OPTICAL INFORMATION PROCESSING AND HOLOGRAPHY

W. THOMAS CATHEY

Optical Information Processing and Holography

W. THOMAS CATHEY
University of Colorado
Denver, Colorado

John Wiley & Sons, New York / London / Sydney / Toronto

Library of Congress Cataloging in Publication Data:

Cathey, W. Thomas, 1937–
Optical information processing and holography.

(Wiley series in pure and applied optics)
"A Wiley-Interscience publication."
Includes bibliographical references.
1. Optical data processing. 2. Optical pattern recognition. 3. Holography. I. Title.

TA1630.C37 621.36 73-14604
ISBN 0-471-14078-3

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

OPTICAL INFORMATION
PROCESSING AND
HOLOGRAPHY

WILEY SERIES IN PURE AND APPLIED OPTICS

Advisory Editor

Stanley S. Ballard, University of Florida

BLOOM • *The Gas Lasers*
CATHEY • *Optical Information Processing and Holography*
CAULFIELD AND LU • *The Applications of Holography*
FRANCON AND MALLICK • *Polarization Interferometers*
GREEN • *The Middle Ultraviolet; Its Science and Technology*
HEARD • *Laser Parameter Measurements Handbook*
HUDSON • *Infrared System Engineering*
LENGYEL • *Introduction to Laser Physics*
LENGYEL • *Lasers*, Second Edition
LEVI • *Applied Optics, A Guide to Optical System Design*, Volume I
LOUISELL • *Quantum Statistical Properties of Radiation*
MOLLER AND ROTHSCHILD • *Far-Infrared Spectroscopy*
PRATT • *Laser Communication Systems*
ROSS • *Laser Receivers*
SHULMAN • *Optical Data Processing*
STEELE • *Optical Lasers in Electronics*
WILLIAMS AND BECKLUND • *Optics*
ZERNIKE AND MIDWINTER • *Applied Nonlinear Optics*

Preface

Holography, pattern recognition, and optical information processing are fields of increasing importance to engineers and scientists. Many are learning through self education, and this book should prove useful to them. The book evolved, however, from notes used in a one-semester class on optical information processing and holography for first-year graduate students in engineering and physics and for well-prepared undergraduates. The course was taught at the Boulder campus of the University of Colorado to primarily younger, full-time students and at the Denver campus to older, part-time students who may have had experience in the field. The necessity of providing a text for such a variety of students has resulted in a book that is larger than anticipated, but contains both basic discussions and detailed descriptions of applications.

The book deals mainly with information conveyed by spatial rather than temporal modulation. The discussion of spatial information processing draws from two areas—information or communication theory and electromagnetic theory. Many of the concepts are borrowed from communication theory, and the existence of information on a wavefront of necessity introduces the problems of propagation and diffraction.

The discussion of spatial information carried on a wavefront and of the means of recording and recovering the information is the basis of holography and spatial filtering. In Chapter 1 scalar diffraction theory is applied to moving fields, emphasizing the propagation of an angular spectrum of plane waves. Chapter 2 introduces the mathematical tools to be used in the processing of spatial information—two-dimensional transforms, convolution, and sampling. Chapter 3 goes into the details of wavefront recording, reconstruction, and modulation. The concept of coherence is introduced in Chapter 4. The degree of coherence required in holography and information processing is discussed, and the coherence properties of many radiating sources are described. In Chapter 5 the imaging and Fourier-transforming properties of a lens are derived from the diffraction formulas.

The general properties of some detectors, recorders, and spatial modulators are presented in Chapter 6. The detection mechanisms and the effects of nonlinearities and finite resolution of recording materials are discussed as they apply to wavefront recording, reconstruction, and modulation.

In communication theory, several useful concepts and techniques have been developed in conjunction with the detection and processing of temporal information. Almost all of these concepts apply with little modification to the detection and processing of spatial information. Detection theory is applied to detection of spatial signals (pattern recognition); spectral analysis and filtering theory are used to analyze and modify the spatial frequency distribution of a spatial signal; and, perhaps the most significant of all, transfer functions are found for the elements operating on the spatial information, enabling us to use the "black box" approach to describe many of the operations.

Chapter 7 discusses spatial filtering and pattern recognition. The first general area covered is that of straightforward spatial filtering by altering the amplitude and phase of the spatial spectrum with separate transparencies. The more powerful technique of complex filtering, made practical by the use of wavefront recording to form a filter, is shown to be useful in pattern recognition. Many of the problems encountered in pattern recognition are discussed in detail.

An analysis of imaging systems appears in Chapter 8. The concept of an angular spectrum of plane waves is used to determine the transfer function of an imaging system in terms of spatial frequencies, and the linear system approach to imaging is discussed.

Chapter 9 deals with holography in general and with types of holograms. The concepts of wavefront recording and reconstruction are examined in Chapter 3, but a more complete description of the holographic process is given in Chapter 9. The topics of magnification and the associated aberrations are covered and some techniques for making holograms with incoherent illumination are presented. An important section of this chapter deals with the effect of the recording medium on the quality of a reconstructed image. Other hologram-generating techniques, such as the computer, are also treated in Chapter 9. Computer generation of spatial modulators is also important in spatial filtering, but because much of the work in this field has been in relation to holography, the subject is placed in Chapter 9.

Applications of holography are dealt with in Chapter 10. This includes holographic interferometry and pulsed laser holography. The application of the concept of holography to wavelengths other than optical ones is discussed and the advantage of using the concept of holography in other fields is pointed out. For example, reconstruction of an image through a phase-distorting medium is an interesting way to describe the operation of adaptive antenna arrays. The vast amount of information that can be stored on a hologram has led to the use of holograms as memory elements in such applications as computer memory and data storage. The techniques used in these applications are discussed briefly. One of the most tantalizing applications is the production of three-dimensional television. Some of the difficulties and some of the proposed solutions are discussed in the last section.

The topics are presented in a logical sequence. No effort was made to preserve the historical order of the developments, but extensive references are given to assist the student in finding background material. The list is not all inclusive, however, and omission of any reference is not intended to reflect on the value of that work. The problem sets are intended to further illustrate the concepts presented in the respective chapters. Hopefully, some of the problems will also encourage reading of the original literature.

One comment concerning notation is in order. The form $e^{-i(\omega t-\gamma z)}$, where γ is the propagation constant, is used to represent a wave traveling in the positive z direction. This notation is generally encountered in diffraction theory and in books on optics. The time variable will usually be dropped, leaving $e^{i\gamma z}$ to represent the wave traveling in a positive direction rather than $e^{-j\gamma z}$ as seen in some engineering texts. To reconcile the formulas, i is to be replaced by $-j$.

I acknowledge the helpful suggestions of my students, the extensive comments of Helmut Lotsch, who carefully read the entire manuscript, and the patience of Judy Price who typed most of the manuscript.

W. THOMAS CATHEY

Denver, Colorado
May 1973

Contents

1

Wavefront Propagation and Scalar Diffraction

In introducing scalar diffraction theory, the phenomenogical arguments of Huygens are given to provide an intuitive understanding of the concepts involved. However, the mathematical relations describing scalar diffraction are derived by the use of an ensemble of plane waves which can be considered as being *equivalent* to the actual wave distribution. The resulting equation is examined for possible simplifications through approximations, leading to a definition of the Fresnel and Fraunhofer diffraction regions. The final section introduces the concept of a transfer function description of diffraction.

1.1 Introduction to Scalar Diffraction

To exactly describe the diffraction of a wave by an aperture, the polarization of the wave must be known. That is, the vector representation of the wave must be used [1-1]. However, the vectorial representation of the wave is neglected in this chapter because (1) the mathematical development is more complicated and (2) the scalar theory gives acceptable results for the cases that we consider. In general, the scalar theory is acceptable if the detail of the diffracting structure and the distance between the aperture and the plane in which the field is to be determined are large with respect to a wavelength. Two examples in which these approximations do not hold are microwave diffraction by small apertures and optical diffraction by a closely spaced grating [1-2].

An expression relating the wave in one plane to the wave in another can be derived using Green's theorem. A summary of this procedure is given by Goodman [1-3]. We shall use a different approach for arriving at the diffraction formula which involves following, from one plane to the next, the angular spectral components of the wavefront. That is, the distribution in one plane is written in terms of an ensemble of plane waves which, when added, in both amplitude and phase, are equivalent to the distribution. The plane waves are allowed to propagate to another reference plane and are added to find the new distribution in the second reference plane. This approach is similar to that of Ratcliffe [1-4]. Before attacking the diffraction problem, let us consider the description of a propagating plane wave.

Propagation of plane waves. A uniform plane wave traveling in an arbitrary direction can be represented by

$$|\mathbf{U}| Re \exp[-i(\omega t - \alpha x - \beta y - \gamma z + \epsilon)] \tag{1-1}$$

where $|\mathbf{U}|$ represents the magnitude of the wave, ϵ denotes the phase, and *Re* denotes real part. We will normally allow a complex $\mathbf{U}$ to include both amplitude and phase. The power density of the wave is given by $\mathbf{UU}^*$. The variable U therefore represents neither the electric nor the magnetic field but could be associated with either, depending on the definition of U. That is, $\mathbf{U} = \mathbf{E}/\sqrt{\eta}$ or $\mathbf{U} = \sqrt{\eta}\,\mathbf{H}$ where η is the intrinsic impedance of the medium, $\mathbf{E}$ represents the electric field, and $\mathbf{H}$ represents the magnetic field. In our scalar work, a choice need not be made. The intensity of a wave is a measure of the energy per unit time per unit area *normal* to the direction of propagation. The variables α, β, and γ determine the direction in which the wave propagates, and are restricted by

$$\alpha^2 + \beta^2 + \gamma^2 = k^2 = \left(\frac{2\pi}{\lambda}\right)^2 \tag{1-2}$$

where λ is the wavelength. The values of α, β, and γ are related to the direction of propagation: that is

$$\alpha = k\cos\theta, \tag{1-3}$$

$$\beta = k\cos\xi, \tag{1-4}$$

$$\gamma = k\cos\chi \tag{1-5}$$

where θ, ξ, and χ, are respectively, the angles between the x, y, and z axes and the direction of propagation (normal to the wave). Note that the dimensions of α, β, and γ must be radians per unit length.

In writing a description of a propagating wave, we shall normally omit the temporal variable and the propagation constant for the z direction. The temporal variation is understood and the value of γ can be found from (1-2). The expression (1-1) can then be written as

$$U\exp\{i2\pi[ux + vy]\} \tag{1-6}$$

where u and v are *spatial* frequencies having the dimensions of cycles per unit length:

$$u=\frac{\alpha}{2\pi}=\frac{\cos\theta}{\lambda}, \qquad v=\frac{\beta}{2\pi}=\frac{\cos\xi}{\lambda}. \tag{1-7}$$

Figure 1-1 illustrates the parameters discussed for a plane wave propagating in the x and z directions. The solid lines are maxima and the dotted lines are minima of the tilted wave. Along the x axis the distance between maxima is given by $\lambda/\cos\theta$. That is, the spatial period in the x direction is $\lambda/\cos\theta$, and the spatial frequency is given by (1-7). A wave propagating partially in the negative x direction, such that the spatial period is again $\lambda/\cos\theta$, has the same magnitude spatial frequency, but the phase regresses $(\cos\theta)\lambda$ cycles per unit length rather than advancing by the same amount. This could be considered as a *negative* progression of spatial phase. Notice, however, that the real part of (1-6) is the same in either case. When two waves represented by $U\exp(i2\pi ux)$ and $U\exp(-i2\pi ux)$ are combined, the resulting amplitude distribution along the x axis is $2U\cos(2\pi ux)$, giving a fringe distribution having a spatial frequency of u cycles per unit length. Consequently, we see an intensity pattern described by $4U^2\cos^2(2\pi ux)$.

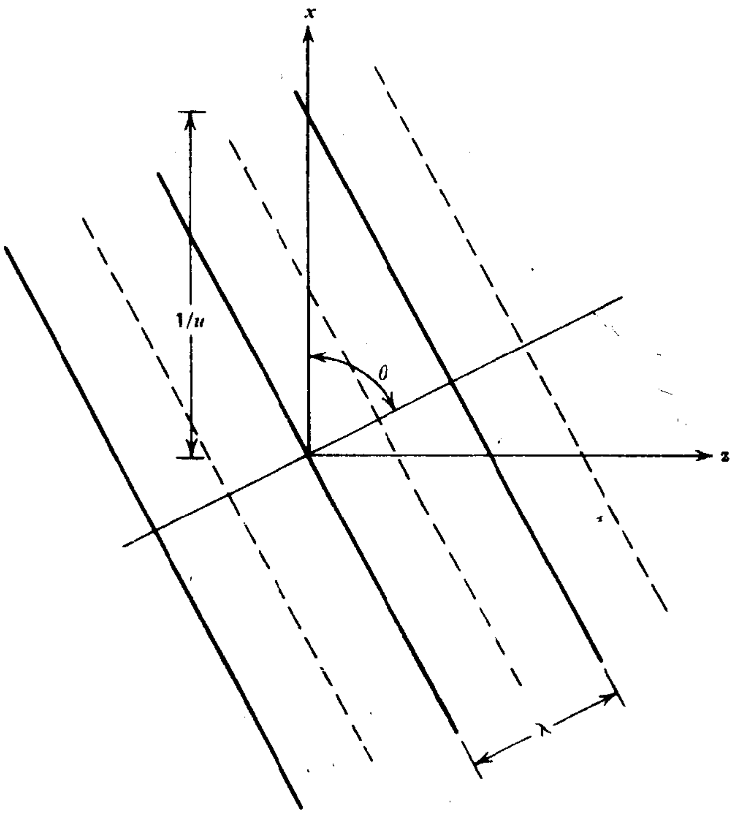

Figure 1-1 A plane wave propagating in the positive x and z directions.

Angular spectrum of plane waves. Let us temporarily leave the discussion of the propagation of a plane wave and consider the Fourier transform representation*

$$\mathbf{A}_1\left(\frac{\cos\theta}{\lambda},\frac{\cos\xi}{\lambda}\right)=\int\int \mathrm{U}_1(x_1,y_1)\exp\left[-i2\pi\left(\frac{x_1\cos\theta}{\lambda}+\frac{y_1\cos\xi}{\lambda}\right)\right]dx_1\,dy_1 \tag{1-8}$$

and the inverse transform

$$\mathrm{U}_1(x,y)=\int\int \mathbf{A}_1\left(\frac{\cos\theta}{\lambda},\frac{\cos\xi}{\lambda}\right)\exp\left[i2\pi\left(\frac{x_1\cos\theta}{\lambda}+\frac{y_1\cos\xi}{\lambda}\right)\right] \times d\left(\frac{\cos\theta}{\lambda}\right)d\left(\frac{\cos\xi}{\lambda}\right) \tag{1-9}$$

where $\mathrm{U}_1(x_1,y_1)$ represents the wave distribution in the x_1-y_1 plane and $\mathbf{A}_1[(\cos\theta)/\lambda,(\cos\xi)/\lambda]$ is the Fourier transform of U_1. We have seen that

$$\exp\left[i2\pi\left(\frac{x_1\cos\theta}{\lambda}+\frac{y_1\cos\xi}{\lambda}\right)\right]=\exp[i2\pi(ux_1+vy_1)] \tag{1-10}$$

represents a plane wave propagating partially in the x and y directions (the $i\omega t$ and $i\gamma z$ terms are dropped). Consequently, (1-9) says that the distribution U_1 can be considered as being made up of plane waves propagating in directions determined by $\cos\theta$ and $\cos\xi$ and having amplitudes and phases as described by $\mathbf{A}_1$. We can therefore say that $\mathbf{A}_1$ describes an angular spectrum of plane waves, just as the Fourier transform of a temporal distribution yields the frequency spectrum of the distribution. The integral of (1-9) extends from minus infinity to positive infinity. The only plane waves that are of interest to us are those having arguments of $\mathbf{A}_1$ from $-1/\lambda$ to $1/\lambda$. Values of the argument outside these limits require that $|\cos|>1$; that is, the angle is imaginary. Waves having such arguments are called *evanescent* or *inhomogeneous* waves. Their directions of propagation are along the positive or negative x axis and decay

*We could have used the Fourier transform pair $\mathbf{A}_1(\cos\theta,\cos\xi)$ and $\mathrm{U}_1(x_1/\lambda,y_1/\lambda)$. That is, the coordinates of the distribution U_1 could have been normalized with respect to the wavelength.

exponentially in the z direction. Investigation of (1-8) shows that these evanescent waves occur only when U has spatial frequency components with periods of less than a wavelength. Even when they occur, they decay rapidly with z so that they contribute to the field only very near the diffracting structure. In most of our work, we shall neglect the effects of evanescent waves.

The propagation of a homogenous plane wave is easy to describe—it remains an infinite, uniform plane wave if the medium is homogenous and isotropic. Only the phase changes. Consequently, we can describe a distribution U_1 in plane x_1-y_1 in terms of its angular spectrum of plane waves A_1, and allow the plane wave components to propagate to plane x_2-y_2 giving us A_2 from which $U_2(x_2,y_2)$ can be found. Figure 1-2 shows the propagation of one of the plane wave components. For simplicity, only the x and z axes are shown. As can be seen from the figure, the tilted waves (not propagating just in the z direction) travel a shorter distance than the nontilted wave. The nontilted wave travels a distance z giving a phase shift of $\exp(ikz)$. In general, the plane wave components travel a distance q giving the phase shift

$$\exp(ikq)=\exp(ikz\cos\chi). \tag{1-11}$$

Because

$$\cos^2\chi+\cos^2\theta+\cos^2\xi=1, \tag{1-12}$$

$$\exp(ikq)=\exp\left(ikz\sqrt{1-\cos^2\theta-\cos^2\xi}\right). \tag{1-13}$$

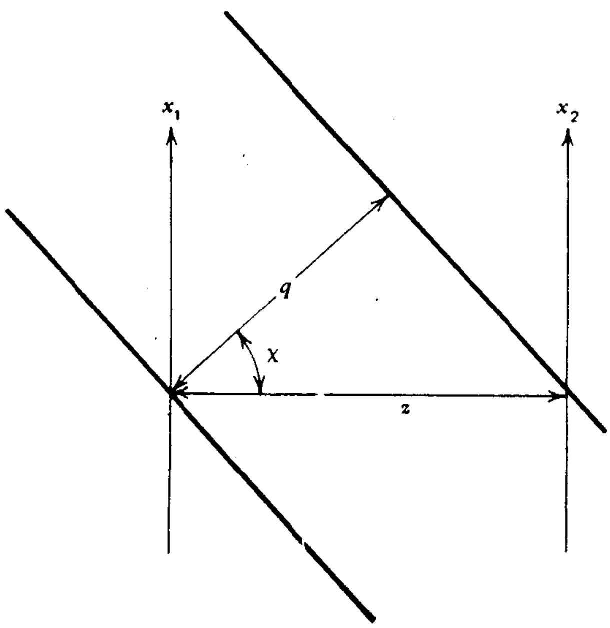

Figure 1-2 Propagation of one plane wave component.