(r) [I_A^{(2)}(\underset{\sim}{r})]^2 + (WT)^{-1} (\lambda Z_0)^4 A^{-3} \left[B^2 I_A^{(2)}(\underset{\sim}{r}) \right.\right. \\ &\qquad \left.\left. + \int \Phi_b(\underset{\sim}{s})\, I_A^{(4)}(\underset{\sim}{r},\underset{\sim}{s})\, d^2\underset{\sim}{s} \right] \right\} d^2\underset{\sim}{r} \\ &\quad - 2\, R\ell \left\{ (\lambda Z_o)^2 A^{-1} \int \Lambda(r)\, \Phi_b(\underset{\sim}{r})\, I_A^{(2)}(\underset{\sim}{r})\, d^2\underset{\sim}{r} \right\} \\ &\quad + \int \Phi_b(\underset{\sim}{r})\, d^2\underset{\sim}{r} \quad . \end{aligned} \tag{1.56}$$

Using the variational calculus, from Eq. (1.56), we find that the linear filter which gives the minimum mean-square error is

$$\Lambda(\underset{\sim}{r}) = \begin{cases} (\lambda Z_o)^2 \, A \, \Phi_b(\underset{\sim}{r}) \, I_A^{(2)}(\underset{\sim}{r}) \, D^{-1}(\underset{\sim}{r}), & r \in A' \\ 0, & r \notin A' \end{cases} \tag{1.57}$$

where

$$D(\underset{\sim}{r}) = \Phi_b(\underset{\sim}{r}) \left[I_A^{(2)}(\underset{\sim}{r})\right]^2 + (WT)^{-1} A^{-1} \left[B^2 I_A^{(2)}(\underset{\sim}{r}) + \int \Phi_b(\underset{\sim}{s}) \, I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s}) \, d^2\underset{\sim}{s}\right], \tag{1.58}$$

and A' is the area where $I_A^{(2)}(\underset{\sim}{r}) \neq 0$, or the so called convolved aperture. The last term in Eq. (1.58) is the noise spectrum. It is clear that the bigger the time bandwidth product WT and the area of the aperture A are, the smaller the noise is.

In our model, we do take into account the effect of rectification, i.e., the measured data is the instantaneous intensity of the in the image plane. The linear filter $\Lambda(\underset{\sim}{r})$ that we have obtained, as shown in Eq. (1.57), is different from the one using a linear model for the formation of the measured data.[7] The first term in the noise spectrum originated from the background noise and the uniform object background B_0; the last term is from the fluctuation of the object radiance function. In the linear model of image formation, only background noise contributes to the noise spectrum.[7]

When $(WT)A(\sigma_B/\bar{B})^2 \gg 1$, $\Lambda(r) \cong (\lambda Z_o)^2 A [I_A^{(2)}(\underset{\sim}{r})]^{-1}$ for most of the convolved aperture A', except for those parts near the edge of the aperture, where $\Lambda(\underset{\sim}{r})$ increases to a finite value $(\lambda Z_o)^2 A^2 (\sigma_B/\bar{B})^2 (WT)$ instead of going to infinity. Since the diffraction-limited system is a bandpass filter, none of the object spectrum outside the passband is recoverable, as can be seen from Eq. (1.57).

Footnotes to Chapter 1

1. Gabor [24].
2. Helstrom [1], p. 13.
3. Reed [25].
4. Mandel [26].
5. Blackman and Tukey [27].
6. Jenkins and Watts [28].
7. Helstrom [5].
8. Mandel and Wolf [29].
9. Helstrom [30].
10. Helstrom [31].
11. Papoulis [32], chapter 11.
12. Helstrom [12].
13. Born and Wolf [33], 484.

Chapter 2

Photon-Counting Statistics and Maximum-Likelihood Estimation of Object Radiance Function

In chapter 1, we have assumed that the measurable datum is the instantaneous intensity of the light field in the image plane. In reality, the instantaneous intensity of the image light field itself is not measurable, and it should interact with a recording medium from which we can take data.[1-6] The recording medium usually used is a photosensitive surface, for example, a photographic film or a photomultiplier tube. We will analyze an imaging system, as shown in Fig. 1.1. in which the light from the object plane is focused on a photosensitive surface from which it ejects photoelectrons. The surface is divided like a mosaic into a large number of small, insulated spots, from each of which the photocurrent can be measured. These measured values of the photocurrent during an observation interval T will constitute the data on which is to be based an estimate of the radiance distribution function of the object plane. When $WT \gg 1$, the numbers of photoelectrons emitted from each spot of the photosensitive recording surface are statistically independent with Poisson distributions whose mean values are proportional to the illuminance at the spots.[2-9]

In this model, the noise arising from the interaction between the light field and the recording medium predominates in causing the fluctuation of the measured data. This kind of noise, which is called shot noise,[10, 11] is signal-dependent. The linear filter for the

least -squares estimation of the object radiance function has the same form as we have derived in chapter 1, Eq. (1.57), except that one more noise term due to the shot noise has been introduced, and in normal circumstance the shot noise dominates. The statistics of the measured data are Poisson and as the noise is signal-dependent, the linear estimator is no longer optimum.[12] The principle of maximum likelihood leads to nonlinear equations for the estimates. The nonlinear estimator is signal-dependent and is asymptotically convergent to the linear estimator, when the received light is strong.

2.1. Photon-Counting Receiver Model and Its Statistics

The photosensitive recording surface in the image plane of Fig. 1.1 is shown in Fig. 2.1. The recording surface is divided into a mosaic of small, insulated spots of area S. The numbers of photoelectrons measured at each spot constitute the data. This receiver may be considered as a matrix of photomultiplier tubes.

The photon counting statistics have been previously discussed extensively in literature.[3-9] We want to present a very brief description of the situation that the incident signal in the image plane $\psi_i(\underset{\sim}{x}, t)$ has large time-bandwidth product, $WT \gg 1$.

To study the statistical properties of the number of photoelectrons, $n(\underset{\sim}{x}_i)$, emitted by each photomultiplier tube centered at $\underset{\sim}{x}_i$ of the image plane, we need to determine the probability $P_{\underset{\sim}{x}_i}(n; T)$ that exactly n photoelectrons are emitted from that photomultiplier tube during an observation interval of T.

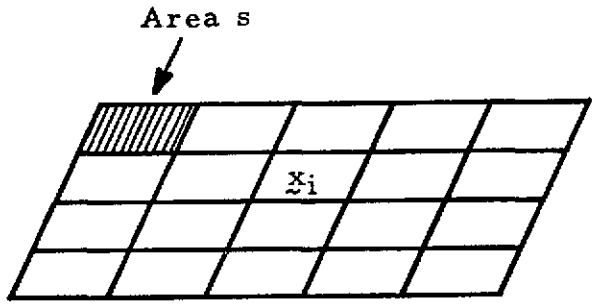


Figure 2.1. A photo-counting receiver model.

It seems reasonable to assume that the probability of emission of a photoelectron during a given interval is statistically independent of the number of photoelectrons emitted previously. We assume that we can write the probability of the emission of a single photoelectron in an infinitesimal interval Δt, $P_{\underset{\sim}{x}_i}(1; \Delta t)$, as

$$P_{\underset{\sim}{x}_i}(1; \Delta t) = \alpha s\, I_i(\underset{\sim}{x}_i, t)\, \Delta t \quad , \tag{2.1}$$

where

$$\alpha = \eta / h\bar{\nu} \quad , \tag{2.2}$$

and η is the quantum efficiency of the photodetector, h is Planck's constant, $\bar{\nu}$ is the mean frequency of the incident light. The intensity of the light field in the image plane, $I_i(x, t)$, as described in chapter 1, is

$$I_i(\underset{\sim}{x}, t) = \frac{1}{2} \left| \psi_i(\underset{\sim}{x}, t) \right|^2 \quad . \tag{2.3}$$

If we add the additional assumption of zero probability of multiple occurrence in the infinitesimal interval Δt $(\Delta t \to 0)$, we can write the conditional probability distribution function[3, 5, 7] of the emission of n photoelectrons in an observation interval of T,

$$P_{\underset{\sim}{x}_i}(n; T \,|\, I_i(\underset{\sim}{x}_i, t)) = \frac{1}{n!} [E(\underset{\sim}{x}_i; T)]^n \exp[-E(\underset{\sim}{x}_i; T)] \quad , \tag{2.4}$$

where

$$E(\underset{\sim}{x}; T) = \alpha s \int_0^T |I_i(\underset{\sim}{x}, t)|^2 \, dt$$

$$= \frac{1}{2} \alpha s \int_0^T |\psi_i(\underset{\sim}{x}, t)|^2 \, dt \ , \qquad (2.5)$$

is the energy incident on the detector of area s centered at $\underset{\sim}{x}$, during an observation interval of duration T.

The probability distribution function described in Eq. (2.4) is the conditional probability distribution function, conditional on a particular realization $I(\underset{\sim}{x}, t)$, $t \in [0, T]$, of a random process whose statistics have been discussed in section 1.3. From Eq. (2.4), the mean value of the number of photoelectrons from the spot centered at $\underset{\sim}{x}$ is

$$\langle n(\underset{\sim}{x}) \rangle = \langle E(\underset{\sim}{x}; T) \rangle$$

$$= \frac{1}{2} \alpha s \int_0^T \langle |\psi_i(\underset{\sim}{x}, t)|^2 \rangle \, dt$$

$$= \alpha s T \ \phi_i(\underset{\sim}{x}, \underset{\sim}{x}) \ , \qquad (2.6)$$

where $\phi_i(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$ is spatial covariance function of the image plane light field, given in Eq. (1.30).

The covariance function of the number of photoelectrons emitted from two different spots can be evaluated as in the semi-classical analysis of the Hanbury Brown-Twiss effect.[13] We use the

equality of the variance of the Poisson distribution of $n(\underset{\sim}{x}_i)$ with its mean value. The correlation function of $n(\underset{\sim}{x}_i)$ and $n(\underset{\sim}{x}_j)$ is

$$\langle n(\underset{\sim}{x}_i)\, n(\underset{\sim}{x}_j)\rangle$$

$$= \langle E(\underset{\sim}{x}_i;\, T)\, E(\underset{\sim}{x}_j;\, T)\rangle + \langle E(\underset{\sim}{x}_i;\, T)\rangle\, \delta_{i,j} \quad , \tag{2.7}$$

where δ_{ij} is the Kronecker delta. Combined with Eqs. (2.5), (2.6) and (2.7), the resulting covariance function is

$$\mathrm{Cov}\{ n(\underset{\sim}{x}_i),\ n(\underset{\sim}{x}_j)\}$$

$$= \langle n(\underset{\sim}{x}_i)\, n(\underset{\sim}{x}_j)\rangle - \langle n(\underset{\sim}{x}_i)\rangle\, \langle n(\underset{\sim}{x}_j)\rangle$$

$$= \mathrm{cov}\{ E(\underset{\sim}{x}_i;\, T),\ E(\underset{\sim}{x}_j;\, T)\} + \langle E(\underset{\sim}{x}_i;\, T)\rangle\, \delta_{ij}$$

$$= (\alpha s)^2 \int_0^T \int_0^T \left|\phi_i(\underset{\sim}{x}_i\ \underset{\sim}{x}_j;\, t_1, t_2)\right|^2 dt_1 dt_2 + \alpha s T\, \phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_i)\, \delta_{ij}$$

$$= (\alpha s)^2 \left|\phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_j)\right|^2 \int_0^T \int_0^T \left|\chi(t_1 - t_2)\right|^2 dt_1\, dt_2 + \alpha s T\, \phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_i)\, \delta_{ij}$$

$$= (\alpha s)^2\, T W^{-1} \left|\phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_j)\right|^2 + \alpha s\, T\, \phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_i)\, \delta_{ij} \quad , \tag{2.8}$$

where W is the bandwidth of the object light field as defined in Eq. (1.6). Comparing Eqs. (2.8) and (1.34), we find that the last term in Eq. (2.8) is an extra term arising from the interaction of the light field with the recording medium. The ratio of the first term to the second term in Eq. (2.8) is

$$
\begin{aligned}
&(\alpha s)\, W^{-1} \left|\phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_j)\right|^2 \left[\phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_i)\right]^{-1} \\
&\leq \frac{\alpha s}{W} \quad \phi_i(\underset{\sim}{x}_i, \underset{\sim}{x}_i) \\
&\approx \alpha s\, W^{-1}\, (\lambda^2/4\pi)\, \bar{B} \\
&= \lambda^2 \bar{B}\, \eta s / 4\pi W h\bar{\nu} \ll 1 \quad .
\end{aligned}
\tag{2.9}
$$

If we consider a scene illuminated by moonlight on a clear night,[1] we have B/W equal to 1.6×10^{-3} W $m^{-2} \mu^{-1}$ at $\lambda = 5150$ Å. The ratio has the value $0.8 \times 10^{-13}\, \eta s$ which will be further diminished by the reflectance of the scene. For illumination by full sunlight at the same wavelength, the ratio is larger than that of moonlight by factor of 10^6. The shot noise arising from the interaction of the light field with the recording medium is thus under most circumstances predominant in the covariance of $n(\underset{\sim}{x})$.

To find the unconditional probability distribution function of $n(\underset{\sim}{x}_i)$, we have to calculate the ensemble average of $P_{\underset{\sim}{x}_i}(n; T \,|\, I_i(\underset{\sim}{x}_i, t))$ with respect to $E(\underset{\sim}{x}_i; T)$. The mean value and variance of $E(\underset{\sim}{x}; T)$, can be calculated from Eqs. (1.36) and (1.38), are

$$
\langle E(\underset{\sim}{x}; T) \rangle = \alpha s T \langle I_i(\underset{\sim}{x}, t) \rangle \quad , \tag{2.10}
$$

and

$$
\mathrm{Var}\,[E(\underset{\sim}{x}; T)] = (\alpha s)^2\, T W^{-1} \left[\langle I_i(\underset{\sim}{x}, t) \rangle\right]^2 . \tag{2.11}
$$

The ratio of the variance to the square of the mean value of $E(\underset{\sim}{x}; T)$ is

$$\frac{\mathrm{Var}\,[E(\underset{\sim}{x}\,;\,T)]}{[\langle E(\underset{\sim}{x}\,;T)\rangle]^2} = (WT)^{-1} \quad . \tag{2.12}$$

If the time-bandwidth product WT is very large and $\langle E(\underset{\sim}{x}\,;\,T)\rangle$ is finite, then we can say that the random variable $E(\underset{\sim}{x}\,;\,T)$ is convergent to its mean in distribution, and

$$P(E(\underset{\sim}{x}\,;\,T)) \rightarrow \delta\,(E(\underset{\sim}{x}\,;\,T) - \langle E(\underset{\sim}{x}\,;\,T)\rangle) \quad . \tag{2.13}$$

This can be explained by the law of large numbers[14, 15]; $E(\underset{\sim}{x}\,;\,T)$ can be considered as a sum of WT independent identical random variables.[16, 17]

From the above argument, we conclude that the unconditional probability distribution is, by (2.4),

$$P_{\underset{\sim}{x}}(n;\,T) = P_{\underset{\sim}{x}}(n;\,T\,|\,\langle I_i(\underset{\sim}{x};\,t)\rangle)$$

$$= \frac{1}{n!}\,[\langle E(\underset{\sim}{x}\,;\,T)\rangle]^n \exp[-\langle E(\underset{\sim}{x}\,;\,T)\rangle] \quad , \tag{2.14}$$

if $WT \gg 1$.

Under the general circumstances the numbers of photoelectrons emitted from two different spots of the photosensitive recording medium are statistically independent and Poisson distributed, if we can neglect the classical fluctuation noise. It is clear from Eqs. (2.8) and (2.9) that the classical fluctuation noise is negligible. So the conditional joint probability distribution function of the numbers of photoelectrons $n_1, n_2, \ldots, n_M$ emitted from the spots centered at

$\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_M$ is

$$P(n_1, n_2, \ldots n_M;\ T \mid I_i(\underset{\sim}{x}_1, t),\ I_i(\underset{\sim}{x}_2, t), \ldots, I_i(\underset{\sim}{x}_M, t))$$

$$= \prod_{j=1}^{M} P_{\underset{\sim}{x}_i}(n_i;\ T \mid I_i(\underset{\sim}{x}_i;\ t)) \quad . \tag{2.15}$$

For $WT >> 1$, $E(\underset{\sim}{x}_i;\ T)$ converges to $\langle E(\underset{\sim}{x}_i;\ T)\rangle$ in probability, and the joint distribution of $E(\underset{\sim}{x}_i;\ T)$, $i = 1, \ldots, M$, will converge to $\prod_{i=1}^{M} \langle E(\underset{\sim}{x}_i;\ T)\rangle$ in probability, so that

$$P(n_1, n_2, \ldots, n_M; T) = \prod_{j=1}^{M} P_{\underset{\sim}{x}_j}(n_j;\ T \mid \langle I_i(\underset{\sim}{x}_i;\ t)\rangle)$$

$$= \prod_{j=1}^{M} \frac{1}{n_i!} [\langle E(\underset{\sim}{x}_i; T)\rangle]^n \exp[-\langle E(\underset{\sim}{x}_i;\ T)\rangle] . \tag{2.16}$$

The numbers of photoelectrons n_i emitted from the spots $\underset{\sim}{x}_i$ are statistically independent and Poisson distributed and have mean values

$$\langle n_i \rangle = \langle n(\underset{\sim}{x}_i)\rangle$$

$$= \langle E(\underset{\sim}{x}_i;\ T)\rangle \quad , \tag{2.17}$$

if $WT >> 1$.

2.2. Linear Least-Squares Estimation of Object Radiance Function Based on the Measured Numbers of Photoelectrons

We have derived a linear estimator for the object radiance function based on the measured image plane intensity in chapter 1. In this section we are going to derive a linear least-squares estimator based on the statistical properties of the measured data, the numbers of photoelectrons emitted from the spots of the image-plane recording surface.[1]

Just as we have divided the photosensitive recording surface into a mosaic of spots, we sample the object plane at a set of points $\underset{\sim}{u}_j = (u_{jx}, u_{jy})$, and the radiance $b(\underset{\sim}{u}_j)$ is to be estimated as a linear combination of the numbers of photoelectrons received,

$$\hat{b}_j = \hat{b}(u_j) = \sum_m L_{jm} n_m + C_o \quad , \tag{2.18}$$

where C_o is a known constant, and the weighting factors L_{jm} introduced by the processing matrix $\underset{\sim}{L}$ are selected to minimize the mean-square error

$$\begin{aligned} \mathcal{E} &= E_b \langle [\hat{b}(\underset{\sim}{u}) - b(\underset{\sim}{u})]^2 \rangle \\ &= \sum_j E_b \langle [\hat{b}(\underset{\sim}{u}_j) - b(\underset{\sim}{u}_j)]^2 \rangle \quad , \end{aligned} \tag{2.19}$$

where E_b denotes the ensemble average over random process $b(\underset{\sim}{u})$

The mean value of the estimator, from Eq. (2.18), (2.6), and (1.30), is

$$\langle b(\underset{\sim}{u}_j)\rangle = \sum_m L_{jm} \langle n_m \rangle + C$$

$$= C_o + \alpha s T \, \phi_i(\underset{\sim}{u}_j, \underset{\sim}{u}_j)$$

$$= C_o + \alpha s T (4\pi Z_i^2)^{-1} A \left\{ \bar{B} + (\lambda Z_o)^{-2} A \int_0 b(\underset{\sim}{v}) \left| \vartheta(\underset{\sim}{v} - \underset{\sim}{u}_m) \right|^2 d^2\underset{\sim}{v} \right\}$$

$$= Cs \sum_m L_{jm} \int b(\underset{\sim}{v}) \left| \vartheta(\underset{\sim}{v} - \underset{\sim}{u}_m) \right|^2 d^2\underset{\sim}{v} \quad , \tag{2.20}$$

where

$$C = \alpha T (4\pi Z_i^2)^{-1} (\lambda Z_o)^{-2} A^2 \quad , \tag{2.21}$$

and

$$C_o = -\alpha s T A (4\pi Z_i^2)^{-1} \bar{B} \sum_m L_{jm} \quad . \tag{2.22}$$

We assume that the area of each element in the image plane s is so small, and the spots are so close together, that the summation in Eqs. (2.20), and (2.22) can be approximated by an integral,

$$\langle \hat{b}(u) \rangle = C \iint L(\underset{\sim}{u} - \underset{\sim}{w}) \left| \vartheta(\underset{\sim}{w}, \underset{\sim}{v}) \right|^2 b(\underset{\sim}{v}) \, d^2\underset{\sim}{v} \, d^2\underset{\sim}{w} \quad ,$$

and

$$C_o = -\alpha T A (4\pi Z_i^2)^{-1} \bar{B} \int_I L(\underset{\sim}{v}) \, d^2\underset{\sim}{v}$$

$$= -\alpha T J_B \quad .$$

Equation (2.21) is identical to Eq. (1.40), except for a constant factor.

The variance of $\hat{b}(\underset{\sim}{u}_j)$ can be evaluated from Eqs. (2.18), (2.8), and (1.30), and is

$$\begin{aligned}
\text{Var}\,\hat{b}(\underset{\sim}{u}_j) &= \sum_{\ell}\sum_{m} L_{j\ell}\, L_{jm}\, \text{cov}\{n_\ell\,,\, n_m\} \\
&= \sum_{\ell}\sum_{m} L_{j\ell}\, L_{jm} (\alpha s)^2 \left|\phi_i(\underset{\sim}{x}_\ell\,, \underset{\sim}{x}_m)\right|^2 T/W + \sum_{\ell} L_{j\ell}^2\, \alpha s T\, \phi_i(\underset{\sim}{x}_\ell\,, \underset{\sim}{x}_\ell) \\
&\to \alpha^2 (T/W) \iint L(\underset{\sim}{u}_j - \underset{\sim}{v}_1)\, L(\underset{\sim}{u}_j - \underset{\sim}{v}_2) \left|\phi_i(\underset{\sim}{v}_1, \underset{\sim}{v}_2)\right|^2 d^2\underset{\sim}{v}_1\, d^2\underset{\sim}{v}_2 \\
&\quad + \alpha T \int L^2(\underset{\sim}{u}_j - \underset{\sim}{v})\, \phi_i(\underset{\sim}{v}, \underset{\sim}{v})\, d^2\underset{\sim}{v}
\end{aligned} \tag{2.23}$$

in the limit $s \to 0$. Comparing Eq. (2.23) and (1.49), we know that the last term in Eq. (2.23) is the extra term in the variance of the estimator when the measured data are the numbers of photoelectrons emitted by the spots of the photosensitive recording surface through the interaction of the light incident on the image plane with the recording medium.

Applying the same analysis as we have used in chapter 1, we find the linear filter which minimizes the mean-square error is

$$\Lambda(\underset{\sim}{r}) = C^{-1}(\lambda Z_o)^2\, A\, \Phi_b(\underset{\sim}{r})\, I_A^{(2)}(\underset{\sim}{r})\, D_1^{-1}(\underset{\sim}{r}) \,, \tag{2.24}$$

where

$$D_1(\underset{\sim}{r}) = D(\underset{\sim}{r}) + 4\pi \bar{B}\, (\lambda^2 \alpha\, T\, A)^{-1} \,. \tag{2.25}$$

and

$$D(\underset{\sim}{r}) = \Phi_b(\underset{\sim}{r})\,[I_A^{(2)}(\underset{\sim}{r})]^2 + A^{-1}(WT)^{-1}\left[\bar{B}^2\, I_A^{(2)}(\underset{\sim}{r}) + \int \Phi_b(\underset{\sim}{s})\, I_A^{(4)}(\underset{\sim}{r},\underset{\sim}{s})\, d^2s\right] .$$

The functions $I_A^{(2)}(\underset{\sim}{r})$ and $I_A^{(4)}(\underset{\sim}{r},\underset{\sim}{s})$ are defined by Eqs. (1.53) and (1.54) and A is the area of the aperture. The optical system that we have discussed here is assumed to be a diffraction-limited system.

Comparing Eqs. (2.24) and (1.57), we find that the only difference between those two linear spatial filters is the shot noise arising from the stochastic nature of the photoelectric emission. Under most circumstances, as we have discussed in section 2.1, the shot noise arising from the quantum nature of light and its interaction with recording medium is predominant in the noise spectral density of the linear filter in Eq. (2.24), so these two linear filters in Eq. (2.24) and (1.57) are basically very different. Equation (2.24) can be approximated as

$$\Lambda(\underset{\sim}{r}) = C^{-1}(\lambda Z_o)^2 A\, \Phi_b(\underset{\sim}{r})\, I_A^{(2)}(\underset{\sim}{r})\,\{\Phi_b(r)[I_A^{(2)}(\underset{\sim}{r})]^2 + 4\pi\bar{B}(\lambda^2 \alpha TA)^{-1}\}$$

$$= C_2\,[I_A^{(2)}(\underset{\sim}{r})]^{-1}\{1 + 4\pi\bar{B}(\lambda^2\alpha TA)^{-1}\,\Phi_b^{-1}(\underset{\sim}{r})\,[I_A^{(2)}(r)]^{-2}\} . \qquad (2.26)$$

The signal-to-noise ratio for this system is

$$\gamma = \lambda^2 \alpha T\,(4\pi\bar{B})^{-1}\,\Phi_b(\underset{\sim}{r})\,[I_A^{(2)}(\underset{\sim}{r})]^2 . \qquad (2.27)$$

2.3. Maximum-Likelihood Estimation

Under most circumstances the shot noise predominates in causing the fluctuations of the numbers n_i of photoelectrons emitted from the spots centered at $\underset{\sim}{x}_i$ of the recording medium; the noise arising from the fluctuation of the light field can be neglected. The object that we discussed radiates incoherently; it is the radiance distribution of the object that is of interest, not the light field it radiates. If the noise arising from the fluctuation of the light field can be neglected, then we can deal directly with the object radiance instead of the field which the object radiates. The transmission of the object radiance function to the image plane can be described as incoherent image formation. The illuminance in the image plane is linearly related to the radiance distribution in the object plane by

$$I(\underset{\sim}{x}) = \int_O K(\underset{\sim}{x}, \underset{\sim}{y})\, B(\underset{\sim}{y})\, d^2\underset{\sim}{y} \quad , \tag{2.28}$$

where $K(\underset{\sim}{x})$ is the incoherent point-spread function. For a diffraction-limited optical system

$$K(\underset{\sim}{x} - \underset{\sim}{y}) = |S(\underset{\sim}{x}, \underset{\sim}{y})|^2 \quad , \tag{2.29}$$

where $S(\underset{\sim}{x}, \underset{\sim}{y})$ is the coherent point-spread function and is defined in Eq. (1.23).

For large time-bandwidth product, when the fluctuation of the light field can be neglected, the numbers n_i are statistically independent, discrete random variables with Poisson distributions and

mean values given by

$$\langle n_i \mid B(\underset{\sim}{y}) \rangle = \alpha s T \; I(\underset{\sim}{x}_i) \quad , \tag{2.30}$$

as has been shown in section 2.1. Under these Poisson statistics, the shot noise predominates in causing random variation of the observed data. The shot noise is signal-dependent. Linear least-squares estimation, which is optimum for the situation when the signal is corrupted by signal-independent additive Gaussian noise, is inadequate for the restoration of images degraded by non-Gaussian-distributed, signal-dependent noise.[1] We will derive a nonlinear estimation scheme based on the principle of maximum likelihood.[18, 19] A nonlinear estimator has to be solved iteratively. The estimator is signal-dependent and varies in each step of the iteration. A digital computer will be used, and we will derive our estimator in digital fashion.

Sampling the object plane at M points U_j, we replace the integral by a sum and write for the mean numbers of photoelectrons,

$$\langle n_j \mid \{ b_i \} \rangle = \alpha s T \; (I_o + V_j) \tag{2.31}$$

where

$$V_i = \sum_{j=1}^{m} K_{ij} b_j \quad , \quad i = 1, 2, \ldots, N \tag{2.32}$$

$$b_j = b(\underset{\sim}{u}_j)$$

$$I_o = B_o \sum_{j=1}^{M} K_{ij} \quad , \quad \forall \; i \quad . \tag{2.33}$$

Here I_o is a uniform, known illuminance, to which background light may contribute, the b_j's are the quantities to be estimated on the basis of the observed numbers n_i of photoelectrons, and K_{ij} is a point-spread matrix derived from the point-spread function $K(\underset{\sim}{x}, y)$. Equations (2.31) and (2.32) can be written in matrix form

$$\langle \underset{\sim}{n} \,|\, \underset{\sim}{b} \rangle = \alpha s T \, (I_o \, \underset{\sim}{1} + \underset{\sim}{V}) \quad , \tag{2.34}$$

and

$$\underset{\sim}{V} = \underset{\sim}{K} \, \underset{\sim}{b} \quad , \tag{2.35}$$

where $\underset{\sim}{n}$ is a $(N \times 1)$ column vector with elements n_i, $\underset{\sim}{V}$ is a $(N \times 1)$ column vector with elements V_i, K is a $(N \times M)$ matrix with elements K_{ij} and $\underset{\sim}{1}$ is the column vector of 1's.

The joint probability distribution function of n_i's is

$$P(\{n_i\} \,|\, \{b_i\}) = \prod_{i=1}^{N} \frac{[\alpha s T \, (I_o + V_i)]^{n_i}}{n_i !} \exp[-\alpha s T \, (I_o + V_i)] \, . \tag{2.36}$$

This joint probability distribution function is conditional on a particular realization $b(\underset{\sim}{u})$, which is a spatially stationary random process with zero mean value and autocovariance function $\phi_b(\underset{\sim}{u})$. The unconditional mean value vector and covariance matrix are

$$\langle \underset{\sim}{n} \rangle = E_b \langle \underset{\sim}{n} \,|\, \underset{\sim}{b} \rangle = \alpha s T \, I_o \, \underset{\sim}{1} \, , \tag{2.37}$$

and

$$\langle (\underset{\sim}{n} - \langle \underset{\sim}{n} \rangle)\ (\underset{\sim}{n} - \langle \underset{\sim}{n} \rangle)^T \rangle$$

$$= (\alpha s T)^2\ \underset{\sim}{K}\ \underset{\sim}{\phi}_b\ \underset{\sim}{K}^T + \alpha s T\ I_o\ \underset{\sim}{I}\ , \tag{2.38}$$

where $\underset{\sim}{I}$ is the identity matrix, and $\underset{\sim}{\phi}_b$ is the covariance matrix $\phi_b = \| \phi_{ij} \|$ with

$$\phi_{ij} = \phi_b (\underset{\sim}{u}_i - \underset{\sim}{u}_j)\ , \tag{2.39}$$

the covariance of the radiance at point $\underset{\sim}{u}_i$ and $\underset{\sim}{u}_j$. From Eq. (2.38), we can see that the first term in the right-hand side is related to the signal and the second term is related to the noise. The signal-to-noise ratio is equivalent to

$$R_S = \alpha s T\ (\sigma_b^2 / I_o)\ . \tag{2.40}$$

The linear estimate, $\hat{\underset{\sim}{b}}$, of $\underset{\sim}{b}$ based on the measured data $\underset{\sim}{n}$ can be easily derived from Eqs. (2.37) and (2.38) and is

$$\hat{\underset{\sim}{b}} = \left(\underset{\sim}{\phi}_b\ \underset{\sim}{K}^T \underset{\sim}{K} + \frac{I_o}{\alpha s T}\ \underset{\sim}{I} \right)^{-1} \underset{\sim}{\phi}_b\ \underset{\sim}{K}^T \left(\frac{1}{\alpha s T}\ \underset{\sim}{n} - I_o\ \underset{\sim}{1} \right)\ . \tag{2.41}$$

The linear estimate $\hat{\underset{\sim}{b}}$ given by Eq. (2.41) is identical to the one given by Eq. (2.24) if the noise arising from the light field fluctuation can be neglected.

In using the method of maximum likelihood, which requires the knowledge of the probability distribution of the estimate, we have already assumed in applying linear estimation that the object radiance function $B(\underset{\sim}{u})$ is a spatially stationary random process with mean value

B_o and autocovariance function $\phi_b(\underset{\sim}{u})$. Here for simplicity we will assume that it is Gaussian distributed. The width of $\phi_b(\underset{\sim}{u})$ as a function of $\underset{\sim}{u}$ represents the size of typical finer details in the object plane, and the ratio $\phi_b(0)/B_o^2$ specifies a mean-square contrast. The size of the typical finer details and the average contrast are the most fundamental (and necessary) information to specify a scene. We are not asserting that each object radiance distribution to be estimated will actually look like the chaotic function by which we commonly picture a Gaussian process. Our assumption is rather that a method that works well for such an ensemble of objects will work well for most objects of similar structure, and structure can seldom be specified in terms less primitive than the size of typical details and the mean-square contrast.[1,20,21]

The joint probability density function[22] of the estimanda b_j is then

$$P(\{b_j\}) = M_o \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mu_{ij}b_i b_j\right), \tag{2.42}$$

where M_o is a normalization constant and the μ_{ij} are elements of a matrix

$$\underset{\sim}{\mu} = \underset{\sim}{\phi}_b^{-1} \tag{2.43}$$

inverse to the covariance matrix $\underset{\sim}{\phi}_b$.

In the absence of a scene in the object plane, $b(\underset{\sim}{u}) = 0$, the illuminance in the image plane is the uniform illuminance, and the joint

probability distribution function of the numbers n_i of photoelectrons is

$$P_o(\{n_i\}) = \prod_{i=1}^{N} \frac{(\alpha s T I_o)^{n_i}}{n_i!} \exp(-\alpha s T I_o) \ . \qquad (2.44)$$

The likelihood ratio of the numbers n_i, i.e., the quotient of their joint probabilities in the presence and absence of an image, is then

$$\Lambda(\{n_i\}|\{b_i\}) = \prod_{i=1}^{N} \left(\frac{I_o + V_i}{I_o} \right)^{n_i} \exp(-\alpha s T V_i) \ . \qquad (2.45)$$

The maximum-likelihood estimate of the true image is given by the set of values b_i for which

$$\Lambda(\{n_i\}) = \Lambda(\{n_i\} \mid \{b_i\}) \, P(\{b_i\}) \qquad (2.46)$$

is maximum. Maximizing $\Lambda(\{n_i\})$ is the same as maximizing its logarithm, since the logarithm is a monotonic function. The maximum-likelihood estimates of the b_j's are the set of values which minimize

$$F(\{b_j\}) = -\ln[\Lambda(\{n_i\})]$$

$$= -\sum_{i=1}^{N} [n_i \ln(I_o + V_i) - \alpha s T (I_o + V_i)] + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \mu_{jk} b_j b_k \ , \qquad (2.47)$$

from which terms independent of the estimanda $\{b_j\}$ have been dropped.

The set of values $\{b_j\}$ which minimize the negative logarithm of the likelihood ratio, $F(\{b_j\})$, can be obtained by solving the set of equations resulting from setting the derivatives of F with respect to each b_j to zero. Differentiating F with respect to b_j, we have

$$(\nabla F)_j = \frac{\partial F}{\partial b_j}$$

$$= \sum_{i=1}^{N} \left(\frac{n_i k_{ij}}{I_o + V_i} - \alpha s T k_{ij} \right) - \sum_{m=1}^{M} \mu_{jm} b_m = 0 \quad . \tag{2.48}$$

Solving these equations for b_m, we obtain

$$b_m = \sum_{j=1}^{M} \sum_{i=1}^{N} \varphi_{mj} k_{ij} \left(\frac{n_i}{I_o + V_i} - \alpha s T \right) \quad , \tag{2.49}$$

where

$$V_i = \sum_{j=1}^{M} k_{ij} b_j \quad .$$

Equation (2.49) can be rewritten by using

$$\frac{1}{I_o + V_i} = \frac{1}{I_o} - \frac{V_i}{I_o^2} + \frac{V_i^2}{I_o^2 (I_o + V_i)}$$

and becomes

$$\sum_{j=1}^{M} \left[\delta_{jm} + \sum_{i=1}^{N} k'_{mi} k_{ij} \frac{n_i}{I_o^2} \right] b_j$$

$$= \sum_{i=1}^{N} \left(\frac{n_i}{I_o} - \alpha s T \right) + \sum_{i=1}^{N} k'_{mi} \frac{n_i V_i^2}{I_o^2 (I_o + V_i)} \quad , \tag{2.50}$$

where

$$k'_{mi} = \sum_{j=1}^{M} \phi_{mj} k_{ij} \quad . \tag{2.51}$$

We can write Eq. (2.50) in matrix form,

$$\hat{\underset{\sim}{b}} = \underset{\sim}{A}^{-1} \underset{\sim}{Y} + \underset{\sim}{A}^{-1} \underset{\sim}{B}(\hat{\underset{\sim}{b}}) \quad , \tag{2.52}$$

where

$\underset{\sim}{A}$ is a $M \times M$ matrix with elements

$$a_{mj} = \delta_{jm} + \sum_{i=1}^{N} k'_{mi} k_{ij} \frac{n_i}{I_o^2} \quad , \tag{2.53}$$

$\underset{\sim}{Y}$ is a $M \times 1$ column vector with elements

$$y_m = \sum_{i=1}^{N} k'_{mi} \left(\frac{n_i}{I_o} - \alpha s T \right) \quad , \tag{2.54}$$

and

$\underset{\sim}{B}(\hat{\underset{\sim}{b}})$ is a $M \times 1$ column vector with elements

$$B_m = \sum_{i=1}^{N} k'_{mi} \frac{n_i V_i^2}{I_o^2 (I_o + V_i)} \quad . \tag{2.55}$$

Examining matrices $\underset{\sim}{A}$, $\underset{\sim}{B}$, and $\underset{\sim}{Y}$, we find both $\underset{\sim}{A}$ and $\underset{\sim}{Y}$ matrices whose elements depend on the data but not the estimanda, and the $\underset{\sim}{B}$ matrix depends on the received data and the estimanda. Matrix $\underset{\sim}{Y}$ can be written as

$$\underset{\sim}{Y} = \underset{\sim}{\phi}_b \underset{\sim}{K}^T \left(\frac{1}{I_o} \underset{\sim}{n} - \alpha s T \underset{\sim}{1} \right) , \tag{2.55}$$

and is very similar to part of the linear estimator in Eq. (2.41). Equation (2.52) can be solved by a perturbation method. The procedure starts with the linear least-squares estimate of the b_j's, which is given by Eq. (2.41). Matrices $\underset{\sim}{A}$ and $\underset{\sim}{Y}$ are fixed for given the measured data; only matrix $\underset{\sim}{B}$ is changing in each step of iteration.

2.4. Comparison of Maximum-Likelihood Estimation and Linear Least-Squares Estimation

The linear least-squares estimator given by Eq. (1.41) is an estimator independent of the estimanda, while the nonlinear estimator is a function of the observed data and the estimanda because of the Poisson distribution of the observed data.

Equation (2.49) can be rewritten by using

$$\frac{1}{I_o + V_i} = \frac{1}{I_o} - \frac{V_i}{I_o (I_o + V_i)}$$

and becomes

$$b_m + \sum_{i=1}^{N} k'_{mi} \frac{n_i V_i}{I_o(I_o+V_i)} = \sum_{i=1}^{N} k'_{mi} \left(\frac{n_i}{I_o} - \alpha s T \right) . \qquad (2.56)$$

If $I_o + V_i$ is very large, then by the law of large numbers, $n_i/(I_o+V_i)$ will converge to $\alpha s T$, and Eq. (2.56) can be approximated as

$$b_m + \sum_{i=1}^{N} k'_{mi} \frac{\alpha s T V_i}{I_o} = \sum_{i=1}^{N} k'_{mi} \left(\frac{n_i}{I_o} - \alpha s T \right), \qquad (2.57)$$

and can be reduced to

$$\left(\underset{\sim}{\phi}_b \underset{\sim}{K}^T \underset{\sim}{K} + \frac{I_o}{\alpha s T} \underset{\sim}{I} \right) \underset{\sim}{b} = \underset{\sim}{\phi}_b \underset{\sim}{K}^T \left(\frac{1}{\alpha s T} \underset{\sim}{n} - I_o \underset{\sim}{1} \right)$$

which is the linear least-square estimate. So for a strong object radiance, the nonlinear estimator will converge to the linear estimator. This can be explained by the fact that the Poisson distribution with large mean value can be well approximated by a Gaussian distribution.

2.5. The Method of Steepest Descents

To find the minimum of $F(\{b_i\})$ as a function of $\{b_i\}$, the method of steepest descents can be used. We will treat the estimanda $\{b_j\}$ as coordinates of a point in an M-dimensional space; points for which $F(\{b_j\})$ is constant make up a level surface of the function to be minimized.

Starting at a trial point $\{b_j^o\}$, one will get to the bottom by always moving in the direction of the steepest slope. This direction is specified by the gradient

$$(\nabla F)_j = \frac{\partial F}{\partial b_j} \quad . \tag{2.58}$$

For point $\{\hat{b}_j\}$ to be a minimum (or maximum) of $F(\{b_j\})$, $(\nabla F)_j$ has to be equal to zero for all components, as is shown in Eq. (2.48). The component $(\nabla F)_j$ is equal to the 'error' by which the j-th equation in Eq. (2.48) remains unsatisfied.

In the neighborhood of the trial point, $F(\{b_j\})$ is expanded as a multivariate Taylor series, carried through quadratic terms. The next trial point is then

$$b_j^1 = b_j^0 - \lambda (\nabla F)_j \quad , \qquad j = 1, 2, \ldots, M \, , \tag{2.59}$$

where the number λ is chosen to minimize the quadratic form

$$F(\{b_j^1\}) = F(\{b_j^0\}) + \sum_{i=1}^{M} e_i (b_i^1 - b_i^0) + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} C_{ij} (b_i^1 - b_i^0)(b_j^1 - b_j^0) \tag{2.60}$$

approximating F in the neighborhood of $\underset{\sim}{b}^o$, where

$$e_i = \left. \frac{\partial F}{\partial b_i} \right|_{\underset{\sim}{b} = \underset{\sim}{b}^o} = (\nabla F)_i \quad ,$$

and

$$c_{ij} = \left. \frac{\partial^2 F}{\partial b_i \partial b_j} \right|_{\underset{\sim}{b} = \underset{\sim}{b}^o} .$$

The λ minimizing Eq. (2.60) is

$$\lambda = \sum_{i=1}^{M} e_i^2 \Big/ \sum_{i=1}^{M} \sum_{j=1}^{M} e_i c_{ij} e_j \quad . \tag{2.61}$$

This procedure is repeated until the values of $\underset{\sim}{b}$ cease to change significantly.

The linear least-squares estimate is usually used as the initial starting point. As radiance is a positive quantity, we want $B_o + b_i \geq 0$, $\forall$ i, and at each stage of the iteration we therefore replace any estimate b_i that is less than $-B_o$ by $-B_o + \epsilon$, where ϵ is a very small positive constant.

Footnotes to Chapter 2

1. Helstrom [12].
2. Helstrom [36].
3. Bar-David [37].
4. Karp and Clark [38].
5. Mandel [39].
6. Helstrom [40].
7. Helstrom [41].
8. Mandel [42].
9. Mandel and Wolf [29].
10. Davenport and Root [43], chapter 7.
11. Pazen [44], chapter 4.
12. Papoulis [32], p. 324.
13. Hanbury-Brown and Twiss [45].
14. Breiman [46], p. 188.
15. Gnedenko and Kolmogorov [47], chapter 3.
16. Helstrom [48].
17. Gallager [49], p. 405.
18. Helstrom [1], p. 256.
19. Van Trees [2], p. 65.
20. Helstrom [5].
21. Helstrom [50].
22. Helstrom [1], p. 50.

Chapter 3

Restoration of Images Based on the Measured Photoelectron Counts - Computer Simulation

The maximum-likelihood estimator of the object radiance function based on the measured numbers n_i of photoelectrons is highly nonlinear. Because of the lack of appropriate mathematical tools, it is difficult to analyze and thus appreciate the nonlinear estimation scheme directly from the estimator itself as shown in Eqs. (2.49) and (2.52). As a result of this that solution tends to be highly specialized. Both the nature of the signal and the nature of the noise affect the validity and usefulness of the estimator. We have tried to make our nonlinear estimator valid for a broad class of signals by specifying only its average contrast, the size of typical finer details, and its Gaussian probability distribution. The significance of the nonlinear estimator could be best demonstrated by computer simulations, which were performed on the Burroughs 6700 computer. Two particular optical system functions considered were spatially uniform motion blur and diffraction-limited image formation. The effects of various parameters were considered.

3.1. Generation of the Object Radiance Function and the Image Plane Illuminance Distribution

In the last chapter, in the derivation of both the linear and maximum-likelihood estimators, we have assumed that the object

radiance function $b(\underset{\sim}{u})$ is a stationary spatial Gaussian random process with mean zero and covariance function $\phi_b(\underset{\sim}{u})$. For mathematical simplicity, we will assume that the object radiance function is generated by a first-order autoregressive process,[1]

$$b_{i+1} = r\, b_i + z_{i+1}, \qquad i = 1, 2, \ldots, M-1 , \tag{3.1}$$

where $\{ z_i \}$ are mutually independent Gaussian random variables with zero mean and variance σ^2. The object has been assumed to be one-dimensional. The constant r is the correlation coefficient of two neighboring b_i's and has absolute value less than 1. It determines the size of typical details of the object radiance function. The larger r is, the smoother the object radiance function. The resulting process $\{ b_i \}$ has a covariance matrix $\underset{\sim}{\phi}_b$ with elements

$$\phi_{ij} = \sigma_b^2 \, r^{|i-j|} , \tag{3.2}$$

where

$$\sigma_b^2 = \frac{\sigma^2}{1 - r^2} \tag{3.3}$$

is the variance of the radiance function. σ_b^2 represents the power of the signal.

To generate identically distributed independent Gaussian random variables $\{ z_i \}$, we first generate two sets of independent random variables $\{ R_{1i} \}$ and $\{ R_{2i} \}$ uniformly distributed between zero and one. By using the random variable transformation

$$z_i = \sigma \cdot [-2 \ln (R_{1i})]^{1/2} \cos(2\pi R_{2i}) , \qquad (3.4)$$

the set of values $\{ z_i \}$ will be identically distributed independent Gaussian random variables with zero mean and variance σ^2.[2]

After we generate the set of values $\{ b_i \}$, we add the uniform background B_o to it to form the total radiance function $\{ B_i \} = \{B_o + b_i\}$. The positivity constraint on $\{B_i\}$ applies here; negative values of B_i will be replaced by zero. In other words, if $b_i < -B_o$, then we set $b_i = -B_o$. A typical object $\{ B_i \}$ is shown as the heavy line in Fig. 3.1. Since this hard-limited nonlinear transformation has been applied in the generation of the scene $\{ b_i \}$, the object is not necessarily a stationary Gaussian random process as described by Eq. (3.1). However, our assumption that the object radiance function is a stationary Gaussian random process is mainly for mathematical tractability. It is most accurate when the contrast is low. The important parameters related to the object radiance function are r, σ_b, and B_o; r defines the size of typical details of the object, B_o defines the uniform background radiance, and σ_b defines the mean deviation of the radiance from its mean value B_o. The value of $\sigma_b^2/(B_o^2 + \sigma_b^2)$ is the mean-square contrast of the object radiance function, and $(\alpha s T)\sigma_b^2/B_o$ is the signal-to-noise ratio as shown in Eq. (2.40).

After we have generated the true image vector $\underset{\sim}{B}$, we can calculate the degraded image

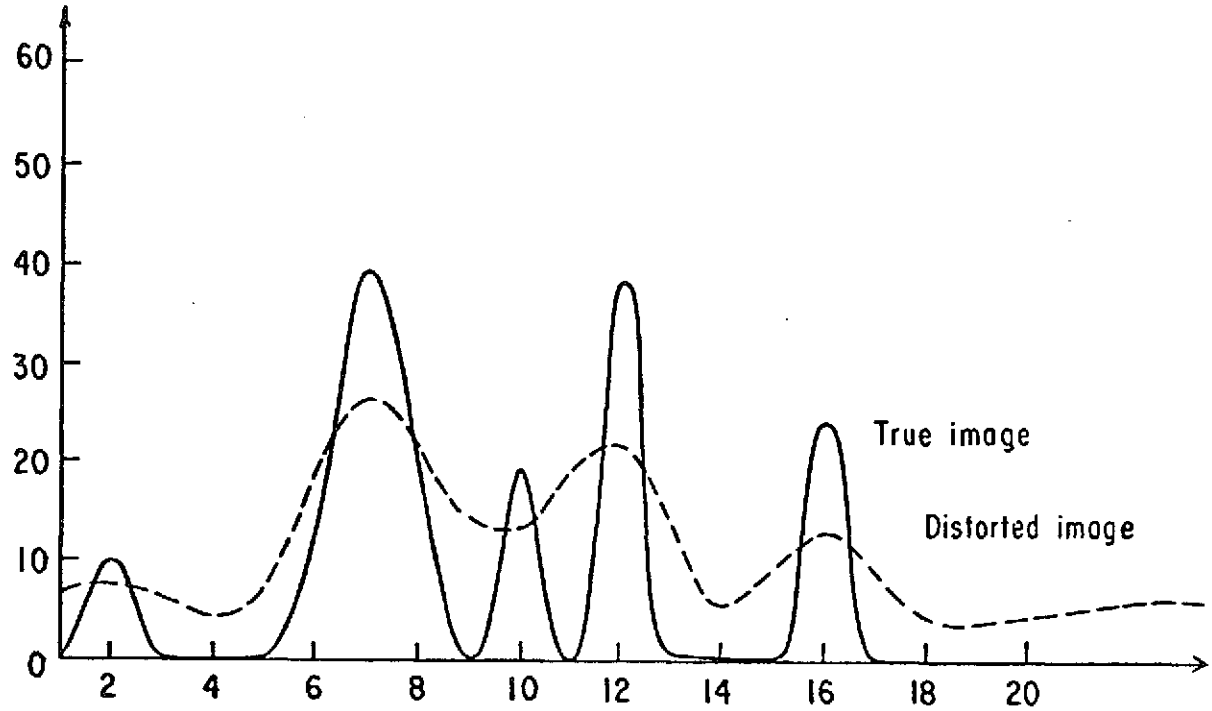


Figure 3.1. Original object (thick line) and its noisy version (thin line) as recorded after blurring by relative motion. The object and data are generated by using random number initiators (347, 5412, 3298, 9547, 2486) and parameters (αs T) = 10, I_o = 3, σ = 25 and r = 0.6.

$$\begin{aligned} \underset{\sim}{I}_i &= I_o \underset{\sim}{1} + \underset{\sim}{V} \\ &= I_o \underset{\sim}{1} + \underset{\sim}{K}\,\underset{\sim}{b} \quad . \qquad (3.5) \end{aligned}$$

Two incoherent point-spread functions will be used, one for motion blur and the other for a diffraction-limited imaging system.

The motion-blur image degradation arises from the relative motion of the object plane and the image recording plane during the recording process. If the imaging system is perfect, and the relative distance of motion between the image plane and the object plane during the interval of recording time T is Δ, then the incoherent point-spread function is$^{(3)}$

$$K(u) = \begin{cases} \Delta^{-1} \; , & 0 \le u \le \Delta \\ 0 \; , & \text{otherwise} \; . \end{cases} \qquad (3.6)$$

Here we have assumed that the relative motion is of constant speed. In the digital system, if Δ is equivalent to D sampling points, then the incoherent point-spread matrix can be rewritten as

$$\underset{\sim}{K} = \frac{1}{D} \times \underbrace{\begin{bmatrix} 1 & 0 & & & 0 \\ 1 & 1 & \ddots & & \\ \vdots & 1 & \ddots & \ddots & \\ 1 & \vdots & \ddots & \ddots & \\ & 1 & \ddots & \ddots & 1 \\ & & \ddots & \ddots & 1 \\ & 0 & & \ddots & \vdots \\ & & & \ddots & 1 \end{bmatrix}}_{M} \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} N = M + D - 1 \qquad (3.7)$$

D

M is the number of points in the object plane and N is the number of points in the image plane.

We have assumed that the image recording plane is large enough, and $N = M + D - 1$, so that all the light emitted by the object has been available for restoration. A close examination of the matrix $\underset{\sim}{K}$ has shown that the sums $\sum_j K_{ij}$ for different indices i are not all equal to each other. This seems to contradict the assumption shown in Eq. (2.33). The fact of the matter is that we really have to extend the $\underset{\sim}{K}$ matrix into

$$\underset{\sim}{K}' = \frac{1}{D} \times \begin{bmatrix} \overleftarrow{\quad D \quad} & & \\ 1 \;\; 1 \;\; \cdots \;\; 1 & & \\ 1 \;\; 1 \;\; \cdots \;\; 1 & 0 & \\ \ddots & \ddots & \\ 0 & 1 \;\; 1 \;\; \cdots \;\; 1 & \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \uparrow \\ \\ \\ \\ \downarrow \end{matrix} N = M + D - 1$$

$$\overleftarrow{\qquad\qquad M + 2D - 2 \qquad\qquad} \tag{3.8}$$

and the vector $\underset{\sim}{b}$ has to be modified into a $(M + 2D - 2)$ dimensional vector $\underset{\sim}{b}'$ with elements

$$b_i' = \begin{cases} b_{i+D-1}, & D \le i \le M + D - 1 \\ 0, & \text{otherwise} \end{cases} . \tag{3.9}$$

The matrix $\underset{\sim}{K}'$ has $\sum_j K_{ij}$ equal to a constant for all different indices i. Since we have the prior information that $b_i' = 0$ for $1 \le i \le D - 1$,

and $M + D \leq i \leq M + 2D - 2$, the whole system can be fully represented by the vector $\underset{\sim}{b}$ and the matrix $\underset{\sim}{K}$.

When the optical imaging system is diffraction-limited, the incoherent point-spread function[4, 5] is

$$K(x - y) = c \left\{ \frac{\sin [W_o(x - y)]}{(x - y)} \right\}^2 , \tag{3.10}$$

where c is a constant, and W_o is the effective width of the aperture of the optical system, or the spatial bandwidth of the optical system. In matrix form $K(x - y)$ can be written as

$$\begin{aligned} K_{ij} &= K_{ji} \\ &= K_{i-j} \\ &= c \left\{ \frac{\sin [W_o(i - j)]}{i - j} \right\}^2 . \end{aligned} \tag{3.11}$$

In the computer simulation we have truncated K_{i-j} at some integer J, so that $K_j = 0$ for $|j| > J$; then the matrix has the form

$$\underset{\sim}{K} = \begin{bmatrix} e_0 & e_1 & \cdots & e_J & & 0 \\ e_1 & e_0 & e_1 & & & \\ \vdots & & & & & e_J \\ e_J & & & & & \vdots \\ & & & & & e_1 \\ 0 & & & e_J & \cdots & e_1 \quad e_0 \end{bmatrix} , \qquad (3.12)$$

where $e_i = e_{-i} = c \left[\frac{\sin(W_o i)}{i} \right]^2 , \quad i = 0, 1, 2, \ldots, J \qquad (3.13)$

and

$$c = \left(\sum_{i=-J}^{J} e_i \right)^{-1} \qquad (3.14)$$

is the normalization constant.

The constant W_o is the spatial bandwidth of the diffraction-limited optical system. The value of $4/W_o$ is equivalent to the number of cells over which the light from each object point will be spread in the image plane. For example, if $W = 0.8$, then $4/W_o = 5$. The elements K_{ij} of matrix $\underset{\sim}{K}$ are $(e_0, e_1 = e_{-1}, e_2 = e_{-2}, \ldots) = (0.268, 0.215, 0.104, 0.021, \ldots)$. We can see that only $e_0, e_1, e_{-1}, e_2, e_{-2}$ are of significance. For $W = 1.2$, $4/W = 3.33$, and $(e_0, e_1, e_2, \ldots) = (0.395, 0.238, 0.031, \ldots)$, so that only e_0, e_1, e_{-1} are of significance.

A typical input object and its image distorted by a diffraction-limited optical system are shown in Fig. 3.1. Note that the small sharp peak between two big peaks has been completely smeared out.

If we assumed that the measurable data are the instantaneous illuminances at points in the image plane, the linear filter for the estimation of the object radiance function is shown in Eq. (1.57). It has been shown that none of the object spectrum outside the passband is recoverable from Eq. (1.57), and the small peak in Fig. 3.1 is not restorable, as will be shown in the computer experiment results.

The argument for modifiying the vector $\underset{\sim}{b}$ and the point-spread matrix $\underset{\sim}{K}$ in the case of motion blur in order to keep $\sum_j K_{ij}$ equal to a constant for all different indices i can be extended to the situation of a diffraction-limited optical system. In motion-blur image degradation, the light from the object plane has been all used in the restoration of the object, but in diffraction-limited image degradation, some light emitted by the object has been wasted. From Eq. (3.12) we can see that the light emitted by b_1 has been wasted, but all the light emitted by b_J has been collected.

3.2. Poisson-Distributed Random Number Generation

The numbers n_i of photoelectrons emitted from the spots of the photoelectrically emissive surface are statistically independent and Poisson distributed with mean values $\lambda_i = \alpha s T (I_o + V_i)$.

After we have generated the image illuminance function $\{ I_o + V_i \}$, we specify the parameter $\alpha s T$ and then start to generate the Poisson-distributed integers m having mean values λ_i. The probability of receiving m photoelectrons is

$$\Pr [m \mid \lambda_i] = \frac{\lambda_i^m}{m!} e^{-\lambda_i} \quad . \tag{3.15}$$

Let $\{ Y_i \}$ be a set of independent random variables uniformly distributed over the interval $[0, 1]$. The number of photoelectrons emitted by the photomultiplier tube centered at point x_i is n_i whenever

$$P_{n_i} \leq Y_i < P_{n_i + 1} \quad , \tag{3.16}$$

where

$$P_n = \Pr \{ j < n \mid \lambda \} = e^{-\lambda} \sum_{j=0}^{n-1} \frac{\lambda^j}{j!} \tag{3.17}$$

is the cumulative Poisson distribution with mean λ. The probability that Y_i will lie in the interval described by Eq. (3.16), and hence n_i photoelectrons will be emitted, is precisely given by Eq. (3.15).

For computation purposes, we rewrite Eq. (3.16) as

$$Q_{n_i - 1} = \sum_{j=0}^{n_i - 1} \lambda_i^j / j! \leq Y_i e^{\lambda_i} < \sum_{j=0}^{n_i} \lambda_i^j / j! = Q_{n_i} \quad . \tag{3.18}$$

We generate the set of numbers $\{ Y_i \}$ and then form $\{ Y_i e^{\lambda_i} \}$, and

begin to generate the sum $\{ Q_{n_i} \}$, starting with $n_i = 0$, $Q_o = 1$, and adding each time $\lambda^n/n!$. The increment $\lambda^n/n!$ can be generated recursively by multiplying the previous value by a factor $\lambda/(n+1)$. As soon as Q_n exceeds $Y_i e^{\lambda_i}$, the computer registers the number n as the number of photoelectrons received.

When the mean value λ_i of the Poisson distribution is large, the Poisson distribution can be well approximated by a discrete version of a Gaussian distribution with mean value λ_i and variance λ_i. Therefore, n_i can be generated by generating a random variable z_i with zero mean and variance λ_i, and then calculating $\lambda_i + Z_i$ and rounding it off to the nearest integer. The Gaussian random variable z_i can be generated by the method described in Eq. (3.4). Two sets of independent random variables $\{ R_3^i \}$ and $\{ R_4^i \}$ uniformly distributed between zero and one were generated and substituted into Eq. (3.4) to generate z_i. We have used $\lambda = 25$ as the threshold for the use of Gaussian distribution.

The generated data, reduced to the scale of the image by dividing by $(\alpha s T)$ are shown as the thin line in Fig. 3.2.

3.3. Methods of Nonlinear Programming

The equations for the maximum-likelihood estimate of the object radiance function $\{ b_i \}$ are highly nonlinear with linear constraints. We can either solve Eq. (2.49) (equivalently Eq. (2.52)) or minimize Eq. (2.47) to obtain the estimates $\{ \hat{b}_i \}$ of $\{ b_i \}$. The set of values

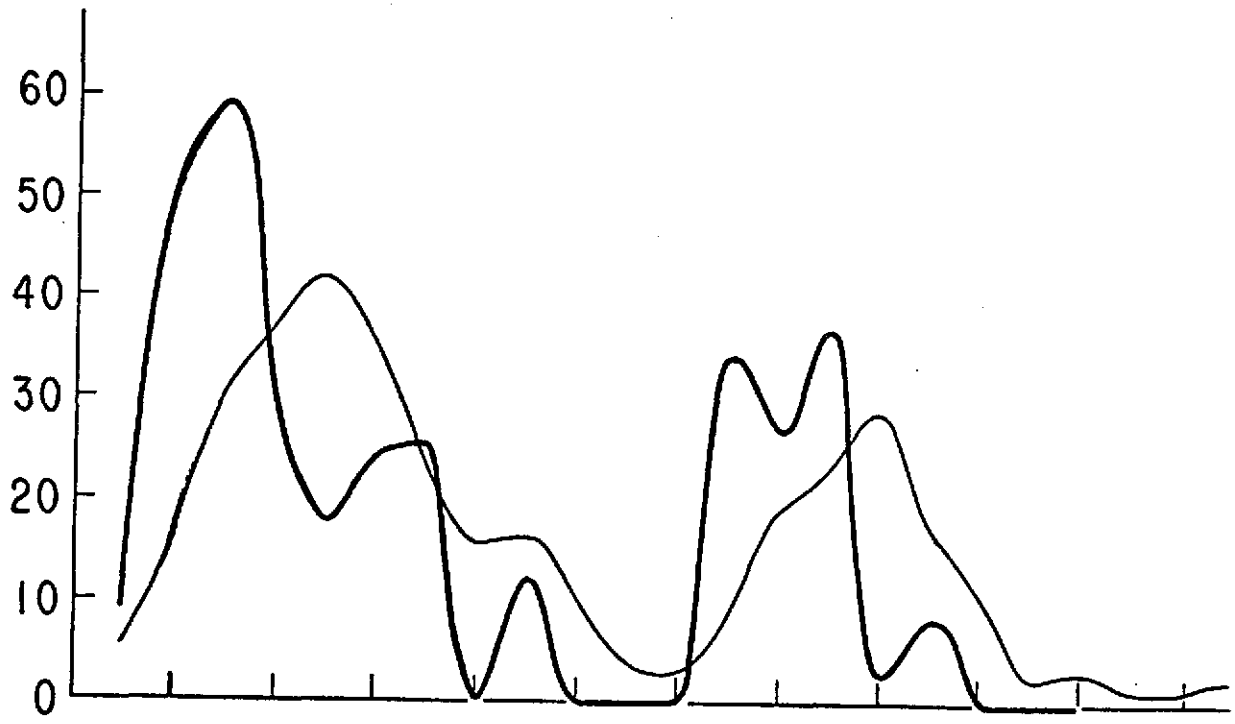


Figure 3.2. Original object (solid line) and its degraded image (dashed line) after distortion by a diffraction-limited optical system. The object and data are generated by using random number initiators (5431, 3214, 5438, 3654, 2495) and other parameters ($\alpha s T$) = 10, I_o = 3, σ = 15 and r = 0.2.

$\{b_i\}$ are known to be subjected to the positivity constraint, i.e., $b_i + B_o \geq 0$ for all i. Several methods such as linear iteration, the method of steepest descents, and the method of conjugate gradients have been coded in Algol computer programs to solve the problem of nonlinear programming with linear constraints.

3.3.1. Linear Iteration

We can solve Eq. (2.52) directly by the method of linear iteration. From Eq. (2.52), we have

$$\hat{\underset{\sim}{b}}^{i+1} = \underset{\sim}{A}^{-1}\underset{\sim}{Y} + \underset{\sim}{A}^{-1}\underset{\sim}{B}(\hat{\underset{\sim}{b}}^{i}) \ ,$$

where $\hat{\underset{\sim}{b}}^{i}$ is the estimate of vector $\underset{\sim}{b}$ in the i-th iteration. Matrices $\underset{\sim}{A}$ and $\underset{\sim}{Y}$ are not functions of $\underset{\sim}{b}$; hence they will not vary in the process of iteration. Only the $\underset{\sim}{B}$ matrix is variable, and it is relatively simple to compute.

The positivity constraint applies at each step in the iteration. Whenever $\hat{b}_i + B_o < 0$ in the iteration, we set $\hat{b}_i = -B_o + \epsilon$, where ϵ is a small positive number. The iteration process stops whenever the increment of $\hat{b}_j$ is smaller than a predesignated value. The convergence of the linear iteration scheme is rather slow, and it took approximately one hundred iterations to converge in typical cases.

3.3.2. The Method of Steepest Descents

The linear iteration method converges rather slowly. In order to speed up the rate of convergence, we have tried second-order methods such as the method of steepest descents.[6,7,8] This method has been described in section 2.5 to minimize $F(\underset{\sim}{b})$ of Eq. (2.47). In this second-order method we need to calculate the gradient vector ∇F and the Hessian matrix $\underset{\sim}{H}$ of function $F(\underset{\sim}{b})$. The elements of the gradient vector and the Hessian matrix are

$$e_i = (\nabla F)_i$$

$$= \frac{\partial F}{\partial b_i}$$

$$= -\sum_{i=1}^{N} \left[\frac{n_i K_{ij}}{I_o + V_i} - \alpha S\dot{T} K_{ij} \right] + \sum_{m=1}^{M} \mu_{jm} b_m , \qquad (3.18)$$

and

$$c_{jk} = H_{jk}$$

$$= \frac{\partial^2 F}{\partial b_j \partial b_k}$$

$$= \sum_{i=1}^{N} \frac{n_i K_{ij} K_{ik}}{(I_o + V_i)^2} + \mu_{jk} . \qquad (3.19)$$

The values for e_i and c_{ij} are then substituted into Eq. (2.61) to evaluate λ, and the new value for $\underset{\sim}{b}$ is then obtained from Eq. (2.59),

$$b_j^{i+1} = b_j^i - \lambda e_j \quad , \quad j = 1, 2, \ldots, M \ .$$

The positivity constraint applies at every step; whenever $b_j^{i+1} < -B_o$, we set $b_j^{i+1} = -B_o + \epsilon$.

In this method, we have to update $\{ e_i \}$ and $\{ c_{ij} \}$ in each iteration step. This takes more effort than the linear iteration method, where we have to update only matrix $\underset{\sim}{B}(\underset{\sim}{b})$. The iteration stops whenever

$$\frac{\sum_{j=1}^{M} | b_j^{i+1} - b_j^i |}{\sum_{j=1}^{M} | b_j^{i+1} |} \leq 10^{-3} \ .$$

The method of steepest descents converges quickly and start to oscillate when approaching the minimum point.[8]

3.3.3. The Method of Conjugate Gradients

To remedy the disadvantage of the method of steepest descents, an attempt to use a more efficient computational method has been made. We have used the method of conjugate gradients. The method of conjugate gradients has been well documented in the literature[10, 11, 12]; for simplicity we will only present the algorithm we have used.

In the method of conjugate gradients, the major steps in the algorithm are:

1. At the initial point $\underset{\sim}{b}^1$ compute the gradient vector $\underset{\sim}{g}_1 = \nabla F(\underset{\sim}{b}^1)$, and set the vector $\underset{\sim}{d}_1 = -\underset{\sim}{g}_1$.

2. For $i = 2, 3, \ldots, M+2$

   (a) set $\underset{\sim}{b}^i = \underset{\sim}{b}^{i-1} + \lambda_{i-1} \underset{\sim}{d}^{i-1}$, where

   (b) λ_{i-1} minimizes $q_{i-1}(\lambda) = F(\underset{\sim}{b}^i)$;

   (c) compute $\underset{\sim}{g}^i = \nabla F(\underset{\sim}{b}^i)$;

   (d) unless $i = M+2$ set

$$\underset{\sim}{d}^i = -\underset{\sim}{g}^i + \gamma \underset{\sim}{d}^{i-1}$$

where

$$\gamma = (\underset{\sim}{g}^i)^T (\underset{\sim}{g}^i) / (\underset{\sim}{g}^{i-1})^T (\underset{\sim}{g}^{i-1}) ,$$

and repeat (a).

3. Replace $\underset{\sim}{b}^1$ by $\underset{\sim}{b}^{M+2}$ and restart the iteration.

We have applied this algorithm to minimize the functional $F(\underset{\sim}{b})$ with $\mu_{jk} \equiv 0$. By letting $\mu_{jk} = 0$, we use no prior knowledge about the object. The elements $\{ e_j^i \}$, $\{ c_{jk}^i \}$ of the gradient vector $\underset{\sim}{g}^i = \nabla F(\underset{\sim}{b}^i)$ and Hessian matrix $\underset{\sim}{H}^i = \left\{ \frac{\partial^2 F(\underset{\sim}{b})}{\partial b_j \partial b_k} \right\}_{\underset{\sim}{b} = \underset{\sim}{b}^i}$ are shown in Eqs. (3.18) and (3.19) except with $\mu_{jk} \equiv 0$.

The above algorithm is for unconstrained minimization. Because our functional is constrained, the algorithm has to be modified. The value of λ in step 2(b) has been evaluated like the one in the method of steepest descents in Eq. (2.61). We have applied the positivity constraint in step 2(a), setting $b_j^i = -B_o + \epsilon$ whenever

$b_j^i < -B_o$ and at the same time changing d_j^i to $(-B_o - b_j^{i-1})/\lambda$. The stopping criterion we have used is similar to the one used in the method of steepest descents.

The positivity constraint presents difficulty for the method of conjugate gradients in our computer programming. The elements of conjugate gradient vector $\underset{\sim}{d}^i$ have been changed whenever the positivity constraint is violated; hence $\underset{\sim}{d}^i$ is no longer a conjugate gradient vector. All the analysis for the convergence for this method is no longer necessarily valid. Up to now this method has not worked well. A better way to handle the positivity constraint such as penalty function method probably will be needed.[13]

3.4. Computer Simulated Experiment Results

In computer simulation, the one-dimensional object can be generated either by the first-order autoregressive process as described before or by just specifying any particular form. The parameters such as the uniform background level B_o, the correlation coefficient r, the signal mean deviation σ_b, and the constant αsT should be given both in the processes of generation and restoration of the object. The parameters r and σ_b used in the restoration process are not necessarily equal to the values used in the generation of the object; usually the prior knowledge about them is not complete. The incoherent point-spread function used was that for either motion-blur image degradation or diffraction-limited distortion as described by

Eqs. (3.7) and (3.12) respectively.

The likelihood achieved by the maximum-likelihood estimate is higher than for the linear least-squares estimate for sure from the definition of maximum-likelihood estimation. If the mean error-squares of the maximum-likelihood estimate are smaller than for the linear estimate, then we can conclude that the maximum-likelihood estimate is superior to the linear estimate.

In the nonlinear estimation scheme, we use both the linear iteration and the method of steepest descents to solve the set of nonlinear equations (2.49). Prior knowledge of the positivity constraint on the object radiance function is used. We use the linear estimate

$$\hat{\underset{\sim}{b}}_L = \left(\underset{\sim}{\phi}_b \underset{\sim}{K}^T \underset{\sim}{K} + \frac{I_o}{\alpha s T} \underset{\sim}{I} \right)^{-1} \underset{\sim}{\phi}_b \underset{\sim}{K} \left(\frac{1}{\alpha s T} \underset{\sim}{n} - I_o \underset{\sim}{I} \right) \quad , \tag{3.20}$$

as the starting point in solving the set of nonlinear equations. When we assume the signal is much larger than the noise, i.e., $\sigma^2 \to \infty$, the linear estimate in Eq. (3.21) will reduce to the inverse filtering estimate

$$\hat{\underset{\sim}{b}}_I = (\underset{\sim}{K}^T \underset{\sim}{K})^{-1} \underset{\sim}{K}^T \left(\frac{1}{\alpha s T} \underset{\sim}{n} - I_o \underset{\sim}{I} \right) \quad . \tag{3.21}$$

The performance of different estimation schemes will be measured by a percentage squared error from the formula

$$\varepsilon = \frac{\sum_{j=1}^{M} (\hat{b}_j - b_j)^2}{\sum_{j=1}^{M} b_j^2} \times 100 \quad , \tag{3.22}$$

where $\hat{b}_j$ is the estimate of b_j. We will compare the performances of the maximum-likelihood estimate, the linear least-squares estimate, and the inverse filtering estimate. For fair comparison the positivity constraint on the object radiance was also applied to linear least-squares and inverse-filtering estimates.

In a typical experiment, we used a set of five numbers (347, 5412, 3298, 9547, 2486) as the initial number for the random number generator computer subroutine to generate random variables $\{ R_1^i, R_2^i, Y_i, R_3^i, R_4^i \}$. The object radiance function was degraded by motion blur with a blur of four cells, i.e., $D = 4$ in Eq. (3.7). The parameters used in the generation of the first object and data were $(\alpha s T) = 10$, $I_o = 3$, $\sigma = 25$, $r = 0.6$. The generated object and the measured data reduced to the scale of the object by dividing by $(\alpha s T)$ are shown as the heavy line and the thin line in Fig. 3.2 respectively. Hard rectification has been applied at six object points. The measured data do not resemble the object at all.

Because of the hard rectification, the object radiance function can no longer be described by a first-order autoregressive process having the covariance matrix as shown in Eq. (3.2). We used the same

values for the parameters r and σ_b in the restoration scheme as in the generation of the data. The result is shown in Fig. 3.3. The signal-to-noise ratio is 5787 >> 1, so the inverse filtering estimate is very similar to the linear least-squares estimate.

In the restoration of the object radiance function, we have used values for parameters r' and σ_b' that are not necessarily equal to the values r and σ_b we have used in the generation of the object radiance function. From Fig. 3.4, we can see that different values used for parameters r' and σ_b' affect the linear least-squares estimate more than the maximum-likelihood estimate. The maximum-likelihood estimator is a function of the measured data; as can be seen from Eq. (2.52), this estimator tends to be adaptive and hence depends less on the prior knowledge of the object radiance function than the linear least-squares estimator. This can be demonstrated by using $\sigma = \infty$ in the restoration process; the restoration uses no prior information whatever about the object, but only the Poisson character of the distribution of the data and the known point-spread function. The result is shown in Fig. 3.5. The percentage error-square for the maximum-likelihood estimate is 2.69%, which is not too far from those results for which the prior knowledge was used. Another interesting point shown in Fig. 3.4 is that if the assumed correlation r is too large, say $r = 0.99$, the result of maximum-likelihood estimation tends to go bad. This happens probably because if we assume that r is close

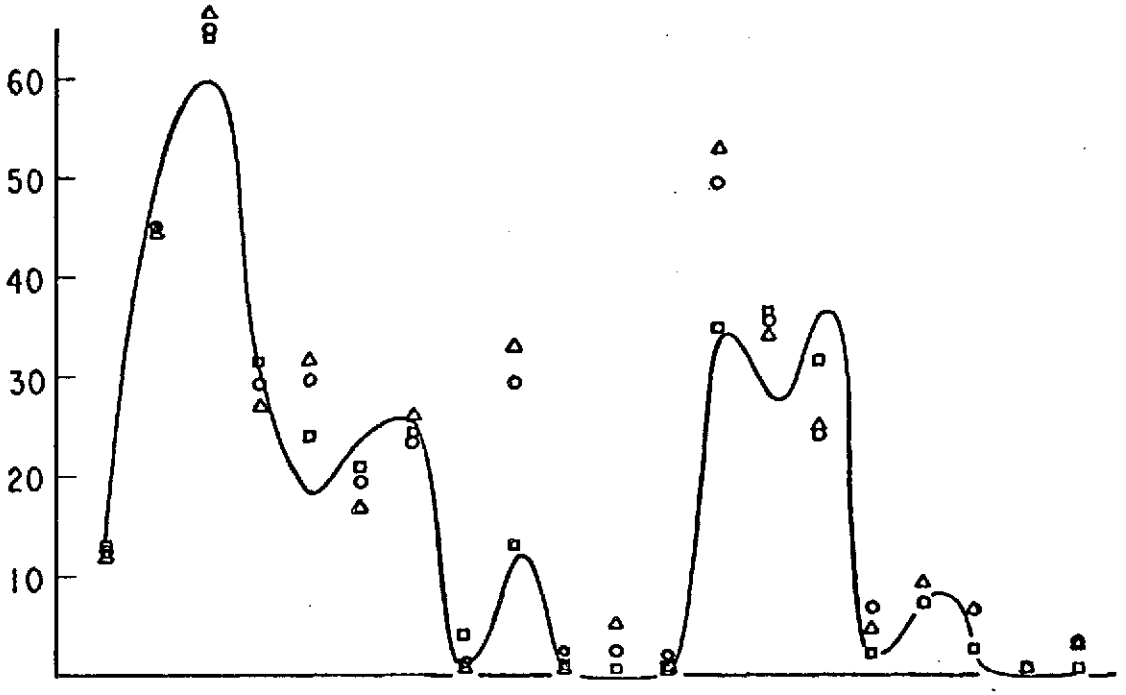


Figure 3.3. Original object and restored values. □ = maximum-likelihood estimate, o = linear estimate, △ = inverse filtering estimate; $r' = 0.5$, $\sigma'_b = 25$.

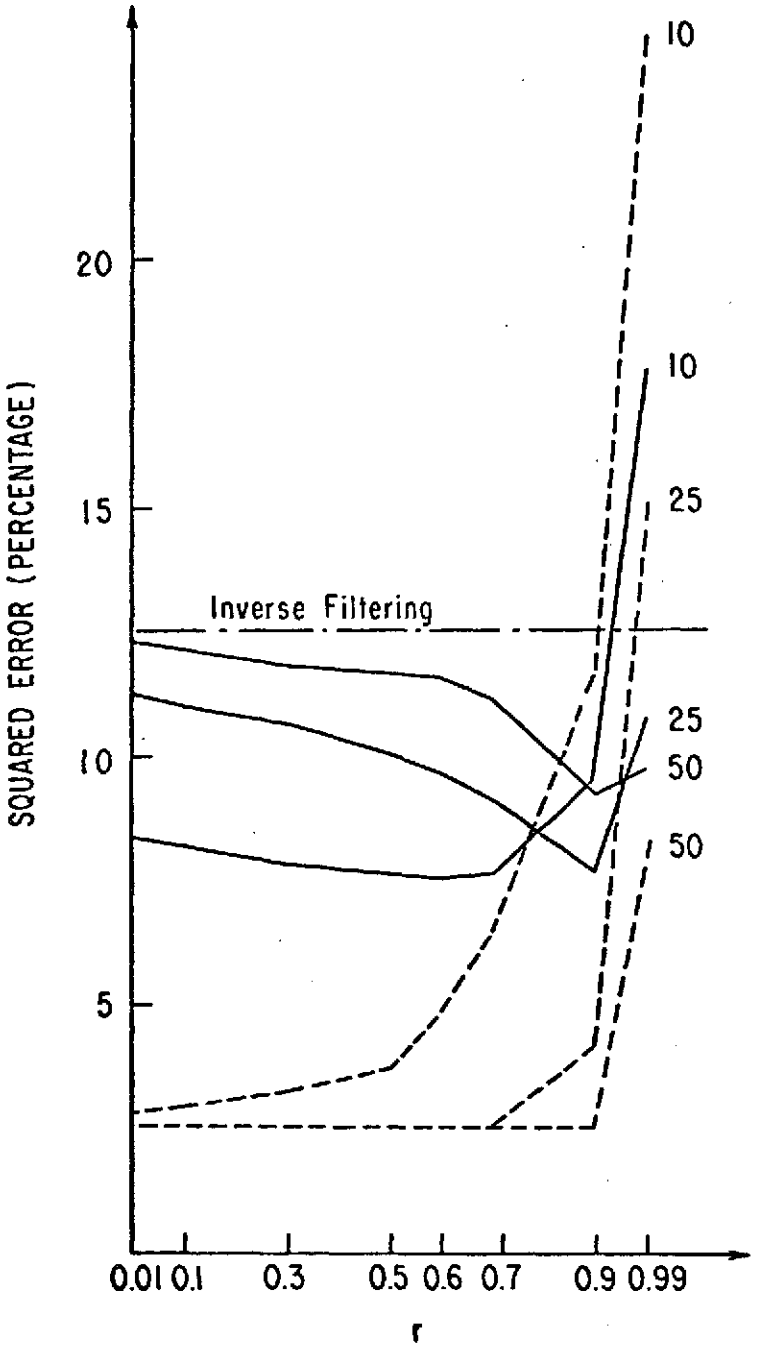


Figure 3.4. Percentage squared error as function of assumed correlation coefficient r ; curves are indexed by assumed standard deviation σ_b . Solid lines - linear estimate; dashed lines - maximum-likelihood estimate.

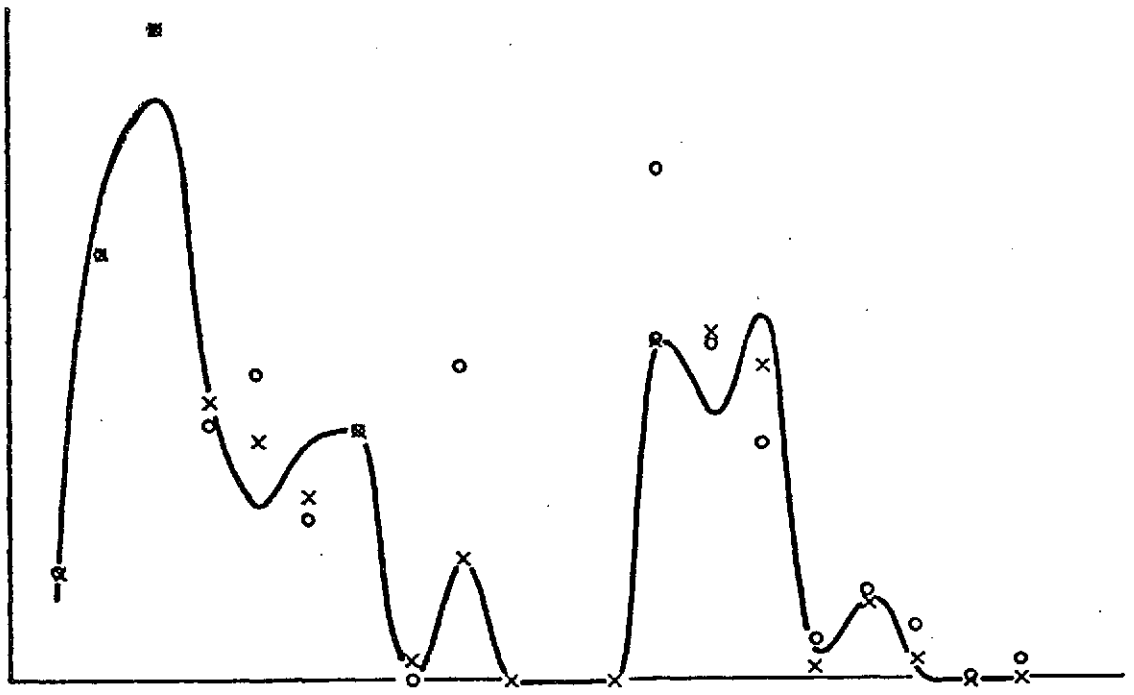


Figure 3.5. Original object and restored values. x = maximum-likelihood estimate, o = linear estimate, $\sigma = \infty$.

to 1 and actually it is not, the maximum-likelihood estimator will tend to closely correlate all the object points. In other words, the estimator has been very confined. In general, over a reasonable range, the bigger σ_b and the smaller r, the better the maximum-likelihood estimator performs. Under this situation, the maximum-likelihood estimator is not too confined and has the ability to adapt itself according to the measured data.

The effect of $\alpha_s T$ is also examined. For an existing optical system, α_s is fixed, but the observation time T can be varied. The result is shown in Fig. 3.6. Both the maximum-likelihood estimate and the linear least-squares estimate perform better for larger αST, since it is related to the signal-to-noise ratio, but the maximum-likelihood estimate is not as sensitive to the change of αST as the linear least-square estimate.

Another input object radiance function tested is shown in Fig. 3.7(a). The parameters used in the generation of the object function are $r = 0.2$, $\sigma = 15$, and $I_o = 3$. The resulting restoration is shown in Fig. 3.7(b). Figures 3.8(a) and (b) show the percentage squared error as functions of various values of assumed correlation coefficient and standard deviation σ. The maximum-likelihood estimate is seen in this case to yield the closer fit to the true object radiance, the greater the assumed value of σ and the smaller the assumed value of r.

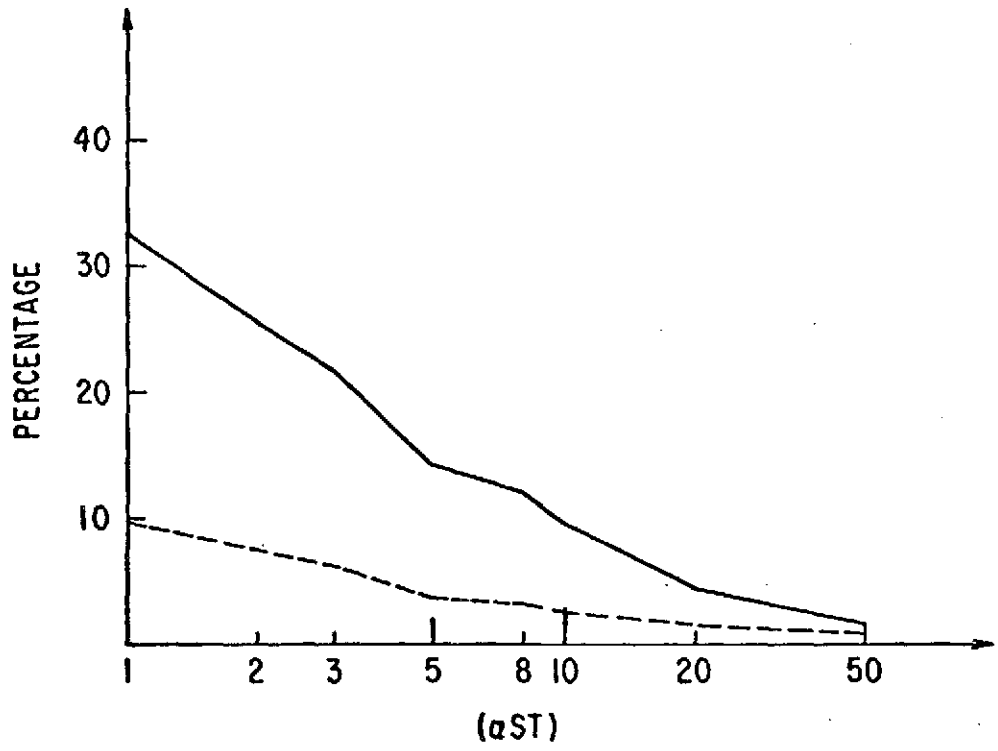


Figure 3.6. Percentage squared error as function of (αsT). Solid line - linear estimate; dashed line - maximum-likelihood estimate.

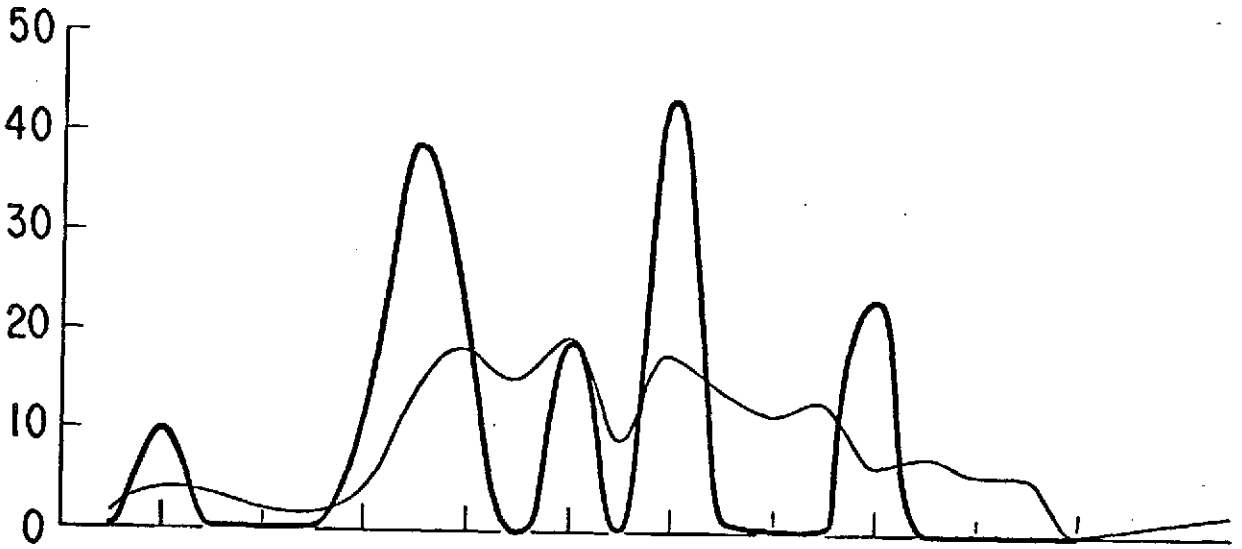


Figure 3.7 (a). Original object (thick line) and its noisy version (thin line) as recorded after blurring by relative motion; (5431, 3214, 5438, 3654, 2495), $I_o = 3$, $\alpha sT = 10$, $r = 0.2$, $\sigma = 15$.

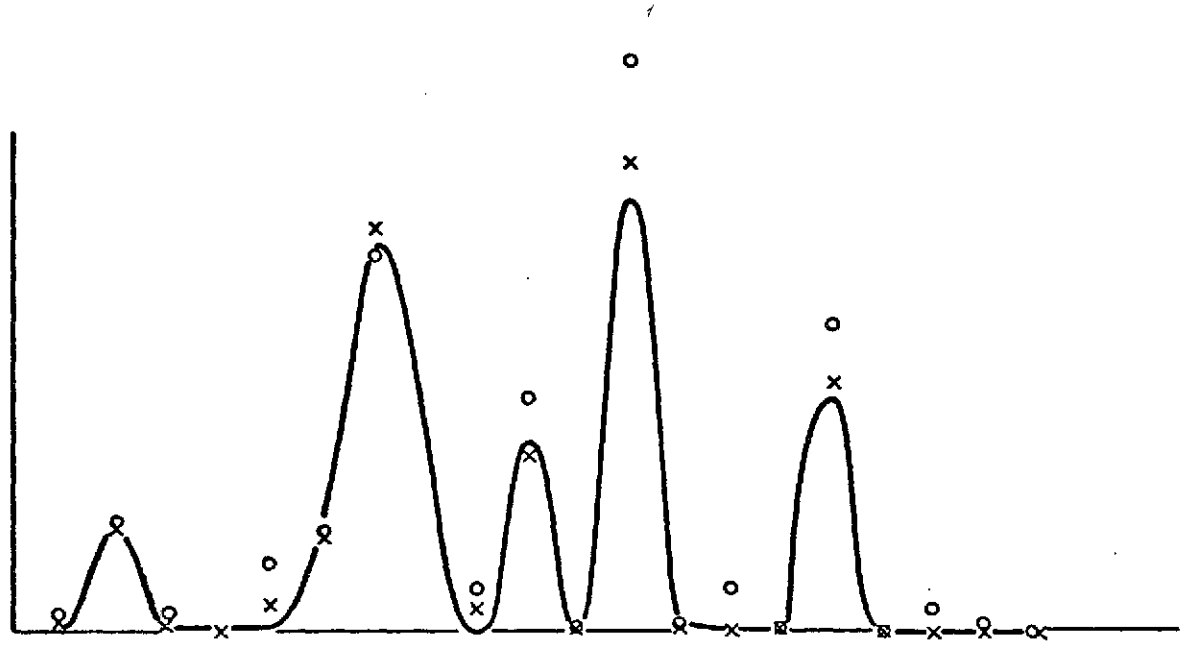


Figure 3.7 (b). Original object and restored values. x = maximum-likelihood estimate, o = linear estimate, $\sigma = \infty$.

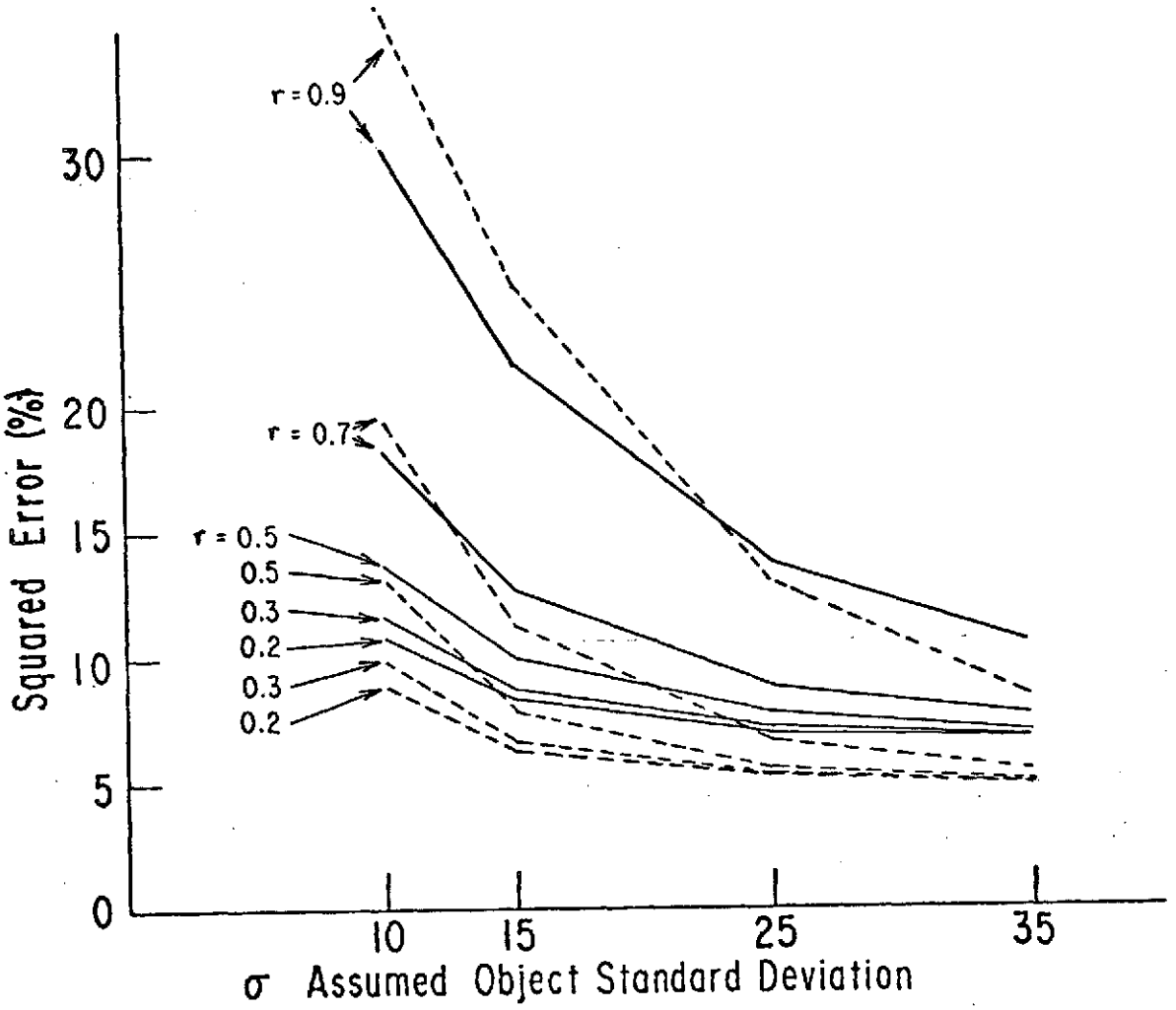


Figure 3.8 (a). Percentage squared error as functions of assumed object standard deviation σ_b ; curves are indexed by the assumed correlation coefficient r. Solid lines - linear estimate, dashed lines - maximum-likelihood estimate.

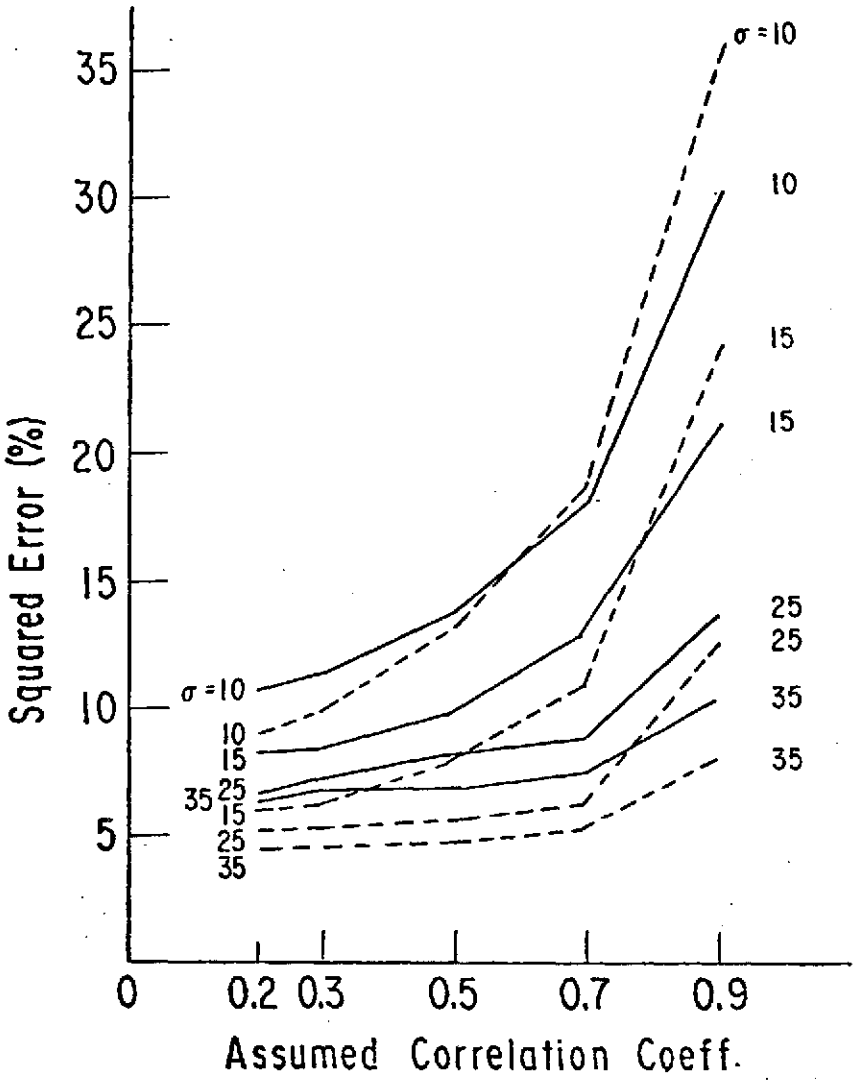


Figure 3.8 (b). Percentage squared error as functions of assumed correlation coefficient r; curves are indexed by assumed standard deviation σ_b. Solid lines - linear estimate, dashed lines - maximum-likelihood estimate.

There is, however, considerable variability in the performance of both the linear and the maximum-likelihood estimators. We should like the estimator to perform well in most of the cases. We took the same blurred object, with the same set of mean values λ_i, generated a number of independent sets of data n_i, and applied our procedure to each, with the resulting percentage errors shown in Table 1. The star indicates the trial shown in the previous figure. In ten of the twelve trials taking $\sigma^2 = \infty$ gives a lower percentage error than when the restoration procedure uses the prior information about how the object is generated. In two trials out of twelve the linear least-squares filtering was more effective than the maximum-likelihood estimation. Most of the time, the maximum-likelihood estimate is considerably closer than the linear one.

Table 2 shows a similar set of trials for a different object; there is much greater variability, but in all but one case the maximum-likelihood estimate was much closer than the linear one. Here in five out of twelve cases using the prior information gave a closer fit to the true object radiance. The starred case is shown in Figs. 3.9(a) and 3.9(b).

Deterministic signals were also used. A typical object radiance function we used in the experiment has the form

$$b_i = a\{\exp[-(i-5)^2/4] + \exp[-(i-9)^2/4] + \exp[-(i-16)^2/4]\}, \qquad (3.23)$$

Table 1

Performances of both linear and maximum-likelihood estimation of an object based on twelve different sets of data. Parameters used to generate the object are $(R_1, R_2) = (5431, 3214)$, $I_o = 3$, $\alpha ST = 10$, $r = 0.2$, $\sigma = 15$.

	Squared Errors (percent)			
	$\sigma = \sigma_s$, $r = R$		$\sigma \to \infty$	
	Linear	Nonlinear	Linear	Nonlinear
1	5.53	1.80	8.81	0.862
2	2.13	2.46	1.02	1.08
3	4.90	2.93	4.76	2.21
4	5.95	4.28	8.20	2.44
5	7.61	5.20	5.17	3.43
6	5.83	5.79	3.43	3.58
7	8.49	4.23	8.25	3.68*
8	6.02	5.64	3.97	3.69
9	5.49	3.67	9.77	4.76
10	9.84	6.22	8.69	5.17
11	7.84	6.36	7.24	5.63
12	12.59	9.39	12.36	9.56
Mean	6.85	4.83	6.81	3.84
Standard Deviation	2.67	2.07	3.18	2.33

*Illustrated in Fig. 3.7.

Table 2

Performances of both linear and maximum-likelihood estimation of an object based on twelve different sets of data. Parameters used to generate the object are $(R_1, R_2) = (842, 3248)$, $I_1 = 5$, $\alpha ST = 10$, $r = 0.3$, $\sigma = 15$.

	Squared Errors (percent)			
	$\sigma = \sigma_s$, $r = R$		$\sigma \to \infty$	
	Linear	Nonlinear	Linear	Nonlinear
1	7.91	6.70	7.74	5.42
2	9.69	7.07	8.11	5.70
3	14.93	8.84	11.82	6.59
4	13.57	7.23	15.79	6.90
5	9.16	5.10	11.55	7.51
6	13.94	8.12	18.44	9.73
7	19.62	11.75	19.91	10.17*
8	9.51	6.86	9.94	10.74
9	29.62	8.62	47.06	10.99
10	16.04	14.55	16.81	14.27
11	23.18	18.07	24.70	17.02
12	22.19	12.78	25.73	17.05
Mean	15.78	9.64	18.13	10.17
Standard Deviation	6.70	3.85	10.90	4.11

*Illustrated in Fig. 3.9.

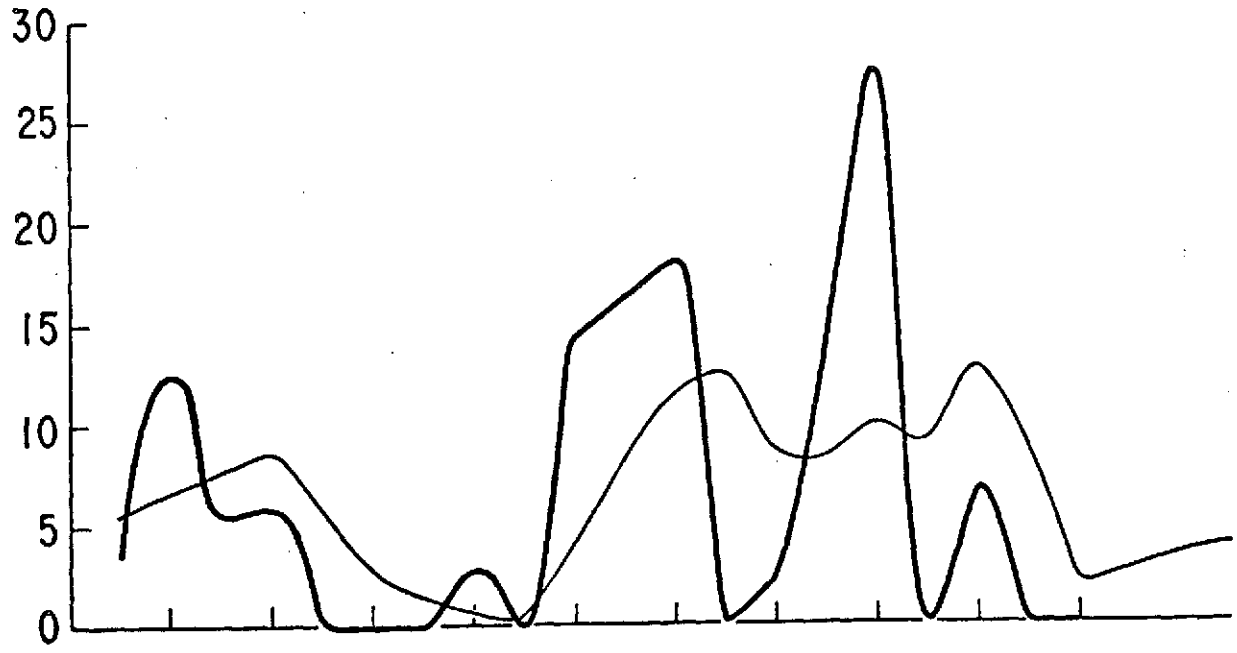


Figure 3.9 (a). Original object (thick line) and its noisy version (thin line) as recorded after blurring by relative motion; (842, 3248, 854, 279, 657). $I_o = 5$, $\alpha sT = 10$, $r = 0.3$, $\sigma = 15$.

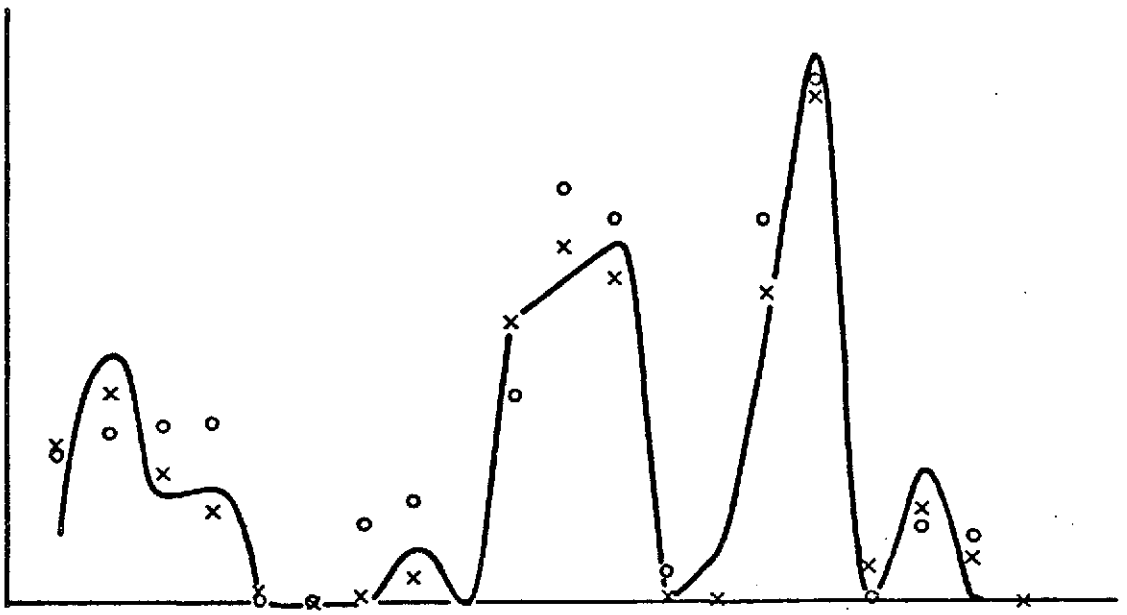


Figure 3.9 (b). Original object and restored values. x = maximum likelihood estimate, o = linear estimate, $\sigma = \infty$.

where a is a positive number which corresponds roughly to the standard deviation of the object. The object radiance function consists of three peaks of height a; two peaks are close together and one peak is more isolated. The width of each peak is about four. For $a = 90$, the object and the smeared image are shown in Fig. 3.10. The object has been smeared by a motion blur of four cells. It is clear from Fig. 3.10 that two peaks have been smeared into a broad peak. The distance between those two peaks is four cells, which is the amount of the motion blur.

A set of data using parameters $\alpha sT = 3$ and $I_o = 5$ was generated for the Gaussian-shaped object with $a = 90$. A typical restoration result using parameters $r' = 0.5$ and $\sigma_s' = 45$ is shown in Fig. 3.11. Two peaks are well resolved for both linear and maximum-likelihood estimate. The performances of both estimators as functions of (αST) for the same object are shown in Fig. 3.12. The effect of using different assumed values of r' and σ_b' is shown in Figs. 3.13 and 3.14. Even though for some values of r' and σ_b', the linear estimate performs better than the maximum-likelihood estimate, in most of the cases the maximum-likelihood estimate performs better than linear estimate. For deterministic signals, all the statistical assumptions about the object are not valid any more; however, we can still have good results. This demonstrates the usefulness of our estimator.

Image degradation by a diffraction-limited optical system has

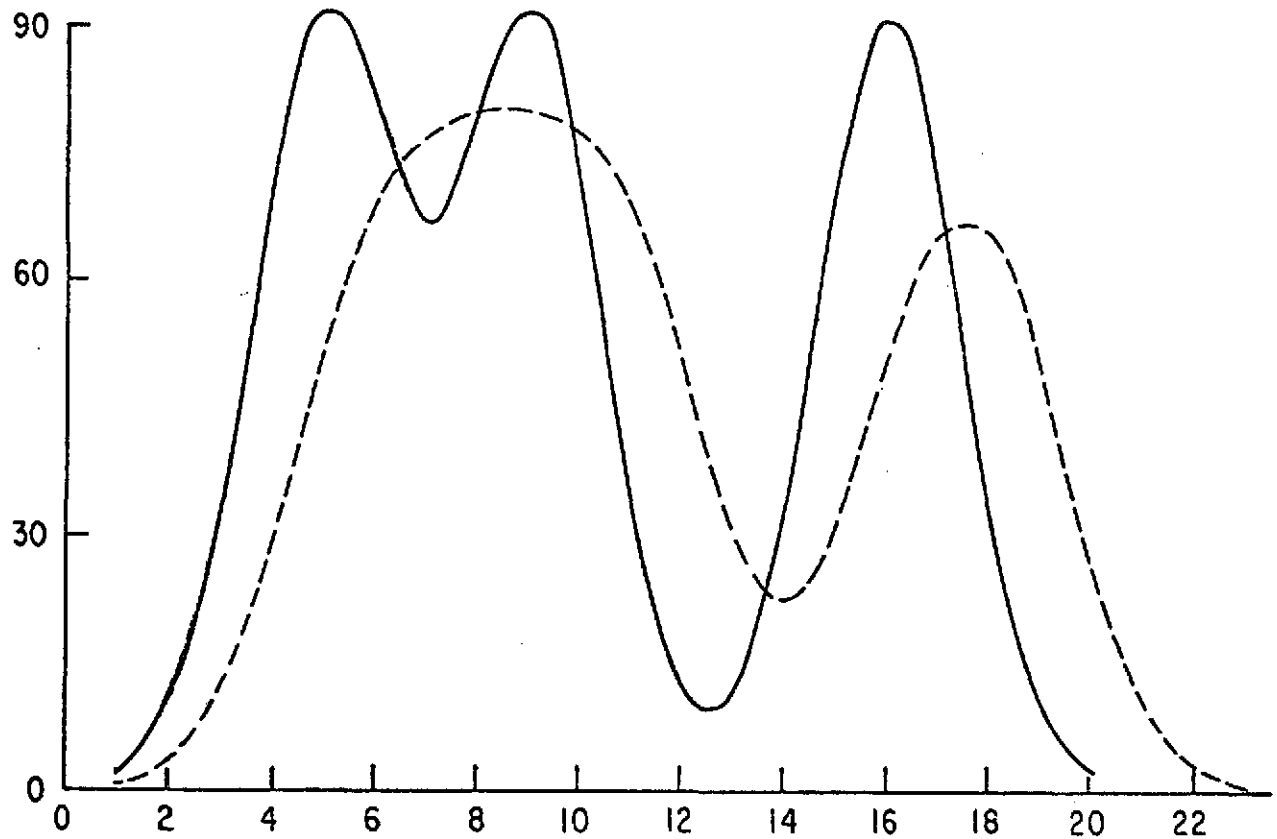


Figure 3.10. Gaussian-shaped object (thick line) and its degraded image (dashed line) after blurring by relative motion; a = 90.

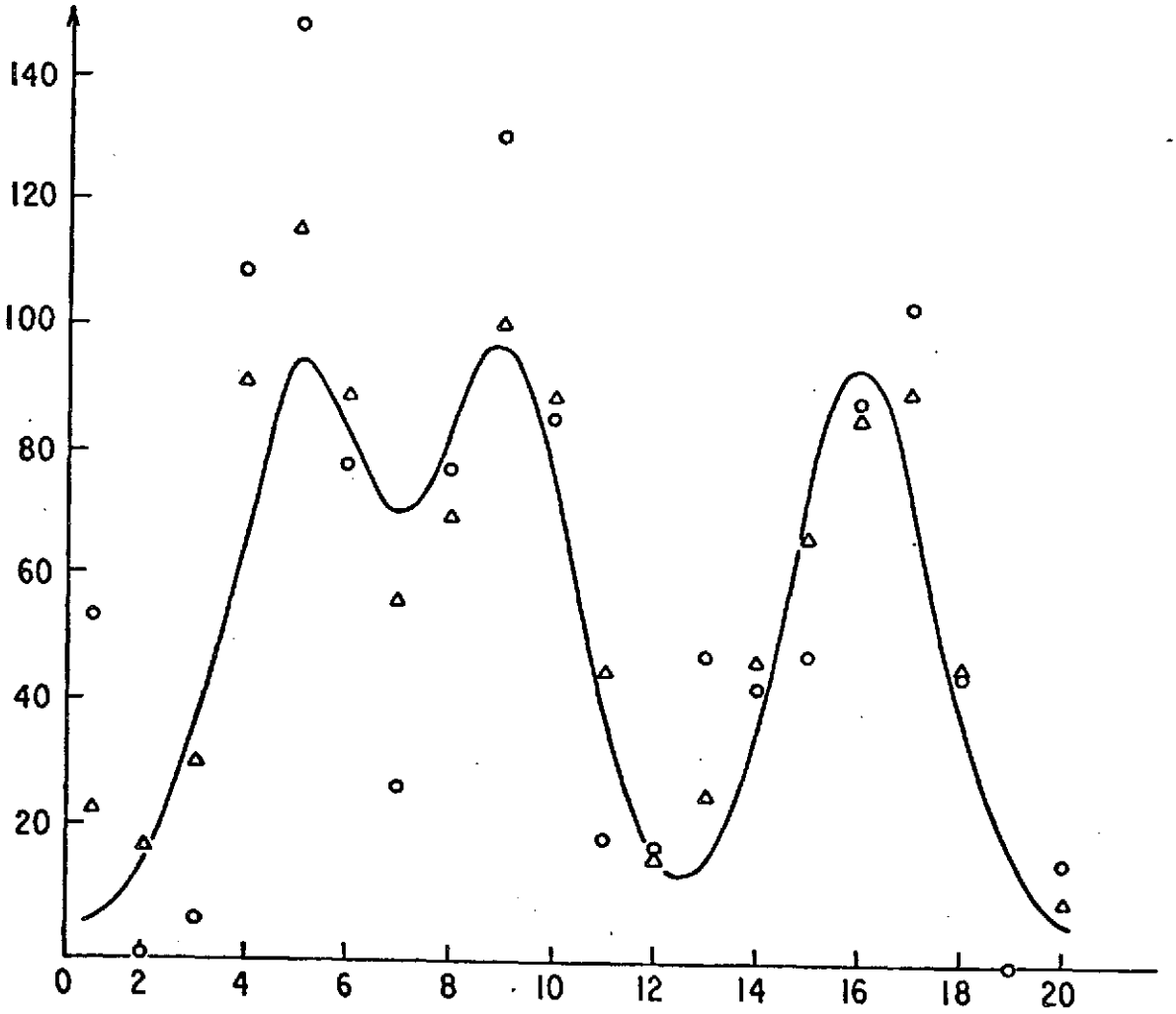


Figure 3.11. Gaussian-shaped object and restored values. Δ = maximum-likelihood estimate, o = linear estimated; $\alpha sT = 3$, $\sigma_b = 45$, $I_o = 5$, $r = 0.5$, $a = 90$.

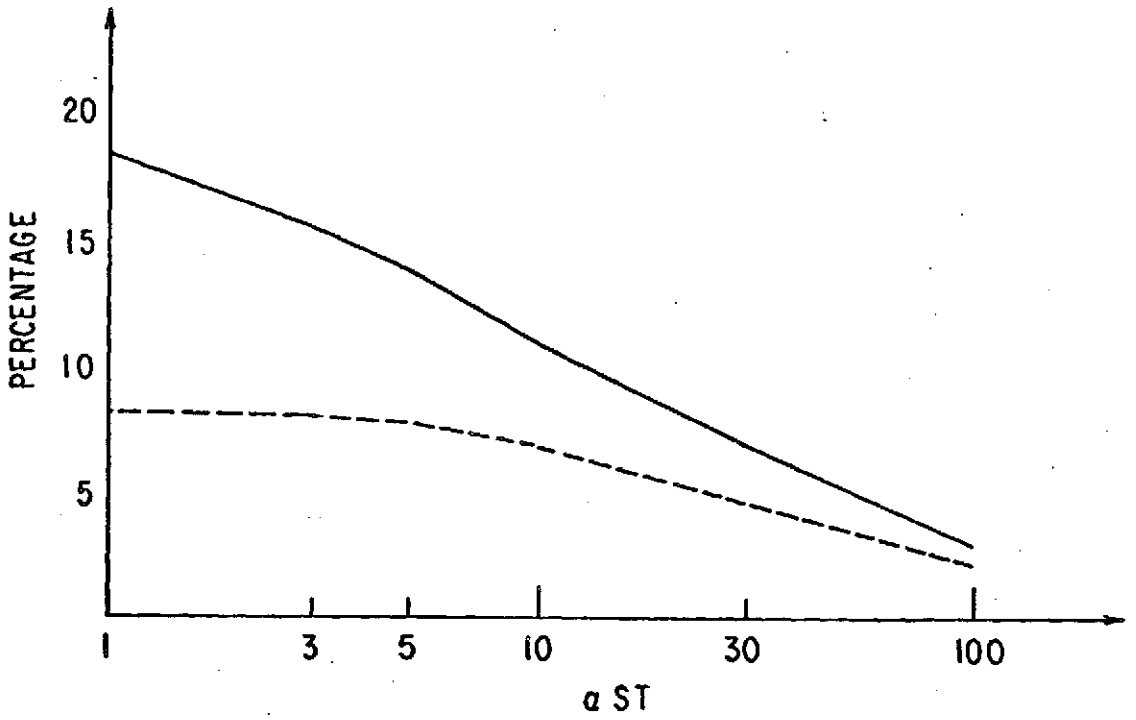


Figure 3.12. Percentage square error as function of (αsT). Solid line - linear estimate; dashed line - maximum-likelihood estimate.

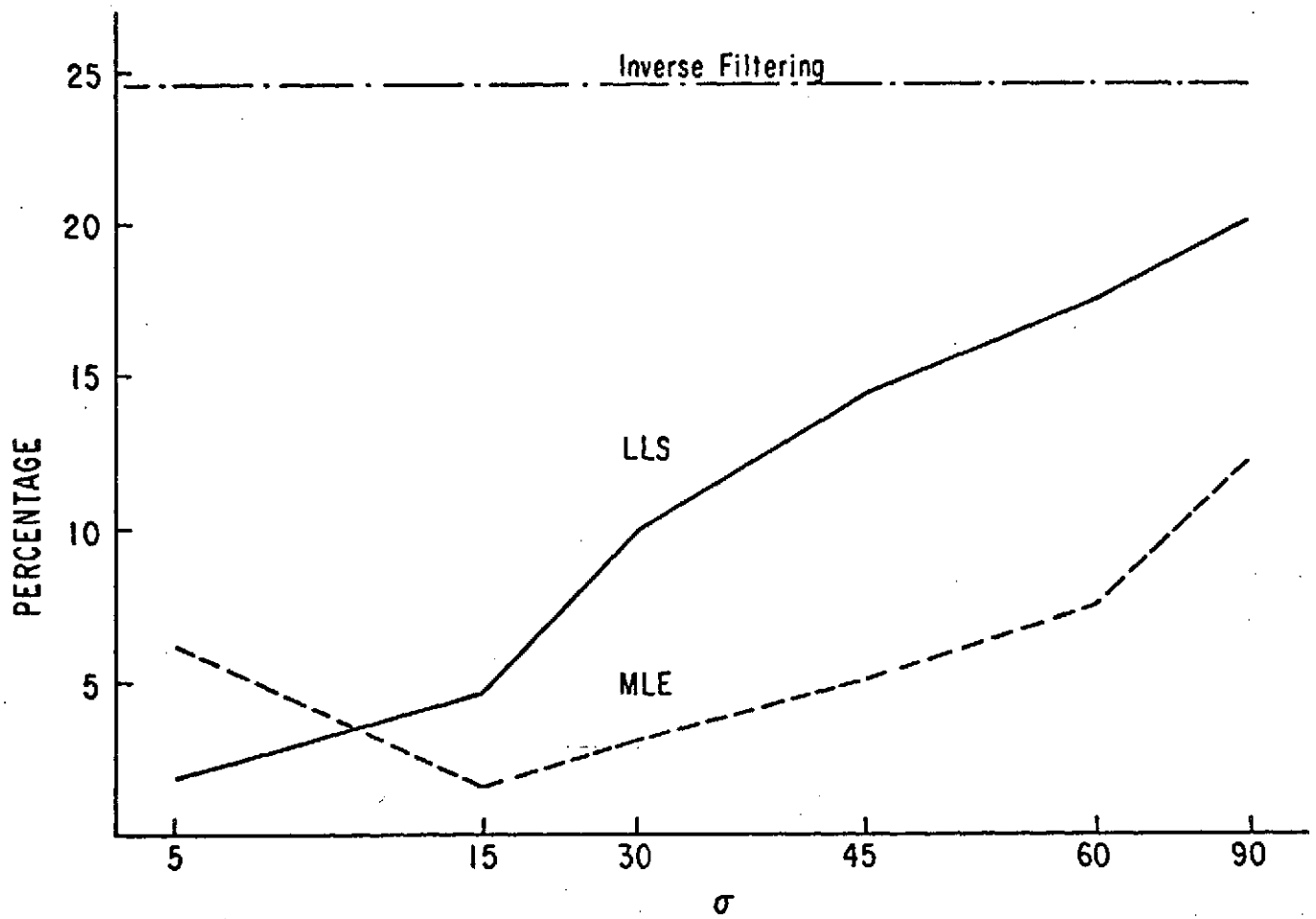


Figure 3.13. Percentage squared error as functions of assumed object standard deviation σ. $\alpha s T = 3$, $I_o = 5$, $r = 0.5$, $a = 90$. Solid line - linear estimate; dashed line - maximum-likelihood estimate.

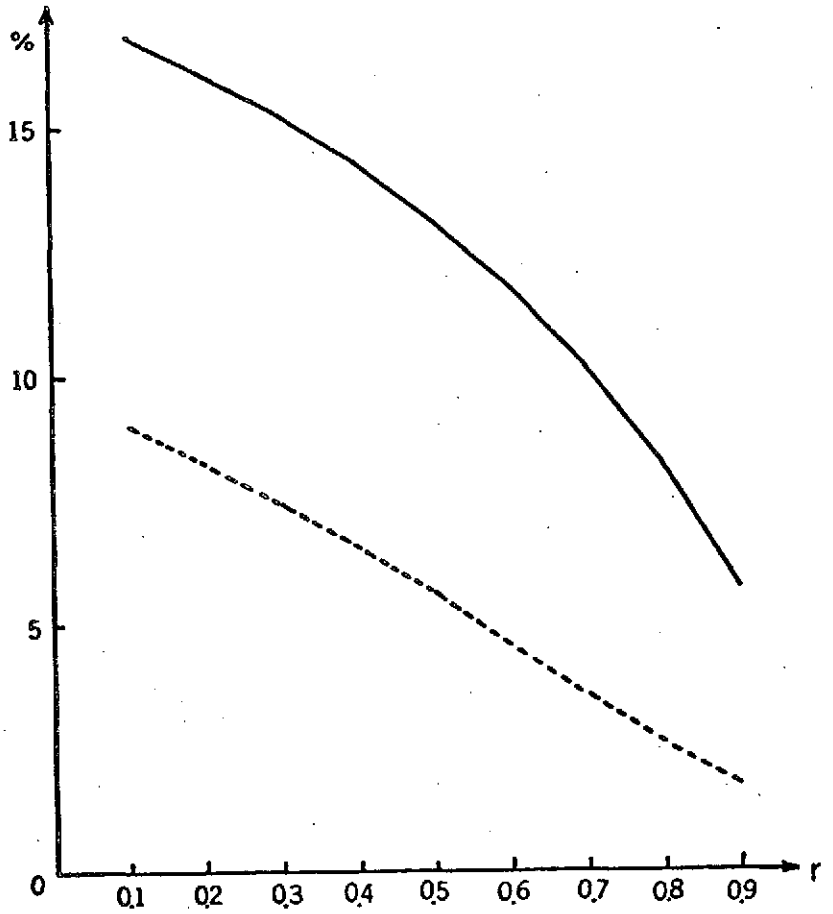


Figure 3.14. Percentage squared error as function of assumed correlation coefficient r . $\alpha sT = 3$, $I_o = 5$, $\sigma = 45$, $a = 90$. Solid line - linear estimate; dashed line - maximum likelihood estimate.

also been considered. Figure 3.15 shows the result of the restoration of a stochastic signal. The original object has been discussed in section 3.1 and shown in Fig. 3.1. The passband width W_o in Eq. (3.13) has been set to 1.2. In this case, the object spectrum outside the passband of the optical system cannot be recovered; hence the small sharp peak between two big peaks is not restored. Figure 3.16 shows the performances of both linear and maximum-likelihood estimators as functions of various assumed values for r and σ_b. For $\sigma_b = 120$, $r = 0.8$, the percentage squared errors are 11.72 and 32.60 for the nonlinear and linear estimators respectively; and for smaller r the performance of the maximum-likelihood estimator deteriorates somehow, but the percentage squared error exceeds 100%. Owing to the adaptive characteristic of the maximum-likelihood estimate, its performance is rather insensitive to the assumed value of the parameters used. We have used $\sigma'_b = \infty$ in our restoration scheme. The percentage squared errors for the nonlinear estimator are 24.52% and 4880% for the linear estimator. 24.52% is about the best the linear estimator can do as we can see from Fig. 3.16. The nonlinear estimator always performs better than the linear estimator in this case.

Figures 3.17 and 3.18 show the performance of the restoration of a deterministic Gaussian-shaped object. The object has been degraded by a diffraction-limited optical system with $W_o = 0.8$. It is clear that maximum-likelihood estimator performs better than the linear estimator.

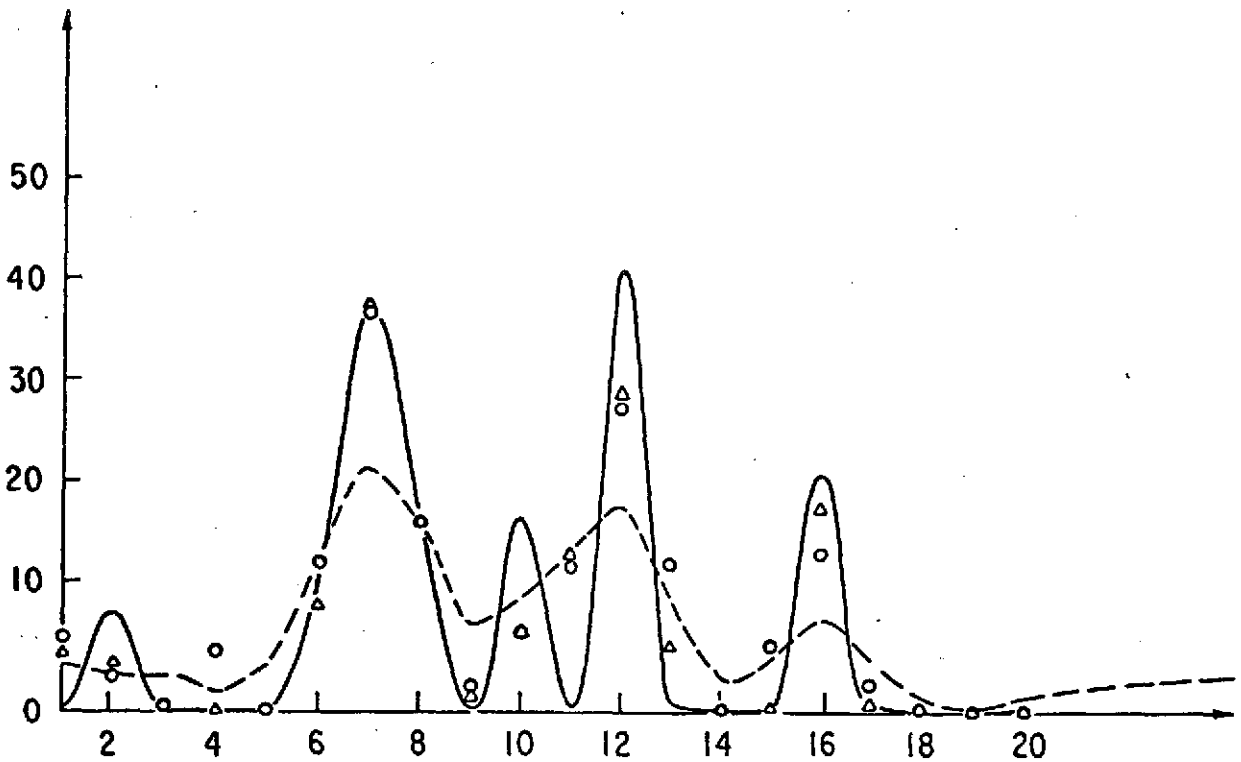


Figure 3.15. Original object and restored values. △ = maximum likelihood estimate, o = linear estimate, r = 0.2, σ = 20, W_o = 1.2.

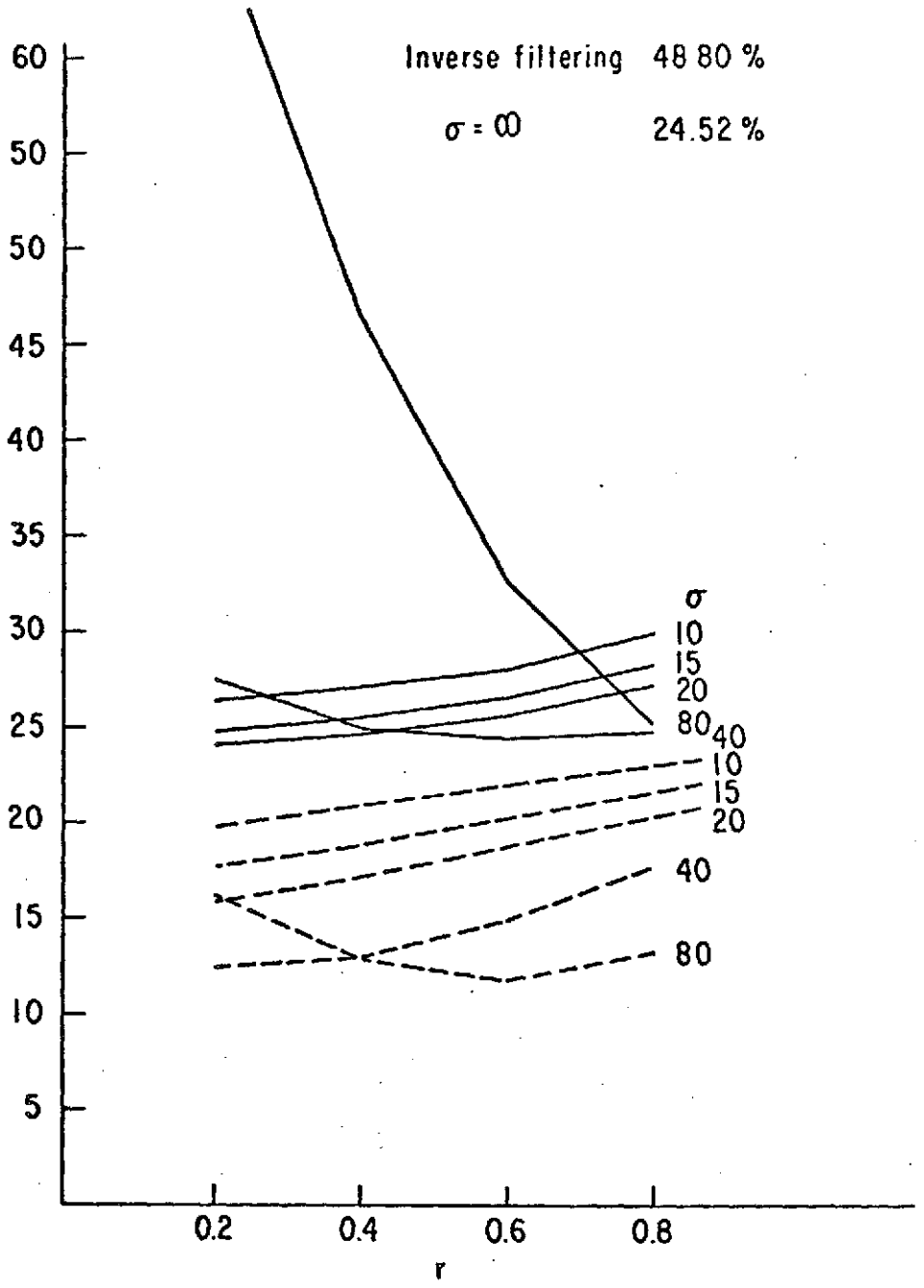


Figure 3.16. Percentage squared error as functions of assumed correlation coefficient r . Curves are indexed by the assumed object standard deviation σ . Solid lines - linear estimate; dashed lines - maximum-likelihood estimate.

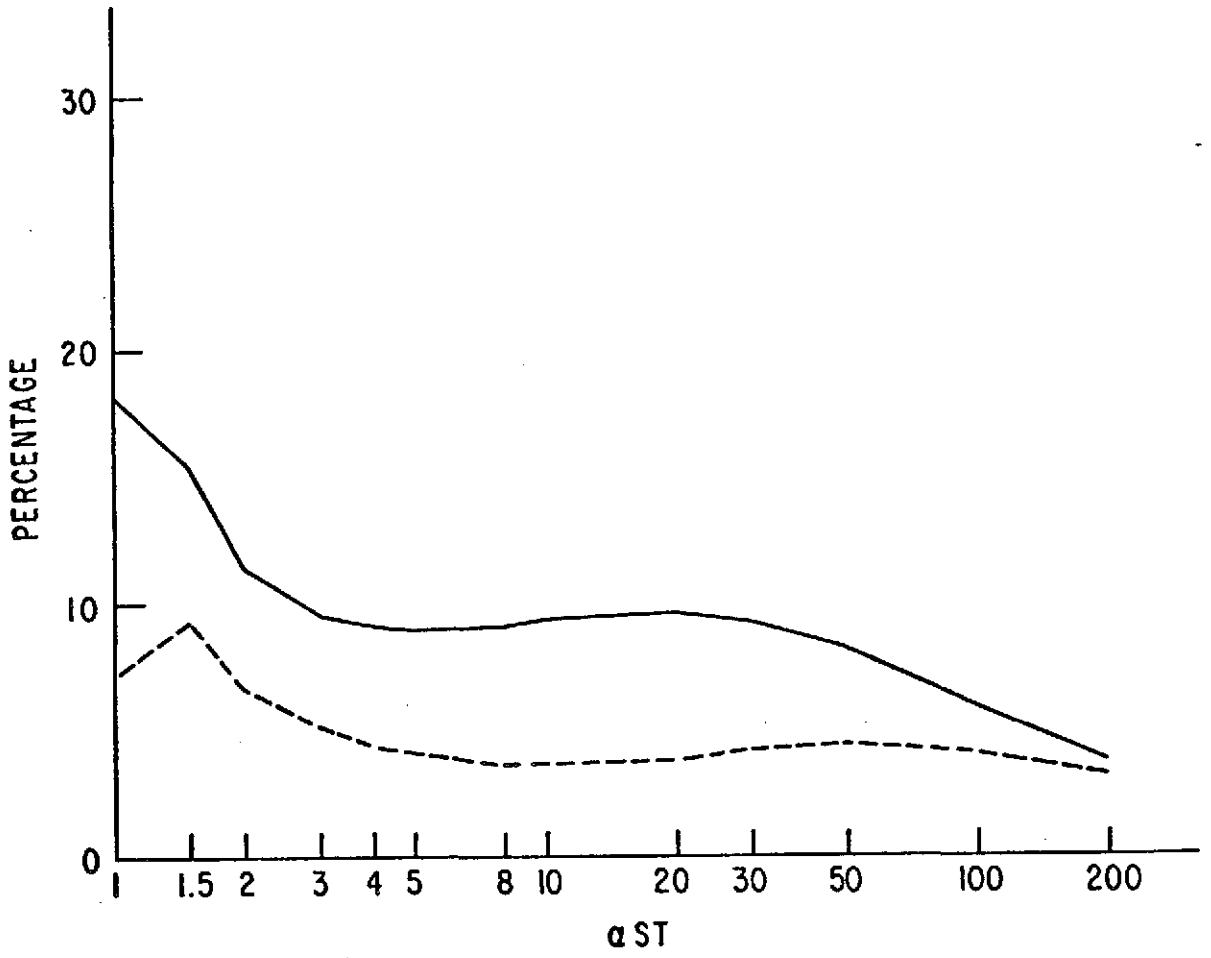


Figure 3.17. Percentage squared error of the restoration of a Gaussian-shaped object as function of $\alpha s T$. Solid line - linear estimate; dashed line - maximum-likelihood estimate; $\sigma = 15$, $I_o = 10$, $r = 0.5$, $W_o = 0.8$.

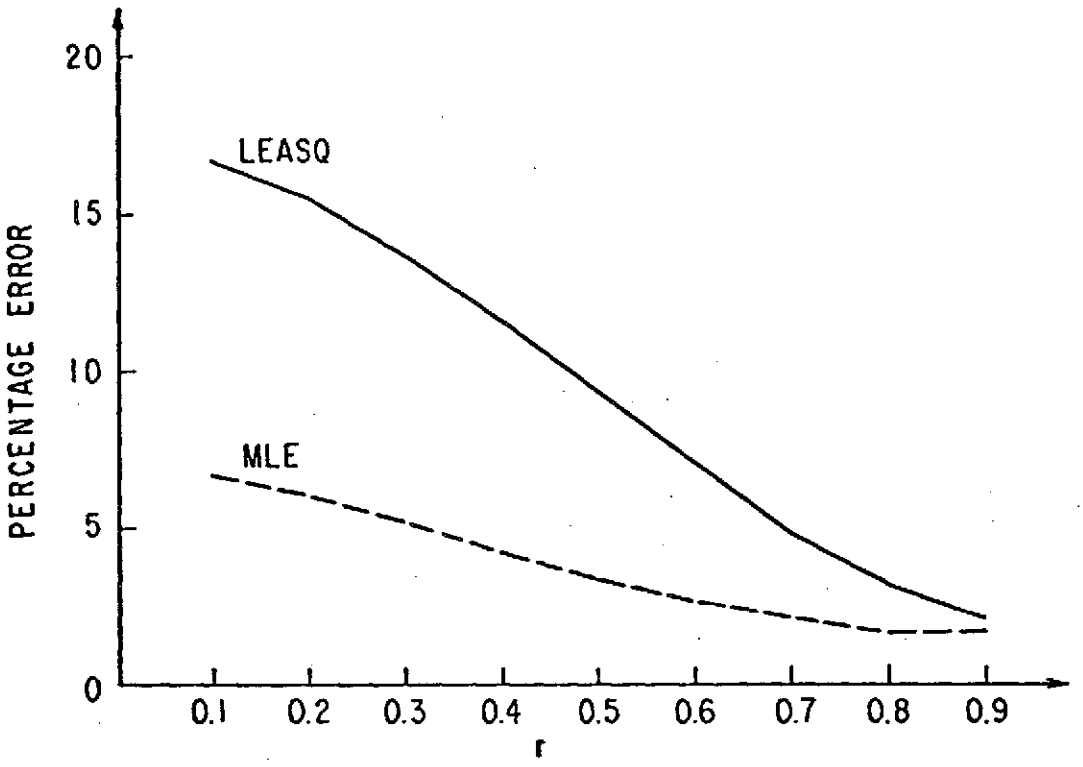


Figure 3.18. Percentage squared error of the restoration of a Gaussian-shaped object as functions of assumed correlated coefficient r. $\alpha sT = 10$, $I_o = 10$, $\sigma = 15$, $W_o = 0.8$.

In summary, we have developed a method of restoring images that takes account of the Poisson character of the recorded image data, instead of making the assumption of Gaussian-distributed data that underlies the usual linear least-squares procedure. The significance of the new restoration procedure has been demonstrated by computer simulations. In the great majority of trials the new method produced a much better fit to the true object radiance than the conventional linear procedure.

Footnotes to Chapter 3

1. Jenkins and Watts [28], p. 162.
2. Parzen [34], p. 334.
3. Helstrom [5].
4. Born and Wolf [33], p. 484.
5. Goodman [35], p. 113.
6. Kowalik and Osborne [87], p. 34.
7. Aoki [86], p. 101.
8. Himmelblau [13], p. 64.
9. Himmelblau [13], p. 67.
10. Kowalik and Osborne [87], p. 41.
11. Himmelblau [13], p. 96.
12. Goldfarb and Lapidus [88].
13. Himmelblau [13], p. 299.

Part II. Optical Signal Processing by Shearing-Interferometric Methods

In part II, we are concerned with signal processing using coherent optical systems.

Communication theory has been successfully applied to optical signal processing, particularly in the areas of holography and spatial filtering. Many optical phenomena can be easily explained using communication theory. Carrier-frequency photography is an amplitude-modulation (AM) process; computer generated holography is just a pulse-width (PWM) and pulse-position modulation (PPM) scheme. Diffraction by a grating is a harmonics generation process, and the moiré effect can be understood by modulation theory. Modulation, demodulation, heterodyning, generation of harmonics, and filtering have been integrated into communication theory. In optics, however, there are as yet few connections between grating diffraction, the moiré effect, carrier-frequency photography, holography, and interferometry. Here we want to integrate these processes and apply them to optical signal processing.

Following an introduction and a brief review of the pertinent details of grating diffraction, Ronchi's interferometer, the moiré effect, carrier-frequency photography, and achromatic holography, two types of interferometers are presented. In these interferometers, diffraction gratings are used at either the spatial-frequency plane or

the object plane of a telecentric image-forming system. In the grating interferometer, gratings are placed at the spatial-frequency plane as beam splitters, and in the carrier-frequency interferometer gratings are placed at the object planes of two cascade telecentric image-forming systems as modulation and demodulation masks. These interferometers can produce variable shearing in both lateral and radial directions. The operation of the grating interferometer is in real time; the operation of the carrier-frequency interferometer can be either in real time or in two steps.

Optical testing, phase-object visualization, spatial differentiation, complex amplitude addition and subtraction, image contrast reversal, image complementary color reversal, two-dimensional photometry, fabrication of filters, pure spatial sinusoidal wave generation, and image multiplexing can all be achieved by using these new interferometers. The methods employing the new interferometers have some definite advantages in terms of simplicity, flexibility, reliability, light efficiency, or real-time operation in comparison with existing methods.

Optics and electronics are means of information processing which have many common features. Communication theory is well developed in electronic signal processing. Recently it has been successfully adapted by optical physicists, especially in the development of holography and spatial filtering. Many optical phenomena can be easily understood from the point of view of communication theory. Modulation, demodulation, heterodyning, harmonics generation, and

filtering have been used successfully in electronic signal processing; we can find their counterparts in optics. In optics, modulation can be achieved by carrier-frequency photography; generation of harmonics by grating diffraction; demodulation and heterodyning by the moiré effect; and filtering by spatial filters. Carrier-frequency photography, diffraction by gratings, and the moiré effect all have their unique functions in optical signal processing. Two new types of interferometers have been developed from the combined use of those concepts.

Diffraction by gratings and the moiré effect have been used extensively in optics. Gratings have been used as beam splitters. Ronchi invented a shearing interferometer using a grating as a beam divider, which has been used extensively for testing optical components such as lenses. This technique is relatively simple and inexpensive. But the use of just one grating results in many drawbacks, such as (a) the fixed amount of shear, (b) the dependence of the fringe contrast on the degree of equality in light flux between different diffraction orders, and (c) the overlap of many diffraction orders.

The drawbacks of Ronchi's interferometer can be overcome by the use of two superimposed gratings instead of just one grating as a beam divider. The amount of shear in Ronchi's interferometer is proportional to the spatial frequency of the diffraction grating used. If we can vary the spatial frequency, we can have variable shear. This can be done by the moiré effect of the two superimposed gratings. When

two gratings are superimposed, owing to the moiré effect a variable spatial frequency is produced by rotating the angle between their axes. An interferometer, called the "grating interferometer," has been developed from two superimposed Ronchi rulings used as beam splitters. All the drawbacks in Ronchi's interferometer are overcome.

Carrier-frequency photography has some similarity to holography. In particular, both are two-step processes, and diffraction methods are generally used in the retrieval stages. Holography has been used very successfully in interferometry. Holographic interferometry can be operated either in real time or in two stages.

A hologram's ability to record both the amplitude and phase of a wavefront is the reason for its success in interferometry. Carrier-frequency photography can register the amplitude as well as the phase variation of objects. When a grating is imaged through a complex object, the resultant image is an amplitude-modulated version of the grating with the slits shifted according to the phase variation of the object. This kind of encoding scheme, involving both amplitude-modulation and pulse-position modulation, called the "detour phase" by A. Lohmann, is the essence of computer-generated holography.[2] It is very difficult to retrieve the information about the phase variation of the object directly from the distorted grating image. But if we "beat" the distorted grating image with a regular grating (or master grating), moiré fringes appear which will reveal the phase variation. A new

type of interferometer, based on carrier-frequency photography and the moiré effect, has been conceived. It is called the carrier-frequency interferometer. It can be operated in real time or in two stages. The system is achromatic and white light can be used.

In chapter 4, we briefly review diffraction by gratings, the moiré effect, carrier-frequency photography, and achromatic holography. This review will serve as the foundation of the following chapters.

The grating interferometer is the subject of chapter 5, and the carrier-frequency interferometer is described in chapter 6. Many kinds of masks can be used as modulation carriers in the carrier-frequency interferometer. We use linear gratings, checkerboard gratings, and circular gratings.

The applications of these two types of interferometers in optical information processing are presented in both chapters 5 and 6 for each interferometer. New methods for optical testing, phase object visualization, image polarity reversal and color complementarity, two-dimensional photometry, spatial differentiation, complex amplitude addition and subtraction, image multiplexing, pure spatial sinusoidal wave generation, fabrication of filters, focusing, and alignment are presented. Some of the applications can be used in optical computing, general image processing, and two-dimensional photometry.

Chapter 4

Review of Diffraction by Gratings, the Ronchi Interferometer, the Moiré Effect, Carrier-Frequency Photography, and Achromatic Holography

Gratings have long been used as beam splitters, particularly in the Ronchi interferometer. The moiré effect has been applied to measurement processes, carrier-frequency photography has been applied to image multiplexing, and holography has been applied in interferometry. They all have their own unique applications to optical information processing. However, few efforts have been made to integrate all these techniques. It is the goal of this part of the thesis to integrate them for use in optical information processing. In this chapter we will review them briefly.

4-1. Gratings, Diffraction by Gratings, and Ronchi Interferometer

One type of grating can be represented as a periodic binary mask. A grating with slits parallel to the y-axis can be represented as

$$M(x) = \begin{cases} 1, & |(x-a)/d - N| \le q/2 \\ 0, & \text{otherwise} \end{cases}, \quad N = 0, 1, 2, \ldots \qquad (4.1)$$

where d is the period of the grating, a is the shift from the origin of symmetry, and q is the ratio of the size of the openings (of unit transmittance) to the period of the grating. For the Ronchi ruling, $q = 1/2$,

so that the size of the opening is equal to half the period of the grating.

M(x) is a periodic function, which can be expressed by means of Fourier series, †

$$M(x) = \sum_n A_n e^{2\pi i n x/d} \quad , \tag{4.2}$$

where $A_n = q\ \mathrm{sinc}(nq)\ \exp(2\pi i na/d)$,

$$\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$$

$$A_n = A^*_{-n}$$

(here * denotes the complex conjugate). The binary grating M(x) and its Fourier series spectrum for the case when $q = 1/2$, are shown in Fig. 4.1. No even harmonics exist.

If we rotate the grating by an angle θ about the origin, its transmittance function becomes

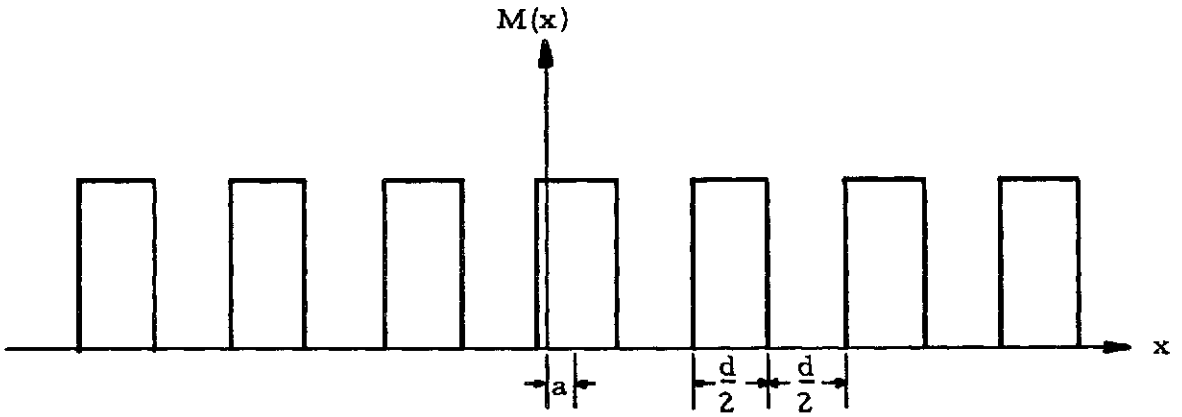
$$M(x, y) = \sum_n A_n \exp[2\pi i\ n(x \cos\theta + y \sin\theta)/d] \ . \tag{4.3}$$

The grating has the frequency $\cos\theta/d$ in the x-coordinate and $\sin\theta/d$ in the y-coordinate.

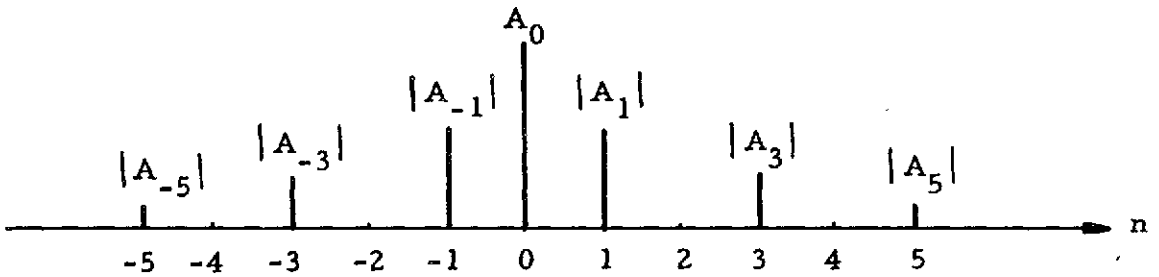
The periodic binary mask M(x) can be generalized to include quasi-periodic masks, e.g. the circular grating,

---

† From now on, $\sum_n$ means summation over n from $-\infty$ to $+\infty$.



(a)



(b)

Figure 4. 1. A Ronchi ruling and its Fourier spectrum .
(a) The transmittance function of a Ronchi ruling.
(b) The Fourier coefficients of the Ronchi ruling; no even harmonics exist.

$$C(r) = \begin{cases} 1, & |r/a - N| \le q/2 \\ 0, & \text{otherwise} \end{cases} \qquad N = 0, 1, 2, \ldots \qquad (4.4)$$

When a monochromatic plane wave in z-direction is used as incident wave, the wavefronts diffracted by a transparent one-dimensional plane grating have their wave vectors in the plane of both the original wave vector and the direction perpendicular to the direction of the slits (or in the direction of grating orientation). The grating splits a wavefront incident on it into many diffraction orders, according to the relation $m\lambda/d = \sin\theta_m$, where m is the diffraction order, λ the wavelength, θ_m the angle between the zeroth order and the m-th order. The amplitude of m-th diffraction order is A_m.

When a transparent one-dimensional plane grating is placed in the spatial-frequency plane of a telecentric image-forming system as shown in Fig. 4.2, the grating acts as an amplitude-modulation mask in the spatial-frequency plane. The grating will generate harmonics of the object. By Fourier analysis, the field propagated through the grating will split into many orders in the image plane. Ronchi invented an interferometer using a grating as beam splitter. Figure 4.2 is a typical setup. A monochromatic plane wave illuminates the object. The field just before the grating, owing to the Fourier transforming property of lens L_1, is

$$U(x, y; 2f_1 - 0) = \tilde{U}_0(x/\lambda f_1, y/\lambda f_1), \qquad (4.5)$$

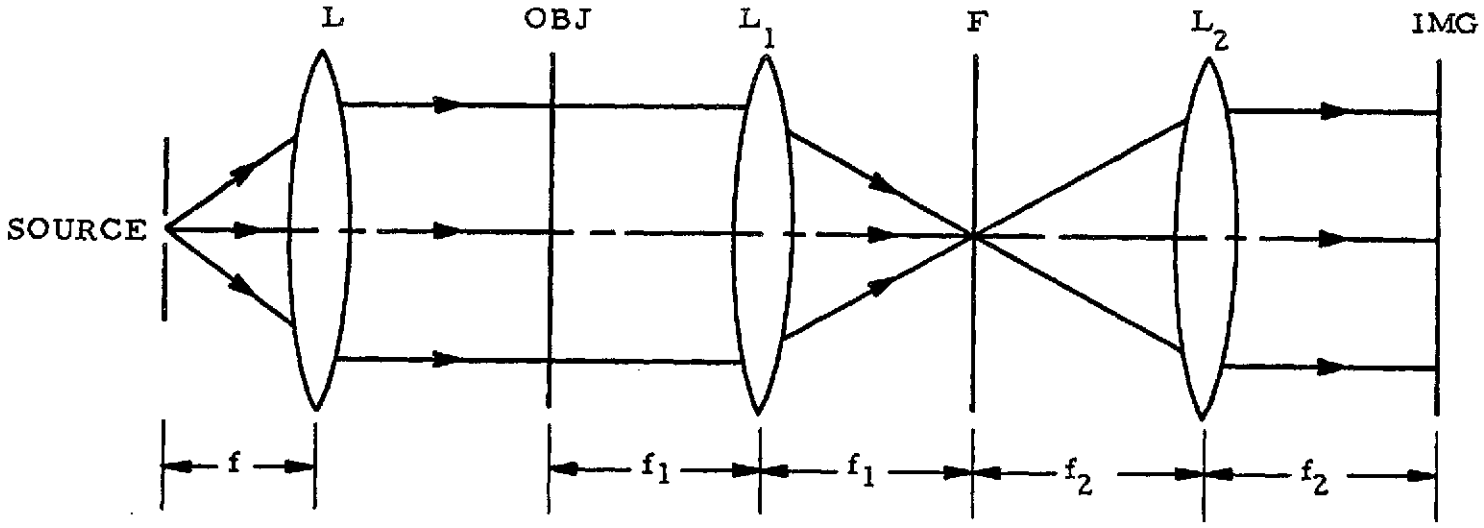


Figure 4.2. A coherent optical signal processor (or telecentric image-forming system). OBJ is the plane where the input object will be located, F is the filter plane where the spatial filter is located, and IMG is the image plane, f_1 and f_2 are the focal length of lenses L_1 and L_2 respectively.

where

$$\tilde{U}_0(x, y) = \iint U_0(x', y') \exp[-2\pi i(xx' + yy')]\, dx'\, dy' . \qquad (4.6)$$

We have omitted an unimportant multiplicative constant. Applying the Kirchhoff boundary condition to the field as it passes through the grating, we have

$$U(x, y; 2f_1 + 0) = U(x, y; 2f_1 - 0)\, G(x, y)$$

$$= \tilde{U}_0(x/\lambda f_1, y/\lambda f_1) \sum_n A_n \exp(2\pi i\, nx/d), \qquad (4.7)$$

where $G(x, y) = \sum_n A_n \exp(2\pi i\, nx/d)$ is the grating transmittance function. The field propagates through lens L_2 and reaches the image plane. Owing to the Fourier transforming property of lens L_2 and the convolution theorem of Fourier transformation, the field at the image plane is

$$U(x, y; 2f_1 + 2f_2) = \tilde{U}(x, y; 2f_1 +)$$

$$= \sum_n A_n U_0(-f_1 f_2^{-1} x - n\lambda f_1 d^{-1}, y) . \qquad (4.8)$$

The object field has been reversed, magnified f_2/f_1 times, and diffracted into many orders with amplitude A_m for the m-th diffraction order, and shifted over a distance $m\lambda f_2/d$.

If the amount of shear, $\lambda f_1/d$, is smaller than the size of the object, fields coming from different diffraction orders will overlap. As all the fields from different diffraction orders are coherent, interference

fringes will appear in the overlapped area. This interferometer has been used extensively in optical testing.

4.2. The Moiré Effect

The moiré effect was described by Lord Rayleigh[1] as long ago as 1874. When two high-spatial-frequency gratings are superimposed, by proper arrangement, a pattern with low-spatial-frequency variation appears.

Two linear gratings $T_1(x, y)$ and $T_2(x, y)$ have transmittance functions

$$T_1(x, y) = \sum_n A_n \exp[2\pi i n(\nu_1 x + \mu_1 y)] ,$$

and (4.9)

$$T_2(x, y) = \sum_m B_m \exp[2\pi i m(\nu_2 x + \mu_2 y)] ,$$

where (ν_1, μ_1) and (ν_2, μ_2) are fundamental spatial-frequencies in (x, y) of gratings T_1 and T_2 respectively. When they are superimposed, the resulting transmittance is

$$T(x, y) = T_1(x, y)\, T_2(x, y)$$

$$= \sum_m \sum_n A_n B_m \exp\{2\pi i [(n\nu_1 + m\nu_2)x + (n\mu_1 + m\mu_2) y]\} . \quad (4.10)$$

Examining the Fourier spectrum of the transmittance functions of T_1, T_2, and T, we see that T_1 has the fundamental frequency $(\nu, \mu) = (\nu_1, \mu_1)$, T_2 has (ν_2, μ_2), and the resulting transmittance T has $(\nu_1 - \nu_2, \mu_1 - \mu_2)$, which is the beat frequency of T_1 and T_2.

4.2.1. Moiré of Two Linear Gratings of Different Period, Parallel Orientation

Two gratings T_1 and T_2 both have their slits parallel to y-axis, and have spatial frequencies ν_1 and ν_2 respectively. The total transmittance of these two gratings superimposed, by eq. (4.10)), is

$$T(x, y) = \sum_m \sum_n A_n B_m e^{2\pi i(n\nu_1 + m\nu_2)x} . \qquad (4.11)$$

When both T_1 and T_2 are Ronchi rulings, $q = 1/2$, no even harmonics exist. For the case that $\nu_1 \approx \nu_2$, and $q = 1/2$, there are only two diffraction orders $(n = 1, m = 0)$ and $(n = 0, m = 1)$ close to $\nu = \frac{1}{2}(\nu_1 + \nu_2)$, as shown in Fig. 4.3.

4.2.2. Moiré of Two Linear Gratings with Same Frequency, Different Orientations

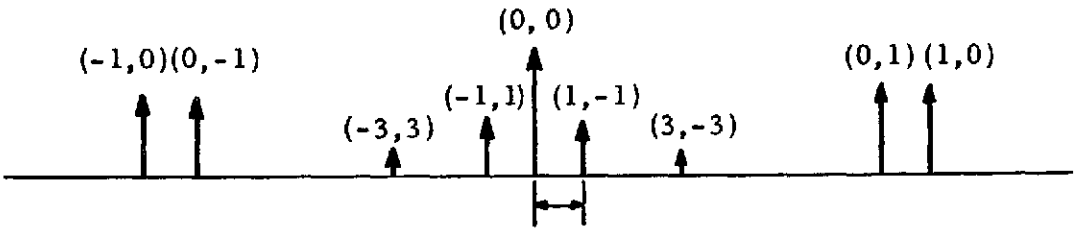
When the two gratings T_1 and T_2 have the same period d, but their orientations are different, one grating having its slits rotated through an angle θ from the x-axis, and the other having its slits rotated through an angle $-\theta$ from the x-axis, the total transmittance is

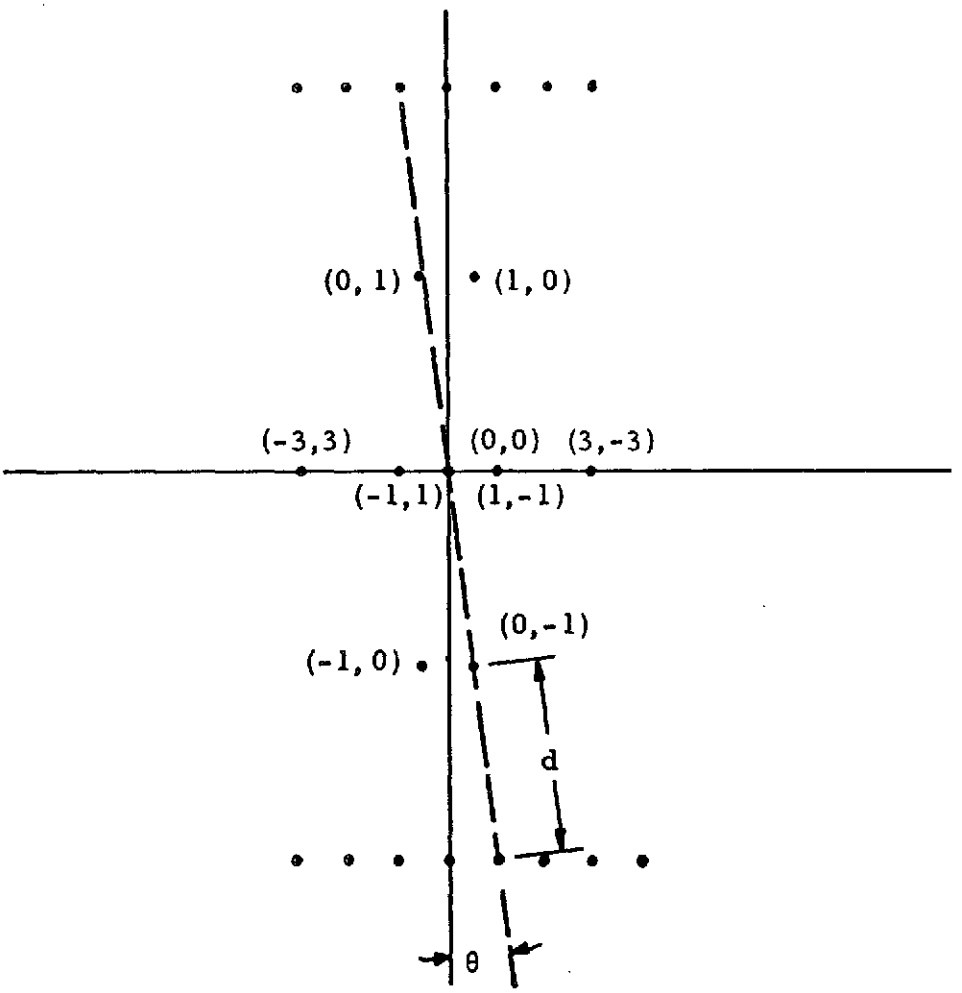
$$T(x, y) = \left\{ \sum_n A_n \exp[2\pi i n(x \sin\theta + y \cos\theta)/d] \right\} \left\{ \sum_m B_m \exp[2\pi i m(-x \sin\theta + y \cos\theta)/d] \right\}$$

$$= \sum_n \sum_m A_n B_m \exp[2\pi i (x\nu_{nm} + y\mu_{nm})] , \qquad (4.12)$$

Figure 4.3. The Fourier spectrum of two superimposed Ronchi gratings of different periods, parallel orientation. There are only two diffraction orders (1, 0) and (0, 1) that are nonzero and satisfy $m + n = 1$.

Figure 4.4. The Miller's indices. For Ronchi rulings there are only two diffraction orders that are nonzero and satisfy $m + n = 1$ and -1.


(0, 0)
(-1,0)(0,-1)
(-1,1)
(1,-1)
(-3,3)
(3,-3)
(0,1) (1,0)


(0, 1)
(1, 0)
(-3,3)
(0,0)
(3,-3)
(-1,1)
(1,-1)
(0,-1)
(-1, 0)
d
θ

where

$$\nu_{nm} = (n - m)\, d^{-1} \sin\theta$$

$$\mu_{nm} = (n + m)\, d^{-1} \cos\theta .$$

The Fourier spectrum of $T(x, y)$ can be visualized in the (ν, μ) domain with the aid of "Miller's indices" (n, m). Figure 4.4 shows the location of different indices in the (ν, μ) plane, for the case that both T_1 and T_2 are Ronchi rulings. The amplitude strength of each diffraction order (n, m) is $A_n B_m$. All the diffraction orders with $n + m = p$ are located at $\mu = pd^{-1} \cos\theta$, with a distance $2d^{-1} \sin\theta$ between the neighboring diffraction orders. If we vary θ, the distance $2d^{-1} \sin\theta$ is varied. The distance $2d^{-1} \sin\theta$ is the beat frequency produced by T_1 and T_2 superimposed. When both T_1 and T_2 are Ronchi rulings, no even harmonics exist, i.e., $A_n B_m = 0$, for n or m an even number. There are only two diffraction orders, $(0, 1)$ and $(1, 0)$, with $m + n = 1$, as shown in Fig. 4.4.

4.2.3. Moiré of One Regular Grating and One Distorted Grating

When a regular grating $G_1(x, y) = a + b \cos(\nu x)$ and a distorted version of that grating $G_2(x, y) = a + b \cos[\nu x + \theta(x, y)]$ are superimposed, the total transmittance is

$$T_p(x, y) = G_1(x, y)\, G_2(x, y)$$

$$= a^2 + \frac{1}{2} b^2 \cos\theta + 2ab \cos\left(\frac{\theta}{2}\right) \cos\left(\nu x + \frac{\theta}{2}\right) + \frac{b^2}{2} \cos(2\nu x + \theta) , \tag{4.13}$$

where $\theta(x, y)$ is related to the distortion of position of the grating slits. The information which we wish to observe, $\theta(x, y)$, can be extracted by spatial filtering if we let pass only the first diffraction order in Eq. (4.13).

If we add the transmittance functions of G_1 and G_2 together, the total transmittance is

$$T_A(x, y) = 2a + \frac{b}{2} \cos\left(\nu x + \frac{\theta}{2}\right) \cos\left(\frac{\theta}{2}\right) \quad . \tag{4.14}$$

The addition of transmittances can be achieved optically by a double-exposure technique. The information about the distortion function appears as the envelope of the carrier wave $\cos(\nu x + \theta/2)$.

4.3. Carrier-Frequency Photography

In communication, modulation theory has been used in connection with the transmission of information. Several signals can be transmitted over a single cable at the same time, by either time multiplexing or frequency multiplexing. In optics, we can do essentially the same thing for multiplex information storage, by which we mean that more than one image is stored in a single film frame. By analogy, optical image multiplexing can be achieved by either spatial multiplexing or spatial-frequency multiplexing. Spatial multiplexing means each single image occupies a certain part of the film frame, which can be done by either a microfilm technique or image interlacing. In optical spatial-frequency multiplexing, each single image is carried by carriers

of different spatial frequencies. As a different orientation of the same grating yields different spatial frequencies, we have a theta-modulation system.[2]

Holography can record both phase and amplitude of the object wavefront and has been successfully applied in interferometry. Carrier-frequency photography has some similarities to holography, but the recording procedure is different. We will apply carrier-frequency photography to interferometry.

The basic recording and reconstruction setups for carrier-frequency photography are shown in Fig. 4.5. In recording, the object is imaged onto a film, and the modulation mask (a Ronchi ruling) can be placed in either the film plane or the object plane. We can build a camera with a grating in contact with the film plane. In reconstruction, coherent light is used to illuminate the film, and a binary filter is placed in the filter plane. This binary filter passes just one of the diffraction orders.

If we want to store more than one signal, e.g. four signals of intensity distribution functions $I_1(x, y)$, $I_2(x, y)$, $I_3(x, y)$ and $I_4(x, y)$, in a single film frame, we can use theta modulation.[3] Four exposures are made, each exposure for a different object. Between each exposure, the grating is rotated through an angle $\theta = 45°$, hence the designation "θ - modulation." The total exposure is

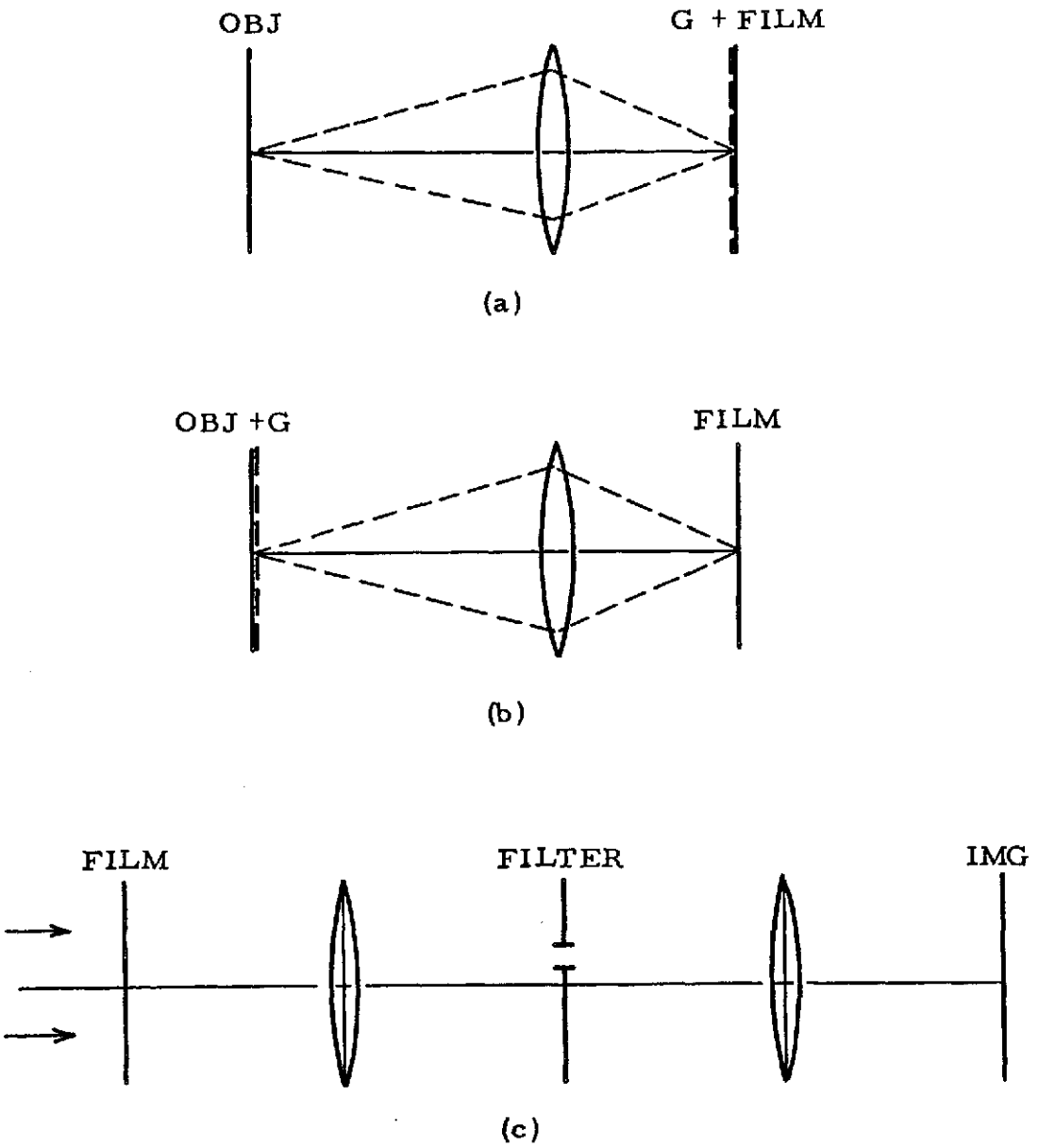


Figure 4.5. The recording and reconstruction setup for carrier-frequency photography. (a) and (b) are for recording and (c) is for reconstruction. The filter in (c) passes only one diffraction order.

$$I(x, y) = I_1(x, y)\, G(x, y; 0) + I_2(x, y)\, G(x, y; \pi/4) + I_3(x, y)\, G(x, y; \pi/2)$$

$$+ I_4(x, y)\, G(x, y; 3\pi/4) \quad , \qquad (4.15)$$

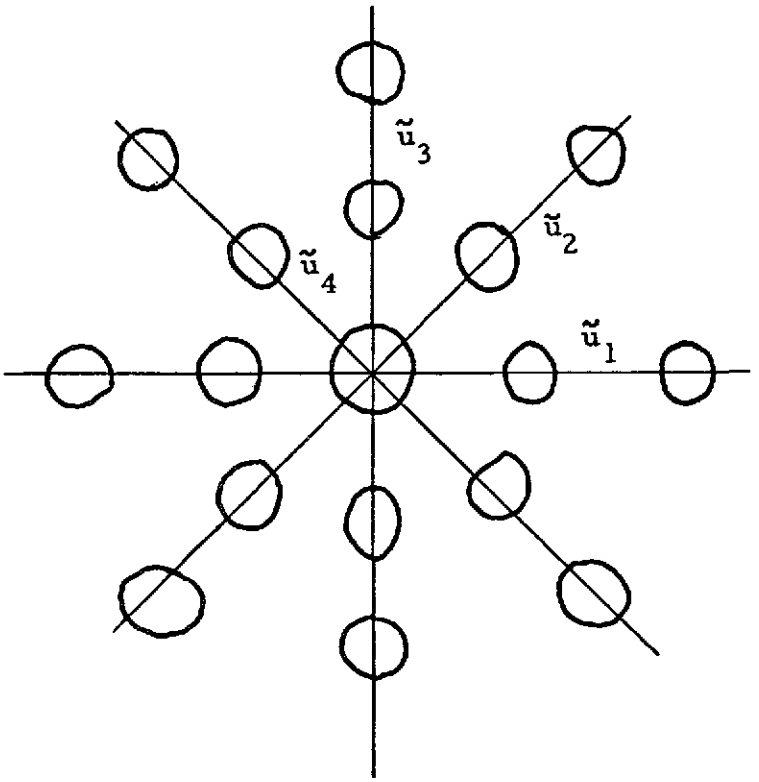
where $G(x, y; \theta)$ is the grating transmittance function rotated through an angle θ. When the film is developed in the linear portion of the characteristic T-E curve, the amplitude transmittance of the film will be proportional to $I(x, y)$.

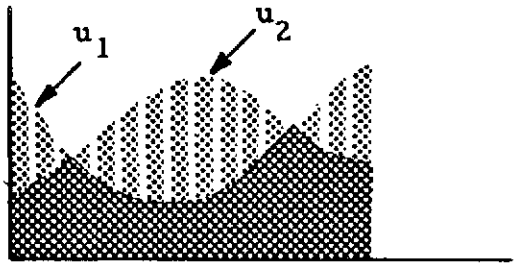
In reconstruction, the developed film is placed in the object plane of a telecentric image forming system as shown in Fig. 4.5(c). The spectrum of $I(x, y)$ in the filter plane will be similar to the one shown in Fig. 4.6. The spectra of different objects occupy lines of different angles θ, each split into many diffraction orders. Each signal can be retrieved separately, or the signals can be combined by using a binary filter to pass only the spectra of the images to be displayed. This technique has been successfully applied to color photography using only black and white film.[3]

Image multiplexing can be achieved by spatial interlacing as well as by theta modulation. In spatial interlacing, the grating is displaced a distance qd between each exposure; q as defined before is the ratio of the slit width to the period of the grating. We can record q^{-1} signals and retrieve them separately. For a Ronchi ruling, $q = 1/2$, we can record two signals in one film frame. The total transmittance is

Figure 4.6. The spectrum of a theta-modulated carrier-frequency photograph.

Figure 4.7. A double-exposure carrier-frequency modulated photograph.


$\tilde{u}_3$
$\tilde{u}_2$
$\tilde{u}_4$
$\tilde{u}_1$


u_1
u_2

$$I(x, y) = I_1(x, y)\, G(x, y) + I_2(x, y)\, G(x - d/2, y)$$

$$= \sum_n A_n [I_1(x, y) + (-1)^n I_2(x, y)] \exp(2\pi i nx/d) . \qquad (4.16)$$

$I(x, y)$ is shown in Fig. 4.7. In reconstruction, we can put the modulation mask in contact with the developed film to pass only either $I_1(x, y)\, G(x, y)$ or $I_2(x, y)\, G(x - d/2, y)$, and retrieve them individually. Or if we use no mask, and if the filter F passes only the first diffraction order, $|I_1(x, y) - I_2(x, y)|^2$ results. We create bipolarity by spatial interlacing. This technique has been used in image subtraction[4] and image polarity reversal.[5]

4.4. Achromatic Holography

Because of the annoying twin image, holography was not very popular after Gabor invented it in 1948. The renaissance of holography came after Leith and Upanieks showed that the twin image can be eliminated by the use of an offset reference wave. The typical setup is shown in Fig. 4.8. A point source of illumination is collimated by the lens L. The beam splitter BS divides the beam into two parts, with one part of the beam striking the object, which is taken to be a transparency with a complex transmittance $O(x, y)$. The second portion of the plane wave strikes a mirror and reflects into the film with an angle θ with respect to the optical axis of the lens L.

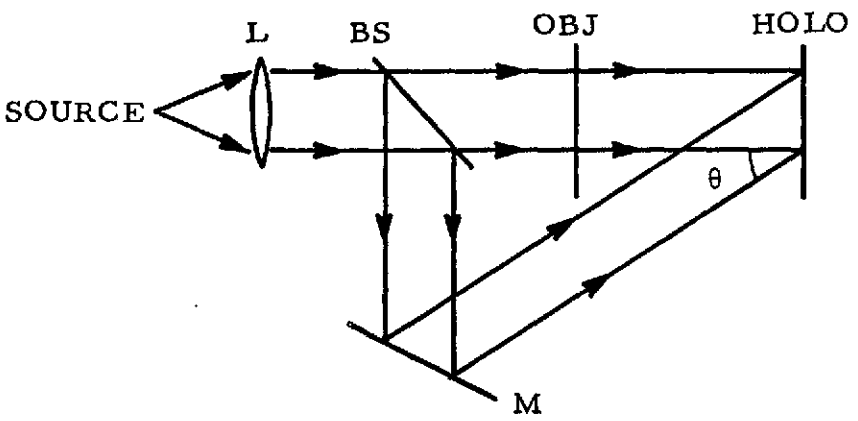


Figure 4.8. A recording setup for holography. BS is beam splitter, HOLO is hologram plane, M is mirror.

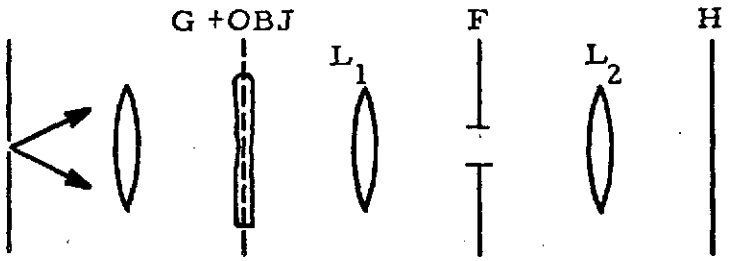


Figure 4.9. A recording setup for achromatic holography. The filter F consists of a pinhole at the center of the zeroth diffraction order and a binary filter at the first diffraction order.

The amplitude distribution across the film can be written as

$$U(x, y) = R\ e^{2\pi i \alpha y} + a(x, y)\ e^{+i\phi(x, y)} \quad ,$$

where R is the amplitude strength of the reference wave, $a(x, y)$ and $\phi(x, y)$ are the amplitude and phase of the wave originating from the object, and $\alpha = \sin\theta/\lambda$ is the spatial frequency of the reference wave. The intensity distribution across the film is

$$E(x, y) = |U(x, y)|^2$$

$$= R^2 + |a(x, y)|^2 + 2R\, a(x, y) \cos[2\pi\alpha y - \phi(x, y)] \quad .$$

Since α (and possibly $a(x, y)$ and $\phi(x, y)$) is dependent on wavelength λ, each wavelength λ produces its own elementary fringe system, and the systems of different wavelengths are shifted with respect to each other. Monochromatic light is necessary.

If the interference fringes on the hologram are achromatic, a broad-band source such as natural white light emitted from a xenon-arc lamp can be used as illuminating source. To produce an achromatic hologram, it is sufficient that (1) the interference fringes on the hologram be achromatic, and (2) the diffraction pattern from the object be fairly independent of wavelength λ in the plane of the hologram.[6] These requirements can be achieved by the optical setup shown in Fig. 4.9.

The object $U_0(x, y)$ is modulated by a grating G. By the

Fourier transforming property of lens L_1, the field at the filter plane F, for each wavelength is

$$U(x, y; \lambda ; 2f - 0) = \sum_n A_n \tilde{U}_0(x - n \lambda f/d, y) , \tag{4.18}$$

where f is the focal length of both lenses L_1 and L_2. If the illuminating source is polychromatic, each diffraction order except the zeroth order will exhibit color dispersion. We can use a binary filter F to pass the +1 diffraction order as the object wave, and a small pinhole in the center of the zeroth diffraction order to create the reference wave. The field after the filter plane is

$$U(x, y; \lambda ; 2f + 0) = R\delta(x) + A_1 \tilde{U}_0(x - \lambda f d^{-1}, y) , \tag{4.19}$$

where R is the strength of the reference wave. The object waves produced by light of different wavelengths are shifted with respect to each other in the filter plane F, but they will be recombined if the wave propagates to the hologram plane H as shown in Fig. 4.9. The total field in the hologram plane H is

$$U(x, y; \lambda ; 4f) = R + A_1 U_0(-x, -y) \exp(2\pi i x/d) , \tag{4.20}$$

which is λ-independent. Thus an achromatic hologram is recorded.

White light can be used in holography. We will use the achromatic holographic technique in a new type of interferometry, which can be used in several ways, especially for real-time image contrast reversal. Positive black- and-white images can be obtained directly from

color negatives in real time without complicated photo-chemical and electronic techniques.

Footnotes to Chapter 4

1. Guild [51], chapter 1.

2. Armitage and Lohmann [18].

3. Mueller [52].

4. Pennington, Will, and Shelton [53].

5. Bryngdahl [54].

6. Bryngdahl and Lohmann [55].

Chapter 5

Variable Shearing Interferometry Based on the Moiré Effect of Gratings - The Grating Interferometer

5.1. Introduction

Shearing interferometry has been used to test instruments and optical components. In this type of interferometry, the wavefront under test is duplicated and displaced. The two resulting wavefronts are made to interfere with each other, yielding a comparison of the wavefront with a shifted image of itself. These interferometers typically used a single beam-dividing element that is simpler than other interferometers.[1] There are two ways of obtaining the shifted wavefronts. The first method is to have the shift sufficiently large so that the displacement between the duplicated wavefronts is larger than the dimension of the wavefront under test. This is called "total shear," since the amount of shift is larger than the size of the object wavefront under test. In the second method the shift is kept smaller than the width of the typical details of the object under observation; this is known as "differential shear," owing to the similarity to differentiation. It is desirable to have an interferometer which can vary the amount of the shift.

The interferometer invented by Ronchi uses a grating as the beam divider. This interferometer is simpler and cheaper than most other methods and has been used extensively. However, as mentioned before, the use of only one grating causes some inconveniences. These

can be overcome by using two gratings as a beam splitter instead of just one. Lohmann and Bryngdahl have described an interferometer using two diffraction gratings in the spatial frequency plane of two cascaded telecentric image-forming systems.[2]

5.2. Principle of the Interferometer

We are proposing a new shearing interferometer based on the moiré effect of two superimposed gratings. The optical arrangement of the grating interferometer is schematically illustrated in Fig. 5.1.

A plane wave falls onto the object plane OBJ, which is imaged via telecentric image-forming lenses L_1 and L_2 onto the image plane IMG. The two superimposed gratings G_1 and G_2 are placed in the spatial-frequency plane. Both gratings G_1 and G_2 are linear and have the same period d; one has its slits rotated through an angle θ with respect to the x-axis; the other has its slits rotated through an angle $-\theta$. The transmittance of the superimposed gratings, as described by Eq. (4.12), is

$$T(x, y) = \sum_m \sum_n A_n B_m \exp[2\pi i (x \nu_{nm} + y \mu_{nm})] , \qquad (5.1)$$

where

$$\nu_{nm} = (n - m)\, d^{-1} \sin\theta$$

$$\mu_{nm} = (n + m)\, d^{-1} \cos\theta ,$$

and A_n and B_m are the Fourier coefficients of gratings G_1 and G_2

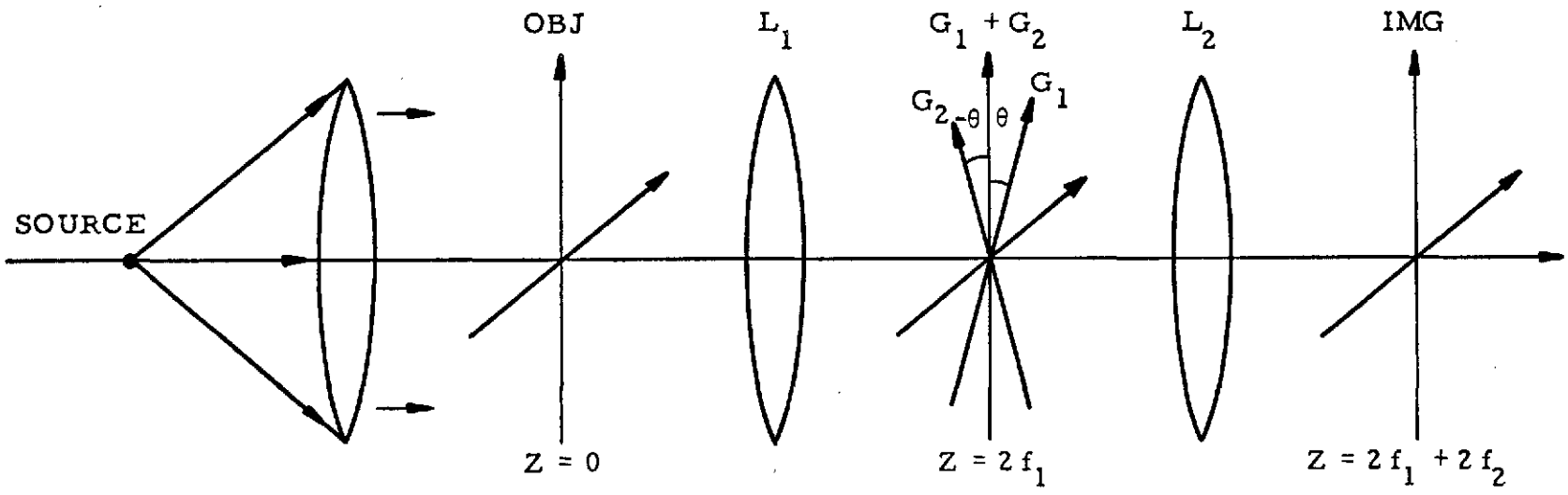


Figure 5.1. The grating interferometer.

respectively. The wavefront diffracted by this composite grating will split into many diffraction orders with the centers of each diffraction order distributed like the "Miller's indices," as shown in Fig. 4.4. Variation of the angle of rotation θ changes the distance between these centers, and hence the amount of shift between the different diffraction orders.

The object $U_0(x, y)$ is illuminated by a collimated plane wave. The field in the plane just preceding the grating, by the Fourier transforming property of lens L_1, is[3]

$$U(x, y; 2f_1 - 0) = \tilde{U}_0(x/\lambda f_1, y/\lambda f_1) \tag{5.2}$$

where f_1 is the focal length of lens L_1. By the application of the Kirchhoff boundary condition, the field just behind the grating is

$$U(x, y; 2f_1 + 0) = \sum_n \sum_m A_n B_m \tilde{U}_0(x/\lambda f_1, y/\lambda f_1) \exp[2\pi i(x\nu_{nm} + y\mu_{nm})]. \tag{5.3}$$

This field will be once again Fourier transformed when it propagates through lens L_2 and arrives at the image plane. By the use of the shift property of Fourier transformations, the field in the image plane can be written as

$$U(x, y; 2f_1 + 2f_2) = \sum_n \sum_m A_n B_m U_0\left(-\frac{f_1}{f_2} x - \nu_{nm}\lambda f_1, -\frac{f_1}{f_2} y - \mu_{nm}\lambda f_1\right), \tag{5.4}$$

where f_2 is the focal length of lens L_2. Equation (5.4) can be written as

$$U(x, y; 2f_1 + 2f_2) = \sum_p U(x, y; p) \quad , \tag{5.5}$$

and

$$U(x, y; p) = \sum_n A_{\frac{n+p}{2}} B_{\frac{p-n}{2}} U_0(-M^{-1}x - n\Delta_x, -M^{-1}y - p\Delta_y) \quad , \tag{5.6}$$

$$M = f_2/f_1 \quad ,$$

$$\Delta_x = \lambda f_1 d^{-1} \sin\theta \quad ,$$

$$\Delta_y = \lambda f_1 d^{-1} \cos\theta \quad ,$$

and only A_i's and B_j's indexed by integers are meaningful; when p is an odd number, n must be an odd number, in order that $A_{\frac{n+p}{2}} B_{\frac{p-n}{2}}$ be meaningful. For diffraction orders located at the same y-coordinate, the distance between each neighboring order is $2M\Delta_x = 2\lambda f_2 d^{-1} \sin\theta$. All the diffraction orders $U(x, y; p)$ appear at the same time in the image plane with their y-coordinate shifted by $-Mp\Delta_y$.

The object field in the image plane has been inverted and magnified by a factor $M = f_2/f_1$, and the center of the diffraction order (n, m) is at $x = (n - m) M\Delta_x$, $y = (n + m)M\Delta_y$. If the dimension δ_y of the object in the y-direction is smaller than Δ_y, all the diffraction orders with $m + n = p$ will appear separately from those with $m + n = p'$, $p' \neq p$. We have achieved total shift in the y-direction.

If we use Ronchi rulings, $q = 1/2$, then no even harmonics exist, i.e., $A_n = B_n = 0$, for n even. For diffraction order $p = n + m = 1$, (n, m) can have values $(1, 0)$, $(0, 1)$, $(2, -1)$, $(-1, 2)$. Only diffraction orders $(1, 0)$ and $(0, 1)$ do not have even harmonics, hence can be nonzero, and

$$U(x, y; p = 1) = A_0 B_1 U_0(-M^{-1}x + \Delta_x, -M^{-1}y - \Delta_y)$$

$$+ A_1 B_0 U_0(-M^{-1}x - \Delta_x, -M^{-1}y - \Delta_y) . \qquad (5.7)$$

The $U(x, y; p = 1)$ diffraction order is centered at $y = -M\Delta_y$. The distance between the $(0, 1)$ and $(1, 0)$ diffraction orders is $2M\Delta_x$, which varies with the angle θ. The new interferometer is thus a variable shearing interferometer with only two diffraction orders overlapping. If gratings G_1 and G_2 are identical, $|A_n| = |B_n|$, hence $|A_0 B_1| = |A_1 B_0|$; the two diffraction orders $(0, 1)$ and $(1, 0)$ have the same strength. The phase difference between A_n and B_n can be adjusted by shifting either G_1 or G_2 in the direction perpendicular to its slits. All the inconveniences of the Ronchi interferometer are overcome.

The total output image field is represented by Eq. (5.5). The $p = 1$ diffraction order represents a two-beam interferometer; so does the $p = -1$ diffraction order. The $p = 1$ diffraction order is centered at $y = M\Delta_y$ (compare to $p = -1$ diffraction order which is centered at $y = -M\Delta_y$) and has two components $(m, n) = (-1, 0)$ and $(0, -1)$. The $p = 0$ diffraction order, centered at $y = 0$, is a multiple-beam interference. All the different diffraction orders appear at the same time in the image

plane with different y-coordinates. The $p = -1, 0, 1$ diffraction orders are shown in Fig. 5.2.

Two gratings of different period can be used. If a composite grating consisting of two gratings of small period difference is oriented along the y-axis, and each individual grating orientation makes an angle θ with the y-axis, but in opposite directions, the field of $p = 1$ diffraction order, as in Eq. (5.7), is

$$U(x, y; p = 1) = A_0 B_1 U_0(-Mx + \lambda f_1 d_1^{-1} \sin\theta, -My - \lambda f_1 d_1^{-1} \cos\theta)$$

$$+ A_1 B_0 U_0(-Mx - \lambda f_1 d_2^{-1} \sin\theta, -My - \lambda f_1 d_2^{-1} \cos\theta), \tag{5.8}$$

where d_1 and d_2 are the periods of gratings G_1 and G_2 respectively. The amount of shift in x and y coordinates is $\lambda f_2(d_1^{-1} + d_2^{-1}) \sin\theta$, and $\lambda f_2(d_1^{-1} - d_2^{-1}) \cos\theta$, respectively.

For the special case, $\theta = 0$, gratings G_1 and G_2 have the same orientation. The amount of shear is $\lambda f_2(d_1^{-1} - d_2^{-1})$ directed along the y-axis; this is the smallest amount that can be achieved by two gratings of different periods.

Two circular gratings of different frequencies having their center points in coincidence can also be used. The result is similar to the case of two linear gratings of different frequencies and has the same orientation discussed above. In this case, the shear is fixed for a pair of circular gratings. It is a circular shearing interferometer with constant radial shear.

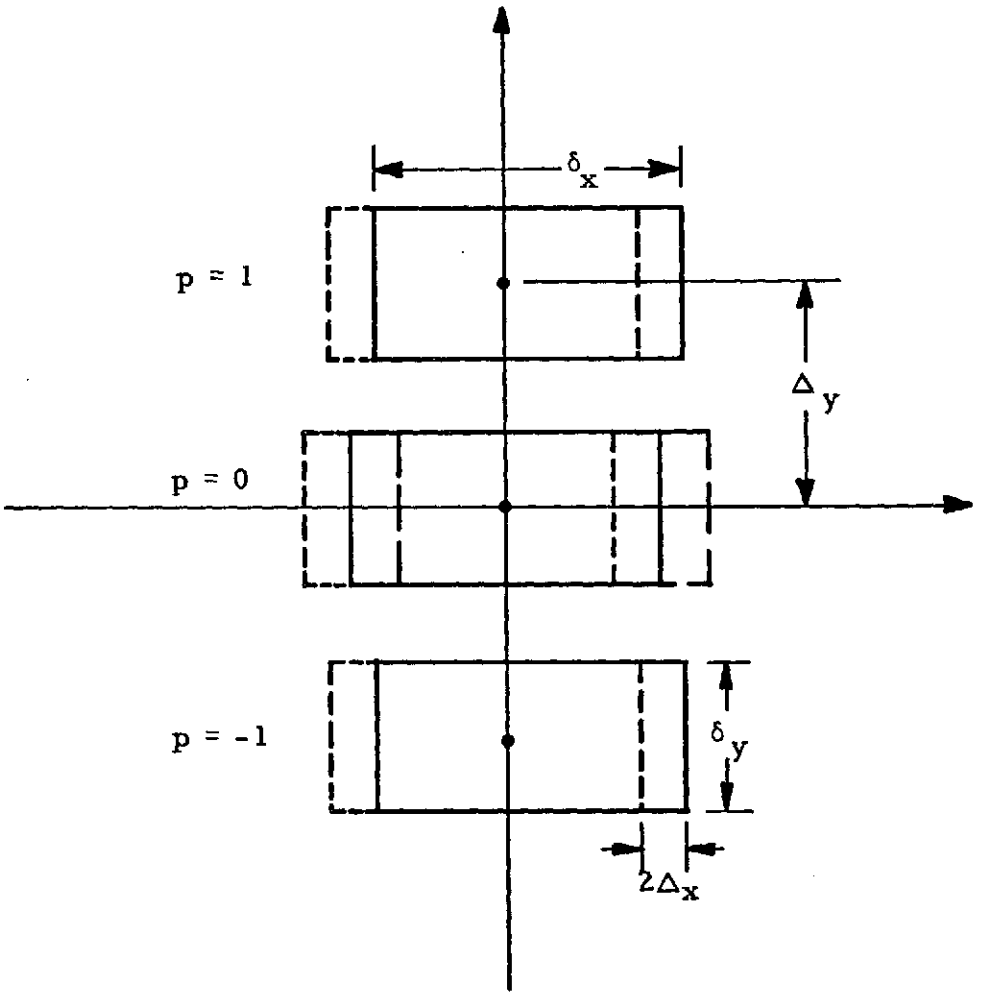


Figure 5.2. The field in the output image plane of the grating interferometer. Here we only show $p = 1, 0, -1$ diffraction orders, and in $p = 0$ diffraction order we only show $(-1, 1)$, $(0, 0)$ and $(1, -1)$ three diffraction orders. There are only two diffraction orders in both $p = 1$ and -1 diffraction orders.

5.3. Variable, Purely Sinusoidal Wave Generator

Purely sinusoidal amplitude gratings are very difficult to construct. By the use of the moiré effect and spatial filtering, we can easily generate a purely sinusoidal wave.

The optical setup is shown in Fig. 5.3. Two superimposed gratings G_1 and G_2 are placed at the object plane of a telecentric image-forming system. The field at the spatial-frequency plane is a collection of peaks* with their locations distributed like the Miller's indices. A binary filter F is placed in the spatial-frequency plane. If the filter passes only two diffraction orders, e.g. (0, 1) and (1, 0), then the field at the output plane is

$$U(x, y;\ z = 4f) = A_0B_1 \exp(4\pi i x d^{-1} \sin\theta) + A_1B_0 \exp(-4\pi i x d^{-1} \sin\theta)$$

$$= |A_0B_1| \cos(4\pi x d^{-1} \sin\theta + \varphi) , \qquad (5.9)$$

if $|A_0B_1| = |A_1B_0|$, or gratings G_1 and G_2 are identical, where φ is a constant. The sinusoidal wave has the spatial frequency $4\, d^{-1} \sin\theta$.

Any two diffraction orders can be used for sinusoidal wave generation. If the absolute amplitudes of those two diffraction orders $|A_mB_n|$ and $|A_{m'}B_{n'}|$ are equal, the generated sinusoidal wave $I(x, y) = |U(x, y;\ 4f)|^2$ has highest contrast. If the absolute amplitude

---

*The area of each peak is roughly $(\lambda f)^2/A$, where A is the area of grating being illuminated by the illuminating plane wave.

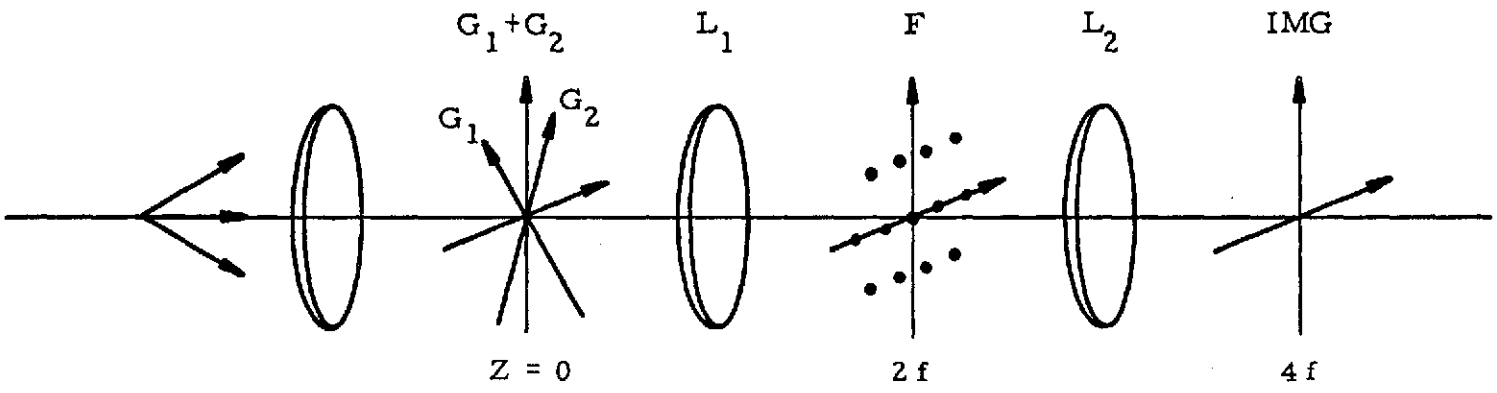


Figure 5.3. Variable, purely sinusoidal wave generator.

strengths are not equal, the contrast will be reduced. We can vary the spatial frequency of the generated output sinusoidal wave either by varying the angle 2θ between gratings G_1 and G_2, or by passing different diffraction orders in filtering plane.

The output image plane is conjugate to the plane where the gratings are placed; hence the system is achromatic, although the spectral width of the light source is limited by separation of the adjacent diffraction orders. A polychromatic light source can be used for illumination.

5.4. Applications

Most methods for optical information processing are based on interference. Interferometry is one of the most important methods in optical information processing. The grating interferometer derived in this chapter can be used in optical signal processing. New methods for optical testing and phase object visualization, image contrast reversal, spatial differentiation, complex amplitude addition and subtraction, image multiplexing, and fabrication of spatial filters will be presented in this section.

5.4.1. Optical Testing and Phase-Object Visualization

The Ronchi interferometer has been used extensively in optical testing, since it is simpler and cheaper than most of the other methods. But the Ronchi interferometer has some drawbacks. To overcome the drawbacks of the Ronchi interferometer[4] is one of the motivations for

our new interferometer. Optical testing and phase object visualization can be achieved by the interference of the tested object wavefront with either a standard wavefront or a replica of the wavefront of the tested object with a shift. The grating interferometer can produce variable shear, either total shear or differential shear, and hence can be used in phase-object visualization. The new method is simpler, cheaper, and much more flexible than most other existing methods.

The setup as shown in Fig. 5.1 will be used in phase object visualization. For simplicity, we will assume that lenses L_1 and L_2 have the same focal length f, and then the magnification factor M is 1.

5.4.1.1. Phase-Object Visualization Using the Grating Interferometer to Produce Differential Shear

The tested phase object $U_0(x, y) = \exp[i\phi(x, y)]$ is placed in the object plane of Fig. 5.1. The two superimposed gratings G_1 and G_2 have the same period d. The $p = 1$ diffraction order, centered at $y = -\Delta_y$, from Eq. (5.7), is

$$\begin{aligned} U(x, y; p = 1) &= A_0 B_1 U_0(-x + \Delta_x, -y - \Delta_y) \\ &\quad + A_1 B_0 U_0(-x - \Delta_x, -y - \Delta_y) \\ &= A_0 B_1 \{ \exp[i\phi(-x + \Delta_x, -y - \Delta_y)] \\ &\quad + \exp[i\phi(-x - \Delta_x, -y - \Delta_y) + i\alpha] \}, \end{aligned} \tag{5.10}$$

if gratings G_1 and G_2 are identical, where α is a real constant depending on the relative shift of gratings G_1 and G_2.

For an object with dimension in y direction $\delta_y \leq \Delta_y$, the diffraction orders p are separated. Equation (5.10) describes a two-beam differential shearing interferometer. The amount of shear is $2\Delta_x = 2\lambda f d^{-1} \sin\theta$.

The intensity of the $p = 1$ diffraction order is

$$I(x, y) = |U(x, y; p = 1)|^2$$

$$= |A_0 B_1|^2 \cos^2 \frac{1}{2}\{\phi(-x + \Delta_x, -y - \Delta_y) - \phi(-x - \Delta_x, -y - \Delta_y) - \alpha\}. \tag{5.11}$$

If Δ_x is smaller than the typical finer details of the phase object tested, we can expand the phase variation function $\phi(x, y)$ in a Taylor series and retain only the first two terms, i.e.,

$$\phi(-x + \Delta_x, -y - \Delta_y) \cong \phi(-x, -y - \Delta_y) + \Delta_x \phi_x(-x, -y - \Delta_y), \tag{5.12}$$

where

$$\phi_x(x, y) = \frac{\partial \phi(x, y)}{\partial x} \tag{5.13}$$

Equation (5.12) can be rewritten as

$$I(x, y) = |A_0 B_1|^2 \cos^2 \left\{\Delta_x \phi_x(-x, -y - \Delta_y) - \frac{\alpha}{2}\right\}$$

$$= |A_0 B_1|^2 \{1 + \cos(2\Delta_x \phi_x(-x, -y - \Delta_y) - \alpha)\}. \tag{5.14}$$

The phase variation of the tested object becomes visible. The interference fringes show the contour of constant derivation of the phase

variation of the phase object along the x-axis. If $\alpha = \pi$, and $\Delta_x \phi_x \ll 1$, Eq. (5.14) can be written as

$$I(x, y) = |A_0 B_1|^2 \, (1 - \cos \, (\Delta_x \phi_x \, (-x, \, -y - \Delta_y))$$

$$\cong \frac{1}{2} \, |A_0 B_1|^2 \, \Delta_x^2 \, \phi_x^2 (-x, \, -y - \Delta_y) \, . \qquad (5.15)$$

The output intensity is proportional to the square of the derivative of the phase object.

If the input phase object to be tested is a perfect lens, $U_0(x, y) = \exp [ia(x^2 + y^2)]$, then the output image intensity from Eq. (5.11) is

$$I(x, y) = |A_0 B_1|^2 \, [1 - \cos (2a \, \Delta_x x)] \, . \qquad (5.16)$$

The interference fringes are straight lines parallel to y-axis. The parameter a can be measured from the period of the inteference fringes. If the input object is not a perfect lens, the interference fringes are not all parallel to y-axis. We can use this method to measure the focal lengths of lenses and test the aberration of the lenses.

5.4.1.2. Phase-Object Visualization Using the Grating Interferometer to Produce Total Shear

If the input object $U_0(x, y)$ consists of two non-overlapping components $U_1(x - \Delta_x, \, y)$ and $U_2(x + \Delta_x, \, y)$,

$$U_0(x, y) = U_1(x - \Delta_x, \, y) + U_2(x + \Delta_x, \, y) \, . \qquad (5.17)$$

The objects $U_1(x, y)$ and $U_2(x, y)$ have dimensions δ_x and δ_y in x

and y directions respectively, and $\delta_x \leq \Delta_x$ and $\delta_y \leq \Delta_y$. The $p = +1$ diffraction order in the image plane is

$$V_1(x, y) = A_0B_1\, U_0(-x + \Delta_x,\ -y - \Delta_y) + A_1B_0\, U_0(-x - \Delta_x,\ -y - \Delta_y)$$

$$= A_0B_1\,[U_1(-x,\ -y - \Delta_y) + U_2(-x + 2\,\Delta_x,\ -y - \Delta_y)]$$

$$+ A_1B_0[U_1(-x - 2\Delta_x,\ -y - \Delta_y) + U_2(-x,\ -y - \Delta_y)]$$

$$= \{\, A_0B_1\, U_1(-x,\ -y - \Delta_y) + A_1B_0\, U_2(-x,\ -y - \Delta_y)\}$$

$$+ A_1B_0U_1(-x - 2\Delta_x,\ -y - \Delta_y) + A_0B_1\, U_2(-x + 2\Delta_x,\ -y - \Delta_y)\,. \tag{5.18}$$

The typical wavefield of $V_1(x, y)$ is shown in Fig. 5.4. The wavefronts of objects $U_1(x, y)$ and $U_2(x, y)$ overlap and are centered at $(x = 0,\ y = -\Delta_y)$ as can be seen from Fig. 5.4. The intensity in the area where U_1 and U_2 overlap, when $A_0B_1 = A_1B_1 \exp(i\alpha)$, is

$$I(x, y) = |\, A_0B_1|^2\ |U_1(-x,\ -y - \Delta_y) + e^{i\alpha}\, U_2(-x,\ -y - \Delta_y)|^2\,. \tag{5.19}$$

If $U_1(x, y) = \exp(i\phi_1(x, y))$ is the phase object under test, and $U_2(x, y) = \exp(i\phi_2(x, y))$ is the standard with which the object $U_1(x, y)$ is to be compared, the intensity $I(x, y)$ will show the deviation of $\phi_1(x, y)$ from $\phi_2(x, y)$. If $\phi_1(x, y)$ equals $\phi_2(x, y)$, $I(x, y)$ is just a uniform field.

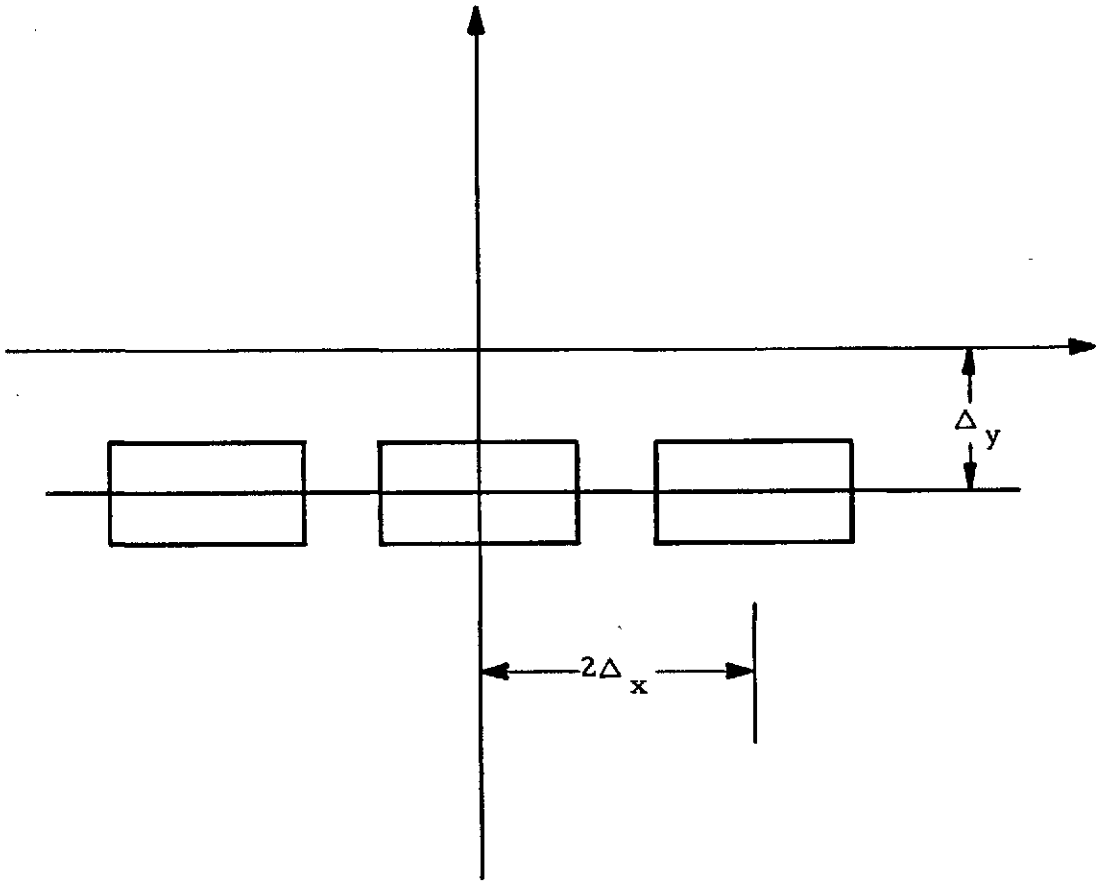


Figure 5.4. The typical wavefield of $V_1(x, y)$.

Usually the standard object $U_2(x, y)$ is an object that creates a uniform field, or a plane wave. The plane wave standard wavefront can be created by simply passing the light through the position of object $U_2(x, y)$. When $U_2(x, y) = 1$, then

$$I(x, y) = |A_0 B_1|^2 \{ 1 - \cos(\phi(-x, -y - \Delta_y)) \} , \tag{5.20}$$

if $\alpha = \pi$. For objects with small phase shift $\phi \ll 1$,

$$I(x, y) \cong \frac{1}{2} |A_0 B_1|^2 \phi^2(-x, -y - \Delta_y) . \tag{5.21}$$

The output image centered at $(0, -\Delta_y)$ is proportional to the square of the phase of the phase object, if the phase object has small phase shift. This is equivalent to what we get from the conventional phase-contrast microscopy.[5]

Total shearing interferometry can be achieved by using only one grating as a beam splitter, i.e., we can use the Ronchi interferometer to produce total shearing for phase object visualization. The tested object $U_1(x, y)$ and the standard wavefront $U_2(x, y)$ are placed in the object plane with their centers separated by a distance $\lambda f / d$ in the direction of grating orientation. The result is the same as we have just discussed.

5.4.2. Spatial Differentiation

Optical spatial differentiation may be desirable in optical signal processing. In the performance of this operation, the principal problem

encountered in a coherent optical data processing system is the fabrication of the spatial filter. A compound filter consisting of an amplitude filter and a phase filter to obtain the required transmittance function $j\omega$ has been reported.[6] One serious difficulty encountered in the fabrication of this compound spatial filter is that the dynamic range of the photographic film is limited between zero and one. The alignment of the amplitude filter and the phase filter might be difficult. Computer-generated holograms[7] and double exposure composite holograms[8] have been used to bypass these difficulties. Birefringement elements such as Savart plates and Wollaston prisms have been used as beam splitters in a differential-shearing interferometer to perform the spatial differentiation.

Differentiation can be approximated by a difference equation with a differential step ϵ in the direction of differentiation, if the differential step ϵ is smaller than the typical finer details of the object to be differentiated,

$$\frac{\partial}{\partial x} U_0(x, y) \cong \frac{1}{\epsilon} [U_0(x + \epsilon, y) - U_0(x, y)] , \tag{5.22}$$

if $\epsilon \leq h_x$, where h_x is the size of typical finer details of the object in x direction. Or equivalently, in the Fourier domain

$$\mathfrak{F}\left[\frac{\partial}{\partial x} U_0(x, y)\right] \cong \frac{j}{\epsilon} \tilde{U}_0(\nu, \mu) \sin\left(\frac{\epsilon}{2} \nu\right) , \tag{5.23}$$

if

$$\tilde{U}_0(\nu, \mu) = 0 , \quad \text{for} \quad |\nu| \geq h_x^{-1} .$$

In the region $|\nu| \leq h_x^{-1}$, $\sin\left(\frac{\Delta}{2}\nu\right) \cong \frac{\Delta}{2}\nu$, so that a sinusoidal grating can be approximately used as the filter for spatial differentiation.

The essence of spatial differentiation is the creation of bipolarity (positivity and negativity) and the shift of the object wavefront. The grating interferometer can produce differential shear and bipolarity, hence can be used in spatial differentiation. Computer-generated holograms and double exposure composite holograms for spatial differentiation all can produce fixed differential step ε. The new method uses the grating interferometer, which can produce variable differential step ε to fit the various kinds of objects.

The optical setup is as shown in Fig. 5.1. The object $U_0(x, y)$ is placed in the object plane. The field of $p=1$ diffraction order, centered at $y = -\Delta_y$, from Eq. (5.10) is

$$U(x, y; p=1) = A_0 B_1 U_0(-x + \Delta_x, -y - \Delta_y)$$

$$+ A_1 B_0 U_0(-x - \Delta_x, -y - \Delta_y)$$

$$\approx A_0 B_1 (2\Delta_x) \frac{\partial}{\partial x} U_0(-x, -y - \Delta_y) , \qquad (5.24)$$

if $2\Delta_x = 2\lambda f d^{-1} \sin\theta$ is smaller than the typical finer details of the object. If we rotate the composite grating $(G_1 + G_2)$ 90°, we can have spatial differentiation along the y-axis, and if the orientation of the composite grating is 45° to the x-axis, we can evaluate $\partial U_0/\partial x - \partial U_0/\partial y$.

5.4.3. Complex Addition and Subtraction

In optics, two functions can be multiplied by using the Kirchhoff boundary condition. When two films are superimposed, the resulting transmittance is the product of their transmittances. Addition of two images can be achieved by double exposure. But subtraction and division are difficult. Our grating interferometer can be used for real-time complex addition and subtraction.

Many methods for subtraction using only optical elements have been already described. They are all based on the interference of the input signals. When they interfere constructively, we have the addition of the input signals; if they interfere destructively, we have the subtraction of the input signals. The first method utilizes interferometry, either coherent holographic interferometry,[9] total shearing interferometry,[10] or incoherent holographic interferometry.[11] The second method uses a previously prepared hologram as a filter in a Fourier hologram recording geometry.[12] The third method uses carrier-frequency photography to encode the polarity of each input signal by shifting the modulation carrier-frequency half a period.[13, 14] The carrier-frequency modulation masks can be regular carrier-frequency masks or random diffusers.

The methods that have been discussed either need very skillful laboratory technique or a 180° phase plate to achieve the bipolarity, or the grating must be shifted between exposures. Most of them are not real-time operations and are inflexible.

One method uses a prefabricated sinusoidal grating as a beam splitter in the Ronchi interferometer.[10] The operation is in real time. The sinusoidal grating can be replaced by a Ronchi ruling. Our grating interferometer can be used as a total shear interferometer, since it can be used in addition and subtraction. By proper arrangement of the signals to be added and subtracted, we can add and subtract four signals at once without using any phase plate, and the operation is in real time.

For the addition and subtraction of two signals, one grating will be used as a beam splitter in the spatial-frequency plane of Fig. 5.1. This method has been discussed in which a sinusoidal grating is used as a beam splitter.[10] The input signals $U_1(x, y)$ and $U_2(x, y)$ are centered at $x = \lambda f/d$ and $x = -\lambda f/d$ respectively, that is

$$U_0(x, y) = U_1(x - \lambda f/d,\ y) + U_2(x + \lambda f/d,\ y) \ . \tag{5.25}$$

The field centered at $x = 0,\ y = 0$ of the output plane is

$$V(x, y) = A_1 U_1(-x, -y) + A_{-1} U_2(-x, -y) \ . \tag{5.26}$$

The grating is real valued, hence $A_1 = A_{-1}^*$. Subtraction is possible if we shift the grating in the x direction, so that $A_{-1} = -A_1$. Then Eq. (5.26) can be written as

$$V(x, y) = A_1 [U_1(-x, -y) - U_2(-x, -y)] \ , \tag{5.27}$$

which is the subtraction of U_2 from U_1. No phase plate has been used, and the operation of addition and subtraction is controlled by the shift of the grating, which is independent of the incident light

wavelength λ.

For more than two signals, more gratings will have to be superimposed for splitting the beams. For four input signals with dimensions in x and y coordinates smaller than $\min(\Delta_x, \Delta_y)$, centered at $(x = \Delta_x, y = \Delta_y)$, $(\Delta_x, -\Delta_y)$, $(-\Delta_x, \Delta_y)$ and $(-\Delta_x, -\Delta_y)$, i.e., the input object $U_0(x, y)$ is

$$U_0(x, y) = U_1(x + \Delta_x, y + \Delta_y) + U_2(x + \Delta_x, y - \Delta_y) + U_3(x - \Delta_x, y + \Delta_y) + U_4(x - \Delta_x, y - \Delta_y), \tag{5.28}$$

the output field centered at $x = 0$, $y = 0$ of the output image plane is

$$V(x, y) = A_0B_{-1}U_1(-x, -y) + A_{-1}B_0U_2(-x, -y) + A_1B_0U_3(-x, -y) + A_0B_1U_4(-x, -y) . \tag{5.29}$$

The relative phases between A_1, A_{-1}, B_1, B_{-1} can be adjusted by shifting the gratings G_1 and G_2 in their orientation directions.

For adding or subtracting just two input signals, the use of two gratings as a beam splitter offers some advantage. From Eq. (5.29), if $U_3 = U_4 = 0$, the output field is

$$V(x, y) = A_0B_{-1}\left[U_1(-x, -y) + \frac{A_{-1}B_0}{A_0B_{-1}} U_2(-x, -y)\right] . \tag{5.30}$$

The distance between the centers of two input signals U_1 and U_2 is $2\Delta_x$, which is adjustable. The ratio $(A_{-1}B_0)/(A_0B_{-1})$ can be adjusted by choosing different space-to-period ratio q of the binary gratings G_1

and G_2, as can be seen from Eq. (4.2).

5.4.4. Real-Time Image Polarity-Reversal and Two-Dimensional Photometry

Several methods for complete or partial reversal of the polarity of images have been proposed. These are alternatives to the traditional photographic method.[15] Image polarity can be reversed in carrier-frequency photography by using the nonlinearities of recording media, a combination of spatially modulated signals, or reconstruction from the zeroth diffraction order.[16] It is also possible to reverse the polarity of an image by using carrier-frequency holography and making the signal-to-reference ratio greater than 1 in the recording of the hologram, or by using the zeroth diffraction order in the reconstruction.[17] All these alternative methods are two-step processes, and they are not superior in either convenience or image quality to conventional photographic polarity-reversal. To my knowledge, there are as yet no real-time optical methods for image polarity reversal.

Image subtraction is the essence of image polarity-reversal. Our grating interferometer can be used in addition and subtraction, hence can be used for image polarity-reversal.

In this method, a Ronchi interferometer is used as a total shearing interferometer. Only one grating is used as a beam splitter in Fig. 5.1. Both the object $U_0(x, y)$ and a uniform field $C\ I(x - \ell\Delta, y)$ of limited extent are placed in the input object plane. The centers of $U_0(x, y)$ and the uniform field are separated by a distance $\ell\Delta = \ell\lambda f/d$

in the x direction, normal to the grating slits, where ℓ is a nonzero integer. C is an adjustable amplitude level of the uniform field, $|C| \leq 1$, and $I(x, y)$ is an indicator function,

$$I(x, y) = \begin{cases} 1 & \text{if } |x| \leq 1/2\ \delta_x \text{, and } |y| \leq 1/2\ \delta_y \\ 0 & \text{otherwise} \end{cases},$$

where δ_x and δ_y are the maximum dimensions of the object in x and y-directions respectively.

The input object is $U_0(x, y) + C\ I(x - \ell\Delta, y)$. The total output field is

$$U(x, y) = \sum_n [A_n U_0(-x-n\Delta, -y) + C\ A_{n-\ell}\ I(-x-n\Delta, -y)]$$

$$= \sum_n V_n(x, y) \quad , \tag{5.31}$$

where

$$V_n(x, y) = A_n U_0(-x-n\Delta, -y) + C\ A_{n-\ell}\ I(-x-n\Delta, -y) \quad . \tag{5.32}$$

If the object size δ_x is smaller than Δ, then for different n's the functions $V_n(x, y)$ appear in the image plane separated in space and centered at $(-n\Delta, 0)$. The different diffraction orders $V_n(x, y)$ have different image-transfer characteristics. Here, the image-transfer characteristic is the relation between the transmittance of the input object $U_0(x, y)$ and the absolute value of the output field amplitude $|V_n(x, y)|$. The image-transfer characteristic has the shape of a "V", and intercepts the U_0-axis at point b (if $b > 0$). The slope is -1 to the left of this point, and +1 to the right. The value b is the polarity

level. For output image $|V_n(x, y)|$ centered at $(-n\Delta, 0)$, from Eq. (5.32), the polarity level b is $-A_n (C\, A_{n-\ell})^{-1}$. From the definition, it is clear that the image-transfer characteristic and hence the polarity level are functions of C and the A_n's. Therefore, the polarity level can be adjusted easily by rotating two superimposed polarizing sheets placed over either the object $U_0(x, y)$ or the reference window $I(x - \ell\Delta, y)$. If the polarity level b is greater than max $U_0(x, y)$, the polarity of the object $U_0(x, y)$ is completely reversed, but if min $U_0(x, y) < b <$ max $U_0(x, y)$, the polarity is only partially reversed. For $0 < b <$ min $U_0(x, y)$, the polarity is not changed, but the contrast has been enhanced.

Typical input and output for $\ell = 2$, $A_n = q \cdot \mathrm{sinc}(nq) \cdot \exp(in\pi/2)$, as shown in Fig. 5.5, are

$$V_n(x, y) = A_n U_0(-x-n\Delta, -y) + C\, A_{n-2}\, I(-x-n\Delta, -y) , \tag{5.33}$$

$$V_0(x, y) = q\,[U_0(-x, -y) - C\ \mathrm{sinc}(2q)\ I(x, y)] , \tag{5.34}$$

$$V_1(x, y) = i\,q\ \mathrm{sinc}(q)\ [U_0(-x-\Delta, -y) - C\ I(-x-\Delta, -y)] , \tag{5.35}$$

$$V_2(x, y) = q\,[-\mathrm{sinc}(2q)\ U_0(-x-2\Delta, -y) + C\ I(-x-2\Delta, -y)] . \tag{5.36}$$

The polarity levels for V_0, V_1, and V_2 are C sinc (2q), C, and C/sinc(2q) respectively. The image-transfer characteristics for V_0, V_1, and V_2 are shown in Fig. 5.5. We have outputs with different polarities at the same time.

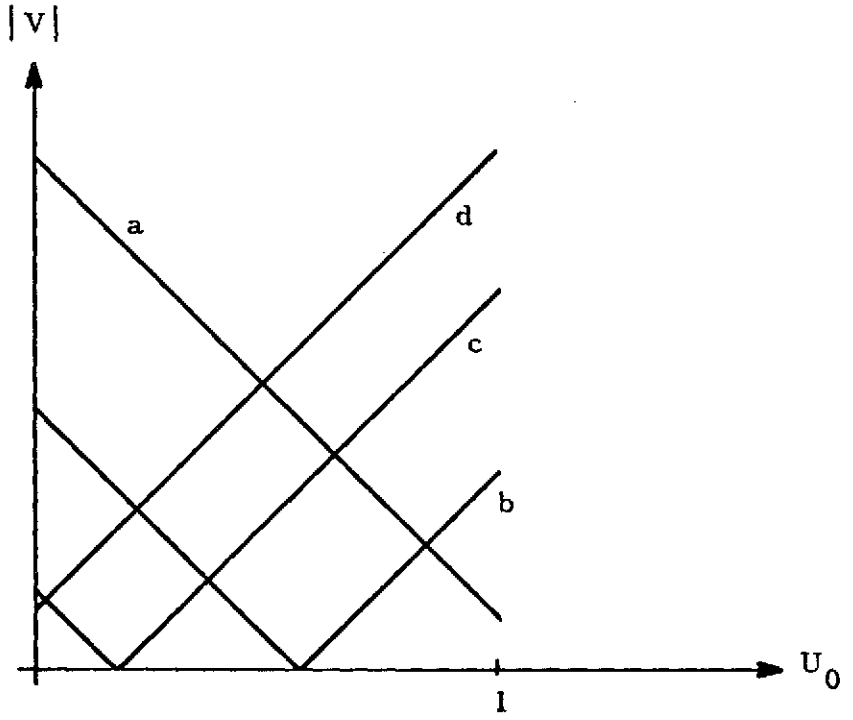


Figure 5.5. The image-transfer characteristics. (a) The polarity level $b > 1$, total image polarity reversal. (b) $0 < b < 1$, partial polarity reversal. (c) $0 < b < \min U_0$, no polarity reversal, contrast enhancement. (d) $b < 0$, contrast reduction.

The input (the object U_0 and the reference window) can be shifted through any desirable transverse displacement. The output image pattern will follow the movement of the inputs.

5.4.5. Fabrication of Filters

We have already shown that the grating interferometer can be modified into a variable, purely sinusoidal wave generator, and a purely sinusoidal grating can be used in spatial differentiation. In this section, we want to use the sinusoidal wave generator to fabricate spatial filters for some optical signal processing.

Spatial filters for spatial differentiation, subtraction, and correlation have been fabricated by the multiple-exposure technique.[8,10,18,19] For example, in the fabrication of the Laplacian operator $\partial^2/\partial x^2 + \partial^2/\partial y^2$, five exposures are made. A phase plate or movement of the recording film between each exposure is needed in order to produce the bipolarity. Film nonlinearity can create troublesome crosstalk, and the mechanical shift between each exposure is undesirable.

All the filters fabricated by the multiple-exposure technique can be fabricated by the use of the sinusoidal wave generator. Only one exposure is needed. The relative sign of each point source in generating the sinusoidal gratings can be controlled easily by the shift of the modulation gratings in their orientation direction. No phase plate is needed, and the mechanical movement between exposures is eliminated.

Several spatial filters for mathematical operations such as $\partial/\partial x$, $\partial^2/\partial x^2$, $\partial^2/\partial x^2 + \partial^2/\partial y^2$, $\partial^2/\partial x \partial y$ can be fabricated. We observe some mathematical identities that

$$\frac{\partial u}{\partial x} = \lim_{h \to 0} \left[\frac{1}{h} U(x, y) * f_1(x, y) \right]$$

where

$$f_1(x, y) = \delta(x + h) - \delta(x) \quad , \tag{5.37}$$

and $*$ is convolution and δ is the Dirac delta function;

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = \lim_{h \to 0} \left[\frac{1}{h} U(x, y) * f_2(x, y) \right] \quad ,$$

where

$$f_2(x, y) = \delta(x + h, y) + \delta(x, y + h) - 2\delta(x, y) \quad ; \tag{5.38}$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \lim_{h \to 0} \left[\frac{1}{h^2} U(x, y) * f_3(x, y) \right] \quad ,$$

where

$$f_3(x, y) = \delta(x + h, y) + \delta(x - h, y) + \delta(x, y + h) + \delta(x, y - h) - 4\delta(x, y) \quad ; \tag{5.39}$$

$$\frac{\partial^2 u}{\partial x \partial y} = \lim_{h \to 0} \frac{1}{h^2} U(x, y) * f_4(x, y) \quad ,$$

where

$$f_4(x, y) = \delta(x + h, y + h) + \delta(x, y) - \delta(x + h, y) - \delta(x, y + h) \quad . \tag{5.40}$$

By the use of the convolution theory of Fourier transform, we know that the spatial filters for mathematical operations $\partial/\partial x$, $\partial/\partial x + \partial/\partial y$,

$\partial^2/\partial x^2 + \partial^2/\partial y^2$, and $\partial^2/\partial x \partial y$ are $\tilde{f}_1$, $\tilde{f}_2$, $\tilde{f}_3$, and $\tilde{f}_4$ respectively, and the filters are composed of sets of gratings.

The spatial filters can be recorded holographically. The optical setup is shown in Fig. 5.6. The left portion of the setup is identical to the sinusoidal wave generator in Fig. 5.3. The field in the filter plane is a set of peaks (approximated to delta functions),

$$U_F(x, y) = \sum_n \sum_m A_n B_m \delta(x - n\Delta_x, y - m\Delta_y) , \qquad (5.41)$$

where $\Delta_x = \lambda f d^{-1} \sin\theta$, and $\Delta_y = \lambda f d^{-1} \cos\theta$ as defined before. We can adjust the angle θ and the orientation direction of the composite grating to obtain the desirable array of peaks. The relative signs of different diffraction orders can be adjusted by shifting the grating along its own orientation direction, and the amplitude $A_n B_m$ can be controlled by using proper space-to-period ratio q of the grating. We use a binary filter that passes only diffraction orders yielding either f_1, f_2, f_3 or f_4, and they will interfere with an offset plane wave to record a Fourier hologram. The recorded Fourier hologram is the spatial filter.

The gradient correlation filter:

If the input to the coherent data processing system is $U(x, y)$, and the output of the system is

$$g(x, y) = \iint \text{grad } U'(x, y) \cdot \text{grad } U^*(x' - x, y' - y)\, dx'\, dy' , \qquad (5.42)$$

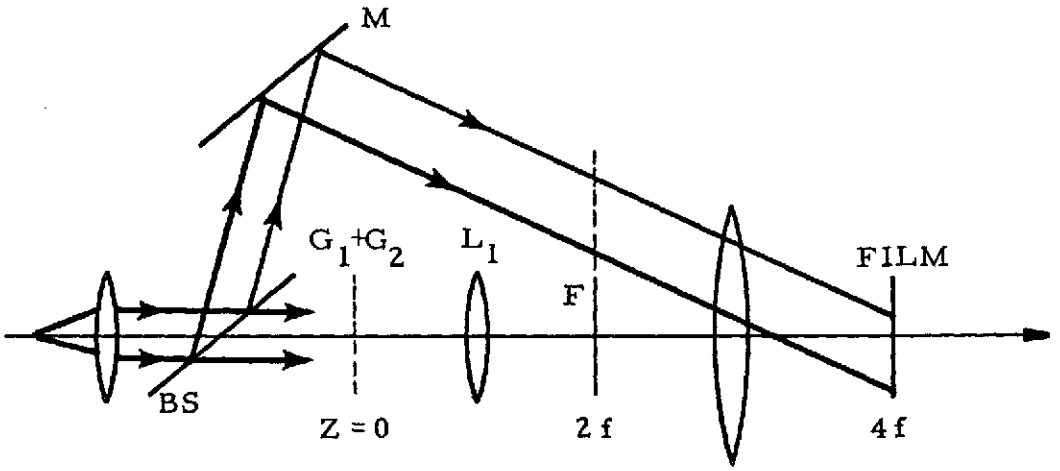


Figure 5.6. Fabrication of filters. The plane indicated by dotted line is the plane where the object $U_0(x, y)$ is located in the fabrication of filter f_5 and f_6. In the fabrication of filters f_1, f_2, f_3 and f_4, the object is replaced by a pinhole.

which is equivalent to

$$g(x, y) = \lim_{h \to 0} \left\{ \frac{1}{h^2} [\delta(x + h, y) + \delta(x - h, y) + \delta(x, y + h) + \delta(x, y - \Delta) - 4\delta(x, y)] * U_0(x, y) * U_0^*(x, y) \right\}$$

$$= \lim_{h \to 0} \left\{ \frac{1}{h^2} f_5(x, y) * U_0(x, y) \right\} ,$$

where

$$f_5(x, y) = [\delta(x + h, y) + \delta(x - h, y) + \delta(x, y + h) + \delta(x, y - h) - 4\delta(x, y)] * U_0(x, y) . \qquad (5.43)$$

This gradient correlation filter is useful in signal discrimination.[21]

The subtraction correlation filter:

If the input is $U_1(x, y)$ and $U_2(x, y)$, the output is $[U_1(x, y) - U_2(x, y)] * U_0(x, y)$. The required spatial filter is

$$f_6(x, y) = [\delta(x, y) - \delta(x + s)] * U_0(x, y) , \qquad (5.44)$$

where s is the separation between the centers of $U_1(x, y)$ and $U_2(x, y)$.

Filters f_5 and f_6 can be fabricated holographically, similarly to the fabrication of filters f_1, f_2, f_3, and f_4. The diffraction orders pass filter F to interfere with the spectrum $U_0(x, y)$ instead of an offset plane wave as shown in Fig. 5.6.

5.4.6. Image Multiplexing

In conventional carrier-frequency image multiplexing, as discussed in section 4.3, a grating is in contact with either the object or the film. The image of the object is an amplitude-modulated version of the object. Between each encoding of each image, the grating was rotated at angle θ. Each object is encoded in different θ-coordinates, the so called "theta modulation." The modulation mask is usually a binary grating. The spectra of different objects occupy lines of different angles θ, each split into many diffraction orders. Only one object can be recorded at that particular angle θ. The number of objects that can be encoded on a single film frame is limited by the angle φ occupied by the spectrum of each object. If we can use a perfect sinusoidal grating, then no high harmonics exist. The spectrum of each object will just occupy part of the space in any θ-coordinate. And many objects can be encoded in the same θ orientation; if we have sinusoidal gratings of different frequencies, then more images can be encoded into a single film frame.

We have already shown that purely sinusoidal gratings of variable frequencies and θ-orientations can be generated easily. This variable, purely sinusoidal grating generator can be used in image multiplexing.

The optical setup is shown in Fig. 5.7. A telecentric image-forming system images the grating plane into the object plane, and the

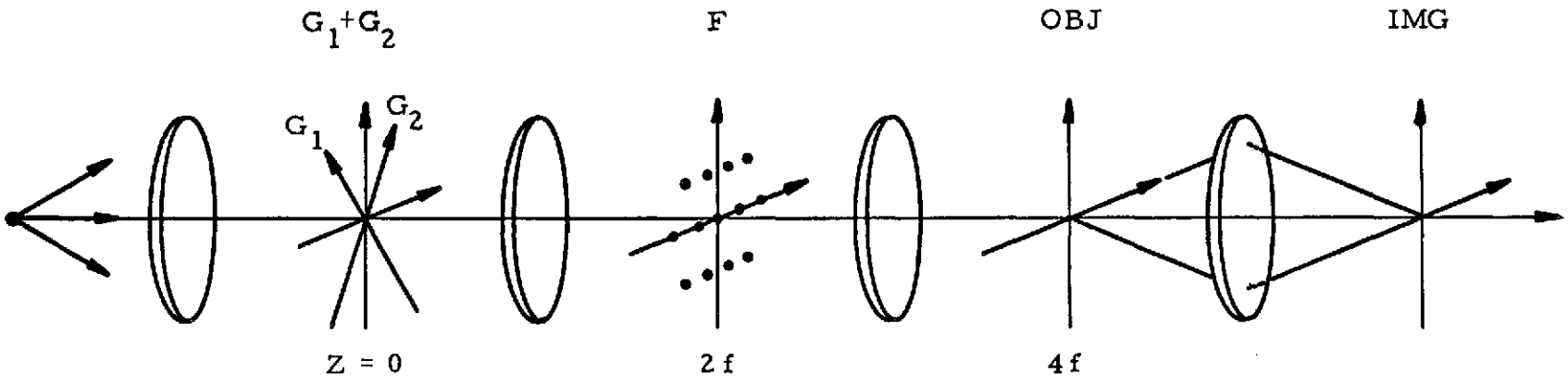


Figure 5.7. An optical arrangement for variable-carrier-frequency photography. The frequency and orientation of the carrier frequency are variable by rotating grating G_1 and G_2, and passing different diffraction orders in filter F.

object plane is imaged by a lens into the image plane. A binary filter is used in the spatial frequency plane to pass only two diffraction orders. A sinusoidal grating will appear in the object plane to modulate the object. The contrast and the frequency of the generated sinusoidal grating can be controlled easily as described in section 5.3. The system is achromatic and broadband source can be used.

In principle, the number of objects that can be encoded into one single film frame is determined by the resolution of the grating and the size of the spatial spectrum of object encoded. We can use the generated sinusoidal gratings of different angles θ and different periods to modulate the encoded images in such a way that the modulated spectra of the images do not overlap and can be retrieved separately. In reality, film nonlinearity and limitations on the dynamic range of the film could be troublesome.

Footnotes to Chapter 5

1. Bryngdahl [56], pp. 39-83.
2. Lohmann and Bryngdahl [57].
3. Goodman [35], p. 83.
4. Ronchi [20].
5. Born and Wolf [33], p. 424.
6. Eguchi and Carlson [58].
7. Lohmann and Paris [59].
8. Yao and Lee [60].
9. Gabor, Strokes, Restrick, Funkhouser and Brumm [61].
10. Lee, Yao, and Milnes [62].
11. Lowenthal and Braat [63].
12. Bromley, Monahan, Bryant, and Thompson [64].
13. Pennington, Will, and Shelton [53].
14. Debrus, Francon, and Grover [65].
15. Lau and Krug [66].
16. Bryngdahl [67].
17. Nishida [68].
18. Yao and Lee [84].
19. Petrosky and Lee [85].
20. Lowenthal and Belvaux [69].

Chapter 6

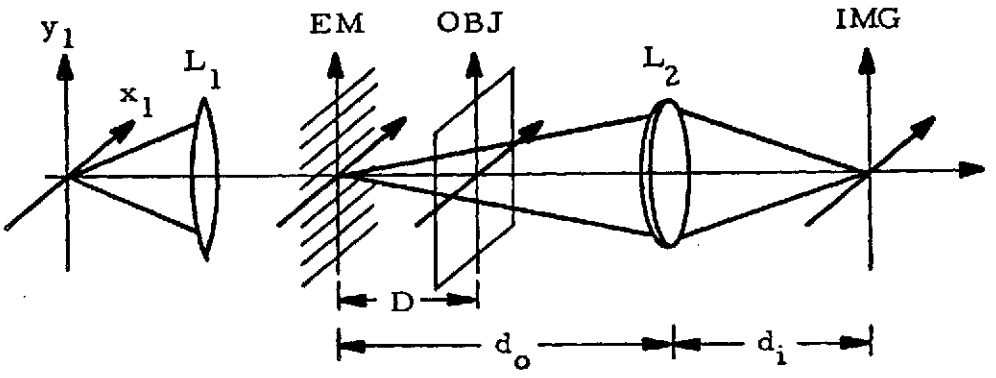
Variable Shearing Interferometry Based on Carrier-Frequency Photography - The Carrier-Frequency Interferometer

6.1. Introduction

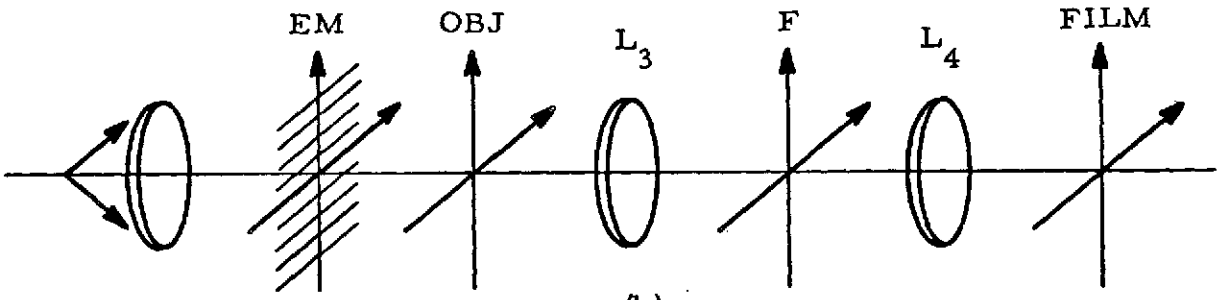
Carrier-frequency modulation and demodulation have been very successfully applied in many branches of electronic signal processing. In communications, the signals to be transmitted are real-valued. Many modulation schemes have been used to transmit real-valued functions, such as pulse-position modulation, pulse-amplitude modulation, and pulse-width modulation. Only one kind of modulation is needed to transmit a real-valued signal. In optics, the signals in general are complex-valued. A modulation scheme for complex signals takes twice the effort of that for real-valued ones. Diffraction gratings can be used as the modulators. For a pulse-modulated complex signal, the amplitude of each pulse is proportional to the amplitude, and the shift of the position of each pulse from the sampling points is related to phase of the signal at the sampled points. It takes both pulse-amplitude modulation and pulse position modulation to register a complex-valued function. This can be done optically by carrier-frequency holography or carrier-frequency photography.

In carrier-frequency photography, a grating is imaged through the object, which is interposed between the grating and the film as shown in Fig. 6.1. The image will register information about both the

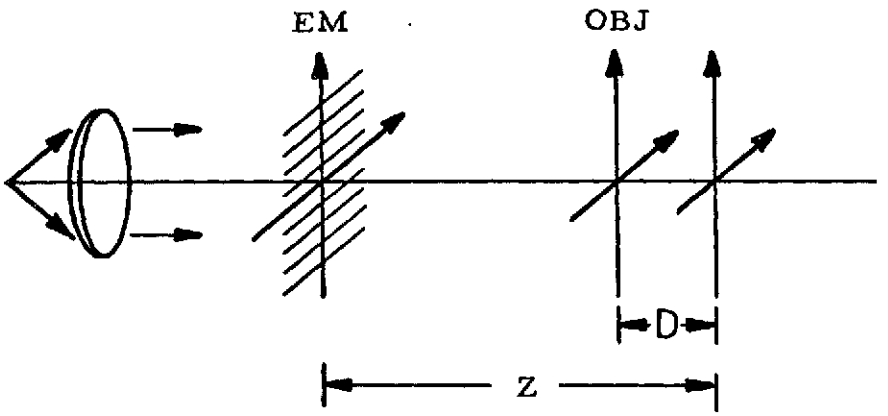
Figure 6.1. Different optical setups for double-exposure carrier-frequency interferometer. (a) Using one lens (L_2), (b) using two lenses (L_3 and L_4), (c) using no lenses.



(a)



(b)



(c)

amplitude and the derivative of the phase of the object along the orientation of the grating. If the object is a pure phase object, the image is a distorted grating, and the distortion is related to the derivative of the phase of the object in the direction of grating orientation. The distortion of the grating is very difficult to detect. But if the distorted grating image is superimposed on a master grating, moiré fringes will appear that reveal the information about the phase variation of the tested object. The encoding and decoding processes are shown in Fig. 6.2.

The effect of a distorted grating has been used by Rowland and Lord Rayleigh[1] in their explanation of the grating ghost, that a lateral displacement of some grating grooves causes a deformation of a diffracted wave, i.e., a phase shift. The so called "detour phase" effect has been used by Leith and Upatnieks[2] in holography and by Brown and Lohmann[3] in computer holograms.

Carrier-frequency photography has been used in image subtraction[4] and phase object visualization,[5,6] both by non-real-time double-exposure techniques. Like holographic interferometry, real-time signal processing can be performed if a reference wavefront is available. If the reference wavefront is the shifted version of the object wavefront, or a plane wave, the reference wavefront can be created by another grating which is identical to the grating used to encode the object wavefront. The second grating is used as a beam splitter and recombiner.

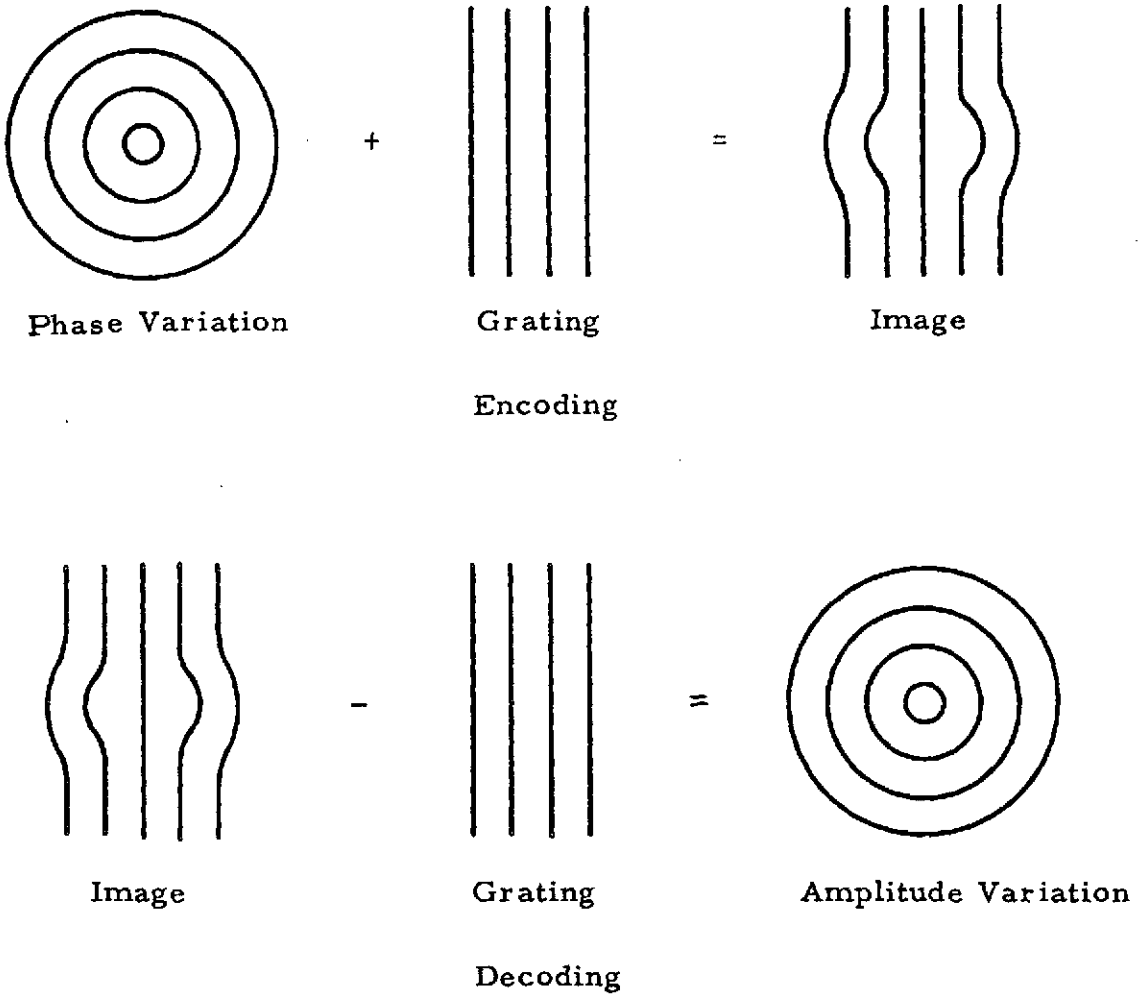


Figure 6.2. Encoding and decoding of phase object.

The encoding process can be done by an image-forming system or by the Talbot effect of regular gratings. Regular linear, checkerboard, or circular gratings can be used in the Talbot interferometer. The Talbot interferometer as described by Lohmann and Silva[7, 8] is for observing the derivative of the tested phase object, but not the phase of the phase object itself. A random diffuser can be used as a modulation mask if an image-forming system is used. The new carrier-frequency interferometer can be used to observe either the phase or the gradient of the phase object tested. Natural white light and a larger source can be used when an image-forming system is used in the encoding process.

The carrier-frequency interferometer can produce variable shear. It can be applied to visualization of a phase object, spatial differentiation, addition and subtraction, image polarity reversal and color complementarity, and equidensitometry.

6.2. Double-Exposure Carrier-Frequency Interferometry

In the last section we have already shown that carrier-frequency photography can register both amplitude and phase information of objects. In particular, phase objects can be visualized by carrier-frequency interferometry. In this section we will show the detailed mathematical analysis of double-exposure carrier-frequency interferometry.

Three different optical setups as shown in Fig. 6.1 can be used in double-exposure carrier-frequency interferometry. A point source located at (x', y') of the front focal plane of lens L_1 will produce a plane wave $\exp[-2\pi i (xx' + yy')/\lambda f]$ to illuminate the carrier-frequency encoding mask (EM). The encoding mask can be a linear grating, a checkerboard grating, a circular grating, or even a random diffuser. The recording film is placed conjugate to the plane in which the encoding grating is located, i.e., the grating will form an image on the film. The image of the encoding mask (grating) is created by the image forming lens L_2 as in Fig. 6.1(a), a system of telecentric image-forming lenses L_3 and L_4 as in Fig. 6.1(b), or the Talbot effect of a regular grating as in Fig. 6.1(c). All the different setups in Fig. 6.1 have certain advantages and disadvantages. Spatial filtering is possible in the setup of Fig. 6.1(b), but two lenses are needed. No lenses are needed when the Talbot-image property is used, but the requirements on the size and spectral purity of the illuminating source are more severe.

The carrier-frequency interferometer is a shearing interferometer. When linear gratings and checkerboard gratings are used as the encoding mask, it is a lateral shearing interferometer; when circular gratings are used, it is a constant-radial-shearing interferometer. A random diffuser is used as the encoding mask because of the convenience when a large size encoding mask is needed.

6.2.1. Linear Carrier-Frequency Interferometry

The optical setup is shown in Fig. 6.1(a). The encoding mask used is a linear grating with transmittance function $G(x) = \sum_n A_n \exp(2\pi i nx/d)$. The film is placed in the plane conjugate to the plane in which we place the encoding mask, i.e., $d_o^{-1} + d_i^{-1} - f^{-1} = 0$, where d_o is the distance between the grating and image-forming lens L_2, d_i is the distance between the film and lens L_2, and f is focal length of both lenses L_1 and L_2. Two exposures will be made. In the first exposure, the object $U_0(x, y)$ is placed at a distance D to the left of the grating.

The field just before the grating is a tilted plane wave,

$$U(x, y; -0) = \exp[2\pi i (\alpha x + \beta y)] \quad , \tag{6.1}$$

where

$$\alpha = -x'/\lambda f$$

$$\beta = -y'/\lambda f \quad ,$$

where x' and y' are the coordinates of the point source in the front focal plane of lens L_1. The reason for the oblique illumination will be clear later. The incident plane wave will be diffracted into many plane waves by the linear grating. The field just before the object $U_1(x, y)$, as shown in Eq. B.1, is

$$U(x, y; D-0) = \sum_n C_n \exp\left[2\pi i x\left(\frac{n}{d} - \frac{x'}{\lambda f}\right)\right] \cdot \exp(-2\pi i yy'/\lambda f) \, , \tag{6.2}$$

where

$$C_n = A_n \exp\left[\frac{i\pi D}{\lambda f^2}(x'^2+y'^2)\right]\exp\left[-i\pi\frac{n^2\lambda D}{d^2}\right]\cdot\exp\left[\frac{2\pi i\, n}{d}\frac{Dx'}{f}\right]. \quad (6.3)$$

The field is made up of many tilted plane waves with direction cosines in x and y $\left(\frac{n}{d}-\frac{x'}{\lambda f}, \frac{-y'}{\lambda f}\right)$ and amplitudes C_n. Those tilted plane waves illuminate the object. The field $V_n(x,y;x',y')$ in the image plane created by the plane wave of direction cosines $\left(\frac{n}{d}-\frac{x'}{\lambda f}, \frac{-y'}{\lambda f}\right)$, according to Eq. (A. 12), is

$$V_n(x,y;x',y') = A_n \exp\left[\frac{i\pi}{\lambda d_i}(1+M^{-1})(x^2+y^2)\right]\cdot\exp\left[\frac{2\pi i}{M}\frac{xx'+y'y}{\lambda f}\right]$$

$$\exp\left(-\frac{2\pi i n}{Md}x\right)\, U_1\left(-\frac{x}{M}+\frac{n\lambda D}{d}-\frac{x'D}{f}, \frac{-y}{M}-\frac{y'D}{f}\right), \quad (6.4)$$

if D is small enough that the integration in Eq. (A. 9), can be performed by the method of stationary phase. $M = d_i/d_o$ is the magnification factor. D has to be smaller than δ^2/λ, where δ is the size of the typical details of the object $U_0(x,y)$.

The total field at the film plane is

$$V(x,y;x',y') = \sum_n A_n \exp\left[\frac{i\pi}{\lambda d_i}(1+M^{-1})(x^2+y^2)\right]\exp\left[\frac{2\pi i}{M}\frac{(xx'+yy')}{\lambda f}\right]$$

$$\exp(-2\pi i nx/Md)\, U_1\left(-\frac{x}{M}+\frac{n\lambda\nu}{d}-\frac{x'D}{f}, \frac{-y}{M}-\frac{y'D}{f}\right). \quad (6.5)$$

Each diffraction order of the grating has produced an image on the film with lateral shift $n\lambda D/d$. The intensity of the first exposure is

$$E_1(x, y; x', y') = |V(x, y; x', y')|^2$$

$$= \sum_n \sum_m A_n A_m^* \exp[-2\pi i(n-m)/Md]\, U_1\left(-\frac{x}{M} + \frac{n\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} - \frac{y'D}{f}\right)$$

$$\times U_1^*\left(-\frac{x}{M} + \frac{m\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} - \frac{y'D}{f}\right). \quad (6.6)$$

In the second exposure, the second object $U_2(x, y)$ is used to replace $U_1(x, y)$, and the grating is shifted laterally by a distance s in the direction of its orientation, i.e., $G_2(x_1) = G(x + s) = \sum B_n \exp(2\pi i nx/d)$, and $B_n = A_n \exp(2\pi i ns/d)$. The intensity of the second exposure is

$$E(x, y, x', y') = \sum_n \sum_m B_n B_m^* \exp[-2\pi i(n-m)/Md]$$

$$\times U_2\left(-\frac{x}{M} + \frac{n\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} - \frac{y'D}{f}\right) U_2^*\left(-\frac{x}{M} + \frac{m\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} - \frac{y'D}{f}\right).$$

(6.7)

The total exposure is

$$E(x, y; x', y') = E_1 + E_2 = \sum_p I_p(x, y) \exp(-2\pi i px/Md), \quad (6.8)$$

where

$$I_p(x,y) = \sum_m A_{m+p} A_m^* \Big[U_1\Big(-\frac{x}{M} + \frac{(m+p)\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} - \frac{y'D}{f}\Big)$$

$$\times U_1^*\Big(-\frac{x}{M} + \frac{m\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} - \frac{y'D}{f}\Big)$$

$$+ \exp[2\pi i ps/d]\, U_2\Big(-\frac{x}{M} + \frac{(m+p)\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} - \frac{y'D}{f}\Big)$$

$$\times U_2^*\Big(-\frac{x}{M} + \frac{m\lambda D}{d} - \frac{x'D}{f}, \frac{-y}{M} + \frac{y'D}{f}\Big)\Big]. \qquad (6.9)$$

The only effect introduced by the oblique plane wave illumination is the shift of the image by $x'D/f$ and $y'D/f$ in the x and y directions respectively. The size of the illuminating point source, Δ_s, has to be small enough that $\Delta_s \ll f\delta/D$. From now on, we will assume normal plane wave illumination, $x' = y' = 0$, and Eq. (6.9) becomes

$$I_p(x,y) = \sum_m A_{m+p} A_m^* \Big[U_1\Big(-\frac{x}{M} + \frac{(m+p)\lambda D}{d}, \frac{-y}{M}\Big)\, U_1^*\Big(-\frac{x}{M} + \frac{m\lambda D}{d}, \frac{-y}{M}\Big)$$

$$+ \exp(2\pi i ps/d)\, U_2\Big(-\frac{x}{M} + \frac{(m+p)\lambda D}{d}, \frac{-y}{M}\Big)\, U_2^*\Big(-\frac{x}{M} + \frac{m\lambda D}{d}, \frac{-y}{M}\Big)\Big].$$

(6.10)

If the film is developed in the linear region of the T-E curve of the film, the transmittance of the developed film is

$$T(x,y) = t_0 + t_1 E(x,y) , \qquad (6.11)$$

where t_0 and t_1 are constants, and $E(x,y) = E(x,y;0,0)$.

To extract the information, the developed film is placed at the input object plane of a telecentric image-forming system. Each diffraction order I_p will form an island in the spatial-frequency plane of the telecentric image-forming system. If we use a binary filter to pass only one diffraction order, e.g. the pth, the field at the output plane is

$$f(x, y) = I_p(-x, -y) \quad . \tag{6.12}$$

The output field is a function of the distance D between the grating and the object. The output field $f(x, y)$ is the interference of object $U_1(x, y)$ and $U_2(x, y)$. The object $U_2(x, y)$ can be a uniform field, $U_2(x, y) = 1$, or a shifted version of $U_1(x, y)$, $U_2(x, y) = U_1(x - \epsilon_x, y - \epsilon_y)$.

The carrier-frequency interferometer can be a two-beam or multiple-beam interferometer, depending on the A_m's. If the setup in Fig. 6.1(b) is used for recording, spatial filtering is possible, and we can eliminate the unwanted diffraction order A_m.

This interferometer can be applied to the visualization of phase objects, spatial differentiation, image polarity reversal, and subtraction of images.

6.2.2. Checkerboard Carrier-Frequency Interferometry

Checkerboard (or cross) gratings can be used to replace the linear grating as encoding masks. In a linear carrier-frequency interferometer, the direction of shear is fixed by the orientation of the grating; the checkerboard grating can give more directions of shear.

A checkerboard grating can be represented as

$$G(x, y) = \sum_n \sum_m A_{nm} \exp(2\pi i\, nx/d_x) \exp(2\pi i\, my/d_y) \quad , \qquad (6.13)$$

where A_{nm} is the Fourier coefficient, d_x and d_y are periods of the grating in x and y directions respectively.

For simplicity, we assume this grating is illuminated by a normal incident plane wave. Two exposures are made, one with object $U_1(x, y)$, the other with object $U_2(x, y)$. By the same calculation as in the last section, the total exposure of the film is

$$E(x, y) = \sum_m \sum_n C_{mn}(x, y) \exp(-2\pi i\, mx/Md_x) \exp(-2\pi i\, ny/Md_y) .$$

where

$$C_{mn} = \sum_\ell \sum_p A_{m+\ell, n+p} A^*_{\ell, p} \left\{ U_1\left(-\frac{x}{M} + \frac{(m+\ell)\lambda D}{d_x}, \frac{-y}{M} + \frac{(n+p)\lambda D}{d_y}\right) \right.$$

$$\times\, U_1^*\left(-\frac{x}{M} + \frac{\ell \lambda D}{d_x}, \frac{-y}{M} + \frac{p\lambda D}{d_y}\right)$$

$$+ \exp\left(\frac{2\pi i\, m\Delta_1}{d_x} + \frac{2\pi i\, n\Delta_2}{d_y}\right) U_2\left(-\frac{x}{M} + \frac{(m+\ell)\lambda D}{d_x}, \frac{-y}{M} + \frac{(n+p)\lambda D}{d_y}\right)$$

$$\left. \times\, U_2^*\left(-\frac{x}{M} + \frac{\ell \lambda D}{d_x}, \frac{-y}{M} + \frac{p\lambda D}{d_y}\right) \right\} \qquad (6.14)$$

where Δ_1 and Δ_2 are the displacements of the grating in x and y direction between two exposures.

The film is developed in the linear region of the T-E curve,

and placed at the object plane of an telecentric image-forming system. A binary filter passes only (n, m)th diffraction order. The field in the image plane, from Eq. (6.14), is $C_{mn}(-x, -y)$. Shearing in both x and y directions can be obtained depending on the diffraction order (n, m) chosen in the filtering plane. Visualization of the gradient of phase objects in many directions and with different sensitivities is possible from one recorded film.

6.2.3. Circular Carrier-Frequency Interferometry

Linear gratings and checkerboard gratings can produce lateral shifts as shown in the last two sections. For circularly symmetric objects, radial shearing is more convenient for the interpretation of interference fringes. When circular gratings are used as the encoding masks, the interferometer is a constant-radial-shearing interferometer.

Circular gratings are very similar to linear gratings, especially at the region far from the center. A circular grating of period a can be written as

$$C(r) = \sum_n C_n e^{2\pi i nr/a} \quad . \tag{6.15}$$

A normally incident plane wave illuminates the circular grating. The field at a distance Z behind the grating, as shown in Appendix A, Eq. (C.8) is

$$U(r, \varphi ; Z) \cong \sum_n C_n \exp(-i\pi \lambda Z n^2/a^2) \exp(2\pi i nr/a) \quad , \tag{6.16}$$

which is very similar to the diffraction of a plane wave by a linear grating, as shown in Eq. (6.2). Using the same calculation as in the last two sections, we find that the field at the image plane is

$$V(r,\varphi) = \sum C_n \exp\left[\frac{i\pi}{\lambda d_i}(1+M^{-1})r^2\right] \exp(2\pi i\, nr/a)\, U_1\left(\frac{r}{M}+\frac{n\lambda D}{a},\ -\varphi\right), \tag{6.17}$$

where U_1 is the object placed at a distance D right after the circular grating.

The total exposure from double exposures, one with object $U_1(r,\varphi)$ and the other with object $U_2(r,\varphi)$, is

$$E(r,\varphi) = \sum_p C'_p(r,\varphi) \exp(2\pi i\, rp/Ma) \quad , \tag{6.18}$$

where

$$C'_p(r,\varphi) = \sum_m C_{m+p} C_m \left\{ U_1\left(\frac{r}{M}+\frac{(m+p)\lambda D}{a},\ -\varphi\right) U_1^*\left(\frac{r}{M}+\frac{m\lambda D}{d},\ -\varphi\right) \right.$$

$$\left. + U_2\left(\frac{r}{M}+\frac{(m+p)\lambda D}{a},\ -\varphi\right) U_2^*\left(\frac{r}{M}+\frac{m\lambda D}{d},\ -\varphi\right) \right\} . \tag{6.19}$$

We can put the developed film in the object plane of a coherent signal processing system, and pass only the diffraction order P. Then in the output image plane the field is $C'_p(r,-\varphi)$. The interferometer is a constant radial-shearing-interferometer. The amount, $p\lambda D/a$, is proportional to D and is variable. It can be used for testing lenses.

6.2.4. Random Carrier-Frequency Interferometry

Random diffusers have been used in holography to illuminate the object when recording holograms. In holography, coherence between object wave and reference wave is essential, but as it introduces strong coherent noise, it is undesirable. Holograms made of diffusely illuminated objects can be scratched, spotted with dirt, and even broken into pieces without serious loss of information. The diffused-object hologram has less coherent noise,[9] but requires larger bandwidth; otherwise the annoying speckle noise will be introduced.[10]

Random carrier-frequency photography has been used in image subtraction, phase object visualization, image multiplexing, and image contrast reversal.[11] Double exposure has also been used. Theoretically, everything random carrier-frequency photography can achieve, can also be achieved by regular carrier-frequency photography, and probably better. But sometimes for convenience, random carrier-frequency is more desirable.

Random carrier-frequency interferometry is especially good for simple objects, e.g. for the measurement of the transverse movement of a rough surface.[13]

6.2.5. Lensless Carrier-Frequency Interferometry

We have shown in the appendix that regular gratings can create self images. In all the interferometry we have discussed in previous sections, lenses are used to image the grating into the recording film.

We can also use the self-imaging property of gratings in carrier-frequency interferometry.

The recording setup for the lensless carrier-frequency interferometry is shown in Fig. 6.1(c). A linear grating is illuminated by a normally incident plane wave, a recording film is placed at a distance Z_o to the right of the grating, and the object is interposed between the grating and the film. The distance between the object and the film is D.

The field at the plane just to the left of the object plane, by Eq. (B.1), is

$$U(x, y\,;\, Z_o\text{-}D\text{-}0) = \sum_n A_n \exp[-i\pi\lambda(Z_o\text{-}D)\, n^2/d^2] \exp(2\pi i\, nx/d)\,, \quad (6.20)$$

By the use of the Kirchhoff boundary condition, the field just behind the object is

$$U(x, y;\, Z_o\text{-}D\text{+}0) = \sum_n A_n \exp[-i\pi\lambda(Z_o\text{-}D) n^2/d^2] \exp(2\pi i\, nx/d)\, U_1(x, y)\,, \quad (6.21)$$

where $U_1(x, y)$ is the transmittance of the first object. When the wave propagates a distance D and arrives at the film plane, the field is

$$U(x, y; Z_o) = \iint U(x_1, y_1; Z_o - D + 0) \exp\left\{\frac{i\pi}{\lambda D}\left[(x_1 - x)^2 + (y_1 - y)^2\right]\right\} dx_1 dy_1 \tag{6.22}$$

$$= \sum_n A_n \exp(-i\pi\lambda Z_o n^2/d^2) \exp(2\pi i nx/d)$$

$$\int U_1(x_1, y_1) \exp\left\{\frac{i\pi}{\lambda D}\left[\left(x_1 - x + \frac{n\lambda D}{d}\right)^2 + (y_1 - y)^2\right]\right\} dx_1 dy_1 \ , \tag{6.23}$$

If the distance D is smaller than δ^2/λ, the method of stationary phase can be applied in the evaluation of Eq. (6.23), and

$$U(x, y; Z_o) = \sum_n A_n \exp(-i\pi\lambda Z_o n^2/d^2) \exp(2\pi i nx/d)\, U_1\left(x - \frac{n\lambda D}{d}, y\right) . \tag{6.24}$$

δ is the size of typical fine details of the object. The intensity of the first exposure is

$$E_1(x, y; Z_o) = \left| U(x, y; Z_o)\right|^2 \tag{6.25}$$

$$= \sum_n \sum_m A_n A_m^* \exp[-i\pi\lambda Z_o (n^2 - m^2)/d^2] \exp[2\pi i (n-m)\, x/d]$$

$$U_1\left(x - \frac{n\lambda D}{d}, y\right) U_1^*\left(x - \frac{m\lambda D}{d}, y\right) \ . \tag{6.26}$$

After the first exposure, the grating is laterally shifted a distance Δ_x in the direction of the grating orientation, and the object $U_2(x, y)$ replaces $U_1(x, y)$ for the second exposure. The total exposure is

$$E(x, y; Z_o) = \sum_p I_p(x, y) \exp(2\pi i px/d) \ , \tag{6.27}$$

where

$$I_p(x,y) = \sum_m A_{m+p} A_m^* \exp\left[i\pi\lambda Z_o p(p+2m)/d^2\right]$$

$$\left[U_1\left(x - \frac{(m+p)\lambda D}{d}, y\right) U_1^*\left(x - \frac{m\lambda D}{d}, y\right)\right.$$

$$\left. + \exp(2\pi i\, p\Delta_1/d)\, U_2\left(x - \frac{(m+p)\lambda D}{d}, y\right)\, U_2^*\left(x - \frac{m\lambda D}{d}, y\right)\right] . \quad (6.28)$$

The total exposure for the lensless carrier-frequency interferometry is very similar to the one when lenses are used, except that the magnification factor is 1, and a phase factor $\exp[i\pi\lambda Z_o(n^2-m^2)/d^2]$.

6.2.6. Discussion

Three different optical arrangements can be used in the double-exposure carrier-frequency. An image forming lens is used in the first setup as shown in Fig. 6.1(a). A telecentric image-forming system is used in the second optical setup as in Fig. 6.1(b), and spatial filtering to eliminate the unwanted diffraction orders is possible. This system is very flexible. The lateral shift of the image due to the tilted plane wave illumination, from Eq. (6.9), is $x'D/f$. If the illumination source is not a point source, every point source will create its own image, and they are shifted with respect to each other, the resulting image will be blurred. For good results, the size of the source B must be less than $(f\delta)/(MD)$, where δ is the size of typical fine details

of the object. If $D = 0$, the images resulting from different point sources will coincide, and a spatially incoherent source can be used.

In the third optical setup, shown in Fig. 6.1(c), where the self-imaging property of the grating is used, the coherence requirement for the Talbot effect has to be met, besides the requirement for using an imaging forming system.

6.3. Real-Time Carrier-Frequency Interferometry

Carrier-frequency interferometry can be performed by a double-exposure technique as shown in section 6.2. Two objects with transmittance functions $U_1(x, y)$ and $U_2(x, y)$ can be registered in a single film frame by two consecutive exposures. The interference pattern between $U_1(x, y)$ and $U_2(x, y)$ can be obtained by reconstruction from the developed film as in double-exposure holographic interferometry. The double-exposure technique is suitable for the situation when the objects $U_1(x, y)$ and $U_2(x, y)$ exist in time sequence.

There are some objections to the double-exposure technique. Film nonlinearities and grain noise may cause some difficulties, and the exact control of all parameters in both exposures is difficult. We have always assumed that the distance D between the object and the modulation mask is small enough for the method of stationary phase to be applicable. All these difficulties and restrictions can be removed by real-time operation.

In the double-exposure process, if the tested object is a pure phase object, the intensity distribution of the first exposure, with the object present, is a distorted version of the grating transmittance function used as modulation mask. To visualize the phase information, we make another exposure on the film with the object removed. The moiré fringes resulting from the sum of the distorted grating image and the undistorted grating image will reveal the phase information of the tested object. This procedure is similar to the one shown in Fig. 6.2. The moiré effect of the product of distorted and undistorted grating images will produce the same moiré fringes in the first diffraction order as the moiré effect of the sum of those two grating images. This is shown in section 4.2. The two grating images can be added by the double-exposure technique, while the superimposition of two grating images (distorted and undistorted) can be implemented in real time and will be shown in this section.

By real-tine operation we mean that there is no time consuming process, e.g. developing a film, involved in the whole process after the system has been set up. We put the tested object in and obtained the result at the same time.

The real-time decoding can be achieved by the use of another grating G_2 at the conjugate plane of the encoding grating G_1 as shown in Fig. 6.3. A telecentric image-forming system added to the right of the decoding grating G will provide the necessary filtering. The observation plane will be the plane conjugate to the plane where the

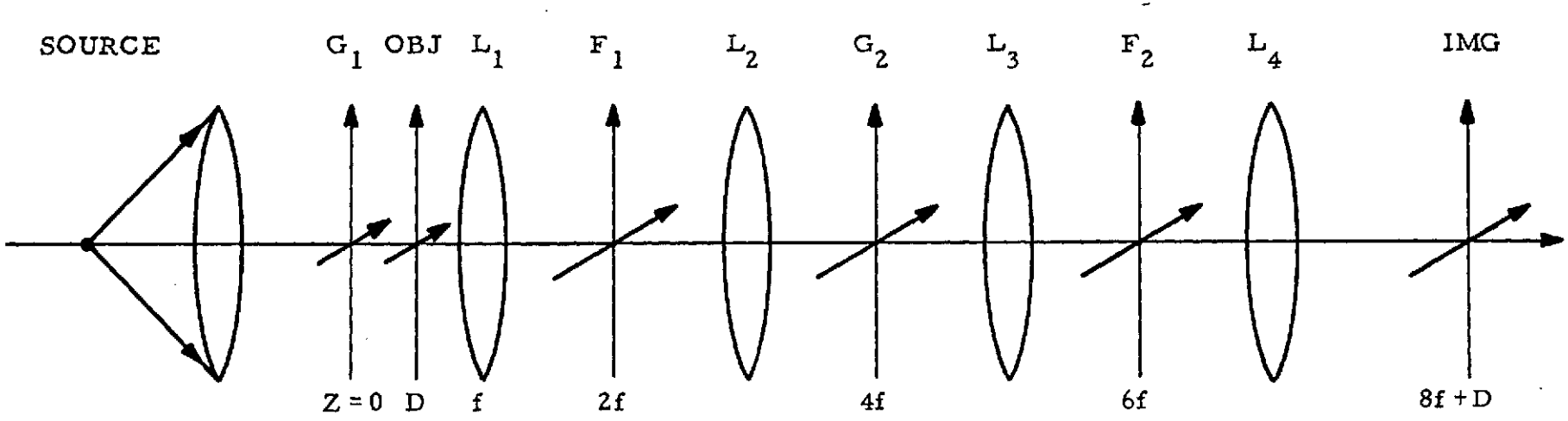


Figure 6.3. The carrier-frequency interferometer.

object is placed.

6.3.1. A Quantitative Analysis

The standard optical setup for the real-time carrier-frequency interferometer is schematically illustrated in Fig. 6.3 For simplicity, we will assume that the focal lengths of all lenses used in Fig. 6.8 are f. The system consists of two cascaded telecentric image-forming systems. The object planes of the first and second telecentric systems are conjugate. The gratings G_1 and G_2 are placed in the first and second object planes, respectively. The object is placed at a distance D to the right of the grating G_1 and forms an image at $Z = 8f + D$. Two spatial filters F_1 and F_2 are used.

A normally incident plane wave illuminates the grating G_1. The light field after the grating G_1 consists of many plane waves, which illuminate the object. At the plane $Z = 4f$ where the grating G_2 is located, the wavefront is the sum of many laterally shifted object wavefronts with a defocus of distance D. Each shifted object wavefront propagates in a different direction. If a filter F_1 is used to pass only finite diffraction orders, then only those orders passed by filter F_1 will reach the plane $Z = 4f$. The grating G_2 further diffracts the incident wavefront into more diffraction orders. Those wavefronts traveling with the same direction cosine can be extracted by a binary filter F_2 located at $Z = 6f$, and the final image is formed at plane $Z = 8f + D$. As the field at the image plane is the sum of

many laterally (or, with circular gratings, radially) shifted replicas of the object wavefronts, this interferometer is a shearing interferometer.

6.3.2. Theory

For mathematical simplicity, we assume that a point source located at the origin of the front focal plane of lens L_0 is used to illuminate the whole optical system. A pair of linear gratings, checkerboard gratings, circular gratings, can be used as modulation and demodulation masks. In this section, a pair of linear gratings is used.

The field just before the object, from Eq. (6.2), is

$$U(x, y;\ D-0) = \sum_n A_n\, e^{-i\pi\lambda D\, n^2/d^2}\ e^{2\pi i\, nx/d} \quad . \tag{6.29}$$

By the Kirchhoff boundary condition, the field just behind the object $U_0(x, y)$ is

$$U(x, y;\ D+0) = U_0(x, y) \sum_n A_n\, e^{-i\pi\lambda D\, n^2/d^2}\ e^{2\pi i\, nx/d} \quad . \tag{6.30}$$

The wave propagates through lens L_1 and arrives at the first filter plane F_1. The field at the plane just before the plane F_1, by the Fourier transforming property of lens L_1, as shown in Appendix A, is

$$U(x, y;\ 2f-0) = \exp\ \frac{i\pi D}{\lambda f^2}\ (x^2+y^2) \sum_n A_n\, \tilde{U}_0\left(\frac{x}{\lambda f} - \frac{n}{d},\ \frac{y}{\lambda f}\right)$$

$$\times\ \exp\ (-i\pi\lambda D\, n^2/d^2) \quad . \tag{6.31}$$

The phase factor in Eq. (6.31) is due to the fact that that object is at a distance D to the right of the front focal plane of lens L_1. If the fundamental spatial frequency of gratings G_1 and G_2, 1/d, is higher than the bandwidth of the object $U_0(x, y)$, $\tilde{U}_0\left(\frac{x}{\lambda f} - \frac{n}{d}, \frac{y}{\lambda f}\right)$ will be separated into separate islands for different diffraction orders n. A filter with transmittance function $F_1(x, y; n)$ is placed at the n-th island of the filter plane F_1. The transmittance function $F_1(x, y; n)$ may be 0 (opaque block), 1 (open window) or a pinhole. The field after the filter F_1 is

$$U(x, y; 2f+0) = \sum_n A_n \exp\left[-\frac{i\pi D}{\lambda f^2}(x^2+y^2)\right]$$

$$\exp(-i\pi\lambda D\, n^2/d^2)\, \tilde{W}_n\left(\frac{x}{\lambda f} - \frac{n}{d}, \frac{y}{\lambda f}\right), \qquad (6.32)$$

where

$$\tilde{W}_n\left(\frac{x}{\lambda f} - \frac{n}{d}, \frac{y}{\lambda f}\right) = \tilde{U}_0\left(\frac{x}{\lambda f} - \frac{n}{d}, \frac{y}{\lambda f}\right) F_1(x, y; n). \qquad (6.33)$$

The field at the plane just before the grating G_2, by the Fourier transforming property of lens L_2, is

$$U(x, y; 4f-0) = \sum_n A_n\, e^{-2\pi i\, nx/d} \int W_n(x_1, y_1)$$

$$\times \exp\left\{-\frac{i\pi}{\lambda D}\left[\left(x_1 + x - \frac{n\lambda D}{d}\right)^2 + (y_1+y)^2\right]\right\} dx_1 dy_1, \qquad (6.34)$$

where

$$W_n(x, y) = \iint \tilde{W}_n(x_2, y_2) \exp\left[2\pi i (xx_2 + yy_2)\right] dx_2 dy_2 \quad . \tag{6.35}$$

The second grating has the transmittance function $G_2(x) = \sum_m B_m \exp(2\pi i\, mx/d)$. The field just after grating G_2 is

$$U(x, y;\, 4f+0) = \sum_n \sum_m A_n B_m\, e^{-2\pi i (n-m)\, x/d}$$

$$\times \iint W_n(x, y) \exp\left\{-\frac{i\pi}{\lambda D}\left[\left(x + x_1 - \frac{n\lambda D}{d}\right)^2 + (y + y_1)^2\right]\right\} dx_1 dy_1 \, . \tag{6.36}$$

From Eq. (6.36), we know that if the wave propagates a distance D it will form an image of $W_n(x, y)$. It is clear from the geometrical optics that the plane $Z = 4f + D$ is the image plane for $Z = D$. The field at $Z = 4f + D$ is

$$U(x, y;\, 4f + D) = \sum_n \sum_m A_n B_m W_n\left(-x + \frac{m\lambda D}{d},\ -y\right)$$

$$\times \exp[-2\pi i (n-m)\, x/d] \cdot \exp\left[\frac{-i\pi\lambda D (n-m)^2}{d^2}\right]$$

$$= \sum_p I_p(x, y)\, e^{-2\pi i\, px/d} \quad , \tag{6.37}$$

where

$$I_p(x, y) = \sum_n A_n B_{n-p} W_n\left(-x + \frac{n-p}{d}\,\lambda D,\ -y\right) e^{-i\pi\lambda D\, p^2/d^2} \, . \tag{6.38}$$

If the second filter F_2 is a binary filter and only passes the p-th diffraction order, then the field in the final output plane $Z = 8f + D$ is

$$V(x, y) = I_p(-x, -y) \tag{6.39}$$

$$= \exp(-i\pi\lambda D\, p^2/d^2) \sum_n A_n B_{n-p} W_n\left(x + \frac{n-p}{d}\lambda D,\ y\right). \tag{6.40}$$

The final output field is a sum of many laterally shifted signals $W_n\left(x + \frac{n-p}{d}\lambda D,\ y\right)$ with amplitude strengths $A_n B_{n-p}$.

6.3.3. Lateral Shearing Interferometer with Variable Shear

The real-time carrier-frequency interferometer can be used as a two-beam or multiple-beam shearing interferometer depending on how many diffraction orders that the filter F_1 has passed, or on how many A_n's are not zero.

If the filter F_1 is a binary filter which passes only the 0-th and 1-st diffraction orders, and filter F_2 is also a binary filter that passes only the diffraction order $p = 1$, the field at the image plane $Z = 8f + D$, from Eq. (6.40), is

$$V_1(x, y) = A_0 B_{-1} U_0\left(x + \frac{\lambda D}{d},\ y\right) + A_1 B_0 U_0(x, y)\ , \tag{6.41}$$

where an unimportant phase factor $\exp(-i\pi\lambda D/d^2)$ has been omitted.

This interferometer is a two-beam shearing interferometer. The amount of shear, $\lambda D/d$, is easily varied in real time. If two

gratings G_1 and G_2 are identical, $|A_n| = |B_n|$ for all n, and $A_1B_0 = A_0B_{-1}\exp(i\alpha)$; the two beams have equal strength. α is a constant depending on the relative lateral shift in the x-direction. It can be adjusted so that $\alpha = \pi$, and then $A_1B_0 = -A_0B_{-1}$.

If the filter F_1 passes only diffraction orders 0, 1, and -1, and F_2 passes only diffraction order $p = 0$, the final output field is

$$V_2(x, y) = A_{-1}B_{-1}U_0\left(x + \frac{\lambda D}{d}, y\right) + A_0B_0U_0(x, y) + A_1B_1U_0\left(x - \frac{\lambda D}{d}, y\right). \tag{6.42}$$

This is a three-beam interferometer. It can be used in high-order differentiation as will be shown in the later section of this thesis.

6.3.4. Real-Time Image Holographic Interferometer

If the filter F_1 passes only the +1 diffraction order and has a pinhole at the center of the island of zeroth diffraction order, and F_2 passes only the p-th diffraction order, the field in the output image plane is

$$V(x, y) = aA_0B_{-p} + A_1B_{1-p}U_0\left(x + \frac{1-p}{d}\lambda D, y\right), \tag{6.43}$$

where

$$a = \iint U_0(x, y)\, dx\, dy$$

is the total transmittance spatially integrated over the whole object. For $p = 1$

$$V(x, y) = aA_0B_{-1} + A_1B_0U_0(x, y). \tag{6.44}$$

This interferometer is a two-beam total shearing interferometer; the output field is independent of parameters λ and D. This optical system is achromatic, and natural white light can be used.

The real-time image holographic interferometer can be applied to image contrast reversal and phase object visualization.

6.3.5. Real-Time Carrier-Frequency Interferometry with Checkerboard Gratings, Circular Gratings, and Random Diffusers as Modulation and Demodulation Masks

All the modulation masks used in the encoding process of double-exposure carrier-frequency interferometry can be used in real-time carrier-frequency interferometry. The optical setup is identical to the one shown in Fig. 6.3 except that the gratings G_1 and G_2 are replaced by checkerboard gratings, circular gratings, or random diffusers. The mathematical analysis and results are similar to those for linear gratings. Here we omit the details of the derivation and just present the results for checkerboard gratings and circular gratings. They have certain applications.

For checkerboard gratings with period d_x in x direction and d_y in y direction the field at the plane $Z = 4f + D$ is

$$U(x, y\,;\, 4f+D) = \sum_m \sum_n \sum_\ell \sum_p A_{mn} B_{\ell p} \exp[-2\pi i(m-\ell)x/d_x - 2\pi i(n-p)y/d_y]$$

$$\exp\left[-\frac{i\pi\lambda D(m-\ell)^2}{d_x^2} - \frac{i\pi\lambda D(n-p)^2}{d_y^2}\right] W_{mn}\left(-x + \frac{\ell\lambda D}{d_x},\ -y + \frac{p\lambda D}{d_y}\right), \tag{6.45}$$

where

$$W_{mn}(x, y) = \iint \tilde{W}_{mn}(x_1, y_1) \exp[2\pi i(xx_1 + yy_1)]\, dx_1 dy_1 \quad , \qquad (6.46)$$

$$\tilde{W}_{mn}(x, y) = \tilde{U}_0(x, y)\, F_1(x, y; m, n) \quad , \qquad (6.47)$$

and A_{mn} and $B_{\ell p}$ are the Fourier coefficients of the transmittance functions of gratings G_1 and G_2, and $F_1(x, y; m, n)$ is the filter transmittance function for the island of index (m, n) in the filter plane F_1. Equation (6.45) can be rewritten as

$$U(x, y; 4f + D) = \sum_{\ell} \sum_{p} I_{\ell, p}(x, y)\, e^{-2\pi i \ell x/d_x} e^{-2\pi i py/d_y} , \qquad (6.48)$$

where

$$I_{\ell, p}(x, y) = \exp\left[-i\pi\lambda D\left(\frac{\ell^2}{d_x^2} + \frac{p^2}{d_y^2}\right)\right] \sum_{m} \sum_{n} A_{mn} B_{m-\ell, n-1} W_{mn}$$

$$\left(-x + \frac{m-\ell}{d_x}\lambda D\ ,\ -y + \frac{n-p}{d_y}\lambda D\right). \qquad (6.49)$$

If the filter F_2 passes only the (ℓ, p) diffraction order, the output field is

$$V_{\ell, p}(x, y) = I_{\ell, p}(-x, -y)$$

$$= e^{i\phi(\ell, p)} \sum_{m} \sum_{n} A_{mn} B_{m-\ell, n-p} W_{mn}\left(x + \frac{m-\ell}{d_x}\lambda D\ ,\ y + \frac{n-p}{d_y}\lambda D\right) . \qquad (6.50)$$

where $\varphi(\ell, p)$ is a function of ℓ and p only.

For circular gratings, if the filter F_2 passes only the p-th diffraction order, the final field is

$$V(r,\varphi;\, p) = \sum_n A_n B_{n-p} W_n \left(r - \frac{n-p}{d} \lambda D \,,\, \varphi \right) \exp(-i\pi\lambda D\, p^2/a^2) \,, \quad (6.51)$$

where a is the period of the circular gratings.

The interferometer using checkerboard gratings can be used as a multiple-beam lateral shearing interferometer, and it becomes a constant radial shearing interferometer when the circular gratings are used. They can be used in spatial differentiation.

6.4. Applications

The carrier-frequency interferometer is a variable shearing interferometer. It can be applied to phase-object visualization, image contrast reversal and color complementarity, spatial differentiation, and complex addition and subtraction.

The spatial frequency of the gratings used in the carrier-frequency interferometer has to be higher than the bandwidth of the object spectrum in order to eliminate the aliasing and satisfy the sampling theorem. The size of the input object is not a limiting factor. But in the grating interferometer, the size of the object is limited to $\lambda f/d$, where d is the period of the grating used. This is the main difference between the grating interferometer and the carrier-frequency interferometer.

6.4.1. Phase-Object Visualization

6.4.1.1 Phase-Object Visualization Using Double-Exposure Carrier-Frequency Interferometry

Double-exposure carrier-frequency photography has been used in phase object visualization. Linear gratings,[5] circular gratings,[6] and even random diffusers[13] have been successfully used as the modulation carriers. In this section we will only present the cases of circular gratings and linear gratings used as modulation masks to encode phase objects.

The optical setup is as shown in Fig. 6.1(a). For simplicity, we assume that $M = 1$, i.e., $d_o = d_i = 2f$. Two exposures are made, one with the phase object under test $U_1(x, y) = \exp(i\phi(x, y))$ present, the other with the standard object $U_2(x, y)$ present. The standard object can be a plane wave, a known wavefront, or the shifted version of the phase object under test.

When the modulation mask is a circular grating, the total exposure, as shown in Eq. (6.18), is

$$E(r, \varphi) = \sum_p C'_p(r, \varphi) \exp(2\pi i\, rp/a) ,$$

where

$$C'_p(r,\varphi) = \sum_m C_{m+p} C_m \left\{ U_1\left(r + \frac{(m+p)\lambda D}{a}, -\varphi\right) U_1^*\left(r + \frac{m\lambda D}{a}, -\varphi\right) \right.$$

$$\left. + U_2\left(r + \frac{(m+p)\lambda D}{a}, -\varphi\right) U_2^*\left(r + \frac{m\lambda D}{a}, -\varphi\right) \right\}. \qquad (6.52)$$

The most often used standard wavefront is a plane wave, i.e., $U_2(x, y) = 1$. Then before the second exposure, the tested phase object is just removed. Equation (6.52) becomes

$$C'_p(r, \varphi) = \sum_m C_{m+p} C_m \left\{ 1 + U_1\left(r + \frac{(m+p)\lambda D}{a}, -\varphi\right) U_1^*\left(r + \frac{m\lambda D}{a}, -\varphi\right) \right\}. \tag{6.53}$$

Let us consider the situation when $\Delta = n\lambda D/a$ is small, and the Taylor series expansion of the tested object $U_1(r - \Delta, \varphi) = e^{i\phi(r-\Delta, -\varphi)} = \exp i[\phi(r, -\varphi) - \Delta\phi_r(r, -\varphi) + \cdots]$ can be well approximated by only the first two terms, where $\phi_r(r, \varphi) = \frac{\partial}{\partial r}\phi(r, \varphi)$. Equation (6.53) becomes

$$C'_p(r,\varphi) = \sum_m C_m C_{m+p} \left(1 + e^{-ip\Delta\phi_r(r, -\varphi)}\right) \tag{6.54}$$

$$= C\left(1 + e^{ip\Delta\phi_r(r, -\varphi)}\right)$$

$$= C \exp\left(\frac{1}{2} p\, \Delta\phi_r(r, -\varphi)\right) \cos\left(\frac{1}{2} p\Delta\phi_r(r, -\varphi)\right) \tag{6.55}$$

where C is a constant. We can place the developed film in a telecentric image-forming system and pass the p-th diffraction by using a binary filter in spatial frequency. The intensity of the output field at the image plane of the telecentric image-forming system is

$$I(r,\varphi) = |C'_p(r, -\varphi)| = c^2 \cos^2\left(\frac{1}{2} p\Delta\phi_r(r, \varphi)\right)$$

$$= c^2 [1 + \cos(p\Delta\phi_r(r, \varphi))] \tag{6.56}$$

The intensity reveals the radial derivative of the phase variation of the tested phase object.

If the modulation mask is a linear grating, and the standard object is $U_2(x, y) = 1$, then the total exposure, from Eq. (6.10), is

$$E(x, y) = \sum_p I_p(x, y) \exp(-2\pi i\, px/d) \quad ,$$

where

$$I_p(x, y) = \sum_m A_{m+p} A_m^* \left[\exp(2\pi i\, ps/d) \right.$$

$$\left. + U_1\left(-x + \frac{(m+p)\lambda D}{d}\,,\ -y\right)\ U_1^*\left(-x + \frac{m\lambda D}{d}\,,\ y\right)\right] \quad . \tag{6.57}$$

We place the developed film in the object plane of a telecentric image-forming system to perform spatial filtering. If the system passes only the +1 diffraction order, the output field is

$$V(x, y) = I_p(-x, -y)$$

$$= A_0 A_1 \left[\exp(2\pi i\, s/d) + U_1\left(x + \frac{\lambda D}{d}\,,\ y\right) U_1^*(x, y)\right]$$

$$+ A_{-1}^* A_0 \left[\exp(2\pi i\, s/d) + U_1(x, y)\ U_1^*\left(x - \frac{\lambda D}{d}\,,\ y\right)\right]$$

$$= A_0 A_1 [\exp(2\pi i\, s/d) + \exp(i2\,\Delta\phi_x(x, y))] \quad , \tag{6.58}$$

if $\Delta = \lambda D/d$ is smaller than the size of typical fine details of the phase variation, where $\phi_x(x, y) = \partial\,\phi(x, y)/\partial x$, and s is the shift of the grating between exposures. The intensity of the output field is

$$I(x, y) = |V(x, y)|^2 = |A_0 A_1|^2 \cos^2(2\Delta\phi_x(x, y) + 2\pi s/d) . \qquad (6.59)$$

The gradient of the phase variation, $\phi_x(x, y)$, is revealed in the interference fringes. The result is similar to Eq. (5.14) of the grating interferometer.

Double-exposure carrier-frequency interferometry can be performed via the use of an image forming system, or by using the Talbot effect. Lensless carrier-frequency interferometry can be used in phase object visualization; the result will be identical to Eq. (6.56), and (6.59) if the modulation mask is a circular grating or a linear grating, except that the contrast is reduced by a factor $\cos(\lambda Z\pi/d^2)$, where Z is the distance between the grating and the recording film. If $Z = 2Nd^2/\lambda$, the Talbot distance, then $\cos(\lambda Z\pi/d^2) = 1$, and there is no loss of contrast; N is an integer.

6.4.1.2. Phase-Object Visualization Using Real-Time Carrier-Frequency Interferometry

The real-time carrier-frequency interferometer can be used as a two-beam interferometer. For phase-object visualization, if the filter F_1 is a binary filter which passes only the zeroth and first diffraction orders and F_2 passes only first diffraction order, the field at the image plane $Z = 8f + D$, as shown in Eq. (6.41), is

$$V(x, y) = A_0 B_{-1} U_0(x + \Delta, y) + A_1 B_0 U_0(x, y) . \qquad (6.60)$$

We have assumed that linear gratings are used. If gratings G_1 and

G_2 are identical, then $|A_n| = |B_n|$, for all n, and $A_0 B_{-1} = A_1 B_0 e^{i\alpha}$, α is a constant. The intensity of the output field is

$$I(x, y) = |V(x, y)|^2 = 2|A_0 B_{-1}|^2 \{1 + \cos[\phi(x + \Delta, y) - \phi(x, y) - \alpha]\}$$

$$\simeq 2|A_0 B_{-1}|^2 \{1 + \cos[\Delta\phi_x(x, y) - \alpha]\} \qquad (6.61)$$

if $\Delta = \lambda D/d$ is smaller than the size of fine details of the phase variation δ. The phase α can be adjusted by laterally shifting either grating G_1 or G_2. For $\alpha = \pi/2$, Eq. (6.61) becomes

$$I(x, y) = |A_0 B_{-1}|^2 \sin^2\left(\frac{\Delta}{2}\phi_x(x, y)\right) .$$

$$\approx \left|A_0 B_{-1} \frac{\Delta}{2}\phi_x(x, y)\right|^2 \qquad (6.62)$$

if $\frac{1}{2}\Delta\phi_x(x, y) \ll 1$. The output image is proportional to the derivative of the phase variation. The interference fringes can be interpreted as the contours of the derivative of the phase object along x-axis.

For observing the phase variation instead of the gradient of the phase variation of phase objects, the real-time image holographic interferometer can be used. If the filter F_1 passes only the first diffraction order and has a pinhole at the center of the zeroth diffraction order, and F_2 passes only the first diffraction order, the field in the output image plane, from Eq. (6.44), is

$$V_a(x, y) = aA_0B_{-1} + A_1 B_0 U_0(x, y) \quad , \tag{6.63}$$

The intensity is

$$I_a(x, y) = |V_a(x, y)|^2$$

$$= |A_0 B_{-1}| \sin^2\left(\frac{1}{2}\phi(x, y)\right) \quad , \tag{6.64}$$

if coefficients A_1 and B_{-1} have relative phase difference π, and $a = 1$. It can be achieved if the images of grating G_1 and grating G_2 are laterally shifted by half a period. The interference fringe is the contour of phase variation. As parameter λ does not appear in Eq. (6.63), the system is achromatic; natural white light can be used. For an object with small phase variation $\phi(x, y) \ll 1$, Eq. (6.64) can be rewritten as

$$I_a(x, y) \simeq \left|\frac{1}{2} a_0 B_{-1} \phi(x, y)\right|^2 \quad .$$

It is similar to phase-contrast microscopy.

In the above discussions, we have always assumed that gratings G_1 and G_2 have the same orientation with their slits parallel to y-axis. For phase-object visualization using carrier-frequency interferometry, the results are shown in Eqs. (6.56), (6.59), (6.62) and (6.64). The interference fringes do not reveal the information about the sign of the phase variation (or gradient of the phase variation). The loss of sign results from the fact that what we observe is intensity, not the amplitude.

If a reference can be provided, the information about the sign of the phase variation (or gradient of the phase variation) can be preserved. We rotate grating G_1 through an angle θ with respect to y-axis, and rotate G_2 through an angle $-\theta$. Following the same analysis as in section 6.3, the field in the final output plane $Z = 8f + D$ of Fig. 6.3, is

$$V_p(x, y) = \exp(-i\pi\lambda D p^2/d^2) \sum A_n B_{n-p}(x + (n-p)\Delta_x;\ y + (n-p)\Delta_y)$$

$$\exp[2\pi i\,(2n-p)\,y \sin\theta/d] \quad , \qquad (6.65)$$

where

$$\Delta_x = \lambda f \cos\theta/d \quad \text{and} \quad \Delta_y = \lambda f \sin\theta/d \quad .$$

If the tested phase object is placed in the input object plane of Fig. 6.3, and if the filter F_1 passes only diffraction orders 0 and +1, and F_2 passes only the +1 diffraction order, the field in the final output plane is

$$V(x, y) = A_0 B_{-1} U_0(x - \Delta_x,\ y - \Delta_y)\, e^{-2\pi i\, y \sin\theta/d}$$

$$+ A_1 B_0 U_0(x, y)\, e^{2\pi i\, y \sin\theta/d} \quad . \qquad (6.66)$$

and the intensity is

$$I(x, y) = |V(x, y)|^2$$

$$\cong |A_0 B_{-1}|^2 \cos^2\left(\pi y \sin\theta/d + \frac{1}{2}\Delta_x \phi_x(x, y) + \frac{1}{2}\Delta_y \phi_y(x, y)\right) \ , \qquad (6.67)$$

if Δ_x and Δ_y are smaller than the fine details of the object. In Eq. (6.67) a reference sinusoidal wave $\cos(2\pi y \sin\theta/d)$ serves as a reference.

When the filter F_1 passes only the +1 diffraction and has a pinhole at the center of the zeroth diffraction order, and F_2 passes only the +1 diffraction order, the field in the image plane is

$$V(x,y) = a_o B_1 e^{-2\pi i y \sin\theta/d} + A_1 B_0 U_0(x,y) e^{2\pi i y \sin\theta/d}; \quad (6.68)$$

it is the interference between the tested object and a tilted plane wave. The intensity is

$$I(x,y) = |a_0 B_1|^2 + |A_1 B_0|^2 + 2|A_1 B_0 a_0 B_1| \cos(4\pi y \sin\theta/d + \phi(x,y) + \alpha), \quad (6.69)$$

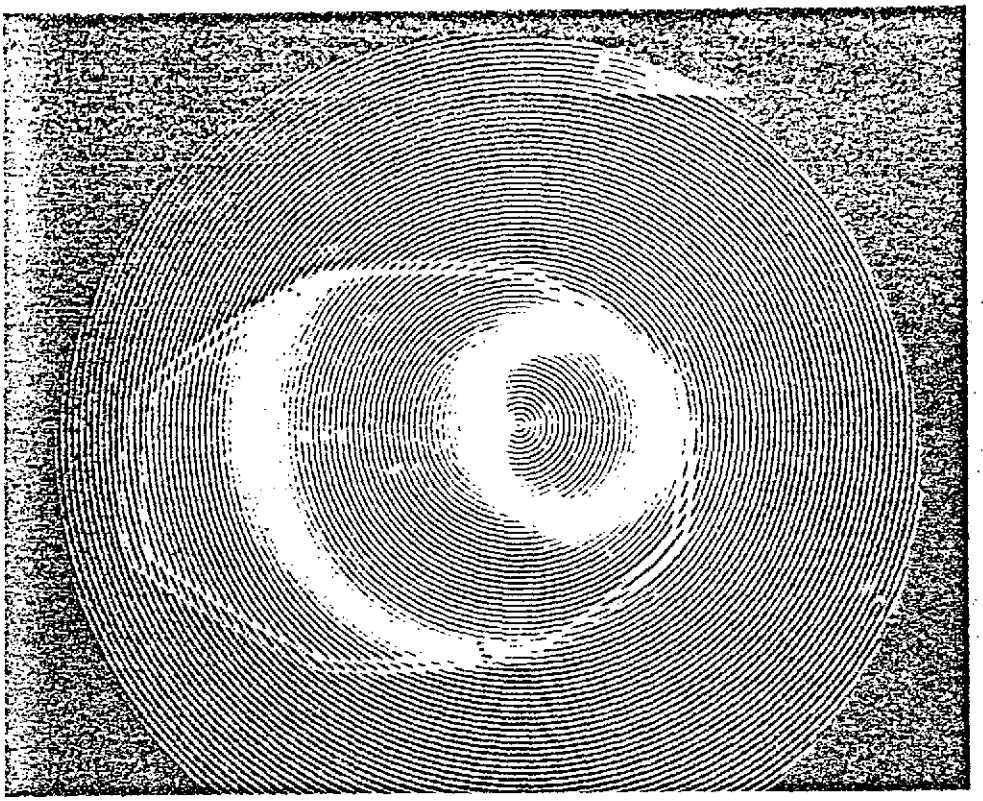
where α is a constant. For $\theta = 0$, Eq. (6.69) is equal to Eq. (6.64).

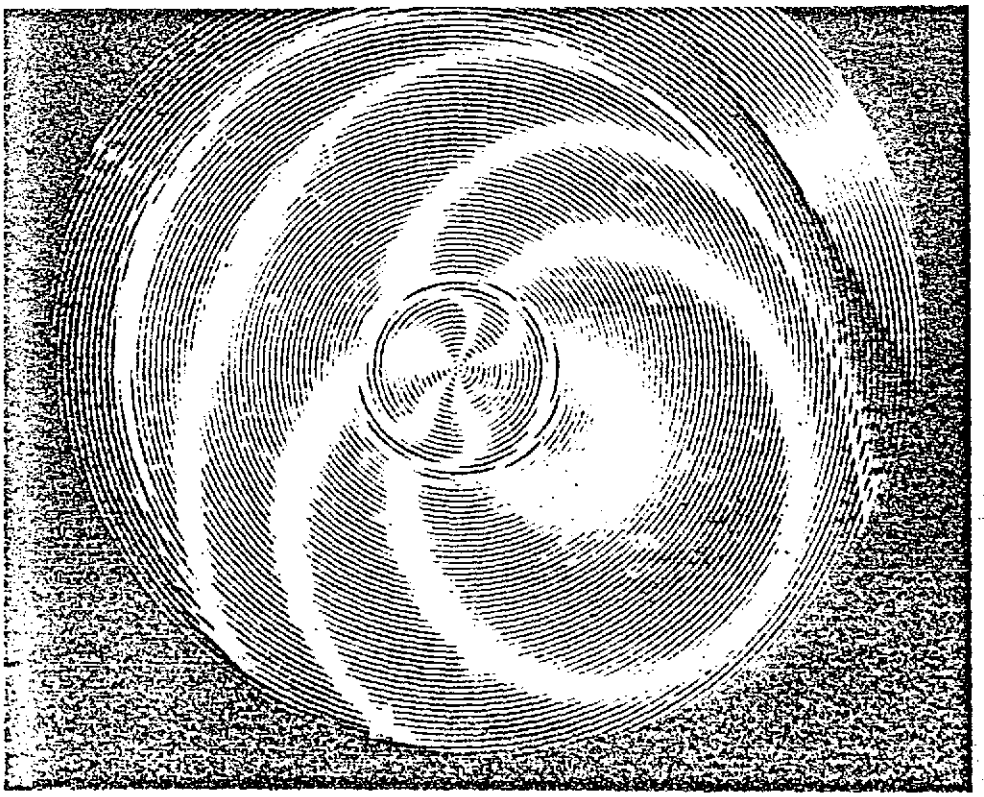
6.4.1.3. Experiment

An experiment on double-exposure circular carrier-frequency interferometry has been performed. A helium-neon laser was used as a source whose beam was expanded to 50 mm. The circular grating used as the modulation mask had a period of 0.25 mm. The object was placed 1 mm behind the grating, and a camera was used to image the grating onto the film. An eyeglass lens was used as a test object. Figure 6.4 shows the result of testing the lens. Two bright fringes in the shape of concentric circles were observed. Another lens with a drop of plastic resin in its center was also tested. Figure 6.5 shows the lens under test.

Figure 6.4. An eyeglass lens under test. The white fringes are fringes of equal radial phase derivative. The fine circular lines belong to the circular grating.

Figure 6.5. An eyeglass lens with a drop of resin in its center is the tested object.





6.4.2. Spatial Differentiation

6.4.2.1. Spatial-Differentiation Using the Double-Exposure Carrier-Frequency Interferometry

Double-exposure carrier-frequency interferometry has been used in image subtraction, since it can produce bipolarity. Bipolarity is essential to subtraction. Differentiation is a special kind of subtraction.

The modulation mask (grating) can be placed either in the object plane or in the film plane of an image-forming system as shown in Fig. 4.5. Incoherent light can be used and is preferred to coherent light for it eliminates the coherent noise. To achieve the bipolarity, the grating has to be shifted half a period between exposures. It is difficult to move it exactly half a period. This mechanical movement of the grating can be eliminated if a polarization grating is used instead of an ordinary grating as the modulation mask. A polarization grating consists of mutually orthogonal stripes of polarizing material as shown in Fig. 6.6. A polarization plate serving as an analyzer is placed in front of the grating. If the input light is unpolarized, natural light, and if the analyzer passes only polarized light in the x-direction, then the light intensity distribution in the film will be a binary grating consisting of black and white stripes. If we rotate the analyzer 90°, the grating image shifts half a period. By rotating the analyzer we can shift the grating half a period and create bipolarity.

Two exposures are made. In the first exposure, we image the object $U_1(x, y)$ onto the film. In the second exposure, we rotate the

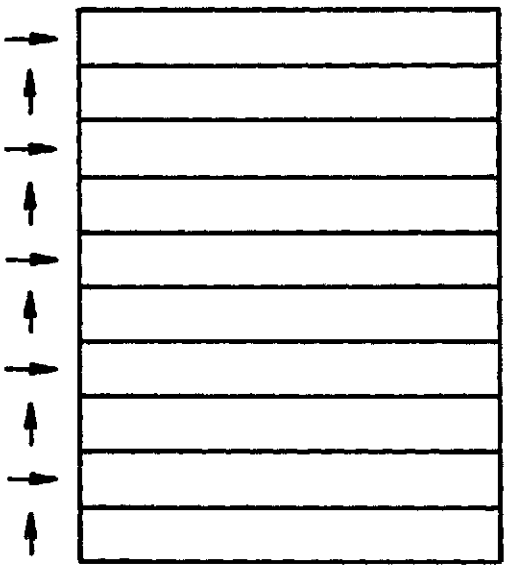


Figure 6.6. A polarized grating. Here ↑ means polarized in y-direction, → mean polarized in x-direction.

analyzer 90° and image the second object $U_2(x, y)$ onto the film. The total exposure is

$$E(x, y) = I_1(x, y)\, G(x) + I_2(x, y)\, G(x - d/2)$$

$$= \sum A_n \exp(2\pi i\, nx/d)\,[I_1(x, y) + e^{-in\pi}\, I_2(x, y)]\ , \qquad (6.70)$$

where $I_1(x, y)$ and $I_2(x, y)$ are intensities of objects U_1 and U_2. We can place the developed film in the input object plane of a coherent signal processing system and extract only the $n = 1$ diffraction order; the final image is proportional to $|I_1(x, y) - I_2(x, y)|$, the difference between I_1 and I_2.

If the second object $U_2(x, y)$ is equal to $U_1(x - \Delta, y)$, the first object with a shift Δ in the x-direction, as when the object $U_1(x, y)$ or the camera set (film and grating) is shifted between exposures, Eq. (6.70) becomes

$$E(x, y) = \sum A_n \exp(2\pi i\, nx/d)\,[I_1(x, y) + e^{-in\pi}\, I_1(x - \Delta, y)]\ . \qquad (6.71)$$

The first diffraction order is proportional to

$$I_1(x, y) - I(x - \Delta, y) \cong \Delta\, \frac{\partial I(x, y)}{\partial x}\ ,$$

if Δ is smaller than the typical fine detail of the object.

Spatial differentiation along another direction can be obtained by shifting the object along that direction between exposures. Higher-order differentiation can be achieved by multiple exposures. The

polarity and intensity strength of each exposure is given in Eqs. (5.38), (5.39), and (5.40) for the evaluation of $\partial I/\partial x + \partial I/\partial y$, $\partial^2 I/\partial x^2 + \partial^2 I/\partial y^2$, and $\partial^2 I/\partial x \partial y$.

6.4.2.2. Spatial Differentiation Using the Real-Time Carrier-Frequency Interferometry

As the real-time carrier-frequency interferometer can produce differential shear, it can be used in spatial differentiation. The optical setup in Fig. 6.3 is used. The object to be differentiated is placed in the object plane OBJ which is at a distance D to the right of grating G_1 . The field in the final output plane $Z = 8f + D$, from Eq. (6.40), is

$$V_p(x, y) = \sum_n A_n B_{n-p} W_n\left(x + \frac{n-p}{d} \lambda D, y\right) , \qquad (6.72)$$

where

$$W_n\left(x + \frac{n-p}{d} \lambda D , y\right) = U_0\left(x + \frac{n-p}{d} \lambda D, y\right)$$

if the n-th diffraction order passes the filter F_1 , and $W_n\left(x + \frac{n-p}{d} \lambda D, y\right) = 0$, if the n-th diffraction order has been blocked by the filter F_1 .

For spatial differentiation, we can vary the filters F_1 and F_2 for a particular kind of spatial differentiation. For example, we can have filter F_1 pass only the diffraction orders $n = 0$ and 1, and F_2 pass only diffraction order $p = 1$. The output field is

$$V_1(x,y) = A_1B_0U_0(x,y) + A_0B_{-1}U(x+\Delta, y)$$

$$\cong A_1B_0\,\Delta\,\frac{\partial}{\partial x}\,U_0(x,y) \quad , \tag{6.73}$$

where $\Delta = \lambda D/d$. We have assumed $A_0B_{-1} = -A_1B_0$, as can be achieved by shifting the image of grating G_1 in the plane of grating G_2 by half a period with respect to grating G_2 . For example, if the gratings G_1 and G_2 are Ronchi rulings, the transparent part of the grating G_1 will image onto the opaque part of the grating G_2 . There is complete darkness after the plane where the grating G_2 is located.

For second order differentiation, the filter F_1 passes only the diffraction orders $n = -1, 0, 1$, and F_2 passes only the diffraction order $p = 0$. Then the field in the image plane is

$$V_2(x,y) = A_{-1}B_{-1}U_0(x+\Delta, y) + A_0B_0U_0(x,y) + A_1B_1U_0(x-\Delta, y) \, . \tag{6.74}$$

$$\simeq A_1B_1\,\Delta^2\,\frac{\partial^2}{\partial x^2}\,U_0(x,y) \quad ,$$

if

$$A_0B_0 = -2A_1B_1 = -2A_{-1}B_{-1} \quad . \tag{6.75}$$

The equality in Eq. (6.75) can be achieved by shifting either the gratings G_1 and G_2 and choosing gratings with the particular ratio q of space to the period of a binary grating that produces $A_0 = \sqrt{2}\,|A_1|$.

For higher-order differentiation, we have to use checkboard gratings as modulation and demodulation masks. The field in the output image plane, from Eq. (6.49), is

$$V_{\ell,p}(x,y) = \sum_m \sum_n A_{m,n} B_{m-\ell,n-p} W_{mn}(x + (p-m)\Delta,\ y + (\ell-n)\Delta)\ , \tag{6.76}$$

if filter F_2 passes only diffraction order (ℓ, p), where a multiplicative constant has been omitted. $W_{mn}(x + (p-m)\Delta,\ y + (\ell-n)\Delta)$ equals $U_0(x + (p-m)\Delta,\ y + (\ell-n)\Delta)$ if diffraction order (m, n) passes filter F_1, and equals zero if diffraction order (m, n) is blocked.

For the evaluation of the Laplacian, the filter F_1 passes diffraction orders $(m, n) = (1, 0)$, $(0, 1)$, $(0, 0)$, $(-1, 0)$ and $(0, -1)$, and F_2 passes only diffraction order $(\ell, p) = (0, 0)$. The output field, from Eq. (6.76), is

$$\begin{aligned} V(x,y) &= A_{10}B_{10}U_0(x - \Delta,\ y) + A_{01}B_{01}U_0(x, y - \Delta) + A_{00}B_{00}U_0(x,y) \\ &\quad + A_{-1,0}B_{-1,0}U_0(x + \Delta,\ y) + A_{0,-1}B_{0,-1}U_0(x,\ y + \Delta) \\ &\simeq A_{10}B_{10}\,\Delta^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) U_0(x,y)\ , \end{aligned} \tag{6.77}$$

if

$$A_{10}B_{10} = A_{01}B_{01} = A_{-1,0}B_{-1,0} = A_{0,-1}B_{0,-1} = -\frac{1}{4} A_{00}B_{00} \tag{6.78}$$

The relation in Eq. (6.78) can be achieved by shifting the gratings and choosing proper grating space-to-period ratio q.

For the evaluation of $\dfrac{\partial^2 U_0}{\partial x \partial y}$, the filter F_1 passes diffraction

orders $(m,n) = (1,1)$, $(1,-1)$, $(-1,1)$, $(-1,-1)$, and F_2 passes only diffraction order $(\ell,p) = (0,0)$.

6.4.3. Real-Time Image Polarity-Reversal, Color-Complementation, and Two-Dimensional Photometry

6.4.3.1. Real-Time Image Polarity-Reversal and Two-Dimensional Photometry

Image polarity reversal is equivalent to subtracting the image $U_0(x,y)$ from a uniform background b to form $b - U_0(x,y)$. The constant b is the image polarity level. We have shown in section 5.4.4 that the grating interferometer can be used for real-time image polarity reversal and two-dimensional photometry. Here we will show that the real-time image holographic interferometer, discussed in section 6.3.4, can be used for image polarity reversal and two-dimensional photometry.

The optical setup is the same as shown in Fig. 6.3 and is discussed in section 6.3.4. We will assume that the illuminating source is monochromatic and the object $U_0(x,y)$ is black, grey, and white. The field in the output image plane, from Eq. (6.44), is

$$F(x,y) = C\, a_0(\lambda)\, A_0 B_{-1} + A_1 B_0 U_0(x,y;\lambda)\ , \tag{6.79}$$

where

$$a_0(\lambda) = \iint U_0(x,\ y\ ;\ \lambda)\, dxdy \tag{6.80}$$

is the measure of the λ spectral amplitude of the whole object (spatially integrated over the whole object). C is the transmittance of the pinhole. $U_0(x,y;\lambda)$ is the λ-amplitude component of the light field at

point (x, y) of the object plane when the object is illuminated by a unit amplitude monochromatic plane wave of wavelength λ. For a black-and-white object (no color), each λ-component $U_0(x, y; \lambda)$ has the same amplitude distribution (as a function of x and y) as the object $U_0(x, y)$ itself, i.e., if we disregard the color, $U_0(x, y; \lambda)$ is identical to $U_0(x, y)$ for every λ. The A_n's and B_n's are the Fourier coefficients of gratings G_1 and G_2. Constants C, $a_0(x)$, A_0 and B_0 are all real and positive (we have assumed the object $U_0(x, y)$ is real and positive), and the phase of A_1 and B_1 can be varied by shifting the gratings G_1 and G_2 along the direction normal to the grating lines. They can easily be adjusted so that A_1 and B_{-1} are out of phase; then Eq. (6.79) becomes

$$V(x, y) = A_1 B_0 [b(\lambda) - U_0(x, y; \lambda)] , \qquad (6.81)$$

where

$$b(\lambda) = \left| \frac{C A_0 B_{-1}}{A_1 B_0} \right| \cdot \iint U_0(x, y ; \lambda) \, dxdy , \qquad (6.82)$$

is the polarity level. For black-and-white objects, $b(\lambda) = B$ for all λ, where B is a positive constant.

It is clear that the output image will have the same color as the illumination monochromatic source. Positive images, negative images, or partial reversal of the grey scale of the object can be obtained, if we control the image polarity level. Image-transfer characteristics similar to those shown in Fig. 5.4 can be obtained.

The constant C can be controlled by placing a neutral-density wedge or two superimposed polarizers in front of the pinhole for real-time adjustment. C is confined in the region between 0 and 1. To increase the amplitude of the reference beam, we can increase a_0 by passing some direct beam through G_1 and $U_0(x,y)$, and focusing it by lens L_1 into the center of the zeroth diffraction order in the filter F_1.

6.4.3.2. Real-Time Image Color-Complementarity and Polarity-Reversal

In the last section we have shown that a monochromatic positive image can be obtained directly from a black-and-white negative film. In this section, we will consider the situation when the input object is a color object and the illumination source is polychromatic.

The real-time image holographic interferometer is achromatic and can be illuminated by a polychromatic point source. Here we will assume a 'white' light spectrum $W(\lambda)$ uniformly distributed with density equal to a constant a between blue and red. Positive color images can be obtained directly from color negatives, if the illuminating source of the system is a white point source, and the color object itself or the modification of it is color-balanced. For a color-balanced object $b(\lambda) = B$, a constant, i.e., the object $U_0(x,y)$ consists of the same amount of different colors. A black-and-white object is color-balanced.

For example, let a color object $U_0(x, y) = U_0(x)$ be one-dimensional and consist of three equal regions of different colors. The first region is cyan (minus red), the second region is magenta (minus green), and the third is yellow (minus blue). The amplitude of the first two regions vary linearly, and the third region is constant as shown in Fig. 6.7, where we have decomposed the object $U_0(x, y)$ into three primary colors, red, green, and blue. The amplitudes $U_0(x, y; \lambda_r)$, $U_0(x, y; \lambda_g)$, and $U_0(x, y; \lambda_b)$ are the amplitude components of the object light field when the object is illuminated by a unit monochromatic source of red, green, or blue light respectively.

If the system is illuminated by a monochromatic light source, the object light field will be $U_0(x, y; \lambda)$, where λ is the wavelength of the monochromatic point source. For example, if the monochromatic point source is red, the object light field is shown in Fig. 6.7(a), if the input object is the object we have just discussed. Following the same analysis as in the last section, the intensity of the light field in the image plane, from Eq. (6.81), is

$$I(x, y; \lambda_r) = |V(x, y; \lambda_r)|^2$$

$$= |A_0 B_1|^2 [b(\lambda_r) - U_0(x, y; \lambda_r)]^2 , \qquad (6.83)$$

where

$$b(\lambda_r) = \left| \frac{C A_0 B_{-1}}{A_1 B_0} \right| \cdot \iint U_0(x, y; \lambda_r) \, dxdy \quad . \qquad (6.84)$$

$b(\lambda)$ is the λ-component of the image polarity level, or the λ-component of the reference beam. From Eq. (6.83) it is clear that polarity reversal for $U_0(x, y; \lambda_r)$ has been achieved. If $b(\lambda_r) = C_1$, the output field $|V(x, y; \lambda_r)|$ is shown in Fig. 6.8(a). The output image appears reddish.

Because different λ-components are incoherent with each other, when the object $U_0(x, y)$ is illuminated by a polychromatic point source, we have to treat each λ-component independently and then sum them up incoherently. Spatial coherence for each λ-component is still maintained, since the source is a point source. If the color object $U_0(x, y)$ is illuminated by a polychromatic source of red and green light, then the image we observe will be $|U_0(x, y; \lambda_r)|^2 + |U_0(x, y; \lambda_g)|^2$, which consists of three regions of green, red, and yellow. The color of the third region, yellow, is the incoherent sum of red and green. The image we observe is different to the original image. If this red and green source is used to illuminate the interferometer, the image we observe will be

$$I(x, y; \lambda_r) + I(x, y; \lambda_g) = |V(x, y; \lambda_r)|^2 + |V(x, y; \lambda_g)|^2$$

$$= |A_0 B_{-1}|^2 \{[b(\lambda_r) - U_0(x, y; \lambda_r)]^2 + [b(\lambda_g) - U_0(x, y; \lambda_g)]^2\} . \quad (6.85)$$

In the example we have, as shown in Fig. 6.8(a) and (b), we will observe an image consisting of three regions of reddish, greenish, and yellowish color. Comparing Fig. 6.7 and Fig. 6.8, we can see that the polarity

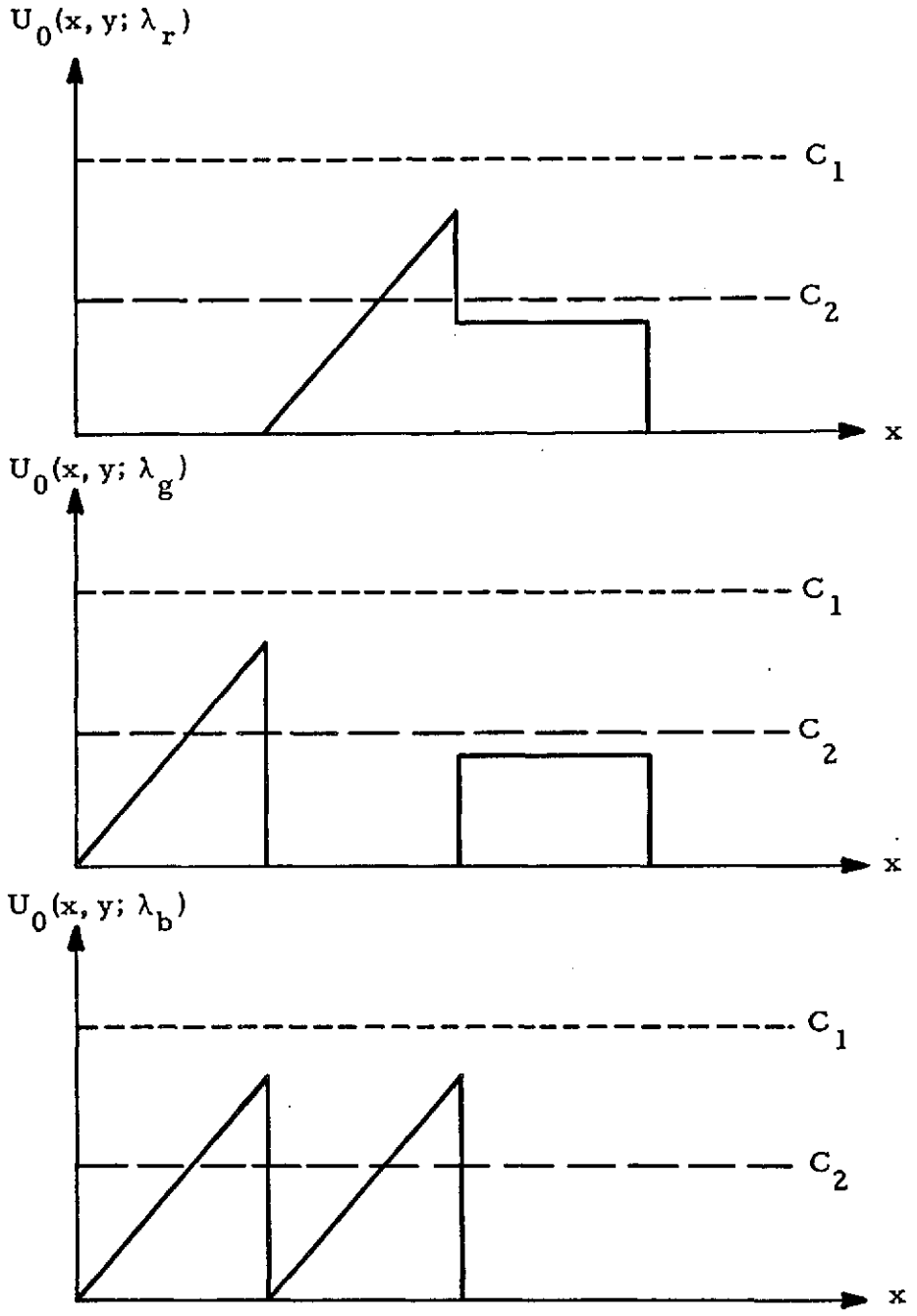


Figure 6.7. The three primary color transmittances of the object. Dashed lines represent the amplitude strength of reference beams C_1 and C_2.

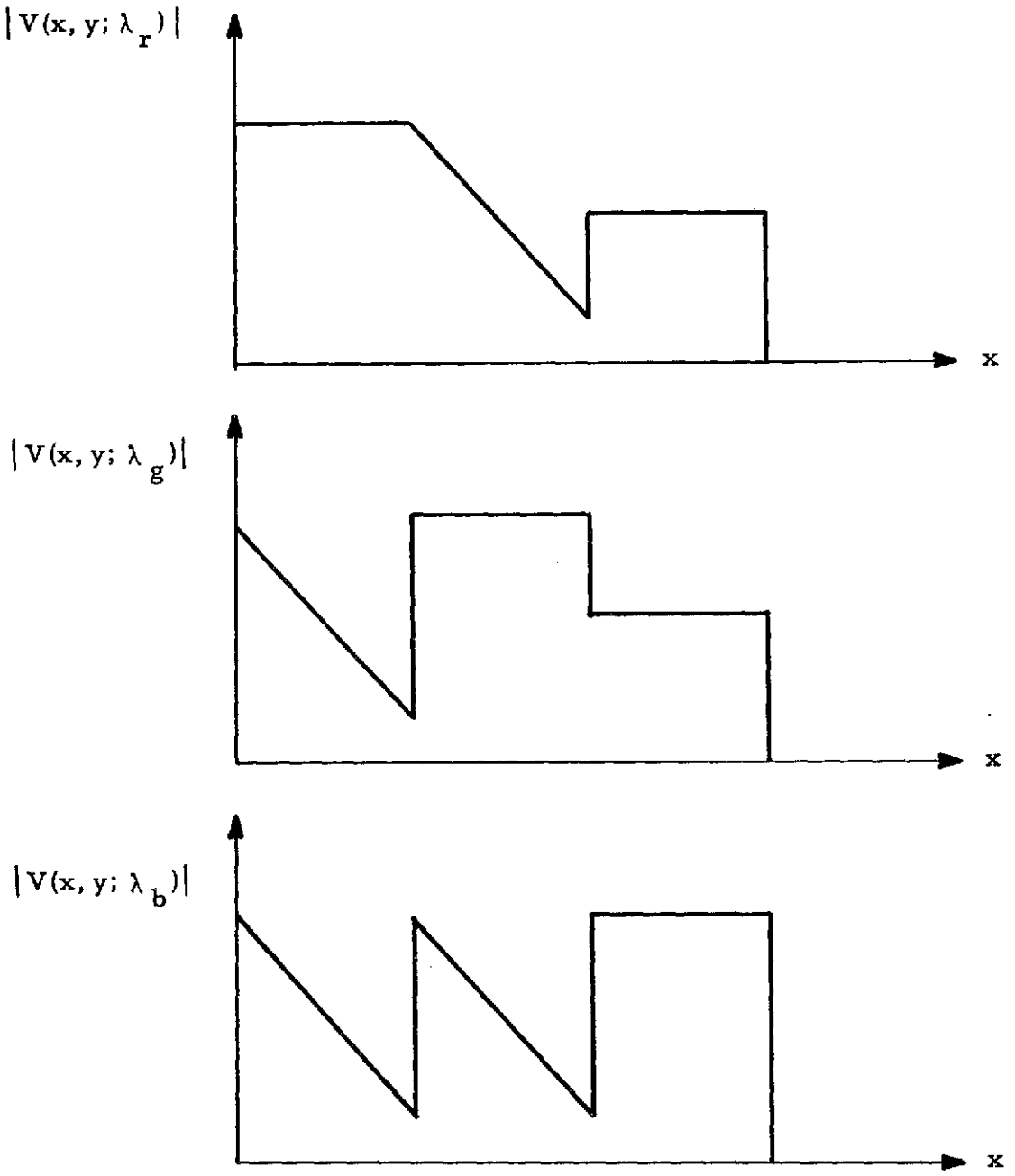


Figure 6.8. The three primary color transmittance of the output field $|V(x)|$ when the reference strength is C_1. Color-complementary and polarity-reversed image is obtained.

of each primary color has been reversed. That is, if at point (x, y) of the color object $U_0(x, y)$ there is no green (or red), then at point (x, y) of the output image plane there will be a lot of green (or red), and vice versa (here we have assumed that the green-component of the reference beam is larger than max $U_0(x, y; \lambda_g)$) .

If the object is illuminated by a polychromatic source of equal amounts of red, green, and blue, the image of the color object is the color object itself as shown in Fig. 6.7. If we use this source to illuminate the interferometer, the output image intensity is

$$I(x, y) = |V(x, y; \lambda_r)|^2 + |V(x, y; \lambda_g)|^2 + |V(x, y; \lambda_b)|^2$$

$$= |A_0 B_{-1}|^2 \{[b(\lambda_r) - U_0(x, y; \lambda_r)]^2 + [b(\lambda_g) - U_0(x, y; \lambda_g)]^2$$

$$+ [b(\lambda_b) - U_0(x, y; \lambda_b)]^2\} . \qquad (6.86)$$

The object we have, $U_0(x, y)$, has equal amounts of red, green, and blue; it is color-balanced, and $b(\lambda) = B$. For if $b(\lambda) = B = C_1$ in Eq. (6.86), the output field $|V(x, y; \lambda)|$ is shown in Fig. 6.8. It consists of three regions of reddish, greenish and bluish color. The color of the output image is the complement of that of the input object $U_0(x, y)$. A color-complemented and polarity-reversed image is thus obtained.

For a polychromatic point source with white spectral density $W(\lambda)$, the λ-component intensity of the image of the interferometer, from Eq. (6.83), is

$$I(x, y; \lambda) = |V(x, y; \lambda)|^2 = |A_0 B_{-1}|^2 S(\lambda) [B_0 - U_0(x, y; \lambda)]^2 ,$$

if the object is color-balanced, and $b(\lambda) = B_0$. For each wavelength λ, the polarity of the point (x, y) at the image plane is reversed. Polarity-reversal in every λ-component is the color-complementarity. If point (x, y) of the object passes only red light, then the image of that point from the interferometer will consist of all other colors except red (minus red), which is complementary to red.

If the object is black-and-white, it is color balanced so the reference wave appears to be white. There is color dispersion in the first diffraction order. The period of the gratings G_1 and G_2 must be small enough to separate different diffraction orders after the color dispersion. The binary filter in the filter plane F_1 passes only the first diffraction order of all λ wavelength components. Positive black-and-white images, or images with partial reversal of the grey scale of the input object can be obtained.

If the object is red, even though the illuminating point source is white, the light after the object will appear red due to the spectral filtering effect of the object. The reference point source at the pinhole of the filter F_1 then appears to be red too. Image polarity reversal can be achieved, but not color-complementarity in this case. The final image is red but with contrast reversal.

To obtain color-complementary images, the reference point source has to be white. In general, some adjustment in the input object

plane will have to be made to achieve this. This can be best understood by an example. Let us consider the object with its three primary color components as shown in Fig. 6.7 but with the third (yellow) region eliminated. This object is not color balanced; blue light is dominant, so the reference point source appears to be bluish. To obtain a white reference source, we have to modify the object by adding a third region of yellow color (minus blue) to achieve color balance. After we add the third region of yellow color, this modified object is then identical to the object shown in Fig. 6.7. Following the same analysis as given before, a color-complemented and polarity-reversed image is thus obtained for the modified object and hence for the input object. Only optical elements, such as gratings and lenses, are used. No complicated and expensive photochemical process or electronics TV technique is required.

The contrast level can be changed independently of the color by varying the strength of the reference beam, $b(\lambda) = B_0$, as explained in the last section. To understand this possibility, let us consider a second example, in which we have adjusted the strength of the reference beam so as to equal C_2 as shown in Fig. 6.7. The resulting image is shown in Fig. 6.9. In this case, the image of the object is color complemented, but only partially polarity reversed.

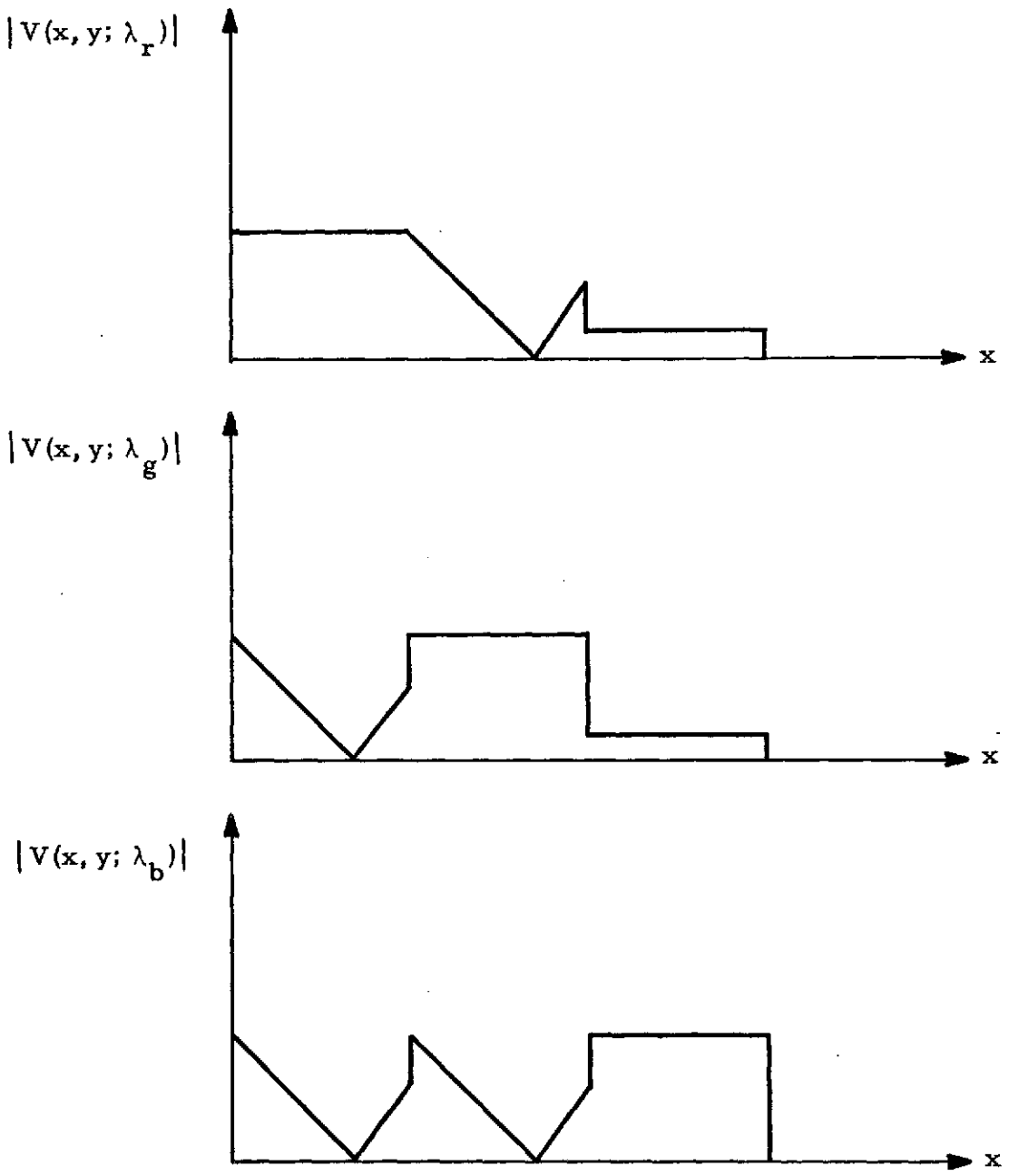


Figure 6.9. The three primary color transmittances of the out field $|V(x)|$ when the strength of the reference beam is C_2. Image color has been complemented, but only partial contrast reversal.

6.4.3.3. Comparison of Various Methods for Image Polarity-Reversal

Two methods for real-time image polarity reversal have been discussed in this thesis, one using the grating interferometer, the other using the real-time carrier-frequency interferometer. In the first method, monochromatic light is necessary, but white light can be used in the second method. Spatially coherent light is necessary. The size of the object should be smaller than $\lambda f/d$ in the first method, and the dimension of the spatial frequency of the spatial bandwidth of the object should be smaller than the frequency of the grating used, d^{-1}, in the second method.

The achromatic property of the second method enables us to obtain a positive black-and-white image directly from a black-and-white negative film, and a positive color image directly from a color negative. Gratings of higher frequency are needed to ensure that the spatial spectra of different diffraction orders do not overlap after the color dispersion effect.

Previous methods[16, 17, 18] using two-stage carrier-frequency photography require a different carrier-frequency photograph to be prepared for each different output image polarity. Good skill in photographic process control, or precise control of grating movement is necessary in those methods. A carrier-frequency photographic negative can be used as the input object in the real-time carrier-frequency interferometer, and it will still provide images with continuously variable polarity from the same carrier-frequency photograph.

Since the polarity level can be easily varied continuously in real time, both methods discussed in this thesis can be used for two-dimensional photometry.[19]

This real-time image-polarity reversal technique is applicable to any image processing such as photography and xerography, photometry, spectral analysis, interferogram analysis, and image evaluation in aerial photography, electron microscopy, and medical diagnosis.

6.5. Focusing and Alignment

In the previous sections, we have often had to place a point source in the front focal plane of a lens, and an object, a grating, or a random diffuser in the back focal plane. It is difficult to place them exactly at the position we want. New methods for focusing and alignment will be presented in this section.

The self-imaging property of a regular grating, or Talbot imaging, is a well known phenomenon. A point source is used via a lens to illuminate the grating as shown in Fig. 6.10. If the grating is illuminated by a plane wave, the Talbot image is identical to the grating itself. Cowley and Moodie[22] have shown that if the wavefront incident on the grating has curvature $1/Z_C$, the position of the self-image of the grating $Z = Z_N$ is given by

$$Z_N^{-1} + Z_C^{-1} = (NZ_T)^{-1} , \tag{6.87}$$

where N is an integer, $Z_T = 2d^2/\lambda$ the Talbot distance. The

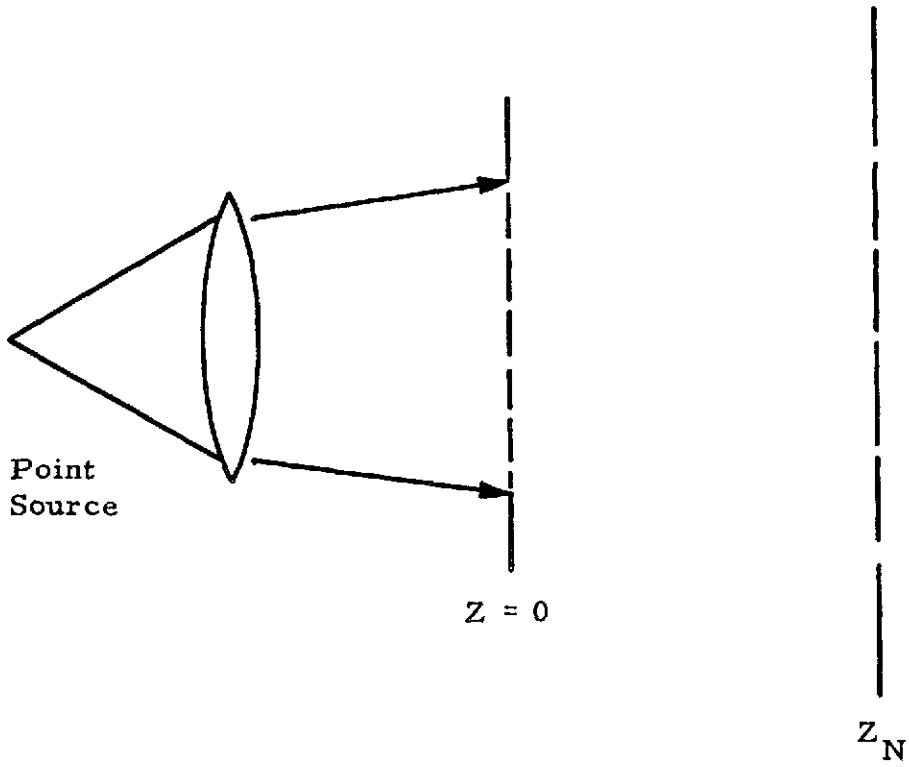


Figure 6.10. A Talbot image forming system for focusing and alignment.

image will be magnified by a factor

$$M_n = 1 + Z_N/Z_C \quad . \tag{6.88}$$

The relation between the curvature of the incident wavefront and the magnification of the self-image has been used to monitor the beam collimation. By beam collimation, we mean placing a point source in the front focal plane of a lens, so that the wavefield after the lens is a plane wave, i.e., $Z_C^{-1} = 0$. To monitor the beam collimation we place two identical linear gratings to the right of the lens, one after the other. If the incident wavefront to the first grating is not a plane wave, the Talbot image of first grating will have a period different from that of the grating itself by a factor M_n , defined by Eq. (6.88). This self-image magnification can be detected by the moiré effect of two gratings of different period. This method of beam collimation and its sensitivity have been analyzed.[20]

We have shown in Appendix C how the Talbot effect occurs for circular gratings. As Eqs. (6.87) and (6.88) also apply to circular gratings, circular gratings can be used in beam collimation. A circular grating can register the position of its center. Owing to this property, circular gratings can be used in alignment. Two circular gratings are placed to the right of a lens as in Fig. 6.10. The coincidence of the centers of the second grating and of the image created by the first grating will show the alignment of two centers of the gratings and the point source. The collimation can be checked by the moiré fringes

created by the second grating and the image of the first image. The sensitivity of this method is similar to what we can get when linear gratings are used.

In beam collimation, we test the flatness of the incident wavefront. We want to place a point source in the front focal plane of a lens to create a plane wave. In certain situations, we want to place some objects, such as gratings, diffusers or spatial filters, in the back focal plane of a lens, i.e., we want to put an object where the wavefront incident to it has infinite curvature, or $Z_C = 0$. If the grating is sufficiently close to the focal point, the area of the grating which is illuminated by the incident beam is about $(B\,\Delta z/f)^2$, where f is the focal length of the length, B the size of the incident plane wave, and Δz is the distance between the focal plane and the plane in which the grating is placed. If the area $(B\,\Delta z/f)^2$ is small, the wavefield illumination of the grating no longer can be approximated by a infinitely wide wavefield. The Talbot effect has somewhat broken down. Equation (6.88) can be explained from the point of view of geometrical optics and will remain valid even when the Talbot effect has broken down.

Experimentally we found that if the finer detail of the object is smaller than the size of the Airy disc $\lambda f/B$, which is the smallest spot size we can focus our incident plane wave into, we can always place the object in the back focal plane. We can identify the focal plane by shifting the object around the focal plane. We can see a speckle pattern at a distance Z from the focal plane, and the closer

the object is placed to the focal plane, the bigger the speckles are. The change of the size of the speckles is approximately inversely proportional to ΔZ .

Footnotes to Chapter 6

1. Lord Rayleigh [70].
2. Leith and Upatnieks [14].
3. Brown and Lohmann [22].
4. Pennington, Will, and Shelton [53].
5. Grover, Mallick, and Roblin [71].
6. Silva and Hong [72].
7. Lohmann and Silva [73].
8. Lohmann and Silva [74].
9. Leith and Upatnieks [75].
10. Gerritsen, Hannan, and Ramberg [76].
11. Francon [77].
12. Arnold, Burch, and Ennos [83].
13. Debrus, Francon, Grover, and May [78].
14. Bryngdahl [79].
15. Nishida [68].
16. Bestenreiner, Deml, and Greis [80].
17. Bryngdahl [54].
18. Bryngdahl [67].
19. Biederman [81].
20. Lau and Krug [66].
21. Silva [82].
22. Cowley and Moodie [89].

List of References

1. Helstrom, C. W., Statistical Theory of Signal Detection. Second edition, Pergamon Press, Oxford, 1968.

2. Van Trees, H. L., Detection, Estimation, and Modulation Theory, Part I. J. Wiley and Sons Inc., New York, 1968.

3. Wiener, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The M. I. T. Press, Cambridge, Mass., 1949.

4. Harris, J. L. Sr., "Image Evaluation and Restoration," J. Opt. Soc. Am., Vol. 56, 569 (1966).

5. Helstrom, C. W., "Image Restoration by the Method of Least Squares," J. Opt. Soc. Am., Vol. 57, 297 (1967).

6. Slepian, D., "Linear Least-Squares Filtering of Distorted Images," J. Opt. Soc. Am., Vol. 57, 918 (1967).

7. Freiden, B. R., "Band-Unlimited Reconstruction of Optical Objects and Spectra," J. Opt. Soc. Am., Vol. 57, 1013 (1967).

8. Rushforth, C. K. and R. W. Harris, "Restoration, Resolution, and Noise," J. Opt. Soc. Am., Vol. 58, 539 (1968).

9. Falk, J. E., "Maximization of Signal-to-Noise Ratio in an Optical Filter," SIAM. J. Appl. Math., Vol. 17, 582 (1969).

10. Helstrom, C. W., "Information Transfer from Incoherently Radiating Objects," J. Opt. Soc. Am., Vol. 60, 1608 (1970).

11. Wozencraft, J. M. and I. M. Jacobs, Principle of Communication Engineering. J. Wiley and Sons, Inc., New York, 1965.

12. Helstrom, C. W., "Linear Restoration of Incoherently Radiating Objects," J. Opt. Soc. Am., Vol. 62, 416 (1972).

13. Himmelblau, D. M., Applied Nonlinear Programming. McGraw-Hill, Inc., New York, 1972.

14. Leith, E. N. and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory," J. Opt. Soc. Am., Vol. 52, 1123 (1962).

15. Elias, P., "Optics and Communication Theory," J. Opt. Soc. Am., Vol. 43, 299 (1953).

16. O'Neill, E. L., "Spatial Filtering in Optics," IRE Trans. IT-2, 56 (1956).

17. Lohmann, A. W. and H. W. Werlich, "Spatial Pulse Modulation," Appl. Opt., Vol. 10, 2743 (1971).

18. Armitage, J. D. and A. W. Lohmann, "Theta Modulation in Optics," Appl. Opt., Vol. 4, 399 (1965).

19. Nishijima, Y. and G. Oster, "Moiré Patterns: Their Applications to Refractive Index and Refractive Index Gradient Measurements," J. Opt. Soc. Am., Vol. 54, 1 (1964).

20. Ronchi, V., "Forty Years of History of a Grating Interferometer," Appl. Opt., Vol. 3, 437 (1964).

21. Collier, R. J., C. B. Burckhardt, and L. H. Lin, Optical Holography, Academic Press, New York, 1971.

22. Brown, B. R. and A. W. Lohmann, "Complex Spatial Filtering with Binary Masks," Appl. Opt., Vol. 5, 967 (1966).

23. Lohmann, A. W. and D. P. Paris, "Binary Fraunhofer Holograms, Generated by Computer," Appl. Opt., Vol. 6, 1739 (1967).

24. Gabor, D., "Theory of Communication," J. I. E. E. 93 (III), 429 (1946).

25. Reed, I. S., "On the Moment Theorem for Complex Gaussian Processes," Trans. IRE, IT-8, 194 (1962).

26. Mandel, L., "Concepts of Cross-Spectral Purity in Coherence Theory," J. Opt. Soc. Am., Vol. 51, 1342 (1961).

27. Blackman, R. B. and J. W. Tukey, The Measurement of Power Spectra from the Point of View of Communications Engineering. Dover Publications, Inc., New York, 1958.

28. Jenkins, G. M. and D. G. Watts, Spectral Analysis and Its Applications. Holden-Day, San Francisco, 1969.

29. Mandel, L. and E. Wolf, "Coherence Properties of Optical Fields," Rev. Mod. Phys., Vol. 37, 231 (1965).

30. Helstrom, C. W., "Detection and Resolution of Incoherent Objects by a Background-Limited Optical System," J. Opt. Soc. Am., Vol. 59, 164 (1969).

31. Helstrom, C. W., "Detection and Resolution of Incoherent Objects Seen through a Turbulent Medium," J. Opt. Soc. Am., Vol. 59, 331 (1969).

32. Papoulis, A., Probability, Random Variables, and Stochastic Processes. McGraw-Hill, Inc., New York, 1968.

33. Born, M. and E. Wolf, Principles of Optics. Fourth Edition, Pergamon Press, London, 1970.

34. Parzen, E., Modern Probability Theory and Its Applications. J. Wiley and Sons, Inc., New York, 1960.

35. Goodman, J. Introduction to Fourier Optics. McGraw-Hill, Inc., New York, 1968.

36. Helstrom, C. W., "The Detection and Resolution of Optical Signals," IEEE Trans. on Information Theory, Vol. 10, 275 (1964).

37. Bar-David, I., "Communication under the Poisson Regime," IEEE Trans. on Information Theory, Vol. IT-15, 31 (1969).

38. Karp, S. and J. Clark, "Photon Counting: A Problem in Classical Noisy Theory," IEEE Trans. on Information Theory, Vol. IT-16, 672 (1970).

39. Mandel, L., "Fluctuation of Photon Beams: The Distribution of the Photo Electrons," Proc. Phys. Soc., Vol. 74, 233 (1959).

40. Helstrom, C. W., "Photoelectric Detection of Coherent Light in Filtered Background Light," IEEE Trans. on Aerospace and Electronics Systems, Vol. 7, 210 (1971).

41. Helstrom, C. W., "The Distribution of Photoelectric Counts from Partially Polarized Gaussian Light," Proc. Phys. Soc., Vol. 83, 777 (1964).

42. Mandel, L., "Intensity Fluctuation of Partially Polarized Light," Proc. Phys. Soc., Vol. 81, 1104 (1963).

43. Davenport, W. B. and W. L. Root, An Introduction to the Theory of Random Signals and Noise. McGraw Hill, Inc., New York, 1958.

44. Parzen, E., Stochastic Processes, Holden-Day, San Francisco, 1962.

45. Hanbury-Brown, R. and R. Q. Twiss, "Interferometry of the Intensity Fluctuations in Light. I. Basic Theory: the Correlation between Photons in Coherent Beams of Radiation," Proc. Roy. Soc. (London), A242, 300 (1957).

46. Breiman, L., Probability, Addison-Wesley Publishing Co., Reading, Mass., 1968.

47. Gnedenko, B. V. and A. V. Kolmogorov, Limit Distribution for Sums of Independent Random Variables. Translated by K. L. Chang, revised edition, Addison-Wiley Pub. Co., Menlo Park, California, 1968.

48. Helstrom, C. W., "Model Decomposition of Aperture Field in Detection and Estimation of Incoherent Objects," J. Opt. Soc. Am., Vol. 60, 521 (1970).

49. Gallager, R., Information Theory and Reliable Communication. J. Wiley and Sons, Inc., New York, 1968.

50. Helstrom, C. W. and Y. M. Hong, "Restoration of Images on Photoelectrically Emissive Surfaces," J. Opt. Soc. Am., Vol. 63, 480 (1973).

51. Guild, J., The Interference Systems of Crossed Diffraction Gratings. Oxford University Press, New York, 1956.

52. Mueller, P. F., "Linear Multiple Image Storage," Appl. Opt., Vol. 8, 267 (1969).

53. Pennington, K. S., P. M. Will, and G. L. Shelton, "Grid Coding: A Technique for Extraction of Difference from Scenes," Opt. Comm., Vol. 4, 113 (1971).

54. Bryngdahl, O., "Image Polarity in Carrier-Frequency Photography," J. Opt. Soc. Am., Vol. 60, 1698 (1970).

55. Bryngdahl, O. and A. W. Lohmann, "Holography in White Light," J. Opt. Soc. Am., Vol. 60, 281 (1970).

56. Bryngdahl, O., "Applications of Shearing Interferometry," Progress in Optics, Vol. IV, ed. by E. Wolf, North-Holland Pub. Co., 1965, pp. 39-83.

57. Lohmann, A. W. and O. Bryngdahl, "A Lateral Wavefront Shearing Interferometer with Variable Shear," Appl. Opt., Vol. 6, 1934 (1967).

58. Eguchi, R. G. and F. P. Carlson, "Linear Vector Operation in Coherent Optical Data Processing Systems," Appl. Opt., Vol. 9, 687 (1970).

59. Lohmann, A. W. and D. P. Paris, "Computer Generated Spatial Filters for Coherent Optical Data Processing," Appl. Opt., Vol. 7, 651 (1968).

60. Yao, S. K. and S. H. Lee, "Spatial Differentiation and Integration by Coherent Optical-Correlation Method," J. Opt. Soc. Am., Vol. 61, 474 (1971).

61. Gabor, D., G. Stroke, R. Restrick, A. Funkhouser, and D. Brumm, "Optical Image Synthesis (Complex Amplitude Addition and Subtraction) by Holographic Fourier Transformation," Phys. Letters, Vol. 18, 116 (1965).

62. Lee, S. H., S. K. Yao, and A. G. Miles, "Optical Image Synthesis (Complex Amplitude Addition and Subtraction) in Real-Time by a Diffraction-Grating Interferometric Method," J. Opt. Soc. Am., Vol. 60, 1037 (1970).

63. Lowenthal, S. and J. Braat, "Subtraction of Intensities by Means of Spatially Incoherent Fourier Holography," Appl. Opt., Vol. 10, 2553 (1971).

64. Bromley, K., M. A. Monahan, J. F. Bryant, and B. J. Thompson, "Holographic Subtraction," Appl. Opt., Vol. 10, 174 (1971).

65. Debrus, S., M. Francon, and C. P. Grover, "Detection of Differences between Two Images," Opt. Comm., Vol. 4, 172 (1971).

66. Lau, E. and W. Krug, Equidensitometry, Focal Pub. Co., London, 1968.

67. Bryngdahl, O., "Image-Transfer Characteristics of Carrier-Frequency Photography," J. Opt. Soc. Am., Vol. 62, 807 (1972).

68. Nishida, N., "Reconstruction of Negative Image in Holography," Appl. Opt., Vol. 7, 1862 (1968).

69. Lowenthal, S. and Y. Belvaux, "Reconnaissance des formes par filtrage des frequences spatiales," Opt. Acta, Vol. 14, 245 (1967).

70. Lord Rayleigh, "On the Manufacture and Theory of Diffraction Gratings," Phil. Mag., pp. 81-93 and pp. 193-205 (1874).

71. Grover, C. P., S. Mallick, and M. L. Roblin, "Observation of a Phase Object Using Carrier-Frequency Photography," Opt. Comm., Vol. 3, 181 (1971).

72. Silva, D. and Y. M. Hong, "Circular Carrier-Frequency Photography for Observing Phase Objects," Opt. Comm., Vol. 7, 283 (1973).

73. Lohmann, A. W. and D. Silva, "An Interferometer Based on the Talbot Effect," Opt. Comm., Vol. 2, 413 (1971).

74. Lohmann, A. W. and D. Silva, "A Talbot Interferometer with Circular Gratings," Opt. Comm., Vol. 4, 326 (1972).

75. Leith, E. N. and J. Upatnieks, "Wavefront Reconstruction with Continuous-tone Objects," J. Opt. Soc. Am., Vol. 53, 1377 (1963).

76. Gerritsen, H., W. Hannan, and E. Ramberg, "Elimination of Speckle Noise in Holograms with Redundancy," Appl. Opt., Vol. 7, 2301 (1968).

77. Francon, M., "New Method of Optical Processing Using a Random Diffuser," Opt. Acta, Vol. 20, 1 (1973).

78. Debrus, S., M. Francon, C. P. Grover, M. May, and M. L. Roblin, "Ground Glass Differential Interferometer," Appl. Opt., Vol. 11, 853 (1972).

79. Bryngdahl, O., "Polarization-Grating Moiré," J. Opt. Soc. Am., Vol. 62, 839 (1972).

80. Bestenreiner, F., R. Deml, and U. Greis, "Negative-Positive Image Transformation by Means of Carrier-Frequency Photography," Optik, Vol. 30, 404 (1970).

81. Biedermann, K., "Tone-Reproduction and Reversal in Carrier-Frequency Photography," J. Opt. Soc. Am., Vol. 61, 1439 (1971).

82. Silva, D., "A Simple Interferometric Method of Beam Collimation," Appl. Opt., Vol. 10, 1980 (1971).

83. Arnold, E., J. M. Burch, and A. E. Ennos, "Recording of In-Plane Surface Displacement by Double-Exposure Speckle Photography," Opt. Acta, Vol. 17, 883 (1970).

84. Yao, S. K. and S. H. Lee, "Synthesis of a Spatial Filter for the Combined Operation of Subtraction and Correlation," Appl. Opt., Vol. 10, 1154 (1971).

85. Petrosky, K. J. and S. H. Lee, "New Methods for Producing Gradient Correlation Filter for Signal Detection," Appl. Opt., Vol. 10, 1958 (1971).

86. Aoki, M. Introduction to Optimization Techniques. The Macmillan Co., New York, 1971.

87. Kowalik, J. and M. R. Osborne, Methods for Unconstrained Optimization Problems. American Elsevier Pub. Co. Inc., New York, 1968.

88. Goldfard, D. and L. Lapidus, "Conjugate Gradient Method for Nonlinear Programming Problems with Linear Constraints," Ind. Eng. Chem. Fundamentals, Vol. 7, 142 (1968).

89. Cowley, J. M. and A. F. Moodie, "Fourier Images: II - The Out-of-Focus Patterns," Proc. Phys. Soc. Lond., Vol. B70, 497 (1957).

Appendix A. Imaging and Fourier Transforming Properties of Lenses

A. 1 General Input, Output Relationship

The object $U_0(x, y)$ is located at a distance Z_0 in front of a positive thin lens L_1 of focal length f. The tilted plane of illumination is created by a point source located at coordinate (x', y') of the front focal plane of lens L_2, as shown in Fig. A. 1.

For simplicity, we will use only one-dimensional notation. It is straightforward to extend this discussion to the two-dimensional case. Coordinates x_1, x_2, and x refer to the object plane, lens plane, and observing plane respectively. The illuminating tilted plane wave can be expressed as

$$U(x_1, y_1; z = -0) = \exp[2\pi i(\alpha x + \beta y)] ,$$

where

$$\alpha = -x'/\lambda f , \quad \beta = -y'/\lambda f \tag{A. 1}$$

By the successive use of the Kirchhoff boundary condition, and the Fresnel diffraction formula, the field at the observation plane is

$$U_i(x) = \iint U_0(x_1) \exp(2\pi i \alpha x_1) \exp\left[\frac{i\pi}{\lambda z_0}(x_1 - x_2)^2\right] \exp\left[-\frac{i\pi}{\lambda f} x_2^2\right]$$

$$\times \exp\left[\frac{i\pi}{\lambda z_i}(x_2 - x)^2\right] dx_1 dx_2 , \tag{A. 2}$$

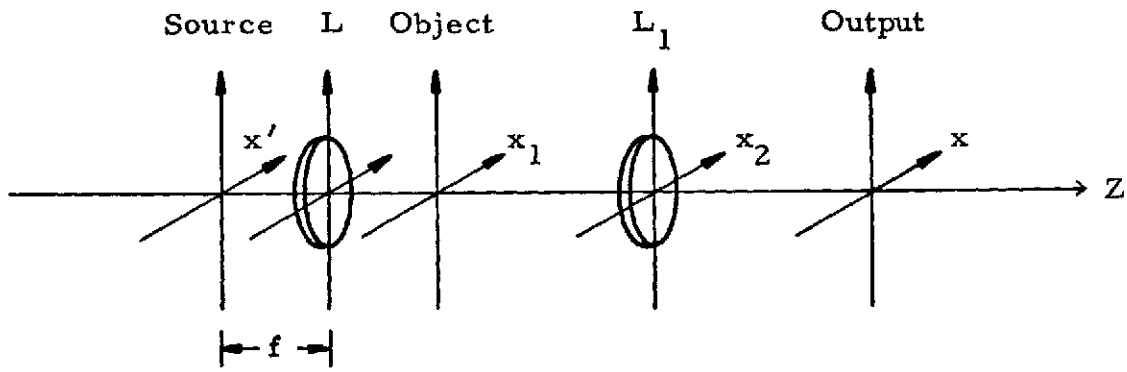


Figure A. 1. An optical system. x' = source plane, x_1 = object plane, x_2 = lens plane, x = output plane.

where a multiplying constant is omitted. The consecutive phase factors in Eq. (A.2) are due to the tilted plane wave illumination, propagation from the object plane to the lens plane, the quadratic phase factor of the positive thin lens, and propagation from the lens plane to the observation plane.

If we assume that the lens aperture is large enough, the integration limits over variable x_2 can be well approximated by $(-\infty, \infty)$. Equation (A.2) can be rewritten as

$$U_i(x) = \exp(i\pi x^2/\lambda z_i)\iint U_0(x_1) \exp\left(2\pi i \alpha x_1 + \frac{i\pi}{\lambda z_0} x_1^2\right) \times \exp\left[x_2^2 + 2x_2\left(\frac{x_1}{z_0} + \frac{x}{z_i}\right)a\right] dx_1 dx_2 \tag{A.3}$$

where

$$\frac{1}{a} = \frac{1}{z_0} + \frac{1}{z_i} - \frac{1}{f}.$$

We can perform the integration over x_2 by completing the square of the quadratic phase factor of x_2. By the Fresnel integration formula, we have

$$U_i(x) = \exp(i\pi x^2/\lambda z_i) \int U_0(x_1) \exp\left(2\pi i \alpha x_1 + \frac{i\pi x_1^2}{\lambda z_0}\right)$$

$$\times \exp\left[\frac{-i\pi a}{\lambda}\left(\frac{x_1}{z_0} + \frac{x}{z_i}\right)^2\right] dx_1$$

$$= \exp\left[\frac{i\pi}{\lambda z}\left(1 - \frac{a}{z_i} - \frac{a^2 b}{z_i z_0^2}\right) x^2\right] \exp\left[\frac{2\pi i ab}{z_0 z_i} \alpha x - i\pi b \lambda \alpha^2\right]$$

$$\times \int U_0(x_1) \exp\left[\frac{i\pi}{\lambda b} x_1 + b\left(\lambda\alpha - \frac{ax}{z_i z_0}\right)\right]^2 dx_1 \qquad \text{(A. 4)}$$

where

$$a^{-1} = z_0^{-1} + z_i^{-1} - f^{-1} \neq 0 \qquad \text{(A. 5)}$$

$$b^{-1} = z_0^{-1} - a z_0^{-2} \quad . \qquad \text{(A. 6)}$$

It can be shown, after some algebra, that

$$\frac{1}{z_i} + \frac{1}{z_0 - b} = \frac{1}{f} \qquad \text{(A. 7)}$$

i. e., if the object moves a distance b closer to the lens, then it will form an image at the observation plane. Or, we can say that the object is out of focus by a distance b.

Equation (A. 4) can be extended into two-dimensional cases easily,

$$U_i(x, y) = \exp \frac{i\pi}{\lambda z_i}\left(1 - \frac{a}{z_i} - \frac{a^2 b}{z_i z_0^2}\right)(x^2 + y^2)$$

$$\exp\left(\frac{2\pi i ab}{z_0 z_i}\right)(\alpha x + \beta y) - i\pi b\lambda(\alpha^2 + \beta^2)$$

$$\int U_0(x_1, y_1)\left\{\exp \frac{i\pi}{\lambda b}\left\{\left[x_1 + b\left(\lambda\alpha - \frac{ax}{z_i z_0}\right)\right]^2 + \left[y_1 + b\left(\lambda\beta - \frac{ax}{z_i z_0}\right)\right]^2\right\}\right\} dx_1 dy_1$$

(A. 8)

If we place our object at $z_0 = d_0 - D$, and the observation plane is the image plane of the plane $z_0 - d_0$, i.e., $z_i = d_i$, and $d_0^{-1} + d_i^{-1} = f^{-1}$, then $a = d_0(d_0 - D)/D$, and $b = -D$. The field at $z_i = d_i$ is

$$\int U_0(x_1, y_1) \exp\left\{-\frac{i\pi}{\lambda D}\left[\left(x_1 + \frac{d_0}{d_i} x - \lambda\alpha D\right)^2 + \left(y_1 + \frac{d_0}{d_i} y - \lambda\alpha D\right)^2\right]\right\} dx_1 dy_1$$

(A. 9)

i.e., if the object field $U_0(x_1, y_1)$ propagates a distance $-D$ from where it is located, then it will form an image at $z_i = d_i$.

There are two interesting special cases: (1) $a^{-1} = 0$, the observation plane is the image plane, and (2) $z_i = f$, the observation plane is the back focal plane of the lens.

A.2 Imaging Property of Lenses

If the observation happens to be at the image plane, then $a^{-1} = 0$. We have no quadratic phase factor in the integration over variable x_2 in

Eq. (A.1.3). Integration over x_2 yields a delta function $\delta\left(\frac{x_1}{d_0} + \frac{x}{d_i}\right)$. Subsequently an integration over x_1 yields, in two-dimensional notation

$$U_i(x, y) = \exp\left[\frac{i\pi}{\lambda d_i}(1 + M^{-1})(x^2 + y^2)\right]\exp[-2\pi i(\alpha x + \beta y)/M]U_0\left(-\frac{x}{M} - \frac{y}{M}\right), \tag{A.11}$$

where

$$M = d_i/d_0 \quad .$$

The field at the image plane has been reversed and magnified M times. The introduction of tilted plane wave illumination yields a phase factor $\exp[-2\pi i(\alpha x + \beta y)/M]$ in the image field. When the image is detected by a square-law detector, such as film, the phase factor is not important. The angle of the illuminating plane wave has no effect on the final image. Consequently, point-source illumination is not necessary, and spatially incoherent illumination can be used.

If the distance D in Eq. (A.9) is so small that λD is much smaller than the typical finer details of object $U_0(x, y)$, then we can apply the method of stationary phase to perform the integration over x_1 and y_1 in Eq. (A.9), and

$$V(x, y) \cong \exp\left[\frac{i\pi}{\lambda d_i}(1 + M^{-1}(x^2 + y^2)\right]\exp\left[-2\pi i(\alpha x + \beta y)/M + i\pi\lambda D(\alpha^2 + \beta^2)\right]$$

$$\times\ U_0\left(-\frac{x}{M} + \lambda\alpha D\ ,\ -\frac{y}{M} + \lambda\beta D\right) \quad . \tag{A.12}$$

The image has been shifted an amount $\lambda\alpha DM$ in the x-direction and an amount $\lambda\beta DM$ in the y-direction, because of tilted plane wave illumination.

A.3 Fourier Transforming Property of Lenses

The other case of special interest is when $z_i = f$ - the observation plane is the back focal plane of the lens.

When $z_i = f$, then $a = z_0$, and Eq. (A.1.8) can be rewritten as

$$U_i(x) = \exp\left[\frac{i\pi}{\lambda f}\left(1 - \frac{z_0}{f}\right)(x^2+y^2)\right]\tilde{U}_0\left(\frac{x}{\lambda f} - \alpha, \frac{y}{\lambda f} - \beta\right), \qquad (A.13)$$

where

$$\tilde{U}_0(x, y) = \int U_0(x_1, y_1) \exp[2\pi i(x_1 x + y_1 y)]\, dx_1 dy_1 \ . \qquad (A.14)$$

The field at point (x, y) of the back focal plane of the lens is related to the Fourier spectrum of the object field at frequency $(x/\lambda f - \alpha, y/\lambda f - \beta)$. There is a quadratic phase factor, which will vanish if $z_0 = f$, i.e., if the object is located at the front focal plane. The tilted plane illumination introduces a shift of amount $(\alpha\lambda f, \beta\lambda f)$ in the observation plane. Spatially coherent illumination is needed for spatial filtering.

Appendix B. Spatial and Temporal Coherence Requirements of Talbot Effect

B.1 Talbot Effect

A monochromatic point source located at point (x', y') of the front focal plane of a lens is used to illuminate the grating located to the right of the lens as shown in Fig. B.1. The grating is illuminated by a tilted plane wave $\exp[-2\pi i (x'x + y'y)/\lambda f]$ which is created by the point source. The incident tilted plane wave will be diffracted by the grating into many plane waves, each of which propagates with different phase angle and phase attenuation. By the use of the Fresnel diffraction formula, the field at distance z beyong the grating is

$$U(x, y; x', y'; z; \lambda) = \sum A_n \exp[-i\pi z(x'^2 + y'^2)/\lambda f^2$$

$$\exp[-2\pi i(xx' + yy')/\lambda f] \exp[-i\pi \lambda z n^2/d^2] \exp(2\pi i n(x/d + zx'/fd)] \qquad (B.1)$$

where $G(x) = \sum_n A_n \exp(2\pi i nx/d)$ is the grating transmittance function, λ the wavelength of the monochromatic source. The first two phase factors are from the tilted incident plane wave, the third phase factor is the phase attenuation of the n-th diffraction order, and the fourth phase factor is the n-th diffraction order with a phase shift zx'/f due to tilted plane wave illumination.

For simplicity, we assume that the grating is a pure sinusoidal grating, i.e., $A_0 = 1$, $A_1 = A_{-1} = \frac{1}{2}$. We are only interested in the

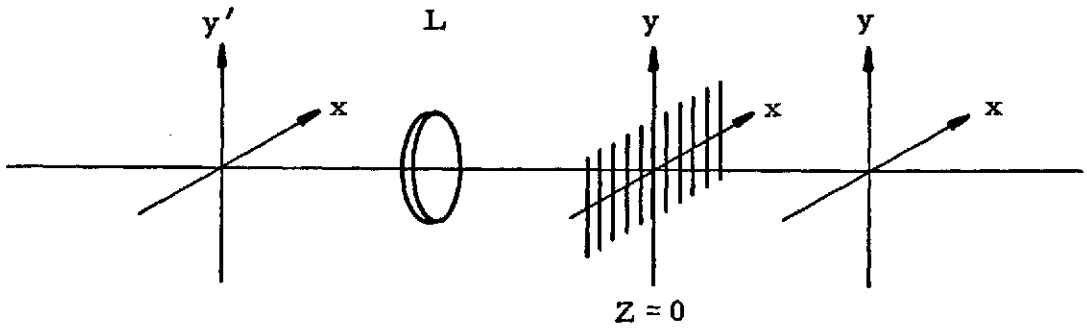


Figure B. 1. A Talbot image forming system.

first harmonics of the image at distance z beyond the grating. The image is

$$I(x, y; x', y'; \lambda ; z) = \frac{3}{2} + 2 \cos(\pi \lambda z \nu^2) \cos[2\pi \nu(x + zx'/f)]$$

$$+ \frac{1}{2} \cos [4\pi \nu (x + zx'/f)] \quad , \tag{B.2}$$

where $\nu = 1/d$. The first harmonic is

$$I_1(x, y; x', y'; \lambda ; z) = 2 \cos (\pi \lambda z \nu^2) \cos[2\pi\nu (x + zx'/f)] . \tag{B.3}$$

The constant factor $\cos(\pi \lambda z \nu^2)$ is the reduction factor of the contrast. Three cases are of interest:

$$Z = \begin{cases} 2 M\nu^2/\lambda & = Z_T \\ (2M+1)\nu^2/\lambda & = Z_C \\ (2M+2)\nu^2/\lambda & = Z_0 \end{cases} \tag{B.4}$$

They correspond to $\cos(Z_T) = 1, -1$, and 0. The image at $Z = Z_T$ is a replica of the grating itself, the image at $Z = Z_C$ has contrast reversal, and the one at $Z = Z_0$ has zero contrast. The image has been shifted by Zx'/f. This interesting feature of sinusoidal gratings holds also for Ronchi rulings.

B.2 Spatial Coherence Requirement - Source Size

If the source is not a single point, then each point of the source will produce a fringe. The fringes produced by each point source are shifted relative to each other, as can be seen from Eq. (B.3). The

contrast of the resulting fringes will be reduced if the source is too big. For a source of size B, the image in the first harmonic is

$$I(x, y; \lambda ; z) = \int_{-B/2}^{B/2} I_1(x, y; x', y'; \lambda ; z)\, dx'$$

$$= 2B\ \mathrm{sinc}(B\nu Z/f) \cos(\pi\lambda Z\nu^2) \cos(2\pi\nu x) . \quad (B.5)$$

The constant $\mathrm{sinc}(B\nu Z/f)$ is the reduction of the fringe contrast due to the broadness of the source. For good fringe contrast, $\mathrm{sinc}(B\nu Z/f) \geq 2/\pi$, the size of the source B should be smaller than $f/2\nu Z$,

$$B \leq f/2\nu Z . \quad (B.6)$$

B.3 Temporal Coherence Requirement - Monochromaticity

For different wavelengths the Talbot distance varies according to Eq. (B.4). If the point source is not monochromatic, then color dispersion will appear. For good contrast, the spectral width of the polychromatic light must be limited.

For a polychromatic source of spectral width $\Delta\lambda$, and mean wavelength λ, the resulting image in the first harmonic is

$$I_1(x, y; x'=0,\ y'=0; z) = 2 \cos(\pi\bar{\lambda}\nu^2 Z)\ \mathrm{sinc}\left(\frac{\Delta\lambda\,\nu^2 Z}{2}\right) \cos(2\pi\nu x) . \quad (B.7)$$

Let $Z = 2\,Md^2/\bar{\lambda}$, the Talbot distance for wavelength $\bar{\lambda}$, then

$$I_1 = 2\,\mathrm{sinc}\,(M\Delta\lambda/\bar{\lambda})\cos(2\pi\nu x). \tag{B.8}$$

For good contrast, $M\Delta\lambda/\bar{\lambda} \leq 1/6$. (B.9)

Appendix C. Talbot Effect of Circular Gratings

A circular grating $C(r) = \sum_n C_n \exp(2\pi i\, nr/a)$ is illuminated by a normal incident plane wave. Using the Fresnel diffraction theory, we find the field at a distance Z behind the circular grating is

$$U(r,\varphi\,;\,Z) = \sum_n C_n \int_0^{2\pi} \int_0^{\infty} \exp(2\pi i n r_1/a)$$

$$\exp\left\{\frac{i\pi}{\lambda Z}\left[r^2 + r_1^2 - 2rr_1\cos(\varphi - \varphi_1)\right]\right\} r_1 dr_1 d\varphi$$

$$= \sum_n C_n \int_0^{\infty} r_1 J_0\left(\frac{2\pi r r_1}{\lambda Z}\right) \exp\left\{\frac{i\pi}{\lambda Z}\left[r^2 + r_1^2 + 2\,\frac{\lambda Z}{a}\, n r_1\right]\right\} dr_1 \,. \tag{C.1}$$

The exact evaluation of Eq. (C.1) is usually impossible, but reasonable approximations can be obtained. The Bessel function can be asymptotically approximated as

$$J_n(x) \cong \begin{cases} \dfrac{x^n}{2^n n!}\left[1 - \dfrac{x^2}{4(n+1)}\right] & \text{for } x \ll 1 \qquad (C.2) \\[2ex] \left(\dfrac{2}{\pi x}\right)^{1/2} \cos\left(x - \dfrac{n\pi}{2} - \dfrac{\pi}{4}\right), & \text{for } x \gg 1 \,. \qquad (C.3) \end{cases}$$

For r in the region such that $\dfrac{2\pi r r_1}{\lambda Z} \ll 1$, i.e., $r \ll \dfrac{\lambda Z}{2\pi r_1}$, we will use Eq. (C.2) in Eq. (C.1), and

$$U(r,\varphi;z) \cong \sum C_n \int r_1 \left[1 - \frac{1}{4}\left(\frac{2\pi r r_1}{\lambda z}\right)^2\right] \exp\left[\frac{i\pi}{\lambda z}\left(r_1^2 + r^2 + \frac{2\lambda Z}{a} n r_1\right)\right] dr_1 \,. \tag{C.4}$$

The point of stationary phase is located at $r_1 = -m\lambda Z/a$ (note that m can be a negative integer). For $r_1 = -m\lambda Z/a$, $r \ll \frac{\lambda Z}{2\pi r_1}$ $= \frac{a}{2\pi |m|} < \frac{a}{6}$, i.e., the integration in Eq. (C.4) only contributed to $r < a/6$. The area $r < a/6$ is so small, that it can be neglected.

Equation (C.1) can be approximated as

$$U(r,\varphi;z) = \sum_n C_n \int_0^\infty \left(\frac{r_1}{r}\right)^{1/2} \exp\left\{\frac{i\pi}{\lambda Z}\left[r^2 + r_1^2 + \frac{2\lambda Z}{a} n r_1\right]\right\} \cos\left(\frac{2\pi r r_1}{\lambda Z} - \frac{\pi}{4}\right) dr_1 \,. \tag{C.5}$$

We can expand the cosine term as the sum of two exponential functions. There are two points of stationary phase, one for each exponential term of the cosine function, one is $r_+ = r - n\lambda Z/a$, the other is $r_- = -r - n\lambda Z/a$. Since $r_1 > 0$, not every diffraction order (n) falls into the region of integration. We can divide r into a set of zones Z(m), given by $Z(m) = [m\lambda Z/a, (m+1)\lambda Z/a]$, $m \geq 0$ integer. For a given zone Z(m), if $m \geq n$ then only r_+ is inside the region of integration, if $1 \leq (m+1) < -n$ then only r_- is inside the region of integration. Using the formula for the method of stationary phase, Eq. (C.5) can be written as

$$U(r \epsilon Z(m), \varphi\,;\, Z) = \sum_{n \le m} C_n \exp[2\pi i\, nr/a - i\pi\lambda\, Zn^2/a^2] \left(1 - \frac{n\lambda Z}{ar}\right)^{1/2}$$

$$+ \sum_{n \le -(m+1)} C_n \exp\left(i\frac{\pi}{2} - 2\pi i\, nr/a - i\pi\lambda\, Zn^2/a^2\right)\left(-1 - \frac{n\lambda Z}{ar}\right)^{1/2} . \qquad (C.6)$$

Changing the index n to $-n$ in the second term of Eq. (C.6), we have

$$U(r, \varphi\,;\, Z) = \sum C_n \left(1 - \frac{n\lambda Z}{ar}\right)^{1/2} \exp(-i\pi\lambda\, Zn^2/a^2) \exp(2\pi i\, nr/a) . \qquad (C.7)$$

For $r \gg \lambda Z/a$, then

$$U(r, \varphi\,;\, Z) = \sum C_n \exp(-i\pi\lambda Zn^2/a^2) \exp(2\pi i\, nr/a) . \qquad (C.8)$$

The result is similar to the Talbot image property of linear gratings as discussed in Appendix B. For $Z = 2Ma^2/\lambda$, the field is the self image of the circular grating. From Eq. (C.8) we can conclude that each spherical wave propagates independently from the other spherical waves and with relative retarded phase $\exp(i\pi\lambda Zn^2/a^2)$ for order n, as in the case of plane waves.