

INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company

73-6847

HSIAO, Stephen Shen-Hsi, 1940-
ANALYSIS ON INFORMATION REDUCTION AND
BEAMWIDTH REDUCTION IN HOLOGRAPHY.

The University of Michigan, Ph.D., 1972
Engineering, electrical

University Microfilms, A XEROX Company, Ann Arbor, Michigan

ANALYSIS ON INFORMATION REDUCTION AND
BEAMWIDTH REDUCTION IN HOLOGRAPHY

by

Stephen Shen-Hsi Hsiao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
1972

Doctoral Committee:

Professor Emmett N. Leith, Chairman
Professor William B. Ribbens
Professor William L. Root
Professor Thomas B. A. Senior
Mr. Juris Upatnieks
Dr. Jerry S. Zelenka

PLEASE NOTE:

Some pages may have
indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company

ACKNOWLEDGMENTS

I am indebted to Professor E. N. Leith, Professor W. B. Ribbens, Professor W. L. Root, Professor T. B. A. Senior, Mr. J. Upatnieks, and Dr. J. S. Zelenka for serving on my doctoral committee. In particular, I wish to thank Professor Leith for serving as the chairman. I am also deeply indebted to Professor Senior and Mr. Upatnieks for their advice, assistance, and encouragement throughout the research.

I wish to express my deepest gratitude to Mr. Robert Lewis who has been a constant source for advice, assistance, encouragement, and many interesting discussions. I wish to thank Mr. Carl Leonard and Mr. Ivan Cindrlich for their constructive comments.

Special thanks goes to my wife, Hsi-yuan, for her understanding and encouragement and for assisting in the typing of my thesis. I am also grateful to Dorothy MacDonald for her editorial services and Gerry Nowak who also assisted in the typing of my thesis.

This work was sponsored in part by the Naval Training Device Center, Orlando, Florida under Contract N61339-69-C-0299.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF ILLUSTRATIONS	vi
LIST OF TABLES	x
CHAPTER 1. INTRODUCTION: THE PROBLEM OF INFORMATION REDUCTION AND BEAMWIDTH REDUCTION IN HOLOGRAPHY ..	1
CHAPTER 2. THE SPACE BANDWIDTH PLOTS AND THE INFORMATION CAPACITY OF A HOLOGRAM	5
2.1. The Concept of a Space Bandwidth (SB) Plot of a Hologram	5
2.2. The Use of SB Plot in a Holographic System	9
2.2.1. The SB Plots for Various Holographic Recording Configurations	9
2.2.2. The Dual Role of the SB Plot of Two Apertures	14
2.3. The Distribution of the Image Resolution and Information Capacity of a Hologram	19
2.3.1. Introduction	19
2.3.2. The Distribution of the Image Resolution and Information Capacity of a Hologram for Which an On-Axis Reference Point Source is Used	20
2.3.2. The Generalization of the System Model and the Use of the Offset Point-Source Reference Beam	35
2.3.4. The Distribution of the Image Resolution and Information Capacity of a Hologram Using a Normal Plane-Wave Reference	48
2.3.5. The Generalization of the System Model and the Use of the Offset Plane-Wave Reference Beam	54
CHAPTER 3. THE INFORMATION REDUCTION METHODS OF A THREE-DIMENSIONAL HOLOGRAPHIC VIEWING SYSTEM	60
3.1. Introduction	60
3.2. The Fundamental Requirements of a Three-Dimensional Viewing System and the Use of a Hologram in Such a System	61
3.3. The Optimum Recording Configuration of a Holographic Viewing System	66
3.4. The Sampling Method for Information Reduction in Holography	68

	Page
3.5. The Scatter-Plate Method for Information Reduction in Holography	73
3.5.1. The Optimization of the Scatter-Plate Method	76
3.5.2. The Periodic Scatter-Plate Approach	83
3.6. The Fly's-Eye Lens Array	91
CHAPTER 4. THE STUDY OF THE FLY'S-EYE LENS ARRAY SYSTEM ..	98
4.1. The Resolution of the Fly's-Eye Lens Array System	98
4.2. The Signal-to-Noise Ratio of the Fly's-Eye Lens Array System	110
4.3. The Effect of the Third-Order Aberrations on the Fly's-Eye Lens Array System	113
4.4. The Change of Wavelength, Hologram Scale Factor and Magnification	124
4.5. The Application of the Fly's-Eye Lens Array to a Simple Three-Dimensional Air-Traffic Control Display System	131
CHAPTER 5. THE BEAMWIDTH REDUCTION METHOD OF A THREE-DIMENSIONAL HOLOGRAPHIC VIEWING SYSTEM	148
5.1. The Single-Lens Movie System and Its SB Characteristics	148
5.2. The Aberrations of the Single-Lens Movie System Caused by the Hologram Aperture	152
5.3. The Aberrations of the Single-Lens Movie System Caused by Other System Parameters	162
5.3.1. Aberrations Caused by ΔX Shift of the Hologram	162
5.3.2. Discussion of Aberrations Caused by Other Means During the Reconstruction Process	167
5.4. Other Engineering Considerations	170
5.4.1. The Pseudoscopic Image Inversion Problems (Including the Use of the Fly's-Eye Lens Array)	170
5.4.2. The Noise Consideration of the Single-Lens Movie System	172
5.4.3. The Production of Holographic Movie Film and the Film-Drive Problem	172
CHAPTER 6. THE EXPERIMENTAL RESULTS AND CONCLUSIONS	174
6.1. The Experimental Results of the Scatter-Plate Methods	174
6.2. The Aberration in a Single-Lens Movie System	181
6.3. The Results of the Pseudoscopic Image Inversion	186
6.4. The Results of the Single-Lens Movie System	186
6.5. Conclusions	190

	Page
Appendix A. DERIVATION OF THE NUMBER OF RESOLVABLE PLANES IN DEPTH FOR THE HUMAN EYE	194
Appendix B. DEFINITION OF A MEAN-SQUARE ERROR CRITERION FOR THE TWO BANDLIMITED PERIODIC FUNCTIONS IN A HILBERT SPACE	196
Appendix C. THE DERIVATION OF EQ. (4-3) AND A DISCUSSION ON ITS VALIDITY	198
REFERENCES	202

LIST OF ILLUSTRATIONS

Figure	Page
2-1. General holographic recording configuration	6
2-2. The SB plot of an object X at the hologram H	8
2-3. Space bandwidth plot of the intensity of a hologram	11
2-4. Space bandwidth plot of a hologram with collimated reference beam	12
2-5. The SB plot of the intensity of a hologram with an on-axis collimated reference beam	13
2-6. The SB plot of a hologram with a point-source reference at the center of the object	15
2-7. The SB plot of the intensity of a hologram with a point- source reference at the center of the object	16
2-8. The SB plot of two apertures H and X	18
2-9. The object space sections of a hologram using an on-axis point source reference beam	22
2-10. The subdivision of the object space regions at a typical object plane for a point-source reference case	24
2-11. The resolution distribution in four different object space sections of a holographic set-up	26
2-12. Resolution distribution along the central axis of a holographic set-up system	32
2-13. The object space sections of a hologram using an off-set point-source reference	36
2-14. Separation of an object space from its $ U ^2$ term by using an off-set point-source reference	39
2-15. The object space sections for the study of separating an object space from its $ U ^2$ term using an off-set point- source reference	41
2-16. Subdivision of the object space regions at a typical plane $z = z_1$ for the case of Fig. 2-15	42
2-17. The object space sections of a hologram using a normal plane-wave reference	49
2-18. The off-set case of Fig. 2-17 and the separation of an object space from its $ U ^2$ terms	55
3-1. The general configuration of a three-dimensional viewing system	62
3-2. The relation between γ and α for different value of β ...	65

Figure	Page
3-3. The optimum recording configuration of a holographic viewing system	67
3-4. The configuration of the sampling method for information reduction in holography	70
3-5. The available object space distance z vs the information reduction factor σ^2 in two dimension	74
3-6. The schematic diagram of a scatter-plate system	77
3-7. The fly's-eye lens array system	92
3-8. The angle β in the study of fly's-eye lens array	95
4-1. System model for the analysis of the fly's-eye lens array	99
4-2. The Gaussian approximation of a sinc function	102
4-3. The single-lens optical system model	114
4-4. The optical system for aberration study	114
4-5. Relation between hologram aperture ρ' and reconstructed beam aperture ρ at the lens aperture for different spherical aberrations deviation $\delta L/R$	119
4-6. The set of ellipse curves in the lens aperture corresponding to different K	123
4-7. Relation between hologram size t and the reconstructed resolution aperture ρ at the lens aperture for coma aberration. The Gaussian optics curve is also included	125
4-8. The model of a simple 3-D air-traffic control display system	133
4-9. The model of the fly's-eye lens array for the study of the axial movement of a point object	134
4-10. The model of the fly's-eye lens array for the study of the vertical movement of a point object	138
4-11. The model of the fly's-eye lens array and the hologram for the study of the air-traffic control system	140
4-12. An alternative configuration of the air-traffic control system	145
5-1. The single-lens system	149
5-2. The characteristic of the SB plot of a single-lens system	151
5-3. The beam pattern of Eq. (5-1).....	154
5-4. The beam pattern for the astigmatism and field curvature when the rays focus to a line $x=0$ or $y=0$	155
5-5. The set of hyperbolic curves in the lens aperture corresponding to different λ	158

Figure		Page
5-6.	The set of curves in the lens aperture corresponding to $x = K$, $y = l$ in the aberration study of coma	159
5-7.	Different hologram configurations and their corresponding beam pattern in the lens aperture (The hatch area)	160
5-8.	Special case of Figure 5-6 with $z = 0$	161
5-9.	The schematic diagram of a single-lens movie system	163
5-10.	The schematic diagram of a single-lens movie system for the study of a rotation aberration	169
5-11.	The system setup for pseudoscopic image inversion	171
6-1.	Resolution measurement of the fly's-eye lens array system vs the object distance z	175
6-2.	The reconstructed 3-D images when a fly's-eye lens array system is used	177
	(a) This picture shows the presence of laser speckle pattern in the reconstructed image	
	(b) This picture shows the speckle pattern is reduced by placing a rotating etched glass in the illuminating beam	
	(c) This picture shows the periodic structure of the fly's-eye lens array has been smoothed out by putting a diffuser at the central plane of the reconstructed image	
6-3.	(a) The spatial spectrum of a scatter plate supplied .. by Holotron	179
	(b) The reconstructed 3-D image using the above scatter plate. The object is diffusely illuminated	
	(c) The reconstructed image of a silhouette object	
6-4.	The reconstructed images of a random scatter-plate system	180
	(a) A transparent type of object	
	(b) A diffusely illuminated 3-D object	
	(c) A silhouette object	
6-5.	Picture of the black circle in the reconstruction beam..	182
6-6.	Spherical aberration causes an interference ring in the reconstruction beam	182
6-7.	Aberrations caused by ΔX shift of the hologram in a single-lens movie system	184
6-8.	Aberrations caused by ΔZ shift of the hologram in a single-lens movie system	184
6-9.	Aberrations caused by wavelength change in a single-lens movie system	185

Figure		Page
6-10.	The experimental setup for pseudoscopic image inversion	187
6-11.	The experimental results of the pseudoscopic image inversion	188
	(a) The original object	
	(b) The pseudoscopic image at $d = 3''$	
	(c) The pseudoscopic image at $d = 8''$	
	(d) The pseudoscopic image at $d = 13''$	
	(e) Same picture of (b) except the camera is focused at the fly's-eye lens array	
	(f) Same picture of (b) except a diffuser is put at the central plane of the image	
6-12.	(a) The reconstructed resolution chart constructed using a Fresnel lens	189
	(b) The reconstructed resolution chart constructed using a condensing lens	189
6-13.	The experimental results of the reconstructed image using liquid gate in a single-lens movie system	191
6-14.	The reconstructed 3-D images of the single-lens movie system	192
A-1.	The viewing configuration for the human eye	194

LIST OF TABLES

Table		Page
1.	Number of resolvable object points in different space regions for the object space using an on-axis point-source reference	33
2.	Number of resolution object points in different space regions for the object space separating from its $ U ^2$ term by an off-set point-source reference	44
3.	Number of resolvable points in each space region for three different cases	47
4.	Number of resolvable object points in different space regions for the object space separating from its $ U ^2$ term by an offset plane wave reference	59

CHAPTER 1

INTRODUCTION: THE PROBLEM OF INFORMATION REDUCTION AND BEAMWIDTH REDUCTION IN HOLOGRAPHY

The theory of holography was first proposed by Dennis Gabor in 1948. He intended to use holography to improve the image in the electron microscope by recording the interference pattern of a specimen object and a reference beam on a photographic plate using a coherent light source. This attempt was not very successful because the reconstructed image could not be separated from its conjugate term. The lack of a suitable coherent light source at that time further restricted the effectiveness of his method. In 1962, E. N. Leith and J. Upatnieks introduced an offset reference-beam method to separate the desired image from its conjugate term. The advent of a highly coherent laser source enabled them to record an object and later to reconstruct a very impressive 3-D image. The images reconstructed by their method are so life-like that a viewer can hardly tell whether he is looking at a real object or an image. Holography has now become a very popular and active research field. In recent years, holography has been applied to several scientific and industrial problems. One of them is the development of a 3-D display system.

A hologram is a two-dimensional storage medium which records the wavefront of a 3-D object by recording the interference pattern of the object beam with a reference beam. After processing, if necessary, the wavefront of the object can be reconstructed by illuminating the hologram with a coherent light beam that is the duplicate of the reference beam. Viewing a hologram is like looking through a window--the observer sees a 3-D image exhibiting parallax and the same field of view that would be obtained if the original object were there. The resolution of the reconstructed image far exceeds the human eye's viewing capability so that the observer perceives no loss of resolution. These properties make the hologram an ideal recording device for a 3-D display system. There are engineering difficulties and system limitations, however, depending on

the system task of a particular type of 3-D display system. In general, for a 3-D display system, the hologram aperture should be reasonably large to permit either several viewers to see the image simultaneously or to permit one person to see a 3-D object from different angles. For a movie system, thousands of holograms have to be made to provide a continuous movie scene. In a TV system, the information content recorded on the hologram must be transmitted from one station to another over a TV communication channel. Each of these two systems has its own characteristics and limitations. The purpose of this dissertation is to study how these two systems might be implemented with the use of a hologram as a recording device.

The information content of a general hologram was first calculated by Leith et al. [12] and was found to be 10^5 times greater than that of a typical TV picture. The information content of a hologram must be reduced before it can be transmitted through a TV channel. Several information-reduction methods have been proposed and studied [14-19]. One method, called the scatter-plate method, appears to be the most practical technique. The method uses a transparent type of phase scatter plate (SP) between the object and the hologram. The SP scatters the wavefront of the object as it travels toward the hologram. A portion of the scattered light is recorded by a small hologram. In the reconstruction process, the reconstructed wavefront propagates back through the SP, and a 3-D image is reconstructed at the original object position. To the observer, the SP appears to have the properties of a hologram; i.e. the image is 3-D, has parallax, and can be seen within the boundaries of the scatter plate. The information content which is recorded by the small hologram is much less than that recorded if a large hologram is used in place of the SP. This method was studied for several years by Haines and Brumm [16,17]. Their principal goal was to determine which SP results in the best system performance. Unfortunately, their works do not specify the basic requirements necessary for the human eye to view a 3-D object. Hence, they lacked a common criterion for comparing the performance of different 3-D display systems and pointing out the upper limits of the information-reduction method.

In this dissertation, Chapter 2 contains discussions of the concept of a space bandwidth plot of a hologram and the capability of a hologram to provide a 3-D image space. Chapter 3 includes a comparison of the human eye's capability for viewing a 3-D object with the capability of the hologram to provide a 3-D object scene. A sampling method model is discussed which can specify the upper limit of the information-reduction method. This model utilizes all the redundancy of the human eye in the holographic system and can thus be regarded as providing the optimum method for information reduction in holography.

In the scatter-plate method, the performance of which should never exceed the sample-method model, the problem of finding an optimum scatter plate was studied by means of rigorous mathematical formulation. In practice, this mathematical analysis is applicable to a two-dimensional transparent object. In the experimental study, a high-quality fly's-eye lens array demonstrates its superiority over other existing scatter plates. Chapter 4 is therefore devoted to the study of the fly's-eye lens array system.

In Chapter 5, a second type of 3-D display system which uses a demagnification process is described. This beamwidth-reduction method does not reduce the information content of the hologram; rather, it reduces the spatial extent of the wavefronts so that they can be recorded by a small hologram. This reduction is essential for a movie system, since the movie strip has to be small enough to be transported by a mechanical film drive. The object-beam reduction is accomplished by a lens system with a large diameter and low f-number (14 in. diameter and $f/0.7$). The reduced object beam is then recorded and later reconstructed by the hologram. The reconstructed wavefront propagates back through the optical imaging system and forms a 3-D image at the original object location. The optical imaging system for this method does not require high-quality optical components because all the aberrations of the system are compensated for during the reconstruction process. Since information is not reduced in this method, the reconstructed images are excellent.

There are other important factors which affect image quality and which are examined in following chapters. One of them is the aberrations of the reconstructed image caused by misalignment of the hologram and the imaging system. This problem is treated in Section 5.3 and experimental verification is presented in Section 6.2. Other engineering problems such as film noise, system stability, movie production, and image inversion are studied in Section 5.4 and the solution of these problems are experimentally performed and reported in Chapter 6. A movie strip of 15 frames was experimentally made in the laboratory and used to demonstrate the feasibility in a 3-D movie system.

CHAPTER 2

THE SPACE BANDWIDTH PLOTS AND THE INFORMATION CAPACITY OF A HOLOGRAM

2.1. THE CONCEPT OF A SPACE BANDWIDTH (SB) PLOT OF A HOLOGRAM

Consider the general holographic recording configuration in Fig.

2-1. An optical signal $U_o(x, y)$ is propagating from the object aperture X toward the hologram aperture H ; z is the distance between X and H . At the object plane, $U_o(x, y)$ can be written as

$$U_o(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_o(f_x, f_y) \exp[j2\pi(f_x x + f_y y)] df_x df_y$$

where $u_o(f_x, f_y)$ is the spatial frequency spectrum of $U_o(x, y)$. This expression is valid for all x, y . As it is well known, the above expression can be physically interpreted as a decomposition of $U_o(x, y)$ into a collection of plane waves, each with the form

$$\exp[j2\pi(f_x x + f_y y)]$$

for all x, y and each weighted by the weighting function $u_o(f_x, f_y)$.

Since $U_o(x, y)$ is zero outside the object aperture X , the physical interpretation of $U_o(x, y)$ can have another form. Instead of $U_o(x, y)$ being viewed as decomposed into a collection of plane waves with expression

$$\exp[j2\pi(f_x x + f_y y)]$$

for all x, y , the plane-wave expression can be such that it is non-zero only for $|x| \leq X, |y| \leq Y$; $U_o(x, y)$ then becomes

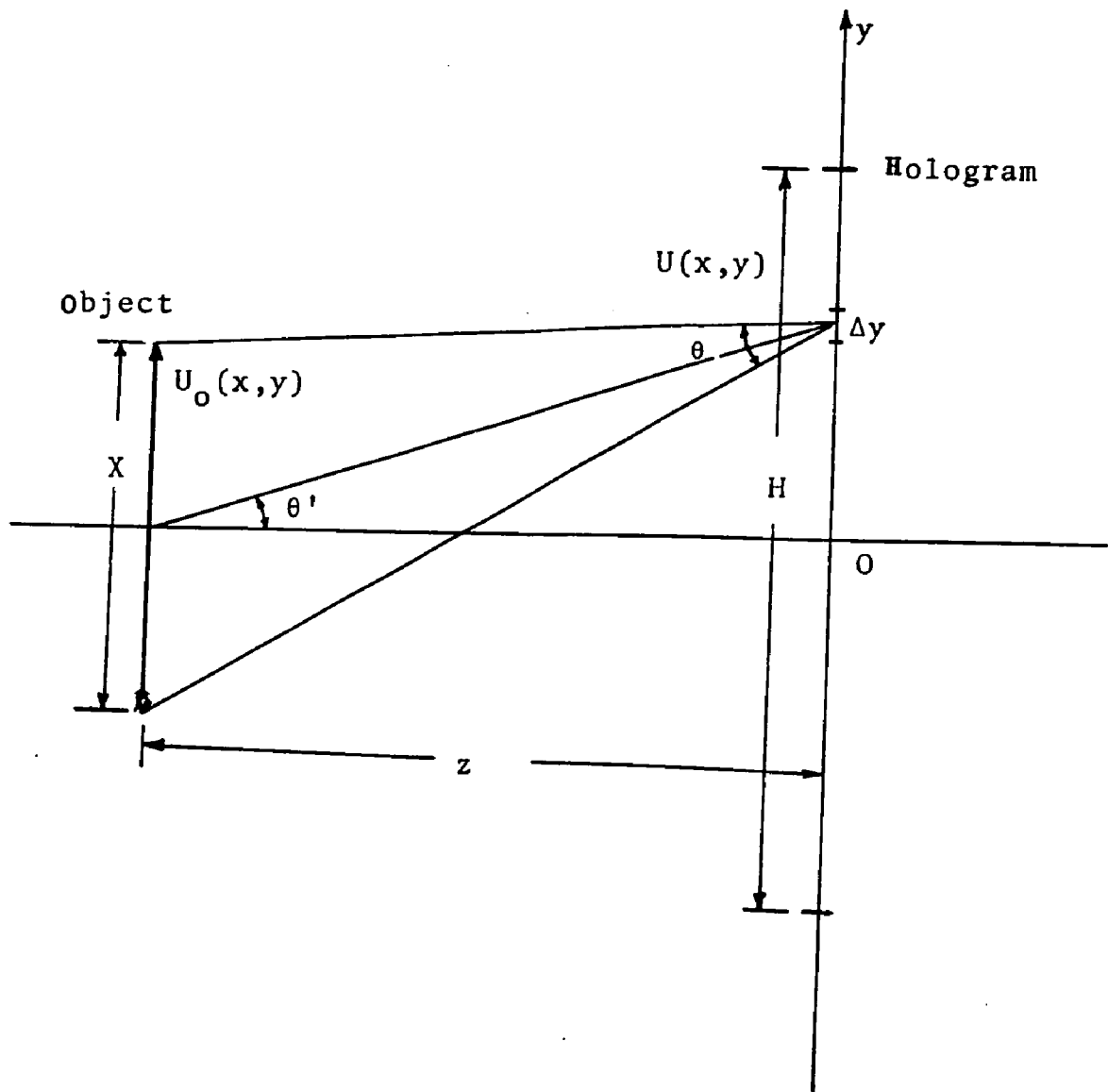


Figure 2-1.
General holographic recording configuration

$$U_0(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(f_x, f_y) h(x, y, f_x, f_y) df_x df_y$$

where

$$h(x, y, f_x, f_y) \triangleq \begin{cases} \exp[j2\pi(f_x x + f_y y)] & |x| \leq X, |y| \leq X \\ 0 & \text{elsewhere} \end{cases}$$

With this interpretation for the optical signal $U_0(x, y)$, we can discuss the propagation of $U_0(x, y)$ from the object aperture to the hologram aperture.

Each space-limited plane wave $h(x, y, f_x, f_y)$ will propagate from aperture X to H according to geometrical optics. The field $U(x, y)$ at any particular point (x, y) in the hologram plane can be expressed as the summation of all the plane-wave components that reach the point (x, y) . For simplicity, only one small element Δy in the hologram plane will be considered from now on. From simple geometry, it can be seen that the highest frequency component and the lowest frequency components of the E-M field $U(y, f_y)$ at Δy have a frequency difference equal to $\sin \theta / \lambda$; θ is the angle subtended by the object aperture X at the hologram H, and λ is the wavelength used in the system. The angle θ at a different element Δy along the hologram plane is a function of y and is given approximately by

$$\theta \doteq \frac{X}{z} \left(1 - \frac{y^2}{z^2} \right).$$

For the case where $z \gg X$ and $z \gg H$, θ is approximately a constant, i.e. the spatial frequency bandwidth at different small recording elements Δy along the y axis in the hologram is a constant. However, the absolute frequency in each Δy element along the y axis relative to the zero frequency will not be constant. It is given by $\sin \theta' / \lambda \approx y / z$ as shown in Fig. 2-1. Using a Cartesian coordinate system, let one axis represent spatial frequency f_y and the other axis represent spatial distance y along the hologram, the so-called "space bandwidth plot" of a hologram can be plotted as shown in Fig. 2-2. At any particular coordinate point y , $\Delta y \rightarrow 0$ and the instantaneous spatial frequency is regarded as a function

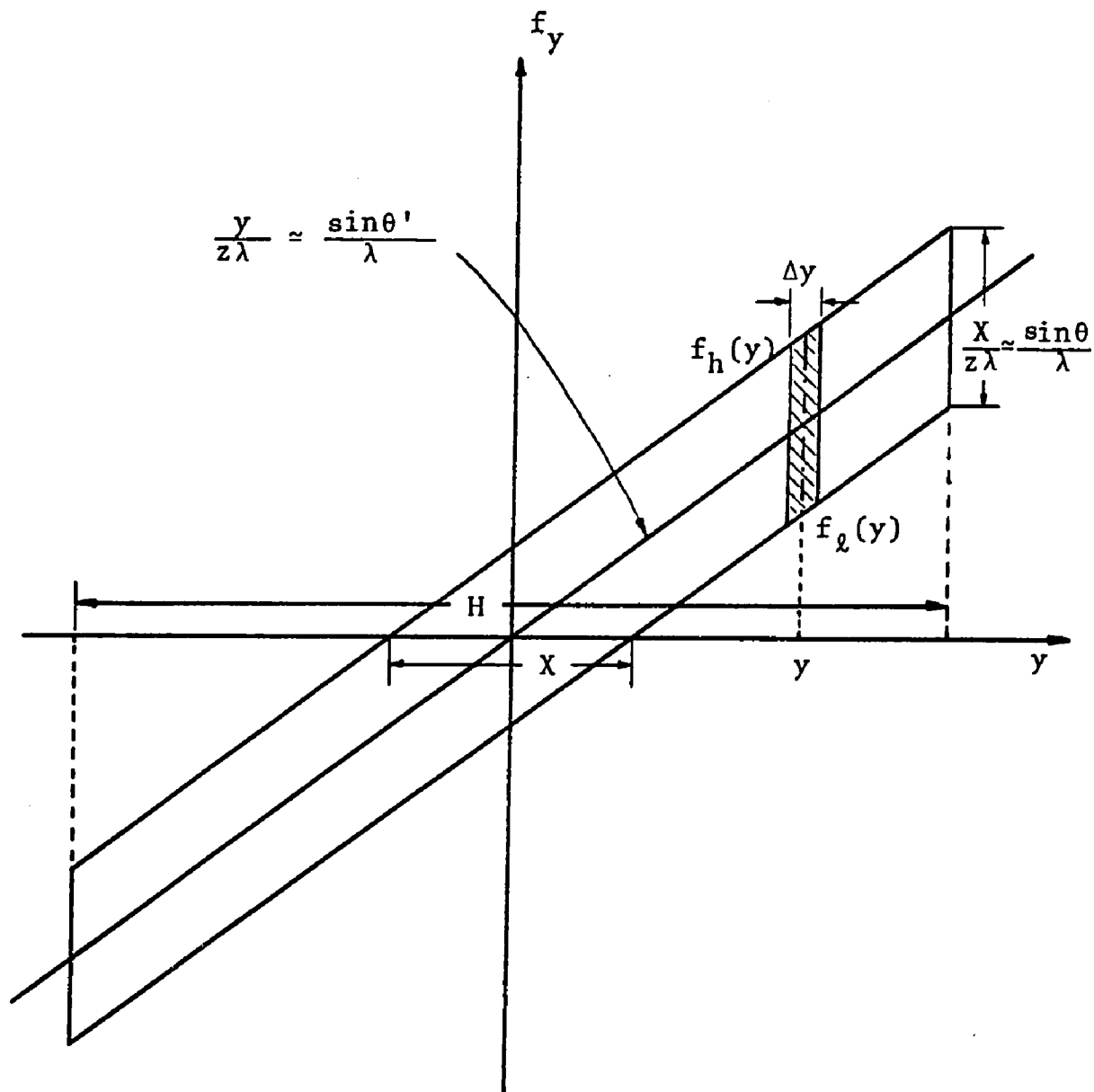


Figure 2-2. The SB plot of an objet X at the hologram H

of spatial distance y . The highest and the lowest instantaneous spatial frequency at the point y is approximately

$$f_h(y) = \frac{1/2(X + y)}{\lambda z}$$

and

$$f_l(y) = \frac{1/2(X - y)}{\lambda z}$$

respectively. The bandwidth at any point y is equal to

$$\frac{\sin \theta}{\lambda} \approx \frac{X}{\lambda z}$$

In order to record the E-M field $U(y, f_y)$ on the hologram, the wavefront should be sampled so that the information content can be stored digitally. The number of samples per unit spatial distance y is equal to the bandwidth of the field $U(y, f_y)$ at any particular point y . Hence, the total number of sampling points in the hologram plane is equal to the product of the spatial bandwidth $X/\lambda z$ and the aperture H of the hologram. This is the so-called space bandwidth product and it is given by $(SW)_H = XH/\lambda z$. In Fig. 2-2, the value $XH/\lambda z$ corresponds to the area bounded by the highest and lowest instantaneous spatial frequency and the hologram size $\pm \frac{1}{2} H$. In the following analysis, the concepts of the space bandwidth (SB) plot and the SB product $(SW)_H$ are of primary importance in understanding the information capacity of a holographic viewing system. In the next section, the SB plot is presented for several cases to show the powerful analytical method which this concept makes available.

2.2. THE USE OF SB PLOT IN A HOLOGRAPHIC SYSTEM

2.2.1. THE SB PLOTS FOR VARIOUS HOLOGRAPHIC RECORDING CONFIGURATIONS

In Section 2.1, the concept of SB plot for an E-M field $U(y, f_y)$ at a hologram H was introduced. In the practical holographic recording system, the intensity rather than the E-M field itself is actually recorded. Moreover, a reference beam is usually needed in the recording process. It is thus highly desirable to investigate the SB plots for

various recording configurations. They are listed below:

- A. The SB plot of the intensity of the E-M field at the hologram plane. The SB plot in Fig. 2-2 is for the field $U(y, f_y)$ at the hologram plane. Since the intensity is equal to the product of the field and its conjugate, the spectrum of intensity at any point y is the autocorrelation function of the field spectrum at y . Hence, from Fig. 2-2, the SB plot for the intensity of $U(y, f_y)$ at the hologram plane is as shown in Fig. 2-3.
- B. The SB plot of the intensity when a collimated reference beam is used. In a general holographic recording system, a reference beam is usually used to separate the desired image from its conjugate terms. If a collimated reference beam is perpendicular to the hologram and the object field (as in Fig. 2-1), the SB plot for the field is like that in Fig. 2-4 where the collimated beam is represented by a δ -function along the y axis. The SB plot for the intensity of the above field is the autocorrelation function of the frequency spectrum at each y . Hence, the SB plot for the intensity is given by Fig. 2-5. In Fig. 2-5, Branch A is the SB plot of the real-image recording; Branch B is the SB plot of the virtual image recording; and Branch C is the SB plot for the autocorrelation term of the object itself. The autocorrelation term of the reference plane wave itself is an δ -function along the y axis. This plot illustrates that the real and virtual images cannot be separated because Branches A and B are inseparable by any simple reconstruction δ -function. If the δ -function in Fig. 2-4 is lowered to a negative frequency so that it does not overlap the object field SB plot, then complete separation between real image and virtual image can be achieved upon reconstruction. In this case, an off-axis collimated beam is used as reference for recording and reconstruction. The exact value of the off-axis angle

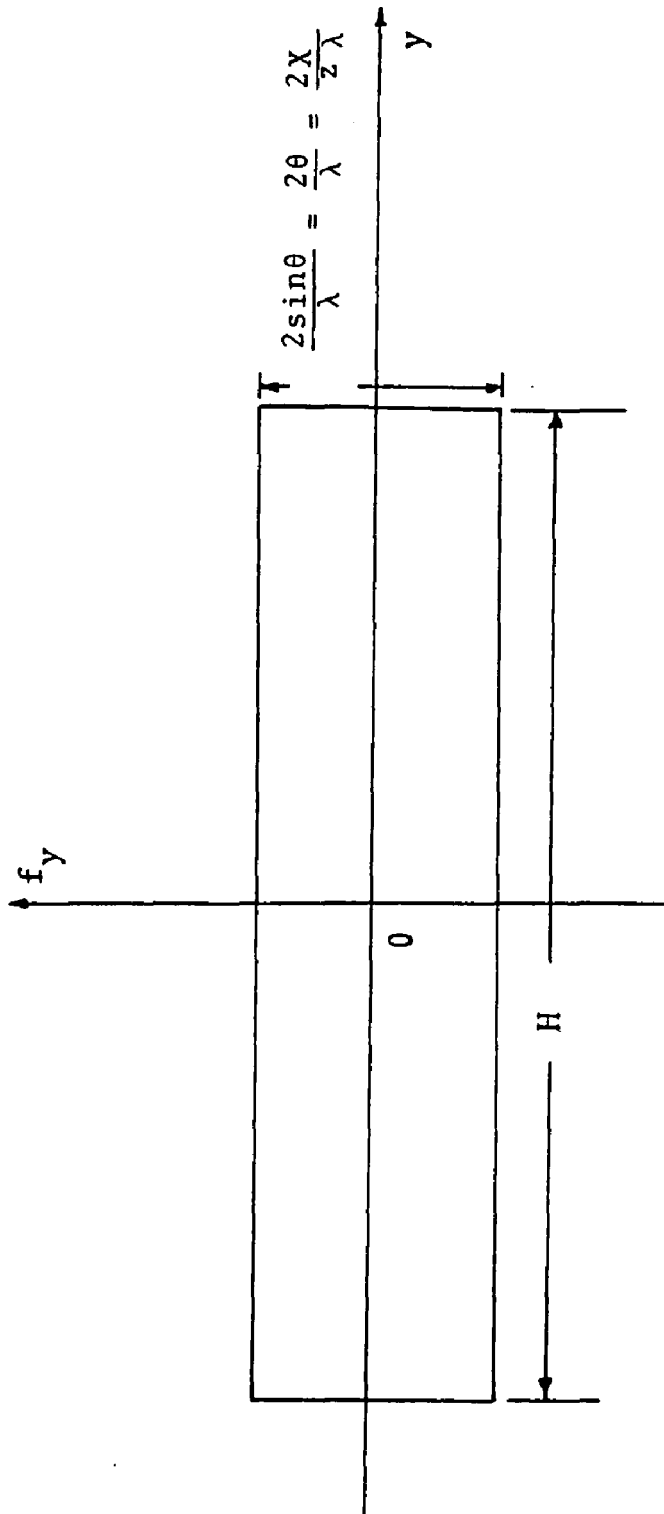


Figure 2-3.

Space bandwidth plot of the intensity of a hologram

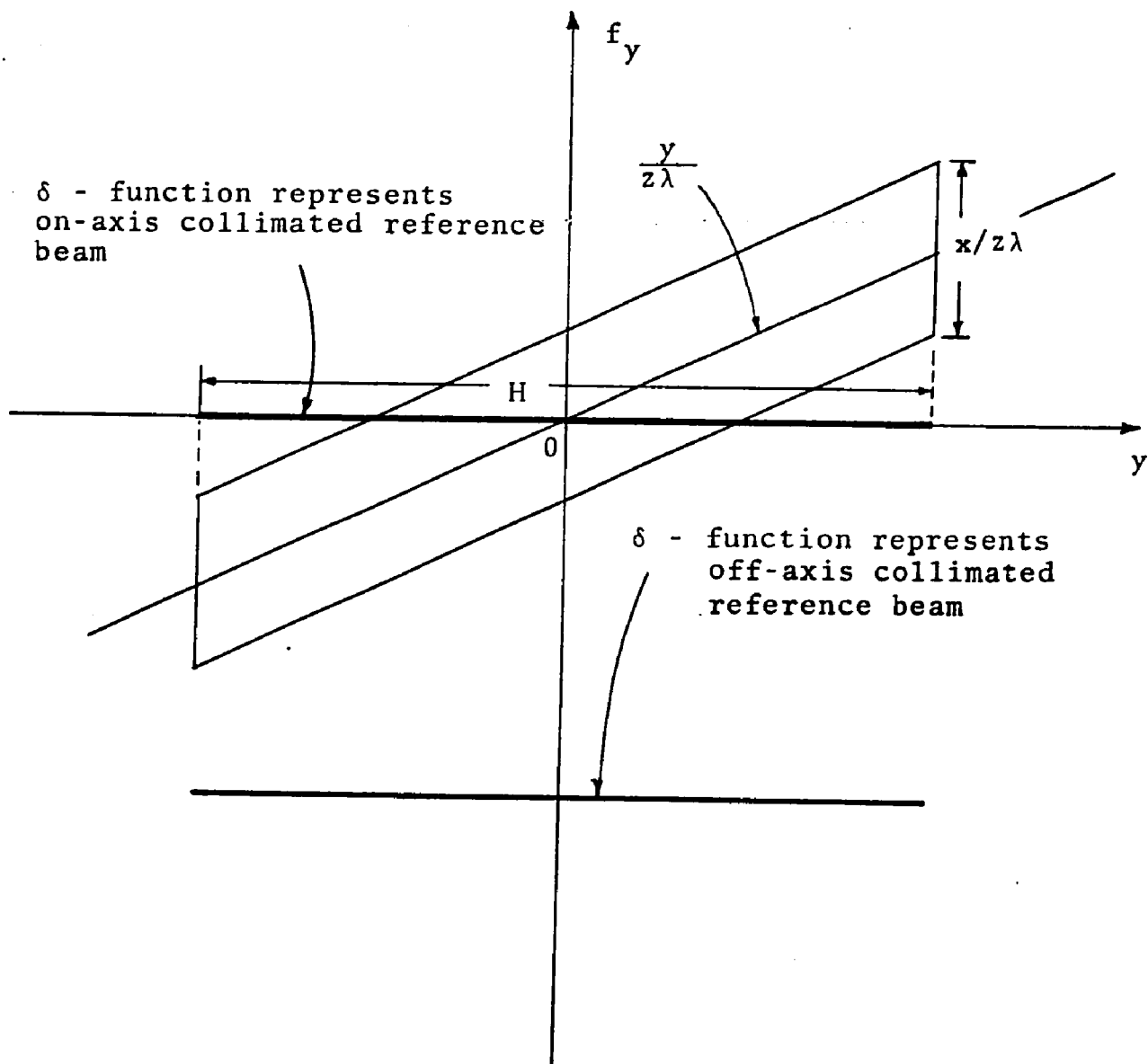


Figure 2-4.

Space bandwidth plot of a hologram with collimated reference beam

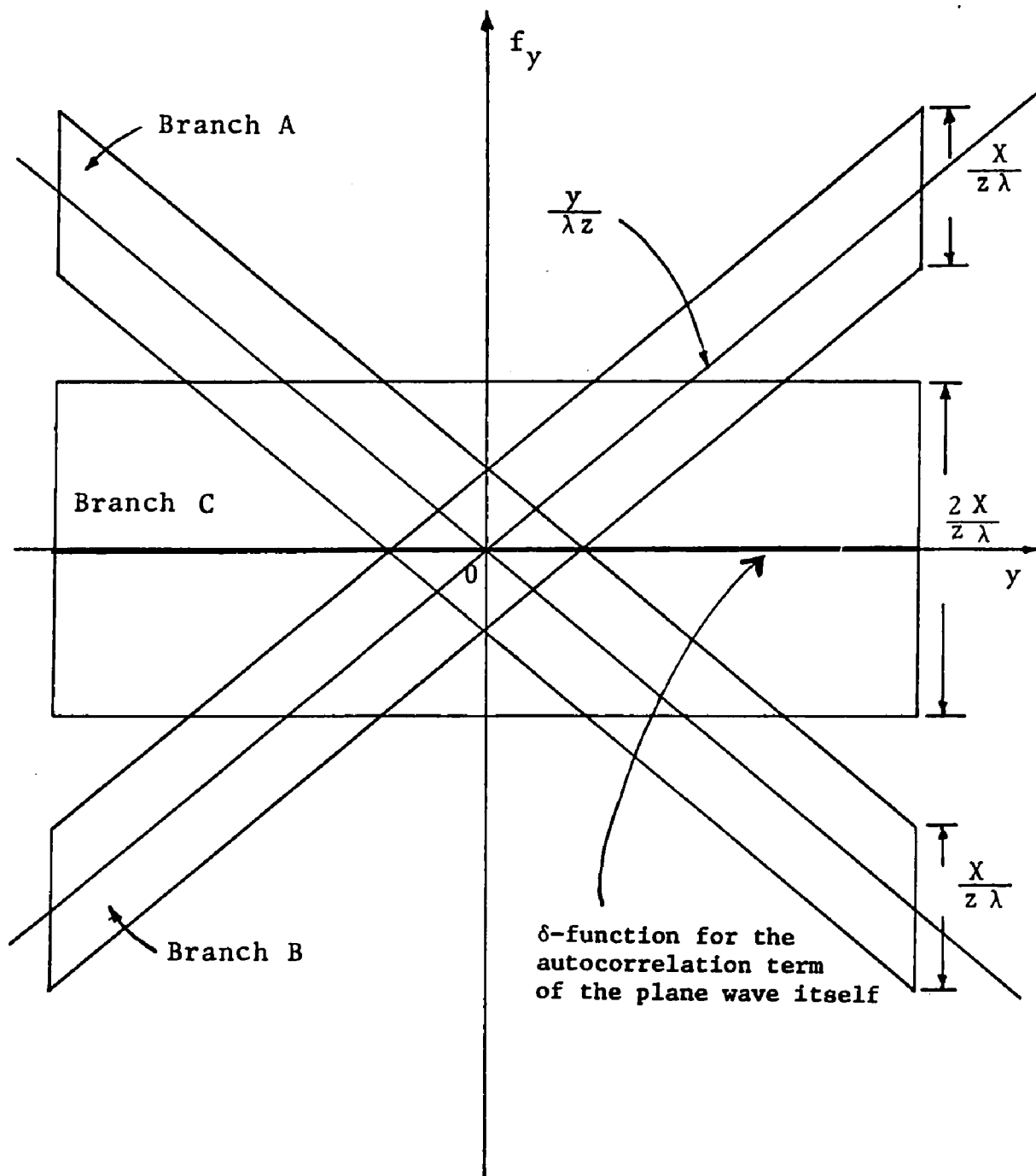


Figure 2-5.

The SB plot of the intensity of a hologram with an on-axis collimated reference beam

required for the separation can be easily obtained by the SB plot in Fig. 2-4.

- C. The SB plot of the intensity when a point-source reference is used. If a point-source reference is used at the center of the object aperture X , the SB plot of the point-source reference is a δ -function located at $y/\lambda z$ for each y (see Fig. 2-6). The SB plot for the intensity of the point-source and the object X should be the autocorrelation between the δ -function and the SB plot of object field in Fig. 2-6. The SB plot for the intensity in this case is shown in Fig. 2-7. The real and virtual image branches are now coincident with each other (Fig. 2-7), but the actual spectrum values for the two branches are different. With the point-source reference at the center of the object, the problem of separating the real image from the virtual image remains. This problem can be solved by putting the point reference away from the object aperture, both in recording and reconstruction. The exact position at which to locate the reference point can be determined from the SB plots of Figs. 2-6 and 2-7.

In summary, the SB plot of an E-M field is different from that of the intensity of an E-M field. The image of the object X can be reconstructed exactly in three dimensions only if the SB plot shown in Fig. 2-2 can be reconstructed exactly. Since the hologram can only record the intensity of an E-M field, a reference beam is usually needed to assure that the E-M field can be reconstructed after recording. A reference beam that is a δ -function in the SB plot is a carrier medium in the recording and reconstruction operation. A reference beam changes only the position or the shape of a SB plot, not its information content, i.e., the space-bandwidth product of an object field in the hologram is preserved although the SB plot varies in the process.

2.2.2. THE DUAL ROLE OF THE SB PLOT OF TWO APERTURES

The information content (i.e., the space-bandwidth product) of an SB plot is preserved when a reference beam is used to change its position or shape in the recording or reconstruction process. Figure 2-2 shows

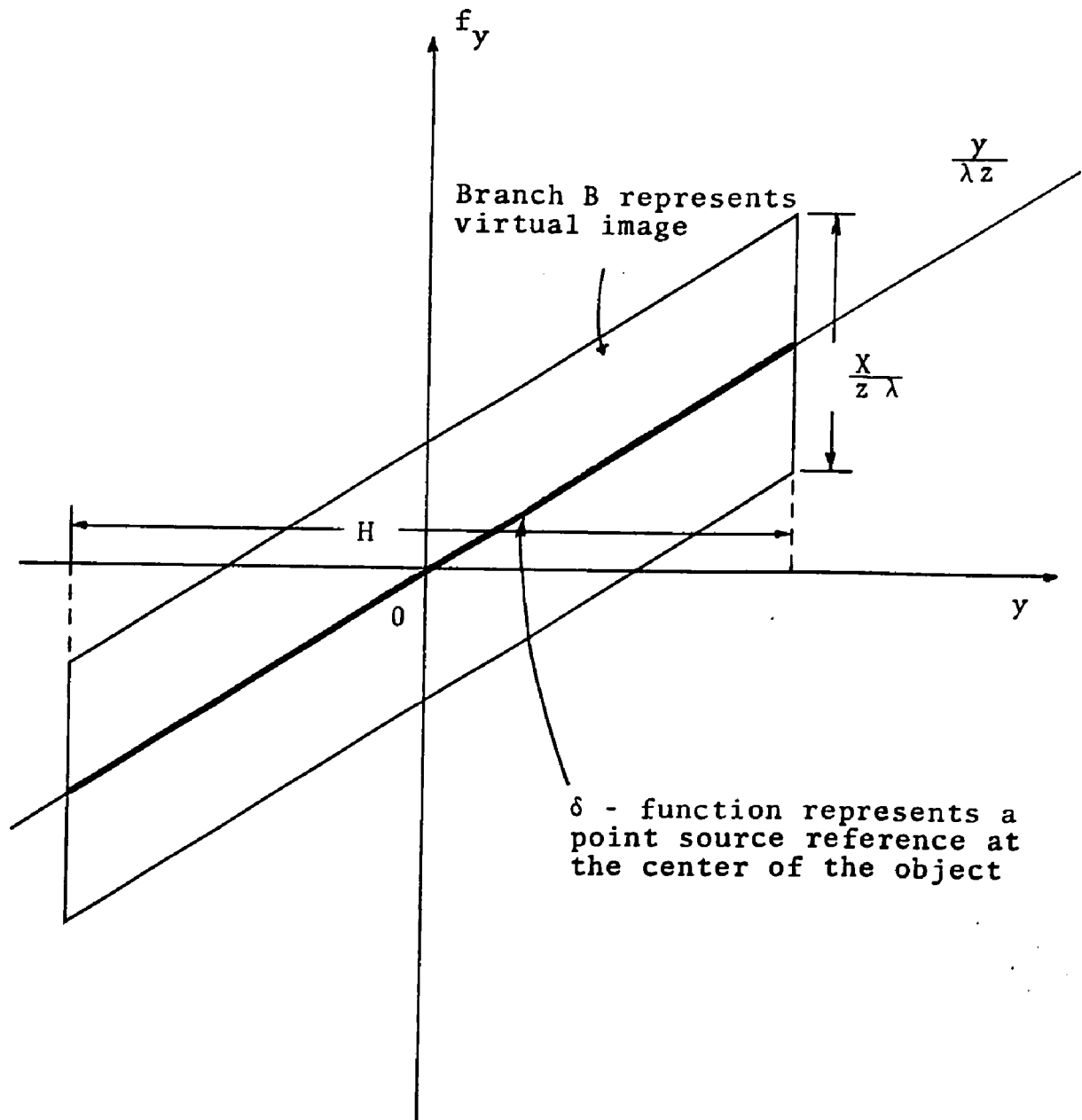


Figure 2-6.

The SB plot of a hologram with a point-source reference at the center of the object

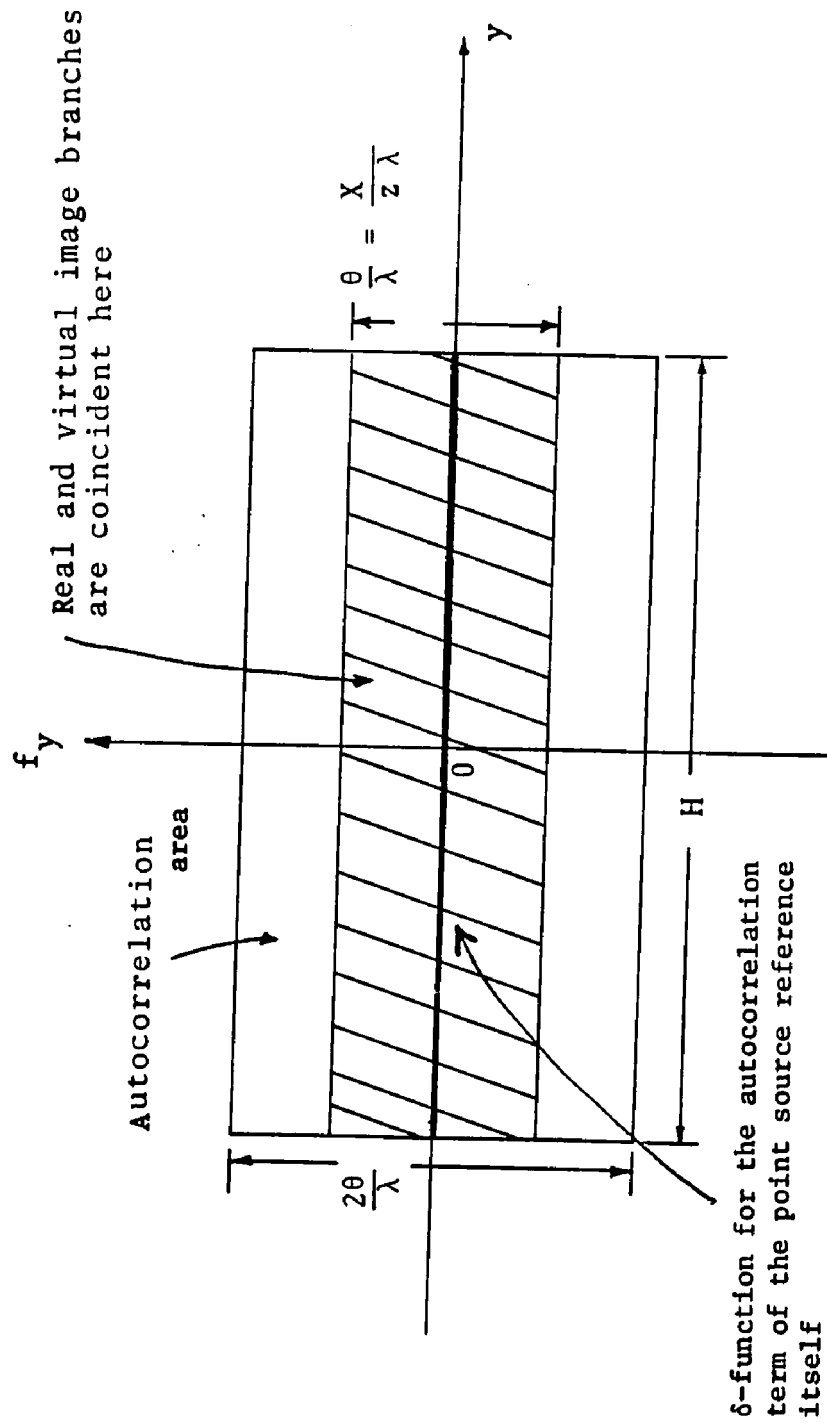


Figure 2-7.

The SB plot of the intensity of a hologram with a point-source reference at the center of the object

the SB plot on an aperture H caused by an optical signal at aperture X. When a hologram is used to record all the information that reaches aperture H and to reconstruct the real image of the original signal at aperture X, we assume that all the space limited plane-wave components of the signal are reconstructed geometrically. The SB plot of the original signal is shown in Fig. 2-1; Fig. 2-8 shows an SB plot of the reconstructed signal. The signals are, of course, conjugates of each other, but a more detailed study reveals the following relations:

1. The area of the two SB plots are equal:

$$\frac{X}{\lambda z} H = \frac{H}{\lambda z} X$$

i.e., the information content has been preserved with the above propagation assumption.

2. The frequency bandwidth of one SB plot is proportional to the size of the other aperture:

$$\text{Frequency bandwidth on H} \triangleq (BW)_X = \frac{X}{\lambda z}$$

$$\text{Frequency bandwidth on X} \triangleq (BW)_H = \frac{H}{\lambda z}$$

3. The central frequency functions are straight lines through the origin with opposite slopes for two SB plots:

$$f_y = \frac{Y}{\lambda z} \text{ for H,} \quad f_y = -\frac{Y}{\lambda z} \text{ for X}$$

4. The intersections of the high-frequency boundary line and the low-frequency boundary line of one SB plot with y axis gives the size of the other aperture.

Recognizing the relationship between the two plots, we can use them to study the effect of increasing one aperture relative to the other or shifting one aperture from one plane to another. In fact, these two SB plots depend only on the size of the two apertures and the distance between them. The plots play a dual role in the SB plot related to information transmission and are very important to the understanding of some sections in this dissertation.

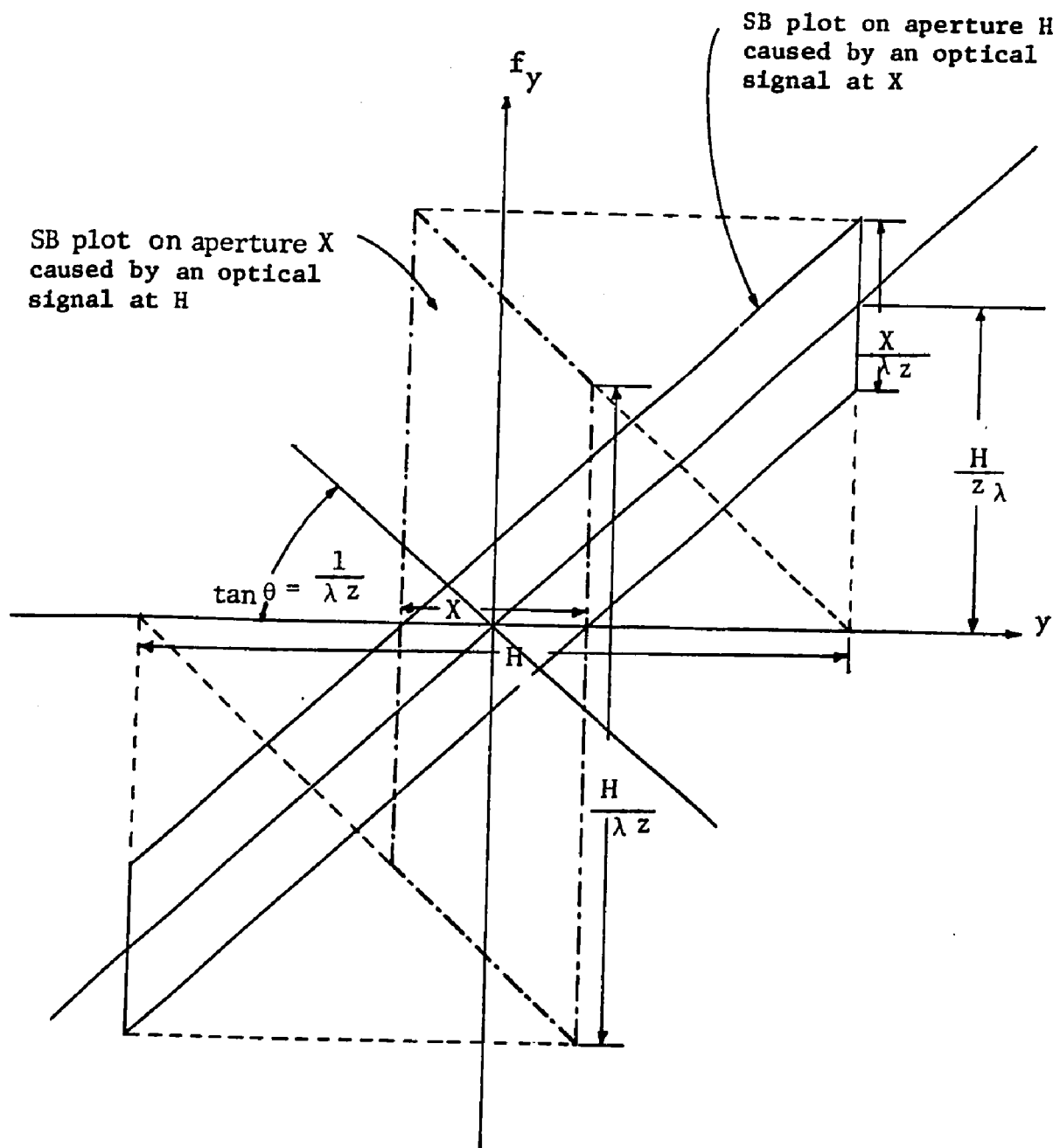


Figure 2-8.

The SB plot of two apertures H and X

2.3. THE DISTRIBUTION OF THE IMAGE RESOLUTION AND INFORMATION CAPACITY OF A HOLOGRAM

2.3.1. INTRODUCTION

Since the advent of holography, there has been much speculation about the possibility of a 3-D movie system. The hologram is superior to other methods for recording three-dimensional imagery if the required number of resolvable points in depth is sufficiently large [1,2] and if the object is viewed by the human eye so that the signal-to-noise ratio at different object planes is not a problem [2]. In a three-dimensional, holographic viewing system, we are concerned about the image resolution and the number of reconstructed image points in space, as well as with the recording data in the hologram plane which provides such a reconstruction.

The resolution and the information capacity of a hologram have been discussed in the literature by various people [3-5]. Studies of resolution have been based on either other considerations [5] or on the assumption that the resolution is strictly limited by the hologram aperture [3,4]. Based on the hologram-limited resolution, A Macovski [3] has calculated the maximum object extent and the number of resolvable object points in the object space. In this section, we intend not only to consider the hologram-limited resolution but also to consider all the object space which can be reconstructed, either by the whole aperture of the hologram or by only part of it. With this approach, the basic capability of a two-dimensional hologram plate to record a three-dimensional object scene is revealed. When a hologram of a finite spatial frequency is used, the object space can be divided into different sections according to the specific value of the spatial frequency, the hologram aperture, and the location of the reference beam in use. If the resolution is limited by the full aperture of the hologram, that space section is denoted as the "H-limited" section; if it is limited by the spatial frequency F_x of the hologram, it is denoted as the " F_x -limited" section; and if it is limited by the hologram aperture on one side and the spatial frequency on the other side, it is denoted as the " HF_x -limited" section. The number of resolvable object points at each section plane is calculated. Then the number of resolvable points for

the volume of each section is calculated. Gabor's on-axis hologram is first considered in detail. Studies are then extended to the offset case in which an object space is separated from its convolution term by an offset reference beam. The object space separated in this way has much less information capacity than the on-axis object space. Both a point-source and a plane-wave reference beam have been used in this study.

Throughout the section, the hologram is assumed to be a square with width W . The spatial frequency of the emulsion is assumed to be the same in both x and y directions and is denoted by F_x . Using the sampling theorem, we define n as the total number of definable elements in each dimension, corresponding to the number of lines in a television system. Thus, n is related to F_x by $n = 2wF_x$ where each cycle defines two independent samples.

2.3.2. THE DISTRIBUTION OF THE IMAGE RESOLUTION AND INFORMATION CAPACITY OF A HOLOGRAM FOR WHICH AN ON-AXIS REFERENCE POINT SOURCE IS USED

Consider a general object point in the object space of a holographic recording system. With the paraxial approximation, the field disturbance at the hologram plane caused by this object point is

$$U(x, y) = U_o \exp \left\{ j \frac{k}{2z_i} \left[(x - x_i)^2 + (y - y_i)^2 \right] \right\}$$

where (x_i, y_i, z_i) are the coordinates of the object point and $(x, y, 0)$ form a general point in the hologram plane. If an on-axis point source is used as a reference, the disturbance at point $(x, y, 0)$ on the hologram resulting from this reference point source $(0, 0, z_o)$ is

$$R = R_o \exp \left[j \frac{k}{2z_o} (x^2 + y^2) \right]$$

The intensity at $(x, y, 0)$ caused by the interference of those two points is thus

$$I = |R|^2 + |U|^2 + U^*R + R^*U = U_o^2 + R_o^2 + 2U_oR_o \cos \delta$$

where

$$\delta = \frac{k}{z_1} \left[x_1 x + y_1 y + \frac{1}{2} \left(\frac{z_1 - z_0}{z_0} \right) (x^2 + y^2) - \frac{1}{2} (x_1^2 + y_1^2) \right]$$

The instantaneous spatial frequencies are given by

$$f_x \triangleq (2\pi)^{-1} \frac{\partial \delta}{\partial x} = \frac{1}{\lambda z_1} \left[x_1 + \left(\frac{z_1 - z_0}{z_0} \right) x \right]$$

$$f_y \triangleq (2\pi)^{-1} \frac{\partial \delta}{\partial y} = \frac{1}{\lambda z_1} \left[y_1 + \left(\frac{z_1 - z_0}{z_0} \right) y \right]$$

where f_x , f_y are functions of both object coordinates and hologram coordinates. If we set the instantaneous spatial frequencies equal to the spatial cut-off frequency F_x of the hologram emulsion, we can find an important relation between object coordinates x_1 and hologram coordinates x

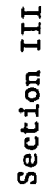
$$x_1 = -(z_1 - z_0) \left(\frac{x}{z_0} \right) + \lambda z_1 F_x \quad (2-1)$$

Equation (2-1) shows that at the plane $z_1 = z_0$, the maximum extent of the object space is limited by the spatial frequency F_x and is independent of the hologram size. In this plane, the object extent in one direction is denoted by X_0 . Hence, $X_0 = \lambda z_0 F_x$ or $F_x = X_0 / \lambda z_0$, as shown in Fig. 2-9. For planes with $z_1 > z_0$, the object space has two important boundaries denoted by X_m and X_b . X_m is the maximum object extent within which the object point forms its maximum spatial frequency at the edge of the hologram which is less than the spatial cut-off frequency F_x and, hence, can be reconstructed by the whole aperture of the hologram. If Eq. (2-1) is used, X_m is given by

$$X_m = -(z_1 - z_0) \left(\frac{W}{2z_0} \right) + \lambda z_1 F_x$$

or

$$X_m = -(z_1 - z_0) \left(\frac{W}{2z_0} \right) + \frac{z_1}{z_0} X_0 \quad (2-2)$$



Section II

Figure 2-9.

where W is the hologram aperture. X_b is the boundary of the object space beyond which no point can be reconstructed by the hologram. Between X_m and X_b , the object point can only be reconstructed by part of the hologram aperture. This is so because the spatial frequencies formed by the object point in part of the hologram is higher than the hologram cut-off frequency and thus cannot be recorded by the hologram. The resolution of the image within this region is HF_x -limited; X_b can be found to be

$$X_b = (z_i - z_o) \left(\frac{W}{2z_o} \right) + \frac{z_i}{z_o} X_o \quad (2-3)$$

In Fig. 2-9, we assume $2X_o < W$. In this case, the object space can be subdivided into four sections: I, II, III, IV as shown in Fig. 2-9, each having boundaries X_m and X_b . The values of X_m and X_b in section I are given by Eqs. (2-2) and (2-3). The corresponding values of X_m and X_b at other sections can be easily obtained with the help of Fig. 2-9. Within each section, consider a general plane $z = z_i$. This plane is subdivided into nine pieces by the two boundaries X_m and X_b in the x and y directions. Because of the symmetry of the system, there are only three different regions. They are labeled by A, B, and C in Fig. 2-10, which shows the subdivision of the plane $z = z_i$. In the analysis which follows, different space regions in each section will be referred to as IA, IIB, and so on.

With the space regions defined and the fundamental boundary equations given above, the resolution distribution and information capacity in each section can be determined. In section I, the boundary equation X_m and X_b is given by Eqs. (2-2) and (2-3). The resolution between $\pm X_m$ is given by $\lambda z_i / W$, which is H-limited. For points between X_m and X_b , the resolution aperture is limited by the hologram on one side and by the spatial frequency on the other. A typical point K in this region is shown in Fig. (2-9). The resolution aperture of this point is

$$A_x = \frac{W}{2} + x'$$

where x' is the value of x obtained from Eq. (2-1) and is given by

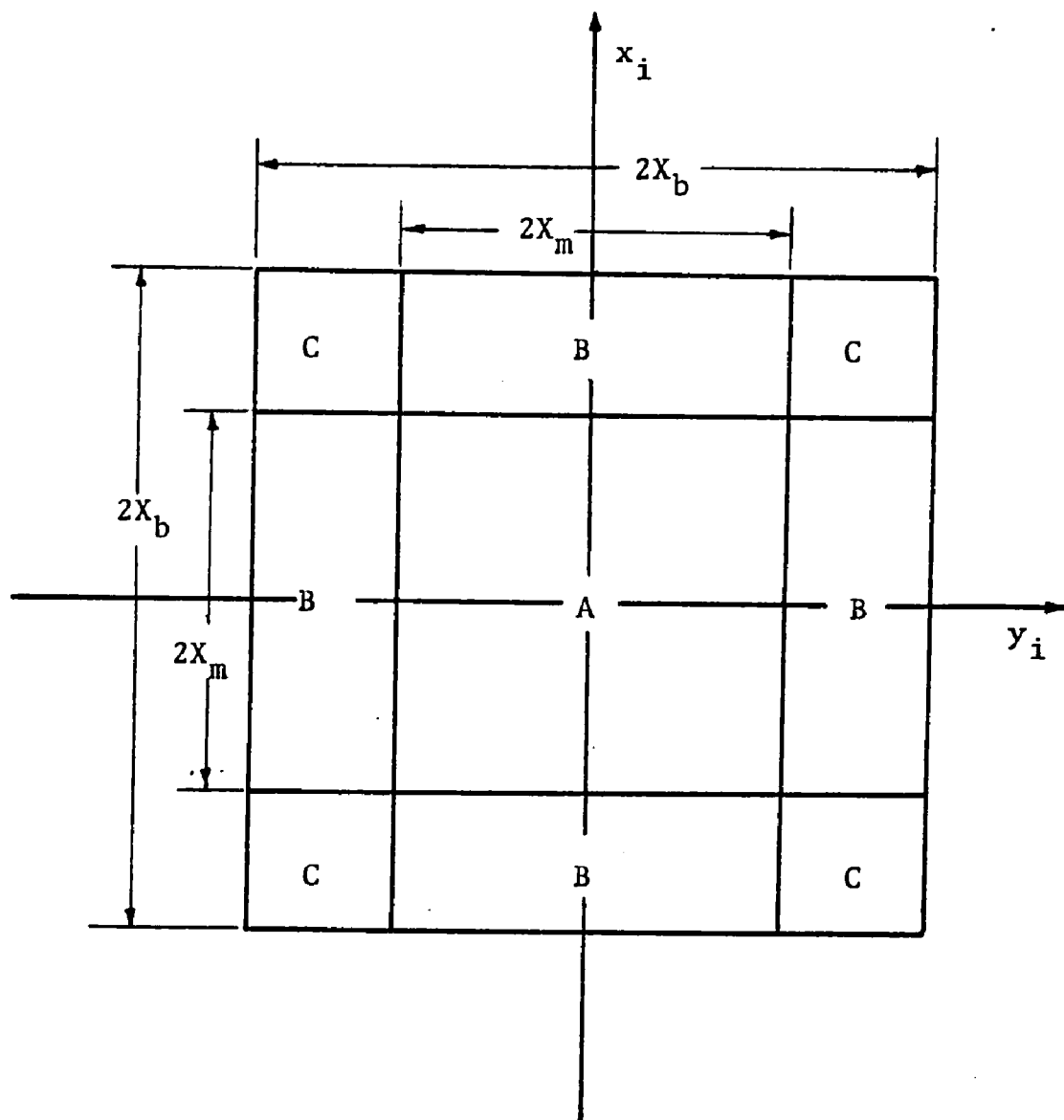


Figure 2-10. The subdivision of the object space regions at a typical object plane for a point-source reference case

$$x' = - \frac{x_1 z_0 - z_1 x_0}{z_1 - z_0}$$

The resolution at point x_1 depends on the coordinates x_1 :

$$R_1 = \frac{\lambda z_1}{A_x} = \frac{\lambda z_1}{\frac{W}{2} + x'} \quad (2-4)$$

This complicated expression can be simplified for a given plane z_1 if we shift the coordinate x_1 to a new coordinate $x_1' \triangleq x_1 - X_b$. In terms of this new coordinate x_1' , R_1 is given by

$$R_1 = \frac{\lambda z_1 (z_0 - z_1)}{z_0 x_1'} \quad (2-5)$$

It is also advantageous to express R_1 in terms of another coordinate $x_1'' \triangleq x_1 - X_m$, which yields

$$R_1 = \frac{\lambda z_1}{W + \frac{x_1'' z_0}{z_0 - z_1}} \quad (2-6)$$

The resolution distribution along the x_1 axis in section I (shown in Fig. 2-11a) can now be plotted with the help of Eq. (2-4) or Eq. (2-6). Within $\pm X_m$, the resolution is H-limited; between X_m and X_b , the resolution is HF_x -limited and is a variable given by Eq. (2-4) or (2-6)

For a typical plane $z = z_1$ in section I, the number of resolvable object points in each region can now be calculated. Between $\pm X_m$, the number of resolvable points is

$$N_x = \int_{-X_m}^{X_m} \frac{dx_1}{\frac{\lambda z_1}{W}} = \frac{2X_m}{\frac{\lambda z_1}{W}}$$

Between X_m and X_b , the number of resolvable points is

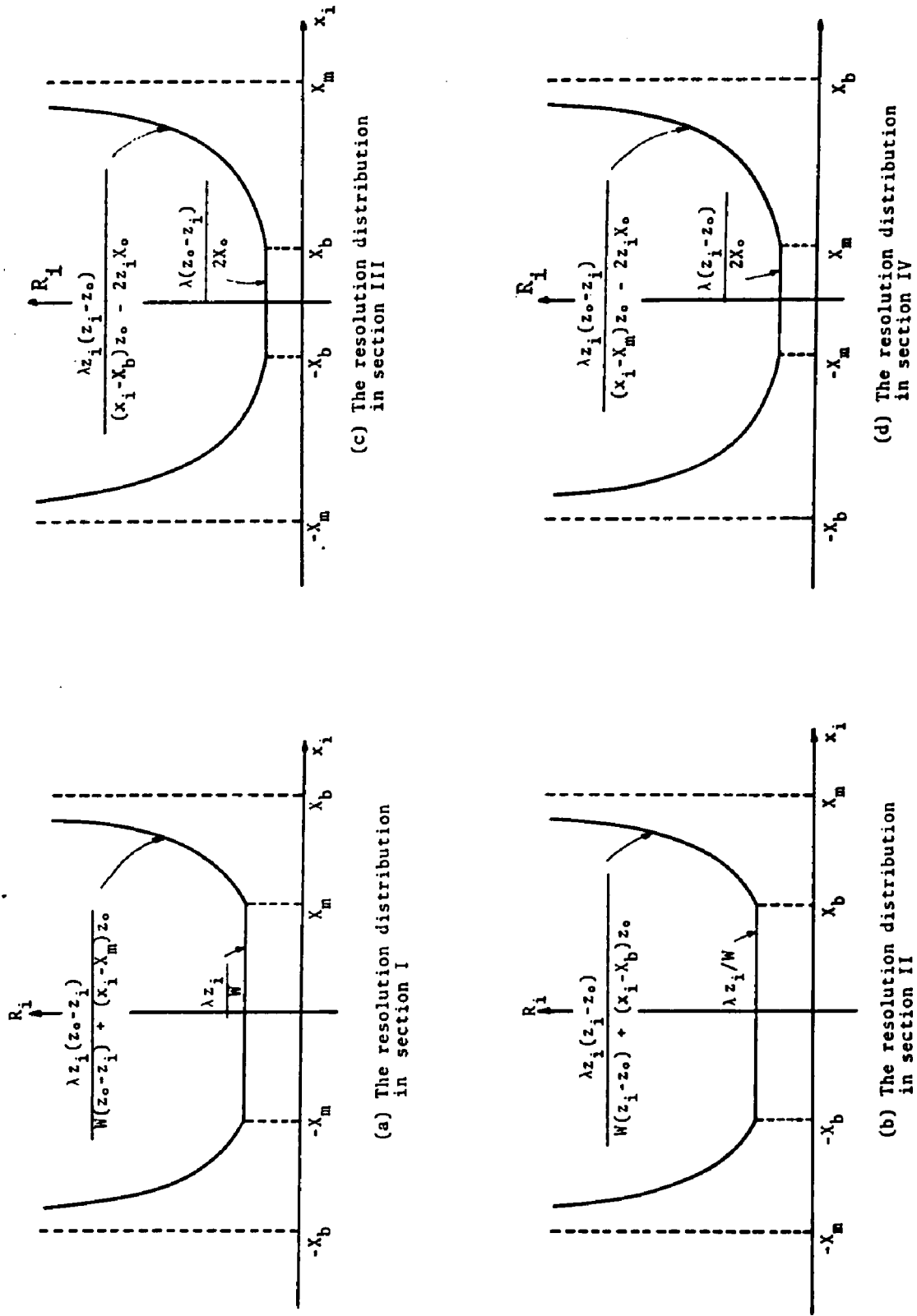


Figure 2-11. The resolution distribution in four different object space sections of a holographic set-up.

$$N'_x = \int_{X_m}^{X_b} \frac{dx_i}{\frac{\lambda z_i}{A_i}} = \int_{-(z_i - z_o)(\frac{W}{z_o})}^0 \frac{\frac{z_o x'_i}{\lambda z_i (z_o - z_i)}}{dx'_i} dx'_i$$

$$= \frac{(z_i - z_o)W^2}{2\lambda z_o z_i}$$

The total number of resolvable points in one dimension is thus

$$N_t = N_x + N'_x = \frac{2X W}{\lambda z_o} = 2F_x W = n \quad (2-7)$$

The total number of resolvable object points is the same as the number of resolvable recording elements in the hologram. If two dimensions are considered, the total number of resolvable points in two dimensions is given by n^2 which is equal to the total number of resolvable recording elements in the hologram. This fact proves a fundamental principle: as the wave propagates from one plane to another, the number of resolvable points in space are preserved. Only the resolution distribution changes. In each section, the total number of resolvable points is n^2 , even though the resolution distribution for each is different. The number of resolvable points in each plane region is given by

$$N(IA) = \left(\frac{2WX_m}{\lambda z_i} \right)^2$$

$$N(IB) = \frac{(z_i - z_o)W^3 X_m}{\lambda^2 z_o z_i^2}$$

$$N(IC) = \left[\frac{(z_i - z_o)W^2}{2\lambda z_o z_i} \right]^2$$

The total number of resolvable points in plane $z = z_i$ is

$$N_t = N(IA) + 4N(IB) + 4N(IC) = n^2$$

which verifies the results of Eq. (2-7). To discuss the number of resolvable elements in the volume of each region in section I we use the same notations (IA or IIB). To find the number of resolvable points in volume IA, we define the resolution in depth by

$$\Delta z = 4\lambda \left(\frac{z_i}{W} \right)^2$$

This is consistent with the definition of lateral resolution, for which the Rayleigh resolution criterion is used [8]. The number of resolvable points in volume IA is thus

$$\begin{aligned} N_v(IA) &= \int_{-X_m}^{X_m} \int_{-X_m}^{X_m} \int_{z_0}^{z_1} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{W} \cdot \frac{\lambda z_i}{W} \cdot 4\lambda \left(\frac{z_i}{W} \right)^2} \\ &= \int_{z_0}^{z_1} \frac{W^4}{4\lambda^3} \frac{(2X_m)^2}{z_i^4} dz_i \end{aligned} \quad (2-8)$$

With the help of the following relations

$$\frac{1}{z_0} - \frac{1}{z_1} = \frac{2X_0}{Wz_0}$$

$$\frac{1}{z_0^2} - \frac{1}{z_1^2} = \frac{4X_0(W - X_0)}{z_0^2 W^2}$$

$$\frac{1}{z_0^3} - \frac{1}{z_1^3} = 3 \left(\frac{1}{z_0} \right)^2 \left(\frac{2X_0}{Wz_0} \right) - 3 \left(\frac{1}{z_0} \right) \left(\frac{2X_0}{Wz_0} \right)^2 + \left(\frac{2X_0}{Wz_0} \right)^3 ,$$

Eq. (2-8) is found to be

$$N_v(IA) = \frac{1}{6} \cdot \frac{1}{2} \left(\frac{2X_0 W}{\lambda z_0} \right)^3 = \frac{1}{6} \cdot \frac{1}{2} \cdot n^3 .$$

Determination of the number of resolvable points in volume IB and IC poses special problems. Since the resolution elements in such regions are not a square, it is difficult to define the resolution in depth. If the resolution in depth is always measured in the x direction, $N_v(IB)$ is given by

$$N_v(IB) = \int_{X_m}^{X_b} \int_{-X_m}^{X_m} \int_{z_o}^{z_1} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{A_x} \cdot \frac{\lambda z_i}{W} \cdot 4\lambda \left(\frac{z_i}{A_x}\right)^2}$$

where $A_x = \frac{W}{2} + x'$ is given in Eq. (2-4).

The above integration can be carried out by changing x_i to $x'_i = x_i - X_b$, which yields

$$N_v(IB) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$$

In the same fashion, $N_v(IC)$ is given as

$$N_v(IC) = \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_{z_o}^{z_1} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{A_x} \cdot \frac{\lambda z_i}{A_y} \cdot 4\lambda \left(\frac{z_i}{A_x}\right)^2}$$

and is found to be

$$N_v(IC) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$$

The total number of resolvable elements in volume I is

$$N_v(I) = N_v(IA) + 4N_v(IB) + 4N_v(IC) = \frac{1}{6} n^3$$

A reasonable way to define the resolution in depth for rectangular resolution elements with length A_x and width A_y is to define the resolution in depth by the largest dimension of the resolution elements, i.e.,

$$\Delta z = 4\lambda \left(\frac{z_1}{A_x} \right)^2 \quad \text{when } A_x > A_y .$$

With this criterion, $N_v(\text{IB})$ and $N_v(\text{IC})$ are given as follows:

$$N_v(\text{IB}) = \int_{X_m}^{X_b} \int_{-X_m}^{X_m} \int_{z_0}^{z_1} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{A_x} \cdot \frac{\lambda z_i}{W} \cdot 4\lambda \left(\frac{z_i}{A_x} \right)^2} = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$$

$$N_v(\text{IC}) = 2 \int_{y_i}^{X_b} \int_{X_m}^{X_b} \int_{z_0}^{z_1} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{A_x} \cdot \frac{\lambda z_i}{A_y} \cdot 4\lambda \left(\frac{z_i}{A_x} \right)^2}$$

$$= \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{12} n^3$$

The $N_v(\text{IB})$ is the same as before, but $N_v(\text{IC})$ must be treated in two parts and the results are slightly different. In this case, $N_v(\text{I})$ is found to be $\frac{1}{6} \cdot \frac{11}{12} n^3$, which is close to $\frac{1}{6} n^3$. Another possible definition of resolution in depth is to define it as

$$4\lambda \left(\frac{z_1^2}{A_x A_y} \right)$$

With this definition

$$N_v(\text{IB}) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{6} n^3$$

$$N_v(\text{IC}) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{9} n^3$$

$$N_v(\text{I}) = \frac{1}{6} \cdot \frac{19}{18} n^3 \quad \text{which is also close to } \frac{1}{6} n^3.$$

We thus conclude that, on the average, the criterion of the resolution in depth in our first approach is reasonable and the total number of resolvable points in section I is $\frac{1}{6} n^3$.

The derivations of the resolution distribution and the information capacity in sections II, III, and IV are quite similar to those of section I. The basic mathematical results are summarized below.

The resolution distribution on a typical object plane for sections II, III, and IV are given in Fig. 2-11(b), (c), and (d), respectively. The shape of each resolution distribution is similar, although the mathematical expressions for each is different. The total number of resolvable points in any object plane $z = z_1$ is found to be n^2 in all four sections proving that the information content is preserved in wave propagation. The resolution distribution along the z axis in Fig. 2-12 can also be plotted. The finest resolution occurs at the point z_2 rather than at the hologram. The resolution at z_2 is given by $\lambda z_2/W$.

The number of resolvable points in each volume region is also calculated and the results are listed in Table 1. The results in section IV are not the same as in the three other sections primarily because the infinite limit in the z_1 integral prevents some of the terms from subtracting out. Since $N_V(\text{IV})$ is smaller than $n^3/6$ if $X_0 > 0$, the total number of resolvable points in all four sections is between $n^3/2$ and $2n^3/3$. This value is more than three times the value of the volume region IA + IIA, which is the H-limited object space defined in Macovski's paper [3].

Thus far, only the case where $2X_0 < W$ has been considered. If a high-resolution film is used, then F_x is large and $2X_0 > W$. In this case, sections I and IV become one section (V), while sections II and III remain separate. The number of resolvable points in each new region is defined and calculated in Table 1. As Table 1 shows, $N_V(\text{V})$ has its minimum value $n^3/6$ when $n = W^2/\lambda z_0$ or $2X_0 = W$. Moreover, no matter how large n is, $N_V(\text{V}) < n^3/3$. Hence, the total number of resolvable points in all object space is between $n^3/2$ to $2n^3/3$, exactly the same range as for the previous case.

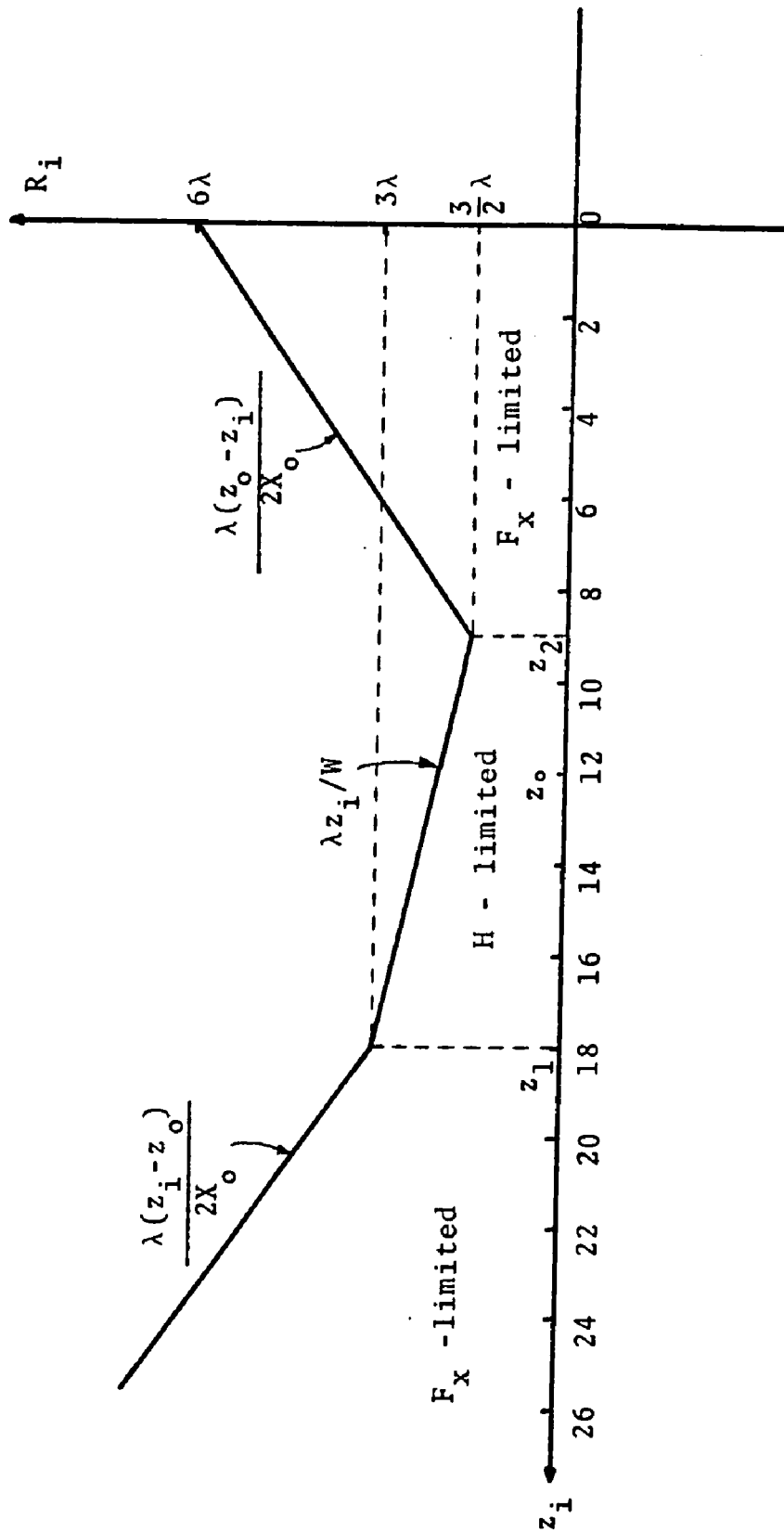


Figure 2-12. Resolution distribution along the central axis of a holographic system.
The system parameters are: $W = 6$, $2X_0 = 2$, $z_0 = 12$, $z_1 = 18$, $z_2 = 9$.

TABLE 1

NUMBER OF RESOLVABLE OBJECT POINTS IN DIFFERENT SPACE REGIONS
FOR THE OBJECT SPACE USING AN ON-AXIS POINT-SOURCE REFERENCE.

space section	regions	definition of the number of resolvable points in volume space regions	results for each space region	total number of resolvable points in each section
I	A	$N_V(IA) \triangleq \int_{-X_m}^{X_m} \int_{-X_m}^{X_m} \int_{z_0}^{z_1} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{W})^2}$	$\frac{1}{6} \cdot \frac{1}{2} n^3$	$N_V(I)$ = $N_V(IA)$ + $4N_V(IB)$ + $4N_V(IC)$ = $\frac{1}{6} n^3$
	B	$N_V(IB) \triangleq \int_{X_m}^{X_b} \int_{-X_m}^{X_m} \int_{z_0}^{z_1} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$	
	C	$N_V(IC) \triangleq \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_{z_0}^{z_1} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$	
II	A	$N_V(IIA) \triangleq \int_{-X_b}^{X_b} \int_{-X_b}^{X_b} \int_{z_2}^{z_0} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{W})^2}$	$\frac{1}{6} \cdot \frac{1}{2} n^3$	$N_V(II)$ = $N_V(IIA)$ + $4N_V(IIB)$ + $4N_V(IIC)$ = $\frac{1}{6} n^3$
	B	$N_V(IIIB) \triangleq \int_{X_b}^{X_m} \int_{-X_b}^{X_b} \int_{z_2}^{z_0} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$	
	C	$N_V(IIC) \triangleq \int_{X_b}^{X_m} \int_{X_b}^{X_m} \int_{z_2}^{z_0} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$	
III	A	$N_V(IIIA) \triangleq \int_{-X_b}^{X_b} \int_{-X_b}^{X_b} \int_0^{z_2} \frac{dx_1 dy_1 dz_1}{\frac{\lambda(z_0 - z_1)}{2X_0} \cdot \frac{\lambda(z_0 - z_1)}{2X_0} \cdot 4\lambda (\frac{z_0 - z_1}{2X_0})^2}$ $= \frac{1}{6} \cdot \frac{1}{2} n^3$		$N_V(III)$ = $N_V(IIIA)$ + $4N_V(IIIB)$ + $4N_V(IIC)$ = $\frac{1}{6} n^3$
	B	$N_V(IIIB) \triangleq \int_{X_b}^{X_m} \int_{-X_b}^{X_b} \int_0^{z_2} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda(z_0 - z_1)}{2X_0} \cdot 4\lambda (\frac{z_1}{A_x})^2} = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$		
	C	$N_V(IIC) \triangleq \int_{X_b}^{X_m} \int_{X_b}^{X_m} \int_0^{z_2} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{A_x})^2} = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} n^3$		

TABLE 1 (continued)

IV	A $N_V(IVA) \triangleq \int_{-X_m}^{X_m} \int_{-X_m}^{X_m} \int_{z_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda(z_1 - z_0)}{2X_0} \cdot \frac{\lambda(z_1 - z_0)}{2X_0} \cdot 4\lambda(\frac{z_1 - z_0}{2X_0})^2}$ $= \frac{1}{6} \cdot \frac{1}{2} \cdot [\frac{(W - 2X_0)2X_0}{\lambda z_0}]^3$ B $N_V(IVB) \triangleq \int_{X_m}^{X_b} \int_{-X_m}^{X_m} \int_{z_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda(z_1 - z_0)}{2X_0} \cdot 4\lambda(\frac{z_1}{A_x})^2}$ $= \frac{1}{16} \cdot \frac{1}{2} (\frac{2X_0}{z_0})^3 (W - 2X_0)^2 W - \frac{1}{6} \cdot \frac{1}{8} [\frac{(W - 2X_0)2X_0}{\lambda z_0}]^3$ C $N_V(IVC) \triangleq \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_{z_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda(\frac{z_1}{A_x})^2}$ $= \frac{1}{16} \cdot \frac{1}{2} (\frac{2X_0}{\lambda z_0})^3 (2X_0 W) (W - 2X_0) + \frac{1}{6} \cdot \frac{1}{16} [\frac{(W - 2X_0)2X_0}{\lambda z_0}]^3$	$N_V(IV)$ $= N_V(IVA)$ $+ 4N_V(IVB)$ $+ 4N_V(IVC)$ $= \frac{1}{24} [n - \frac{(2X_0)^2}{\lambda z_0}]^3$ $+ \frac{1}{8} n^2 [n - \frac{(2X_0)^2}{\lambda z_0}]$
V	A $N_V(VA) \triangleq \int_{-X_m}^{X_m} \int_{-X_m}^{X_m} \int_{z_0}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda(\frac{z_1}{W})^2}$ $= \frac{1}{12} (\frac{W^2}{\lambda z_0} - n)^3 + \frac{1}{12} n^3$ B $N_V(VB) \triangleq \int_{X_m}^{X_b} \int_{-X_m}^{X_b} \int_{z_0}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda(\frac{z_1}{A_x})^2}$ $= \frac{W^5 X_0}{16(\lambda z_0)^3} - \frac{1}{6} \cdot \frac{1}{8} (\frac{W^2}{\lambda z_0})^3$ C $N_V(VC) \triangleq \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_{z_0}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda(\frac{z_1}{A_x})^2}$ $= \frac{1}{6} \cdot \frac{1}{16} (\frac{W^2}{\lambda z_0})^3$	$N_V(V)$ $= N_V(VA)$ $+ 4N_V(VB)$ $+ 4N_V(VC)$ $= \frac{1}{24} \{ (\frac{W^2}{\lambda z_0}) \cdot [(\frac{W^2}{\lambda z_0})^2 + 3n^3] + 3n(n - \frac{W^2}{\lambda z_0}) \cdot (\frac{W^2}{\lambda z_0}) \}$

2.3.3. THE GENERALIZATION OF THE SYSTEM MODEL AND THE USE OF THE OFFSET POINT-SOURCE REFERENCE BEAM

In the previous section, we have analyzed the resolution distribution and calculated the information capacity in different object space regions of a hologram using an on-axis point-source reference. The analysis can be extended to a more general case in which the point-source reference is not in the center of a hologram. In practice, an offset reference is usually required to separate the desired image term from its undesirable $|U|^2$ term. The results obtained in the on-axis case can be applied equally well to the more general case, in which the reference is offset from the central axis. In Fig. 2-13, a point-source reference is located a distance s from the central axis. The field disturbance at the hologram plane thus becomes

$$R = R_o \exp \left[j \frac{k}{2z_o} (2xs + x^2 + y^2) \right]$$

Following the same derivation we used before, we get

$$f_x = \frac{1}{\lambda z_i} \left[\left(x_i + \frac{z_i}{z_o} s \right) + \left(\frac{z_i - z_o}{z_o} \right) x \right] \quad (2-9)$$

$$f_y = \frac{1}{\lambda z_i} \left[y_i + \frac{(z_i - z_o)}{z_o} y \right]$$

Considering the x direction only and setting $z_i = z_o$ in Eq. (2-9), we have

$$f_x = \frac{1}{\lambda z_i} \left(x_i + \frac{z_i}{z_o} s \right) \quad (2-10)$$

If $f_x = \pm F_x$ and $F_x = \frac{X_o}{2\lambda z_o}$ when $s = 0$, we solve Eq. (2-10) for x_i :

$$x_i = -s \pm \frac{X_o}{2} \quad (2-11)$$

Equation (2-11) indicates that the object extent in the plane $z = z_o$ is $2X_o$ wide but is lowered a distance s away from the central axis. The corresponding curve for X_m and X_b in Fig. 2-13 can be obtained from the following two equations by setting $x = \pm W/2$:

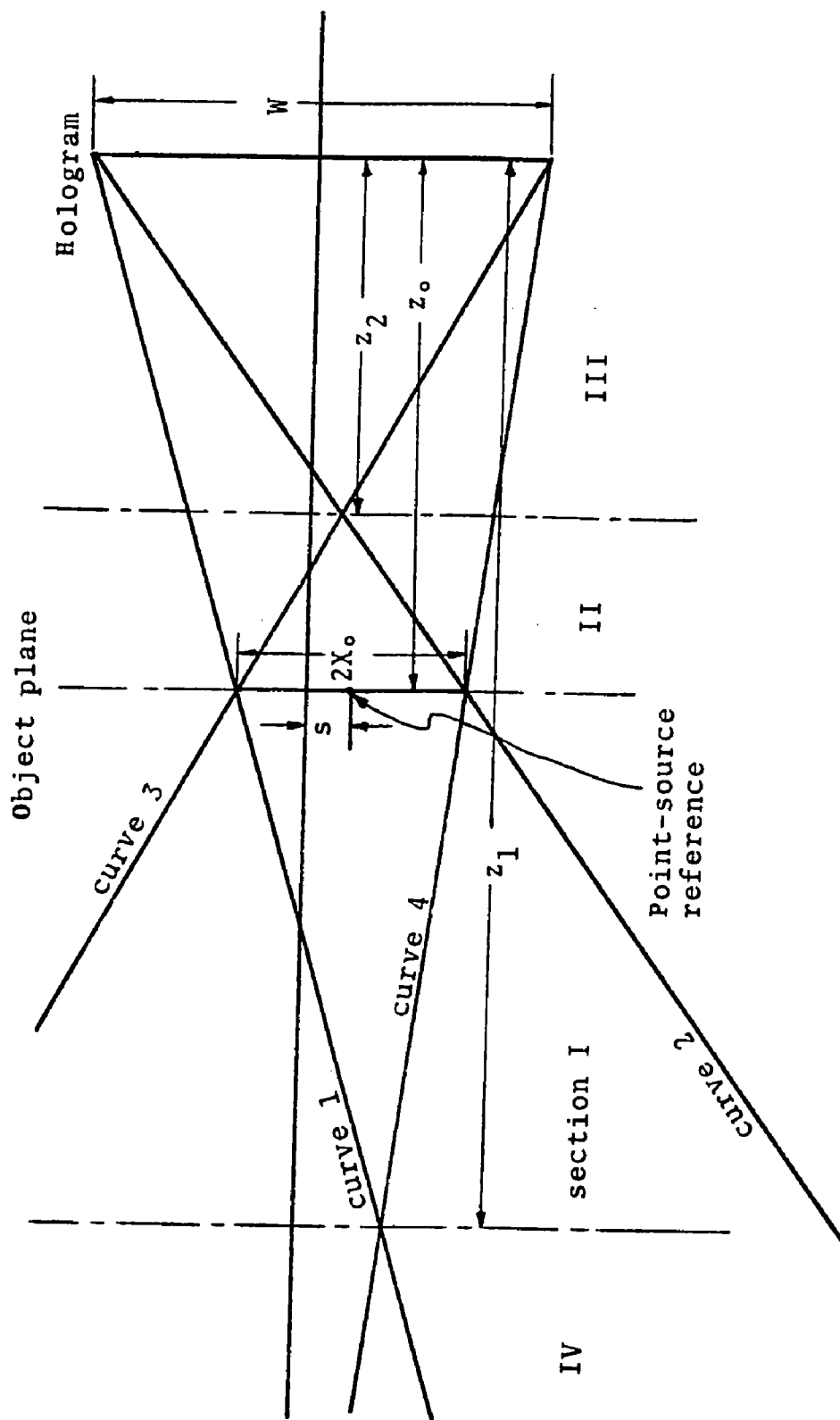


Figure 2-13. The object space sections of a hologram using an off-set point-source reference.

$$x_1 = -(z_1 - z_0) \frac{x}{z_0} + \frac{z_1}{z_0} X_0 - \frac{z_1}{z_0} s \quad (2-12)$$

$$x_1 = -(z_1 - z_0) \frac{x}{z_0} - \frac{z_1}{z_0} X_0 - \frac{z_1}{z_0} s \quad (2-13)$$

Equations (2-12) and (2-13) are obtained from Eq. (2-9) by setting

$$f_x = F_x = \frac{z_0}{2\lambda z_0}$$

A comparison of Eq. (2-12) to Eq. (2-1) reveals that the only difference between an offset reference and an on-axis reference is a $z_1 s/z_0$ shift in coordinates x_1 . At a given plane $z = z_1$, all the coordinates x_1 are shifted by a constant $z_1 s/z_0$. Hence, the resolution distribution at any given plane $z = z_1$ is the same in both on-axis and off-axis cases. The number of resolvable points in both two-dimensional planes and volume regions should also remain the same. The location and the configuration of the space region is changed slightly according to the curves shown in Fig. 2-13. The curves in Fig. 2-13 can be obtained with the following equations for X_m and X_b when $x = \pm W/2$ in Eq. (2-12):

$$\text{Curve 1. } X_m = -(z_1 - z_0) \frac{W}{2z_0} + \frac{z_1}{z_0} (X_0 - s)$$

$$\text{Curve 2. } X_b = -(z_1 - z_0) \frac{W}{2z_0} - \frac{z_1}{z_0} (X_0 - s)$$

$$\text{Curve 3. } X_b = (z_1 + z_0) \frac{W}{2z_0} + \frac{z_1}{z_0} (X_0 - s)$$

$$\text{Curve 4. } X_m = (z_1 + z_0) \frac{W}{2z_0} - \frac{z_1}{z_0} (X_0 - s)$$

In practice, however, these curves can be obtained by drawing lines between four corresponding points $(\pm W/2, 0; \pm X_0, z_0)$, as shown in Fig. 2-13. Once the curves are drawn in the space, the resolution distribution and the information capacity in different space regions can be readily understood.

We now turn our attention to the problem of separating an object space from its $|U|^2$ term. Separating a plane object in a plane $z = z_0$ from its $|U|^2$ term is different than separating a solid object from its $|U|^2$ term. Consider a plane object with width L in the plane $z = z_0$ in Fig. 2-14. To separate this object from its $|U|^2$ term with the use of a point-source reference at the same plane $z = z_0$, the point-source reference is put at a distance $3L/2$ from the center of the object. However, for a solid object, such as the one shown by the dotted line in Fig. 2-14, the maximum frequency f_x formed by this object depends on the shape and the orientation of the object space. Moreover, for any reasonable object space, the maximum spatial frequency usually occurs at the edge of the hologram. Assuming that the dotted object in Fig. 2-14 is rectangular and is symmetrically centered on the central axis of the hologram, the maximum spatial frequency of the object space is contributed by the two points at the corner of the object and subtends a maximum angle at the edge of the hologram. If a point-source reference is used to record the object and the reconstructed image is to be separated from its $|U|^2$ term, the reference source has to be at the plane $z = z_0$ and $3L'/2$ distance away from a new plane object L' , which is larger than L (see Fig. 2-14). The object space, reconstructed in this way has a resolution limited only by the hologram aperture. In fact, the whole region bounded by the four points a, b, c, d, in Fig. 2-14 can be reconstructed and separated from its $|U|^2$ terms with a resolution limited only by the hologram aperture. Moreover, the object points outside this region can also be reconstructed separately from their $|U|^2$ terms although poorer resolution results. In Fig. 2-14, $z_i > z_0$, any point above curve A can be separated from its $|U|^2$ term; when $z_i < z_0$, all the points above curve B can be separated from their $|U|^2$ term.

When we say those points in regions outside a, b, c, d, and above curve A for $z_i > z_0$ and curve B for $z_i < z_0$ can be separated from their $|U|^2$ term with poorer resolution, we don't mean that we can practically put the object points in such regions in the recording process. In the recording process, an object point in these outside regions has to have a poorer resolution to begin with. The object points between curve A and B for either $z_i > z_0$ or $z_i < z_0$ cannot be separated from their $|U|^2$

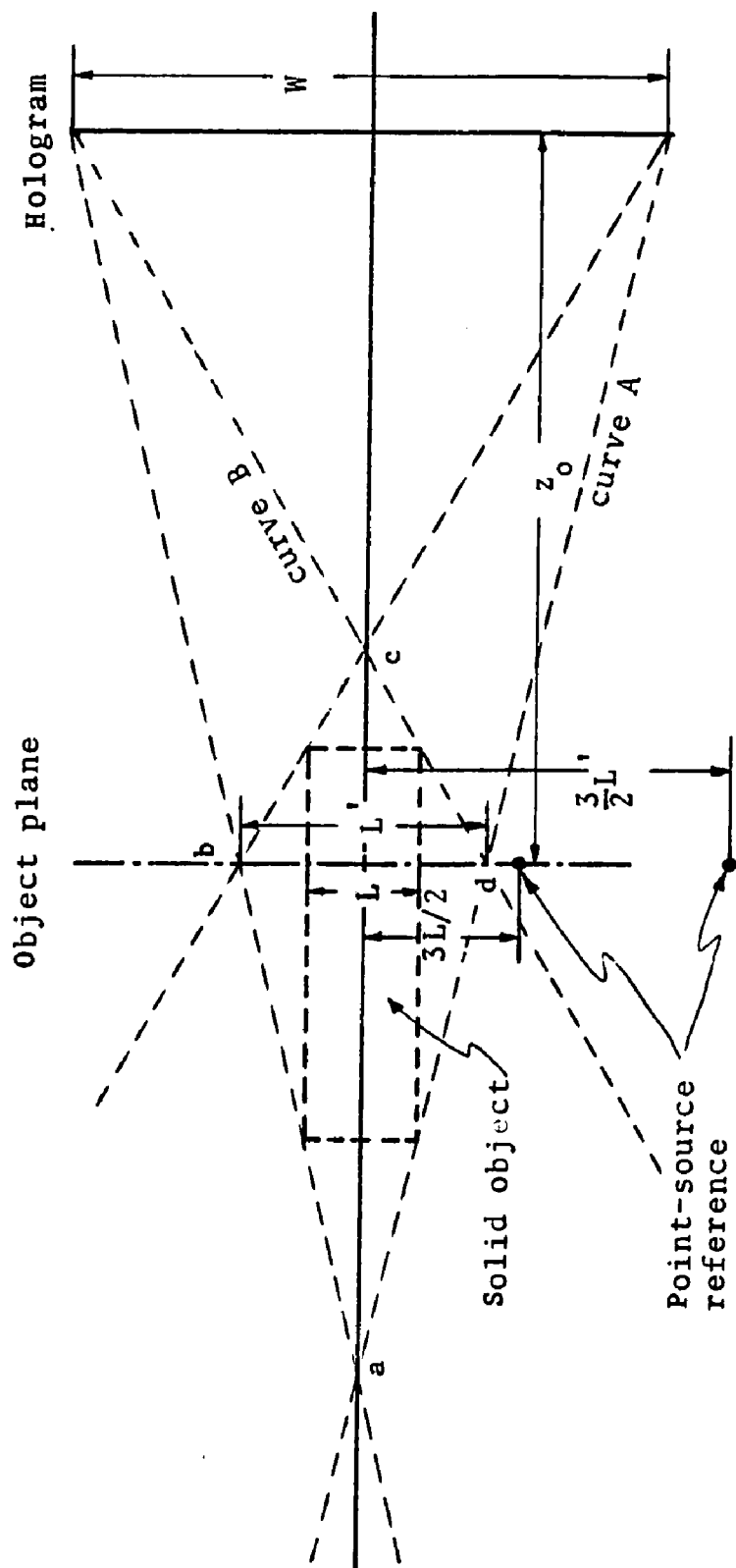


Figure 2-14. Separation of an object space from its $|u|^2$ term by using an off-set point-source reference

terms (even though we can tolerate poorer resolution for these points) because the rays of the $|U|^2$ term propagate through the aperture $2L'$ centered at the reference point and will interfere with the object point in this region. Clearly, for given holographic recording data in terms of F_x , W , and z_o , when a point-source reference beam is used, there is an object space which can be separated from its $|U|^2$ term. This object space is the shaded area in Fig. 2-15 which is a duplication of Fig. 2-14, except the reference point source is at the central axis of the hologram.

Now, the total number of resolvable points in such an object space, as well as the number of resolvable points in each section region, will be found. The object aperture L' is now equal to $X_o/2$ centered at $3X_o/4$ in Fig. 2-15. Curves A and B are also shown in Fig. 2-15. As we have just pointed out, only the shaded space region can be separated from its $|U|^2$ term after reconstruction in the set-up of Fig. 2-15. Our goal here is to calculate the number of resolvable points in each space region in the shaded area. Since the offset is only in the x direction, the space region in the y direction is still the same as the on-axis case. Because of the existence of two additional curves labeled A and B, the space region for the shaded areas in x direction has to be subdivided into six different regions for the calculation. These space regions are denoted by I', II', III', IV', V', and VI', as shown in Fig. 2-15. The equations for curves A and B are given by

$$\text{Curve A: } X'_b = (z_1 - z_o) \left(\frac{W}{2z_o} \right) + \frac{z_1}{z_o} \cdot \frac{X_o}{2}$$

$$\text{Curve B: } X'_m = -(z_1 - z_o) \left(\frac{W}{2z_o} \right) + \frac{z_1}{z_o} \cdot \frac{X_o}{2}$$

The curves for X_m and X_b are still defined as before. At a typical plane $z = z_1$ in section I' and II', the six divisions of the plane in the shaded area have been made as indicated in Fig. 2-16(a) and four of them have to be treated separately. These are denoted by B', C', D', E' in Fig. 2-16(a). In sections III', IV', V', VI', the space has been divided into three regions as shown in Fig. 2-16(b); two of them, denoted by B' and C', have to be treated separately.

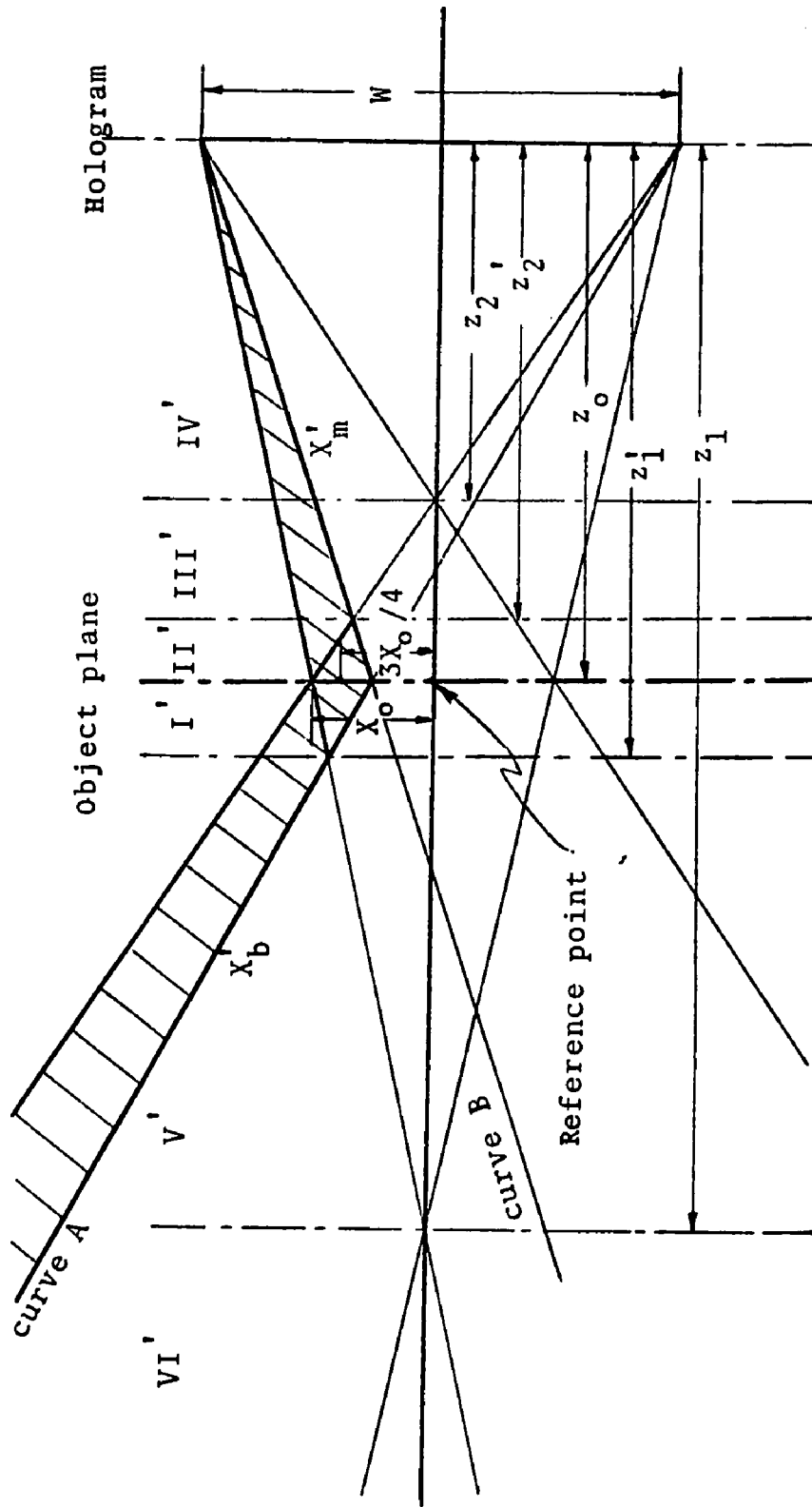
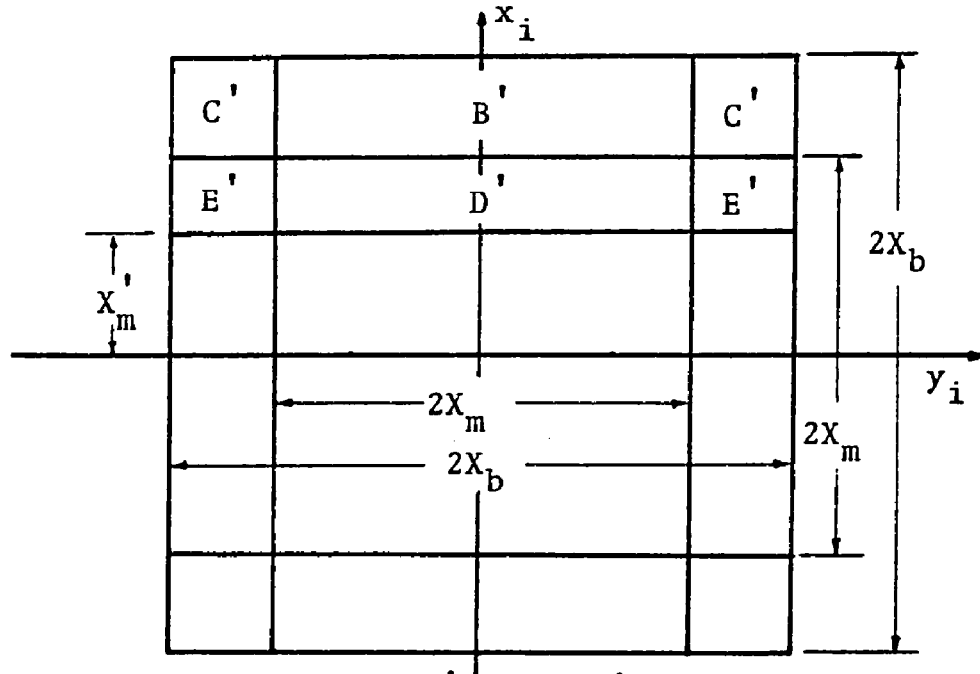
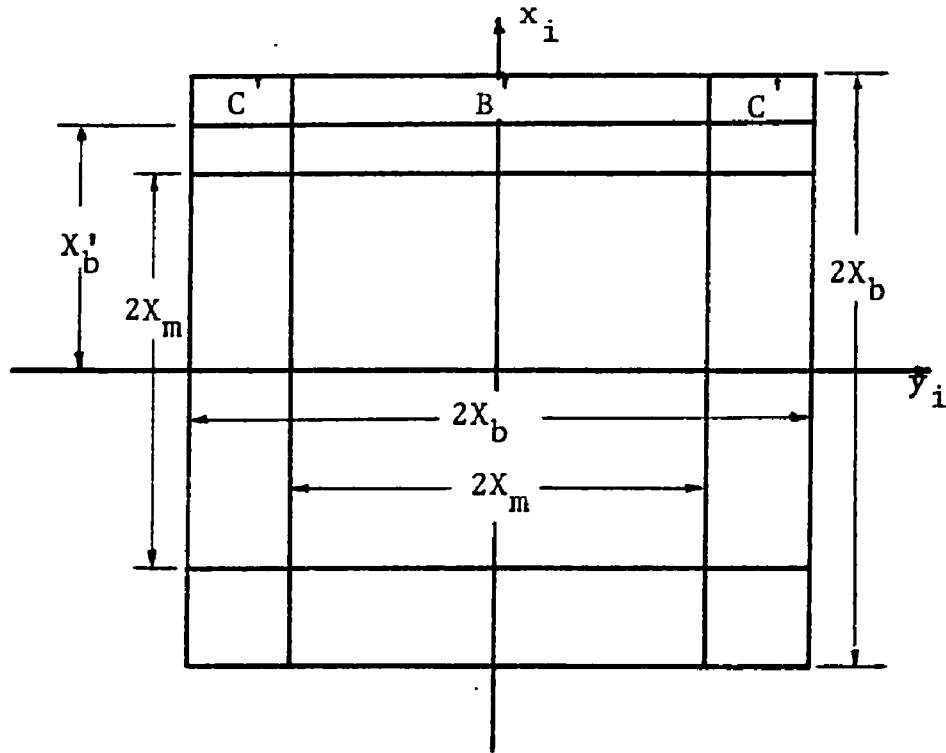


Figure 2-15. The object space sections for the study of separating an object space from its $|U|^2$ term using an off-set point-source reference



(a) for space sections I' and II'



(b) for space sections III', IV', V', IV'.

Figure 2-16. Subdivision of the object space regions at a typical plane $z = z_i$ for the case of Fig. 2-15

The detailed treatment is not given here but the results are listed in Table 2, where the integral expression for each space volume is defined. A comparison of the results is also given in Table 3. In Table 3, for the case $2X_0 < W$, the number of resolvable points decreases very rapidly from region I' or II' to III' or V' (about 5 times) and from III' or V' to IV' or VI' (about 16 times). In the space region VI', the integral is from z_1 to infinity, so the result does not end with a simple expression in terms of n^3 . The result is given by

$$N_V(VI') = \frac{1}{16^3} \cdot \frac{1}{2} n \left\{ n^2 - \left[\frac{(2X_0)^2}{\lambda z_0} \right]^2 \right\}$$

which is also shown in both Tables 2 and 3.

When $2X_0 > W$, the value of $N_V(VI')$ is negative which indicates that the space region VI' does not exist when $2X_0 > W$. In fact, when $2X_0 > W$, space sections V' and VI' become one space section. The number of resolvable points in this case can also be found and the results are shown in Table 2 as section VII' $\triangleq$ V' + VI'. The total number of resolvable points in section VII' is

$$N_V(VII') = \frac{1}{2} \cdot \frac{1}{16} n \left[n^2 - \left(\frac{X_0^2}{\lambda z_0} \right)^2 \right] \quad (2-14)$$

This equation gives the same value as $N_V(V')$ when $2X_0 = W$, as it should, since in this case, $N_V(VI') = 0$. Moreover, from Eq. (2-14), $N_V(VII')$ will be a negative value if $X_0/2 > W$. This is the case when section VII' and section I' become one (section VIII'). The number of resolvable points in section VIII' is also listed in Table 2 and the total number of resolvable points in section VIII' is given by

$$N_V(VIII') = \frac{1}{16} \cdot \frac{1}{2} n \left(\frac{W^2}{\lambda z_0} \right) \left[2n - 3 \left(\frac{W^2}{\lambda z_0} \right) \right] \quad (2-15)$$

To check the validity of this expression, we let $X_0 = 2W$ in Eq. (2-15) and verify that it is equal to

$$N_V(I') = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{8} \cdot \frac{30}{32} n^3$$

TABLE 2
NUMBER OF RESOLUTION OBJECT POINTS IN DIFFERENT SPACE REGIONS
FOR THE OBJECT SPACE SEPARATING FROM ITS $|U|^2$ TERM BY
AN OFF-SET POINT-SOURCE REFERENCE

space section	regions	definition of the number of resolvable points in volume space regions	results for each space region	total number of resolvable points in each section
I'	B'	$N_V(I'B') \triangleq \int_{X_m}^{X_b} \int_{-X_m}^{X_m} \int_{z_0}^{z_1'} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{5}{32} n^3$	$N_V(I')$ $= N_V(I'B')$ $+ 2N_V(I'E')$ $+ 2N_V(I'C')$ $+ N_V(I'D')$ $= \frac{1}{6} \cdot \frac{1}{16} \cdot \frac{30}{32} n^3$ $+ \frac{1}{6} \cdot \frac{1}{16} n^3$
	C'	$N_V(I'C') \triangleq \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_{z_0}^{z_1'} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{1}{64} n^3$	
	D'	$N_V(I'D') \triangleq \int_{X_b'}^{X_m} \int_{-X_m}^{X_m} \int_{z_0}^{z_1'} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{W})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{11}{16} n^3$	
	E'	$N_V(I'E') \triangleq \int_{X_b'}^{X_m} \int_{X_m}^{X_b} \int_{z_0}^{z_1'} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{W})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{1}{32} n^3$	
II'	B'	$N_V(II'B') \triangleq \int_{X_b}^{X_m} \int_{-X_b}^{X_b} \int_{z_2'}^{z_0} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{5}{32} n^3$	$N_V(II')$ $= N_V(II'B')$ $+ 2N_V(II'C')$ $+ N_V(II'D')$ $+ 2N_V(II'E')$ $= \frac{1}{6} \cdot \frac{1}{16} \cdot \frac{30}{32} n^3$ $+ \frac{1}{6} \cdot \frac{1}{16} n^3$
	C'	$N_V(II'C') \triangleq \int_{X_b}^{X_m} \int_{X_b}^{X_m} \int_{z_2'}^{z_0} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{1}{64} n^3$	
	D'	$N_V(II'D') \triangleq \int_{X_m'}^{X_b} \int_{-X_b}^{X_b} \int_{z_2'}^{z_0} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{W})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{11}{16} n^3$	
	E'	$N_V(II'E') \triangleq \int_{X_m'}^{X_b} \int_{X_b}^{X_m} \int_{z_2'}^{z_0} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{W})^2}$	$\frac{1}{6} \cdot \frac{1}{16} \cdot \frac{1}{32} n^3$	
III'	B'	$N_V(III'B') \triangleq \int_{X_m'}^{X_m} \int_{-X_b}^{X_b} \int_{z_2}^{z_2'} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{2} \cdot \frac{9}{16^3} n^3$	$N_V(III')$ $= N_V(III'B')$ $+ 2N_V(III'C')$ $= \frac{1}{2} \cdot \frac{15}{16^3} n^3$
	C'	$N_V(III'C') \triangleq \int_{X_m'}^{X_m} \int_{X_b}^{X_m} \int_{z_2}^{z_2'} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda (\frac{z_1}{A_x})^2}$	$\frac{1}{2} \cdot \frac{3}{16^3} n^3$	

TABLE 2 (continued)

TABLE 2 (continued)

IV'	B'	$N_V(IV'B') \triangleq \int_{X'_m}^{X_m} \int_{-X_b}^{X_b} \int_0^{z_2} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda(z_0 - z_1)}{2X_0} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{6} \cdot \frac{1}{16^3} n^3$	$N_V(IV')$ $= N_V(IV'B')$ $+ 2N_V(IV'C')$ $= \frac{1}{2} \cdot \frac{1}{16^3} n^3$
	C'	$N_V(IV'C') \triangleq \int_{X'_m}^{X_m} \int_{X_b}^{X_m} \int_0^{z_2} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{6} \cdot \frac{1}{16^3} n^3$	
V'	B'	$N_V(V'B') \triangleq \int_{X'_b}^{X_b} \int_{-X_m}^{X_m} \int_{z'_1}^{z_1} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{W} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{2} \cdot \frac{9}{16^3} n^3$	$N_V(V')$ $= N_V(V'B')$ $+ 2N_V(V'C')$ $= \frac{1}{2} \cdot \frac{15}{16^3} n^3$
	C'	$N_V(V'C') \triangleq \int_{X'_b}^{X_b} \int_{X_m}^{X_b} \int_{z'_1}^{z_1} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{2} \cdot \frac{3}{16^3} n^3$	
VI'	B'	$N_V(VI'B') = \int_{X'_b}^{X_b} \int_{-X_m}^{X_m} \int_{z_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda(z_1 - z_0)}{2X_0} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$ $= \frac{1}{16^3} \left\{ \frac{1}{16} \cdot \frac{1}{2} \left(\frac{2X_0}{\lambda z_0} \right)^3 W(W - 2X_0)^2 - \frac{1}{6} \cdot \frac{1}{8} \left[\frac{(W - 2X_0) 2X_0}{\lambda z_0} \right]^3 \right\}$		$N_V(VI')$ $= N_V(VI'B')$ $+ 2N_V(VI'C')$ $= \frac{1}{16^3} \cdot \frac{1}{2} n^2 \left\{ \right.$ $\left. - \left[\left(\frac{2X_0}{\lambda z_0} \right)^2 \right]^2 \right\}$
	C'	$N_V(VI'C') \triangleq \int_{X'_b}^{X_b} \int_{X_m}^{X_b} \int_{z_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$ $= \frac{1}{16^2} \left\{ \frac{(2X_0)^4 W(W - 2X_0)}{16 \cdot 2 \cdot \lambda^3 z_0^3} + \frac{1}{6} \cdot \frac{1}{16} \left[\frac{(W - 2X_0) 2X_0}{\lambda z_0} \right]^3 \right\}$		
VII'	B'	$N_V(VII'B') \triangleq \int_{X'_b}^{X_b} \int_{-X_m}^{X_m} \int_{z'_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{W} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$ $= \frac{1}{16^2} \left(\frac{X_0}{\lambda z_0} \right)^3 W(2W - X_0)(X_0 + W)$		$N_V(VII')$ $= N_V(VII'B')$ $+ 2N_V(VII'C')$ $= \frac{1}{2} \cdot \frac{1}{16^2} n^2 \left\{ \right.$ $\left. - \left[\left(\frac{X_0}{\lambda z_0} \right)^2 \right]^2 \right\}$
	C'	$N_V(VII'C') \triangleq \int_{X'_b}^{X_b} \int_{X_m}^{X_b} \int_{z'_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$ $= \frac{1}{16^2} \left(\frac{X_0}{\lambda z_0} \right)^3 \frac{W^2(2W - X_0)}{2}$		

TABLE 2 (continued)

VIII'	B' $N_V(VIII'B')$	$\Delta \equiv \int_{X_m}^{X_b} \int_{-X_m}^{X_m} \int_{z_0}^{\infty} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{A})^2} = \frac{W^5(3X_0 - W)}{16(\lambda z_0)^3 \cdot 3}$	$N_V(VIII') \\ = N_V(VIII'B') \\ + 2N_V(VIII'C') \\ + N_V(VIII'D') \\ + 2N_V(VIII'E') \\ = \frac{1}{16} \cdot \frac{1}{2} n \left(\frac{W}{\lambda z_0} \right)^2 \\ [2n - 3 \left(\frac{W}{\lambda z_0} \right)^2]$
	C' $N_V(VIII'C')$	$\Delta \equiv \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_{z_0}^{\infty} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{A} \cdot 4\lambda (\frac{z_1}{A})^2} = \frac{1}{6} \cdot \frac{1}{16} \cdot \frac{W^6}{(\lambda z_0)^3}$	
	D' $N_V(VIII'D')$	$\Delta \equiv \int_{X_m}^{X_m} \int_{X_b'}^{X_m} \int_{-X_m}^{X_m} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{W})^2}$	
		$= \frac{1}{4 \cdot 12} \left(\frac{W^4}{3\lambda z_0} \right) (4W^2 - 15X_0W + 12X_0^2)$	
	E' $N_V(VIII'E')$	$\Delta \equiv \int_{X_b'}^{X_m} \int_{X_m}^{X_b} \int_{z_0}^{\infty} \frac{dx_1 dy_1 dz_1}{\lambda z_1 \cdot \frac{\lambda z_1}{W} \cdot 4\lambda (\frac{z_1}{W})^2}$	
		$= \frac{W^5}{8 \cdot 12(z_0\lambda)^3} (3X_0 - 4W)$	

TABLE 3
NUMBER OF RESOLVABLE POINTS IN EACH SPACE REGION FOR
THREE DIFFERENT CASES

Cases	Space Regions						
	VI'	V'	I'	II'	III'	IV'	
$2X_o < W$	$\frac{1}{16^3} \cdot \frac{1}{2} \cdot n \left\{ n^2 - \left[\frac{(2X_o)^2}{\lambda z_o} \right]^2 \right\}$	$\frac{1}{16^3} \cdot \frac{15}{2} n^3$	$\frac{1}{16^2} \cdot \frac{5}{2} n^3$	$\frac{1}{16^2} \cdot \frac{5}{2} n^3$	$\frac{1}{16^3} \cdot \frac{15}{2} n^3$	$\frac{1}{16^3} \cdot \frac{1}{2} n^3$	
	VII'		I'	II'	III'	IV'	
$4W > 2X_o > W$	$\frac{1}{2} \cdot \frac{1}{16} n \left[n^2 - \left(\frac{X_o^2}{\lambda z_o} \right)^2 \right]$		"	"	"	"	
	VIII'			II'	III'	IV'	
$X_o > 2W$ or $2X_o > 4W$	$\frac{1}{16} \cdot \frac{1}{2} n \left(\frac{W^2}{\lambda z_o} \right) \left[2n - 3 \left(\frac{W^2}{\lambda z_o} \right) \right]$			"	"	"	

The results for the number of resolvable points in each section for three different cases are listed in Table 3. In no case, however, is the number of resolvable points in this offset case larger than $(1/16^2) \cdot (19n^3/3)$. If this case is compared with an on-axis case which has between $n^3/2$ and $2n^3/3$, the total number of resolvable points in space for the latter is about 20 times greater. With the number of resolvable points defined in each space region and the results obtained in Table 2, we can compare various regions in various cases.

2.3.4. THE DISTRIBUTION OF THE IMAGE RESOLUTION AND INFORMATION CAPACITY OF A HOLOGRAM USING A NORMAL PLANE-WAVE REFERENCE

When a plane wave is used as reference, the analysis can be carried out in about the same way as for a point source, except that

$$R = R_0 e^{-jk\beta x}$$

where β is related to the incident angle θ by $\beta = \sin \theta$. The intensity at the hologram plane caused by R and a general object point is

$$I = R_0 + U_0 + 2R_0 U_0 \cos \delta \text{ where}$$

$$\delta = \frac{\pi}{\lambda z_i} [(x - x_i)^2 + (y - y_i)^2 + 2\beta z_i x]$$

The instantaneous spatial frequency is given by

$$f_x = (2\pi)^{-1} \frac{\partial \delta}{\partial x} = \frac{1}{\lambda z_i} (x - x_i + \beta z_i) \quad (2-16)$$

$$f_y = (2\pi)^{-1} \frac{\partial \delta}{\partial y} = \frac{1}{\lambda z_i} (y - y_i)$$

If we first consider the case in which the plane-wave reference is normally incident on the hologram, we then have a symmetrical configuration in both the x and y directions. The structure is plotted in Fig. 2-17 and can be compared with the point-source case shown in Fig. 2-9. If the f_x in Eq. (2-16) is set equal to the spatial cut-off frequency of the hologram F_x and if $\beta = 0$, the relation between x_i and x is found to be

$$x_i = f_x \lambda z_i - x = F_x \lambda z_i - x = \frac{n \lambda z_i}{4|x|} - x \quad (2-17)$$

X_m and X_b can be found from Eq. (2-17)

$$X_m = \frac{n \lambda z_i}{2W} - \frac{W}{2}$$

$$X_b = \frac{n \lambda z_i}{2W} + \frac{W}{2}$$

From Fig. 2-17, it can be seen that we have two sections I and II divided by plane $z = z_1$; z_1 is determined by F_x and is given by

$$z_1 \div \frac{W}{2 \sin \alpha} = \frac{W}{2 \lambda F_x}$$

The resolution distribution and the information capacity in section I are now found. The resolution in the space region between $\pm X_m$ is H-limited and is given by $\lambda z_i / W$. The resolution in the region between X_m and X_b is HF_x-limited:

$$R' = \frac{\lambda z_i}{\frac{W}{2} - x'} = \frac{\lambda z_i}{-x_i + X_b}$$

$$\text{where } x' = x_i - F_x \lambda z_i = \frac{W}{2} + x_i - X_b.$$

The number of resolvable points in the space region between $\pm X_m$ is

$$N_x = \frac{2X_m}{\frac{\lambda z_i}{W}} = n - \frac{W^2}{\lambda z_i}$$

The number of resolvable points in the space regions between X_m and X_b is

$$N'_x = \int_{X_m}^{X_b} \frac{dx_i}{\frac{\lambda z_i}{-x_i + X_b}} = \int_{X_m - X_b}^0 - \frac{x_i'}{\lambda z_i} dx_i' = \frac{W^2}{2 \lambda z_i}$$

The total number of resolvable points in one dimension is thus

$$N_t = N_x + 2N'_x = n$$

and the total number of resolvable points in two dimensions is n^2 , the number of recording elements in the hologram. Thus, the information theory laws hold for the plane-wave case. The results for three different plane regions at a typical plane $z = z_i$ are found to be

$$N(IA) = \left(n - \frac{W^2}{\lambda z_i}\right)^2$$

$$N(IB) = \frac{W^2}{2\lambda z_i} \left(n - \frac{W^2}{\lambda z_i}\right)$$

$$N(IC) = \frac{W^2}{2\lambda z_i}$$

The number of resolvable points in the volume regions are

$$N_v(IA) = \int_{-X_m}^{X_m} \int_{-X_m}^{X_m} \int_{z_1}^{\infty} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{W} \cdot \frac{\lambda z_i}{W} \cdot 4\lambda \left(\frac{z_i}{W}\right)^2}$$

$$= \frac{1}{6} \cdot \frac{1}{2} \cdot n^3$$

$$N_v(IB) = \int_{+X_m}^{X_b} \int_{-X_m}^{X_m} \int_{z_1}^{\infty} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{A_x} \cdot \frac{\lambda z_i}{W} \cdot 4\lambda \left(\frac{z_i}{A_x}\right)^2}$$

$$= \frac{1}{6} \cdot \frac{1}{16} \cdot n^3$$

$$N_v(IC) = \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_{z_1}^{\infty} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{A_x} \cdot \frac{\lambda z_i}{A_y} \cdot 4\lambda \left(\frac{z_i}{A_x}\right)^2}$$

$$= \frac{1}{6} \cdot \frac{1}{16} \cdot n^3$$

The total number of resolvable points in section I is thus

$$N_v(I) = \frac{1}{6} n^3 .$$

For section II, the resolution is F_x -limited between $\pm X_m$ and is HF_x -limited between X_m and X_b ; X_m and X_b are given by

$$X_m = \frac{W}{2} - F_x \lambda z_i$$

$$X_b = \frac{W}{2} + F_x \lambda z_i$$

The resolution between $\pm X_m$ is

$$\frac{\lambda z_i}{A_x} = \frac{\lambda z_i}{2\theta z_i} = \frac{\lambda z_i}{2\lambda F_x z_i} = \frac{1}{2F_x}$$

which is independent of the coordinator z_i . As in section I, the resolution between X_m and X_b is $\frac{\lambda z_i}{-x_i + X_b}$. The number of resolvable points in one dimension is given by the following:

$$N_x = \frac{2X_m}{\frac{\lambda z_i}{A_x}} = 4F_x X_m$$

$$N'_x = \int_{X_m}^{X_b} \frac{dx_i}{\frac{\lambda z_i}{-x_i + X_b}} = 2F_x^2 \lambda z_i$$

$$N_t = N_x + 2N'_x = 2F_x W = n$$

The number of resolvable points in three plane regions are:

$$N(IA) = (4F_x X_m)^2$$

$$N(IB) = 2F_x^2 \lambda z_i \cdot 4F_x X_m$$

$$N(IC) = (2F_x^2 \lambda z_i)^2$$

The number of resolvable points in the volume region are:

$$\begin{aligned} N_v(IIA) &= \int_{-X_m}^{Z_m} \int_{-X_m}^{X_m} \int_0^{z_1} \frac{dx_i dy_i dz_i}{\frac{1}{2F_x} \cdot \frac{1}{2F_x} \cdot 4\lambda \left(\frac{1}{2F_x \lambda}\right)^2} \\ &= \frac{1}{6} \cdot \frac{1}{2} \cdot n^3 \end{aligned}$$

$$\begin{aligned} N_v(IIB) &= \int_{+X_m}^{X_b} \int_{-X_m}^{X_m} \int_0^{z_1} \frac{dx_i dy_i dz_i}{\frac{z_i}{A_x} \cdot \frac{1}{2F_x} \cdot 4\lambda \left(\frac{z_i}{A_x}\right)} \\ &= \frac{1}{6} \cdot \frac{1}{16} \cdot n^3 \end{aligned}$$

$$\begin{aligned} N_v(IIC) &= \int_{X_m}^{X_b} \int_{X_m}^{X_b} \int_0^{z_1} \frac{dx_i dy_i dz_i}{\frac{\lambda z_i}{A_x} \cdot \frac{\lambda z_i}{A_y} \cdot 4\lambda \left(\frac{z_i}{A_x}\right)^2} \\ &= \frac{1}{2} \cdot \frac{1}{16} \cdot n^3 \end{aligned}$$

The total number of resolvable points in section II is

$$N_v(II) = \frac{1}{6} n^3$$

Hence, the total number of resolvable points in all object space is $\frac{1}{3} n^3$, which is less than for the point-source reference case

In conclusion, the volume region IA here is similar to the volume region IVA for the point-source reference case; the resolution is determined by the maximum aperture $2X_o$ in the point-source case, while it is determined by the hologram aperture in the plane-wave reference case. The geometry of both regions relative to the resolution aperture $2X_o$ and W are quite similar in that an angle 2α is subtended by the aperture. The same number of resolvable points should be obtained in each case. The number of resolvable points in IVA is given by

$$N_v(VIA) = \frac{1}{6} \cdot \frac{1}{2} \cdot \left[\frac{(W - 2X_o) 2X_o}{\lambda z_o} \right]^3$$

Since $\frac{W - 2X_o}{z_o} = 2\alpha$, we have

$$N_v(IVA) = \frac{1}{6} \cdot \frac{1}{2} \cdot \left(\frac{2X_o \cdot 2\alpha}{\lambda} \right)^3 \quad (2-18)$$

The number of resolvable points in IA is

$$N_v(IA) = \frac{1}{6} \cdot \frac{1}{2} \cdot \left(\frac{W^2}{\lambda z_1} \right)^3 = \frac{1}{6} \cdot \frac{1}{2} \cdot \left(\frac{W \cdot 2\alpha}{\lambda} \right)^3$$

This result is essentially equivalent to Eq. (2-18) when W is replaced by $2X_o$, and verifies that equation.

2.3.5. THE GENERALIZATION OF THE SYSTEM MODEL AND THE USE OF THE OFFSET PLANE-WAVE REFERENCE BEAM

Figure 2-10 illustrates the general case in which the reference plane wave forms an incident angle θ with the normal of the hologram. Again, the whole object space can also be divided into two sections by the plane $z = z_1$. Moreover, all the analysis and the calculated results in the case of normal incident can be applied to this general case, except that the geometric location and the shape of each space region are slightly changed. This will be true later when the expressions are found for the corresponding curves X'_m and X'_b for this general case. The new set of boundary equations differ from the old ones at any given

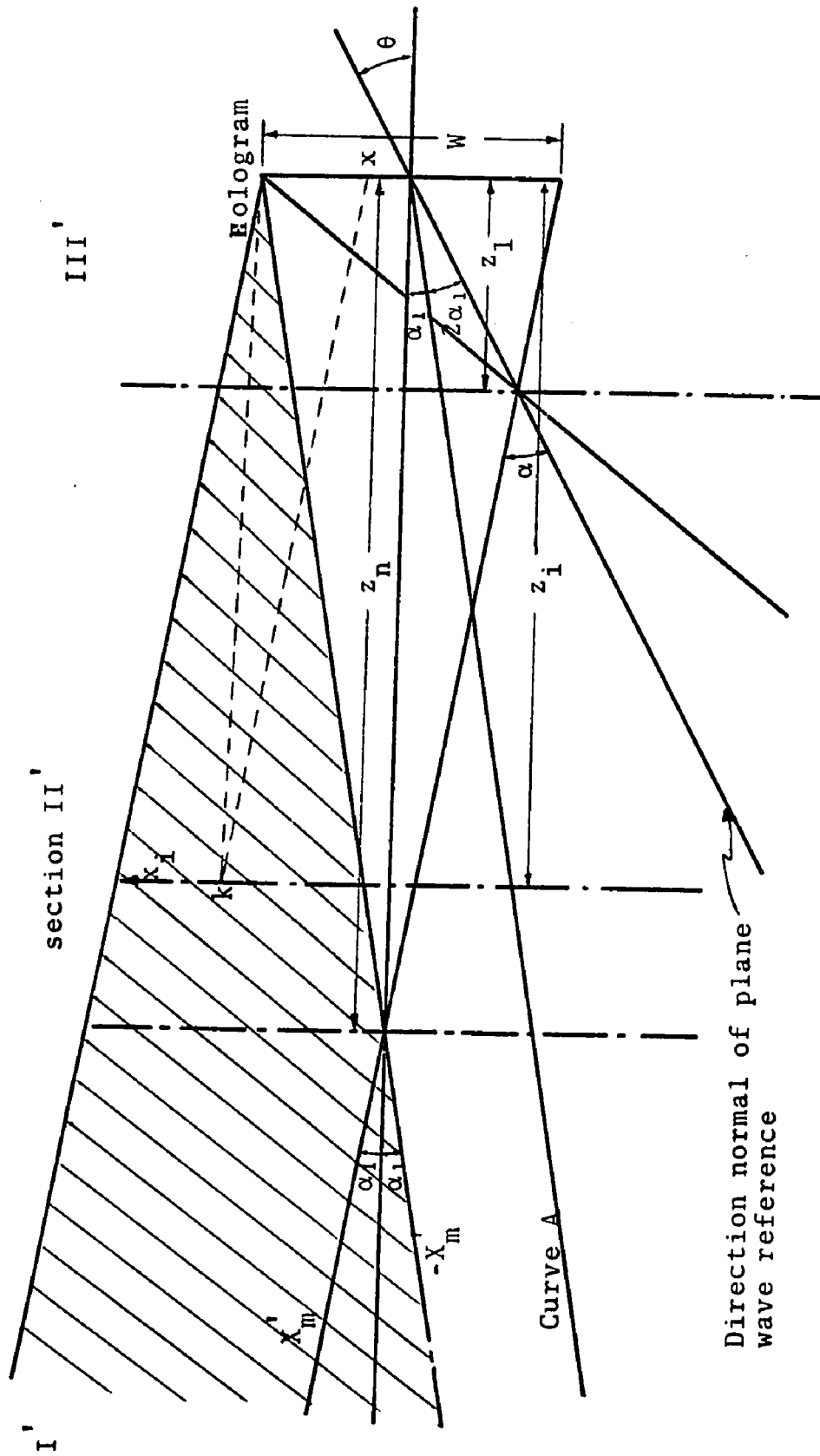


Figure 2-18. The off-set case of Fig. 2-17 and the separation of an object space from its $|U|^2$ terms

plane $z = z_1$ only by a constant βz_1 . From Eq. (2-16) with $\beta > 0$, the maximum spatial frequency $f_x = f_{\max}$ occurs when x_1 is negative and x is positive. If we set $f_{\max} = F_x$, this leads to the relation

$$\begin{aligned} x_1 &= f_x \lambda z_1 - x - \beta z_1 = F_x \lambda z_1 - x - \beta z_1 \\ &= \frac{n \lambda z_1}{4|x|} - x - \beta z_1 \end{aligned} \quad (2-19)$$

where x_1 is a positive value and x is the hologram coordinate, which can be positive or negative. For a given hologram aperture $|x|$, F_x is related to n by $n = 2F_x|x|$. With the hologram aperture given by W , Eq. (2-19) generates two boundary Eqs. X'_m and X'_b as

$$X'_m = \frac{n \lambda z_1}{2W} - \frac{W}{2} - \beta z_1 \quad (2-20)$$

$$X'_b = \frac{n \lambda z_1}{2W} + \frac{W}{2} - \beta z_1 \quad (2-21)$$

A comparison of Eqs. (2-20) and (2-21) with the corresponding equations for the normal incidence case, reveals that they differ only by a constant βz_1 for a fixed z_1 . The number of resolvable points in either plane region or volume region will be the same as the normal incident case, since a shift of the integral boundary as well as the integrand itself will not change the final results.

To consider the separation of an object space with its $|U|^2$ term, we must choose a suitable object space. The object space can be any shape and each different shape yields a different result. However, we want to choose an object space that can be separated from its $|U|^2$ term and at the same time have the highest number of resolvable points when compared with another chosen object space for a given recording capacity.

In fact, for a plane-wave reference case, the object space which has the highest number of resolvable points for a given recording data is the object space formed by the configuration of Fig. 2-17. If we

choose such a configuration in Fig. 2-18 by the two curves x'_m and $-x'_m$ and consider the object space bounded by x'_m and $-x'_m$ for $z_i > z_n$, we can proceed to find the minimum offset angle θ or β necessary for the separation of this object space (I') from its $|U|^2$ term. Macovski proved that the minimum β in this case is $3(n\lambda/W)/8$ [3]. He also calculated that the number of resolvable points in I'A is $11n^3/768$.

Our task here is to extend this object space so that the object points can also be separated from their $|U|^2$ term, although with poorer resolution. Figure 2-18 shows that the maximum spatial frequency of the $|U|^2$ term for the object space I'A is $2\alpha_1/\lambda$, while the object space I'A itself has a spatial frequency band $2\alpha_1/\lambda$ which is $\pm\alpha_1/\lambda$ from the zero frequency. This object space is separated from its $|U|^2$ term if $\theta = 3\alpha_1$. Any object point outside I'A can also be separated from the $|U|^2$ term if the maximum frequency formed by it is within the spatial frequency band $2\alpha_1/\lambda$ of the object space I'A itself. This can be achieved for all the object points above curve $-x'_m$ or x'_m , if they can tolerate a poorer resolution so that the maximum spatial frequency formed by them is α_1/λ . A typical point k with poorer resolution is shown in Fig. 2-18. The resolution of point k is the same as if it were reconstructed by a hologram with maximum spatial frequency $\alpha_1/\lambda = W/2z_n \lambda = F_x/4$ and a normal plane-wave reference. The object points below curve $-x'_m$ cannot be separated from their $|U|^2$ terms, even if a poorer resolution for them can be tolerated. This is true because the $|U|^2$ term occupies a space within an angle $2\alpha_1$ from the normal of the reference wave which makes an angle $\theta = 3\alpha_1$ with the hologram normal. If we plot a curve A parallel to the curve $-x'_m$ in Fig. 2-18, this curve is the boundary of $|U|^2$ term. Since this relation holds for all the points in the hologram up to the edge $W/2$, curve A will coincide with curve $-x'_m$ for the point at the edge of the hologram. This makes any point below $-x'_m$ interfere with the $|U|^2$ term.

The number of resolvable points in all the space regions above the curve $-x'_m$ will now be calculated. The configuration of Fig. 2-18 is in the x direction only and the configuration in y direction is the same as the normal incident case. The whole shaded object space can be divided into three sections, I', II', and III'; and the number of resolvable

points in each section region is defined in Table 4. In order to evaluate the integrals in Table 4, the expressions for X'_m and X'_b can be obtained by substituting $\beta = 3(n\lambda/W)/8$ into Eq. (2-20) and Eq. (2-21):

$$X'_m = \frac{1}{4} F_x \lambda z_i - \frac{W}{2}$$

$$X'_b = \frac{1}{4} F_x \lambda z_i + \frac{W}{2}$$

The distribution of the number of resolvable points in the three sections indicates five-sixths of the information capacity is in section I'. Region $N_v(I'D')$ is the object space investigated by Macovski; the value is about two-thirds of the total number of resolvable points in the object space which can be separated from its $|U|^2$ term. The total number of resolvable points in the whole shaded object-space region is $3n^3/16^2$, which is about half the number obtained for the shaded area in Fig. 2-15 by a point-source reference case. Comparing this value with the total number of resolvable points in the whole object space obtained by a collimated reference beam, it is only 1/28 of the total object points in space.

TABLE 4
NUMBER OF RESOLVABLE OBJECT POINTS IN DIFFERENT
SPACE REGIONS FOR THE OBJECT SPACE SEPARATING
FROM ITS $|u|^2$ TERM BY AN OFFSET PLANE-WAVE REFERENCE

space section	reg-ions	definition of the number of resolvable points in volume space region	results for each space region	total number of resolvable points in each section
I'	B'	$N_V(I'B') \triangleq \int_{X'_m}^{X'_b} \int_{-X'_m}^{X'_m} \int_{z_n}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{64} \cdot \frac{5}{48} n^3$	$N_V(I')$ $= 2N_V(I'B')$ $+ 2N_V(I'C')$ $+ N_V(I'D')$ $+ 2N_V(I'E')$ $= \frac{1}{16^2} \cdot \frac{5}{2} n^3$
	C'	$N_V(I'C') \triangleq \int_{X'_m}^{X'_b} \int_{X'_m}^{X'_b} \int_{z_n}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{64} \cdot \frac{1}{48} \cdot \frac{1}{2} n^3$	
	D'	$N_V(I'D') \triangleq \int_{-X'_m}^{X'_m} \int_{-X'_m}^{X'_m} \int_{z_n}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda \left(\frac{z_1}{W}\right)^2}$	$\frac{1}{64} \cdot \frac{11}{24} n^3$	
	E'	$N_V(I'E') \triangleq \int_{-X'_m}^{X'_m} \int_{X'_m}^{X'_b} \int_{z_n}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{W} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{W}\right)^2}$	$\frac{1}{64} \cdot \frac{1}{48} n^3$	
II'	B'	$N_V(II'B') \triangleq \int_{-X'_m}^{X'_b} \int_{-X'_m}^{X'_m} \int_{z_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{W} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{16^2} \cdot \frac{9}{32} n^3$	$N_V(II')$ $= N_V(II'B')$ $+ 2N_V(II'C')$ $= \frac{1}{16^2} \cdot \frac{15}{32} n^3$
	C'	$N_V(II'C') \triangleq \int_{-X'_m}^{X'_b} \int_{X'_m}^{X'_b} \int_{z_1}^{\infty} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{16^2} \cdot \frac{3}{32} n^3$	
III'	B'	$N_V(III'B') \triangleq \int_{-X'_m}^{X'_b} \int_{-X'_m}^{X'_m} \int_0^{z_1} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{1}{2F_x} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{16^3} \cdot \frac{2}{3} n^3$	$N_V(III')$ $= N_V(III'B')$ $+ 2N_V(III'C')$ $= \frac{1}{16^3} n^3$
	C'	$N_V(III'C') \triangleq \int_{-X'_m}^{X'_b} \int_{X'_m}^{X'_b} \int_0^{z_1} \frac{dx_1 dy_1 dz_1}{\frac{\lambda z_1}{A_x} \cdot \frac{\lambda z_1}{A_y} \cdot 4\lambda \left(\frac{z_1}{A_x}\right)^2}$	$\frac{1}{16^3} \cdot \frac{1}{6} n^3$	

CHAPTER 3

THE INFORMATION REDUCTION METHODS OF A THREE-DIMENSIONAL HOLOGRAPHIC VIEWING SYSTEM

3.1. INTRODUCTION

The use of a hologram as a recording device to provide a three-dimensional viewing system has been studied for many years [1,12-16]. The fundamental problem of such a system is the large amount of information that has to be recorded in the hologram plane to achieve a reasonable viewing angle [12,16]. Several methods which have been proposed to reduce the amount of information sacrifice other aspects necessary for a satisfactory, realistic 3-D viewing. Those aspects include resolution, signal-to-noise ratio, vertical parallax, and depth of field. In this chapter, the problem will be approached theroretically. In Section 3.2, a model for a basic three-dimensional viewing system is described. The fundamental requirements of a 3-D viewing system are given and the conditions under which these requirements can be met for a holographic system which employs information reduction are stated. Section 3.3 discusses the optimum recording configuration for a holographic viewing system. In Section 3.4, a method is proposed for further reducing the information required in such a system. This method maintains the aspects of a 3-D viewing system required for satisfactory viewing by a human observer. Section 3.5 is devoted to the mathematical study of the scatter-plate method [16,17]. A mathematical formulation has been used to clarify the scatter-plate optimization problem and the difficulties encountered in this approach have been stated. Except for a very special case, this approach is not very fruitful. It is the author's belief that an experimental approach to optimizing the scatter-plate method requires that new techniques be developed for fabricating high quality, sophisticated, scatter plates.

Years of trial-and-error experiments are necessary to improve such a system. Because of the limitations in the present state of the art, only a limited number of existing scatter plates have been employed in experimental systems by the author. These scatter plates were not originally designed for the scatter-plate system and their surface quality was usually quite poor. One of the commercially available scatter plates, the fly's-eye lens array, has better surface quality and yields better results than the other scatter plates that were tried. The performance and the limitations of the fly's-eye lens array can be analyzed by geometric optics and diffraction theory. Included in Section 3.6 is a study of the fly's-eye lens array for which geometric optics is used. The treatment of the fly's-eye lens array based on diffraction theory is presented in Chapter 4.

3.2. THE FUNDAMENTAL REQUIREMENTS OF A THREE-DIMENSIONAL VIEWING SYSTEM AND THE USE OF A HOLOGRAM IN SUCH A SYSTEM

Consider the general three-dimensional viewing system shown in Fig. 3-1, in which the human eye is modeled as two small lenses, each with aperture e separated by a distance E , in a large viewing aperture A . The object is a distance d away from the viewing aperture and has a lateral dimension L and a depth D . In general, $d \gg D$ and $d \gg L$. If a two-dimensional object is considered, the number of object points in the object space that can be resolved by the human eye can be calculated. This number is also the minimum number of object points in the object space required for a reconstruction that gives the human eye an ungraded perception of a vivid 3-D viewing. Since $d \gg D$, we assume that the size of each resolution element that can be resolved by the eye in any lateral plane in the object space is given by

$$\frac{\lambda d}{e}$$

The number of resolvable points in one lateral plane is thus given by

$$\frac{Le}{\lambda d}$$

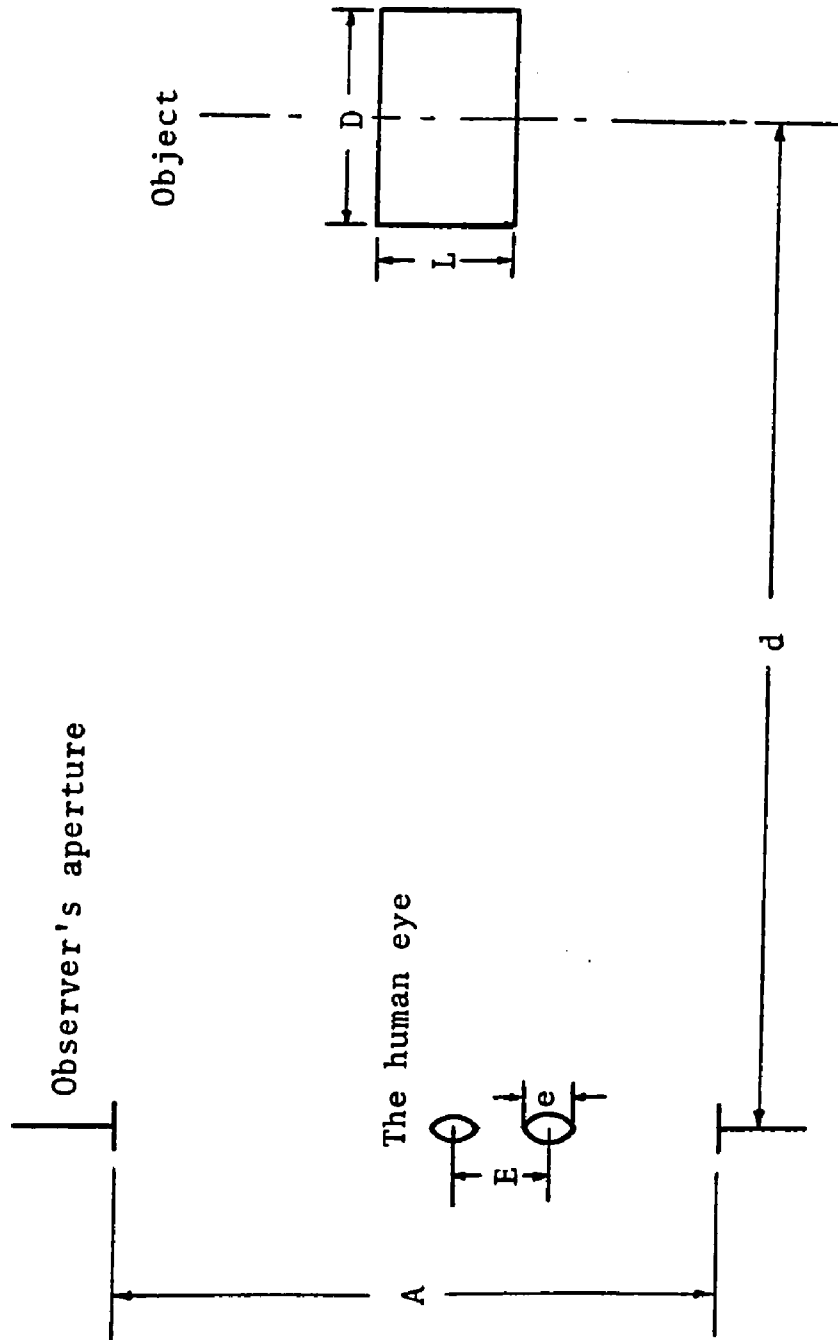


Figure 3-1. The general configuration of a three-dimensional viewing system

The number of resolvable planes in depth for the human eye is given by

$$\frac{DeE}{\lambda d^2}$$

The deviation of this expression is shown in Appendix A. The total number of resolvable object points in the object space is thus

$$N \approx \frac{LDe^2E}{\lambda^2 d^3}$$

If an observer views an object of size $L \times D$ from a distance d with $d \gg D$, $d \gg L$, the eye cannot tell whether the object is a real, solid object or whether it is just a set of N object points positioned in space in such a way that the eye can only just resolve them. This number N is thus the minimum number needed to reconstruct a 3-D object in space and to provide undegraded three-dimensional viewing. In this reconstruction, the eye cannot be closer to the set of N object points than the distance d and still enjoy the realistic 3-D viewing. When the eye is closer, the N point object looks granular. As long as the eye is more than d away from the N point object, an observer experiences the same comfortable 3-D viewing obtainable from looking at a solid object. Up to this point, no comment was made on how to record and reconstruct such a N point object in space. In fact, it is not important if the N point object is reconstructed in space or not. The concern is that there is a recording method which makes the observer see the wavefront of such an object while viewing it through a screen. Holography seems to be the best recording method to provide such a 3-D viewing. Other methods are discussed in Ref. 1. Moreover, a hologram recording may require less information content than the actual reconstruction of N object points in space. The conditions that this is true are derived below.

A large hologram with a size A is put at the viewing aperture A so that the eye can see the object anywhere within the viewing aperture. The space-bandwidth of the hologram for an object with dimension $L \times D$ and a distance d away is approximately

$$(SW)_H = \frac{AL}{\lambda d},$$

if it is assumed that $d \gg D$ and $d \gg L$. The ratio of the hologram recording data, $(SW)_H$, to the N is

$$\gamma = \frac{\Delta}{N} \frac{(SW)_H}{\lambda d} = \frac{LA}{\lambda d} \cdot \frac{\lambda^2 d^3}{LDe^2 E} = \frac{\frac{A}{e}}{\frac{D}{\lambda} \cdot \frac{eE}{d^2}} = \frac{\Delta}{\alpha} \frac{\beta}{\alpha}$$

where

$\alpha = \frac{\Delta}{\lambda} \cdot \frac{eE}{d^2}$ is the number of resolvable planes in depth that can be resolved by the human eye, and

$\beta = \frac{A}{e}$ is the ratio of viewing aperture to the eye lens aperture.

A plot of γ versus α for different β is shown in Fig. 3-2. It can be seen that if $\alpha > \beta$, i.e., if the number of resolvable planes in depth is larger than the aperture ratio, then $\gamma < 1$. This means that the recording data for the hologram is less than N . In other words, the use of a hologram as a recording device is preferable to a method which directly records the number of resolvable points N in space if the required number of resolvable planes in depth is larger than the aperture ratio. Moreover, it is clear that since the aperture size of the hologram is much larger than the eye aperture and the eye separation distance E , the number of resolvable planes in depth provided by the hologram should be larger than the number that can be resolved by the human eye.

In Chapter 2, we have calculated the number of resolvable object points in the whole object volume space to be in the range of $n^3/2$ to $2n^3/3$ for a point source reference, where n is the space-bandwidth product in one dimension. In the configuration of Fig. 3-1, if we assume the object space is just the space section I + II in Chapter 2, the total number of resolvable object points in section I + II is $n^3/3$. Since the number of resolvable points in any object plane $z = z_1$ is a constant and is equal to n^2 , the number of resolvable planes in depth

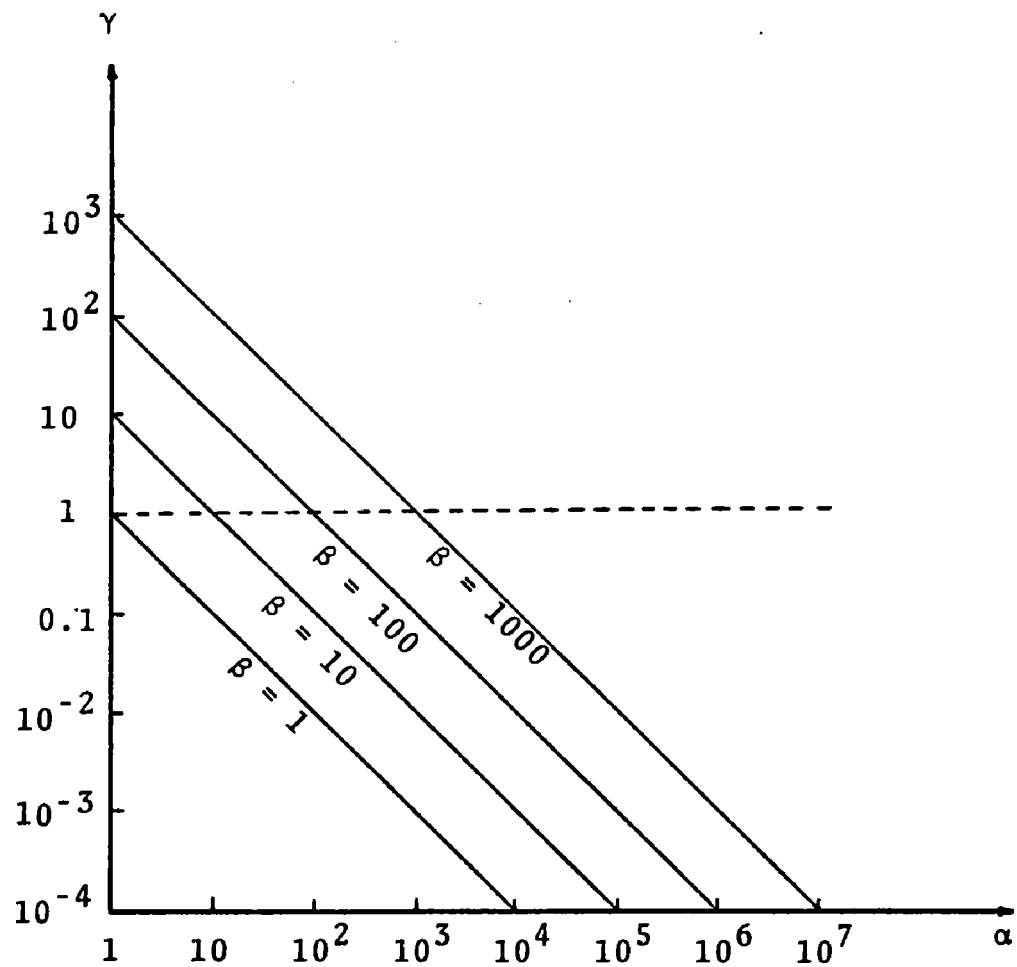


Figure 3-2. The relation between γ and α for different value of β

in section I + II is equal to $n/3$. Comparing $n/3$ with $DeE/\lambda d^2$, we have

$$\frac{\frac{n}{3}}{\frac{DeE}{\lambda d^2}} = \frac{LA}{3\lambda d} \cdot \frac{\lambda d^2}{DeE} = \frac{1}{3} \cdot \frac{d}{D} \cdot \frac{A}{E} \cdot \frac{L}{e}$$

In a general viewing configuration, $d \gg D$, $A > E$, and $L > e$. Hence, $n/3$ is much larger than $DeE/\lambda d^2$, i.e., the hologram provides a much larger number of resolvable planes in depth than required by a human eye. It is clear that the hologram does have the capability to provide the necessary 3-D viewing and, under certain condition ($\alpha > \beta$), the recording data $(SW)_H$ can be even less than that required for directly recording N object points. In practice, however, because of the large size of the hologram aperture necessary to fit the viewing aperture, the hologram is placed somewhere near the object. This makes the viewing system practical to build and viewing the hologram is similar to viewing a TV screen. This subject will be discussed in the next section.

3.3. THE OPTIMUM RECORDING CONFIGURATION OF A HOLOGRAPHIC VIEWING SYSTEM

It was pointed out in the previous section that a hologram has to be rather close to the object for a practical three-dimensional holographic viewing system. This situation is shown in Fig. 3-3, where H is the size of the hologram, A is the observer's viewing aperture, and D_0 is the distance of the hologram from the viewing aperture. In general, $D_0 > H$. From the geometry, it is clear that for a given hologram aperture in Fig. 3-3, an object point has to be within the shaded area to be seen by an observer anywhere within the viewing aperture.

Consideration will now be given to the shaded object space in front of the hologram, i.e., the shaded area between A and H . An effort will be made to determine the space bandwidth product of the hologram and the object space for this space bandwidth product. The possibility of various locations for a point source reference beam in the practical recording configuration must be made. Suppose that for given system parameters A , H , and D_0 in Fig. 3-3, all the shaded area in front of the hologram is to be reconstructed by the hologram so that it can be viewed from anywhere in A . Then the point-source reference has to be

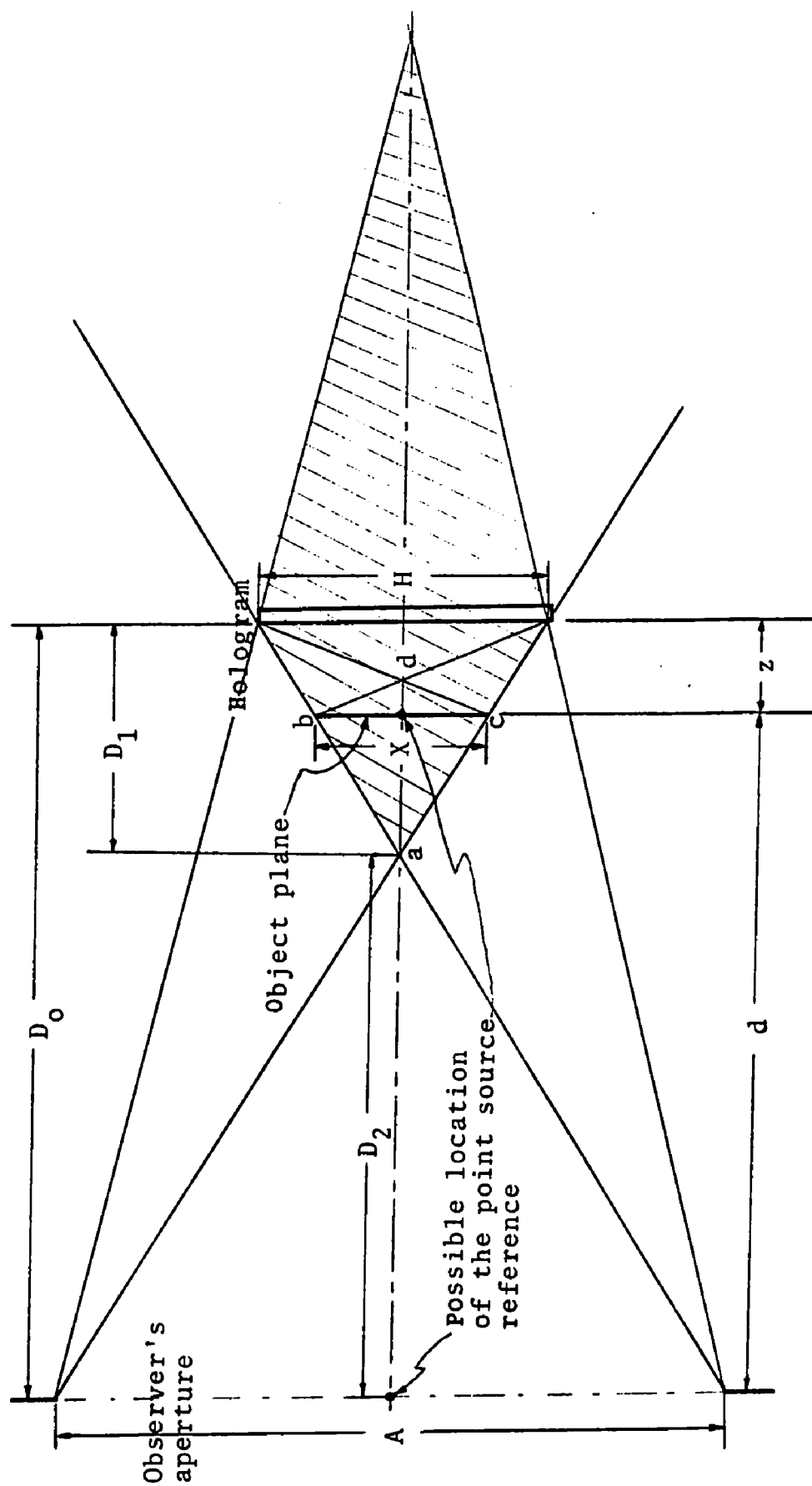


Figure 3-3. The optimum recording configuration of a holographic viewing system

at the center of A, and the space-bandwidth product in this case is $AH/\lambda D_o$, the optimum recording configuration for the system if all the shaded area is to be reconstructed. However, if the object will fit in a smaller object space, it might be possible to record less information. Consider an object aperture X in Fig. 3-3, located a distance d away from the aperture A and a distance z away from the hologram aperture H. If a reference point source is put at the center of the object aperture X and X is recorded by the hologram, the space bandwidth product is $XH/\lambda z$. The object space provided in this recording configuration is the area bounded by four points--a, b, c, and d in Fig. 3-3. It is clear that $XH/\lambda z \leq AH/\lambda D_o$ only if $X/z \leq A/D_o$. In other words, if the angle subtended by the object aperture is smaller than the angle subtended by the viewing aperture toward the hologram, the space-bandwidth product can be reduced by putting the reference point source at the center of the object aperture rather than at the center of the viewing aperture. When $X/z = A/D_o$, $z \triangleq z_o = D_o H / (H + 2A)$. The corresponding object aperture is found to be $X \triangleq X_o = AH / (H + 2A)$. Thus, if the required object aperture in the object space in Fig. 3-3 is larger than X_o , the optimum recording configuration for the system can be achieved by putting the reference point source at the center of the viewing aperture. The space-bandwidth product for this case is $AH/\lambda D_o$. On the other hand, when the required object aperture is smaller than X_o in Fig. 3-3, less information is required if the reference point is placed at the center of the object aperture. In this case, the space-bandwidth product is $XH/\lambda z$. In general, the maximum information for the system is $AH/\lambda D_o$ if all the shaded object space in front of the hologram in Fig. 3-3 is used. However, further reduction of information is desirable. The following section describes a method for accomplishing this reduction.

3.4. THE SAMPLING METHOD FOR INFORMATION REDUCTION IN HOLOGRAPHY

In the previous investigation, the holographic viewing system shown in Fig. 3-3 provided an object space with a resolution better than the human eye requires. Since the human eye requires less resolution, more information reduction can be achieved by further degrading the resolution of the object to the standard required by the human eyes.

The method described below is proposed to achieve this reduction by sampling the hologram recording plate.

In Fig. 3-4, an eye-lens aperture e is viewing an object plane reconstructed by the hologram plate behind it. The eye cannot resolve any two points in the object plane if the angle subtended by those points at the eye is less than one minute of arc. Thus, only a finite number of points need be reconstructed by the hologram. The angular separation between adjacent points can be one minute of arc, which suggests that we can sample the recording hologram plate using a sample spacing m which subtends one minute of an arc at the eye. The angle with arc length m formed at the eye cannot be larger than one minute of arc, but it can be smaller. We thus have

$$\frac{m}{D_o} \leq 1' = \frac{1}{3600} \quad (3-1)$$

Each object point in the object plane should form a cone at the eye large enough to fill the whole aperture of the eye. In this way, the eye views the object with diffraction-limited resolution. To provide such a cone, the size of the recording hologram element ℓ should satisfy the following condition

$$\frac{\ell}{z} \leq \frac{e}{d} \quad (3-2)$$

From Eqs. (3-1) and (3-2) above, the separation m between each recording sample and the size ℓ of each recording sample have the following inequalities:

$$m \leq \frac{D_o}{3600} \quad (3-3)$$

$$\ell \geq \frac{ez}{d} \quad (3-4)$$

The information can be reduced only if $m > \ell$. As a measure of the information reduction, an information-reduction factor σ is defined as

$$\sigma = \frac{m}{\ell}$$

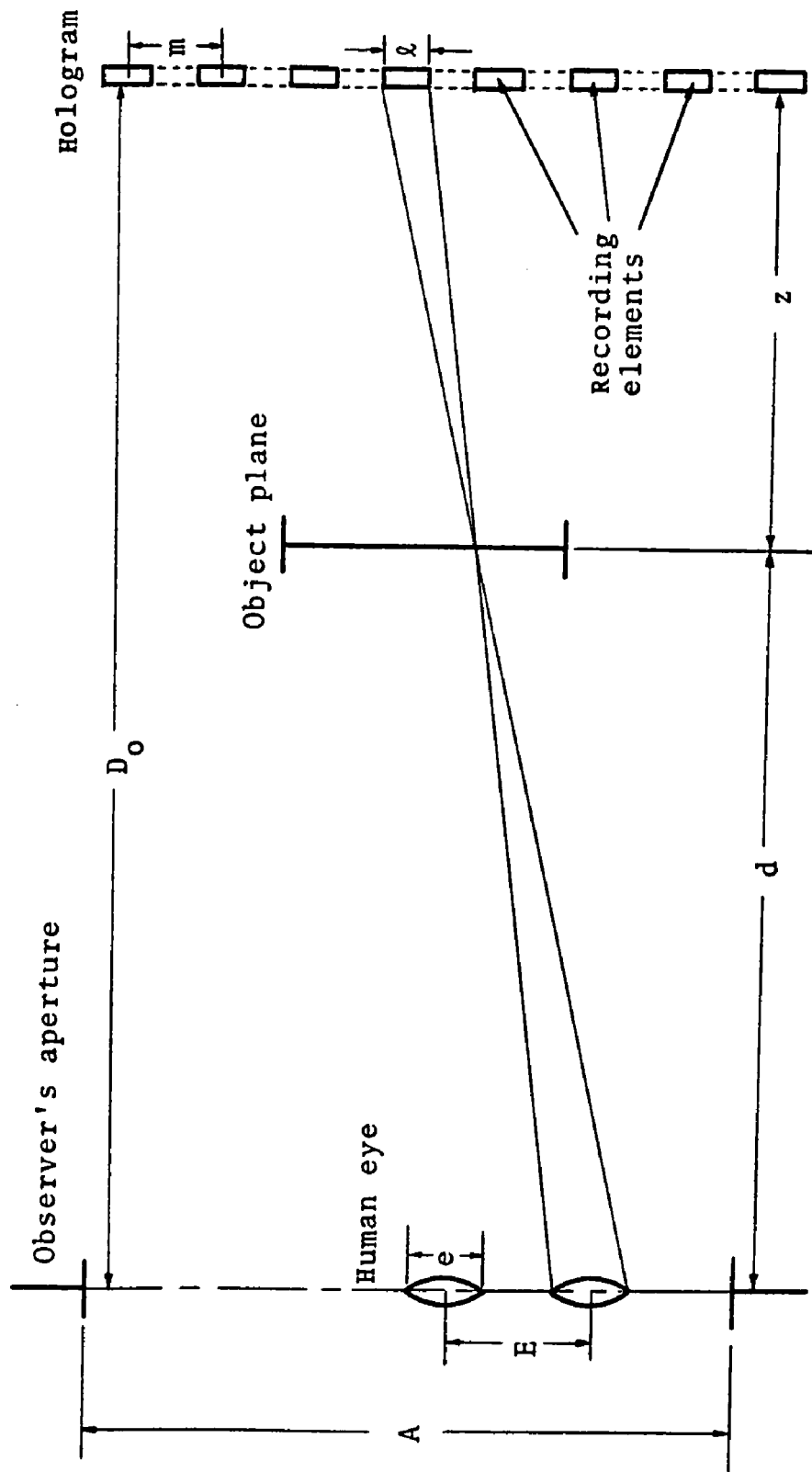


Figure 3-4. The configuration of the sampling method for information reduction in holography

With this definition and the relation of Eqs. (3-3) and (3-4), the following inequalities are obtained:

$$\sigma \leq \frac{D_o}{3600e} \cdot \frac{d}{z} \quad \text{or} \quad \frac{z}{d} \leq \frac{D_o}{3600e\sigma} \quad (3-5)$$

The physical significance of Eq. (3-5) is very important. Given an information-reduction factor σ , Eq. (3-5) says that the object plane should be close to the hologram to the extent that z/d is less than $D_o/3600e\sigma$. The larger the information-reduction factor σ is, the smaller z must be, and the smaller the available object space becomes. Equation (3-5), however, may not be satisfied in all cases. For a given σ , suppose m is chosen to satisfy Eq. (3-3). There is a minimum z , so Eq. (3-5) is satisfied. Those object planes farther away from the hologram than this minimum z does not conform to Eq. (3-4). This means that the image resolution at those planes is poor. On the other hand, if ℓ is made large enough to satisfy Eq. (3-4) for those object planes larger than the minimum, then Eq. (3-3) is not satisfied for the whole object space; and the whole object space will have a granular appearance. Equation (3-5) therefore relates the information-reduction factor σ to the location of the object plane satisfying the standard required by the human eye without degrading the image quality.

The curve of z versus σ or the curve of the available object space volume against σ will be plotted. First, however, the maximum value of σ must be determined. Theoretically, σ can be extremely large if z can be extremely small. Of course if $\sigma \rightarrow \infty$, $z \rightarrow 0$. In practice, however, there is an obvious limitation for σ in the recording material itself. As Fig. 3-3 shows, the minimum resolvable recording-element size in the hologram plane is given by $\lambda D_o/A$. Definitely, the size of the recording sample elements cannot be smaller than the size of the minimum resolvable elements which the emulsion of the hologram can record. Hence, $\ell \geq \lambda D_o/A$. Using the definition $\sigma \triangleq m/\ell$, it follows that

$$\sigma \leq \frac{mA}{\lambda D_o} \leq \frac{A}{3600\lambda} = \frac{A}{e} \quad (3-6)$$

where the relation $\lambda/e = 1/3600$ for the normal eye aperture is used. Equation (3-6) says that the maximum information-reduction factor cannot be larger than the ratio of the observer's viewing aperture to the human eye's aperture. Intuitively, this conclusion seems to be true. In reality, when $\sigma = A/e$ and $l = D_o/A$, the size of each recording sample element is just equal to the size of one intensity fringe. Such a recording cannot diffract a reasonable object wavefront in the reconstruction process and make three-dimensional viewing possible. In fact, when the information-reduction factor is close to the value of A/e , the object plane distance z will be several millimeters from the hologram. In this case, the eye cannot perceive any depth of field when looking at the hologram plate. The hologram looks like a TV screen and only a two-dimensional image can be seen. As the information-reduction factor gets smaller, the viewer sees some depth of field in front of the hologram screen. However, he is never able to see through the hologram plane, since each of the sampling elements is so small that the eye can just resolve them and each one of them acts like a single resolvable object point. For this reason, the main concern in the previous section was with the object space in front of the hologram. The whole situation can be visualized if the reader has had the experience of seeing an image reconstructed by the scatter-plate method with a ground glass as the scatter plate. Because of its fine structure, the eye cannot see through the ground glass, but it can see a three-dimensional image in front of the ground glass.

To conclude this section, the relation between z and σ is derived and a numerical example given to indicate the degree of information reduction that can be practically achieved with this sampling method. If Eq. (3-3) is used, the value of m is a constant for a constant D_o , i.e., $m = D_o/3600$. For a given σ , l can be determined from $l = m/\sigma$. With geometry, $d/z = l/e$ and $D_o = d + z$, and z becomes

$$z = \left(\frac{l}{l + e} \right) D_o = \left(\frac{\frac{m}{\sigma}}{\frac{m}{\sigma} + e} \right) D_o = \left(\frac{m}{m + \sigma e} \right) D_o = \left(\frac{\frac{m}{e}}{\frac{m}{e} + \sigma} \right) D_o \quad (3-7)$$

Equation (3-7) is the relation between z and σ ; m/e in Eq. (3-7) is a function of D_o and λ , i.e.:

$$m/e = \frac{1}{(3600)^2} \frac{D_o}{\lambda} .$$

A numerical example is now given below. The system parameters are chosen as follows:

$$A = 1 \text{ meter} = 1000 \text{ mm}$$

$$D_o = 3 \text{ meters} = 3000 \text{ mm}$$

$$e = 2.16 \text{ mm}$$

$$\lambda = 0.6 \times 10^{-3} \text{ mm}$$

$$\frac{\lambda}{e} = \frac{1}{3600}$$

Now,

$$m = \frac{3000}{3600} = 0.833 \text{ mm}$$

$$\frac{m}{e} = \frac{0.833}{2.16} = 0.385$$

Equation (3-7) becomes

$$z = \left(\frac{0.385}{0.385 + \sigma} \right) \times 3000$$

Figure 3-5 is the plot of the above equation, where σ^2 instead σ is used as one coordinate; σ^2 corresponds to the information-reduction factor in two dimensions. The maximum $\sigma = A/e = 1000/2.16 = 463$ and the minimum $\sigma = 1$. The minimum resolvable fringe in the hologram is

$$\frac{D_o}{A} = \frac{0.6 \times 10^{-3} \times 3}{1} = 1.8 \times 10^{-3} \text{ mm}.$$

3.5. THE SCATTER-PLATE METHOD FOR INFORMATION REDUCTION IN HOLOGRAPHY

One of the information-reduction methods for a holographic three-dimensional viewing system is the "scatter-plate method" first described by E. N. Leith [13] and later studied by K. Haines and D. B. Brumm [17,19]. The fundamental goal of their experimental and theoretical studies was

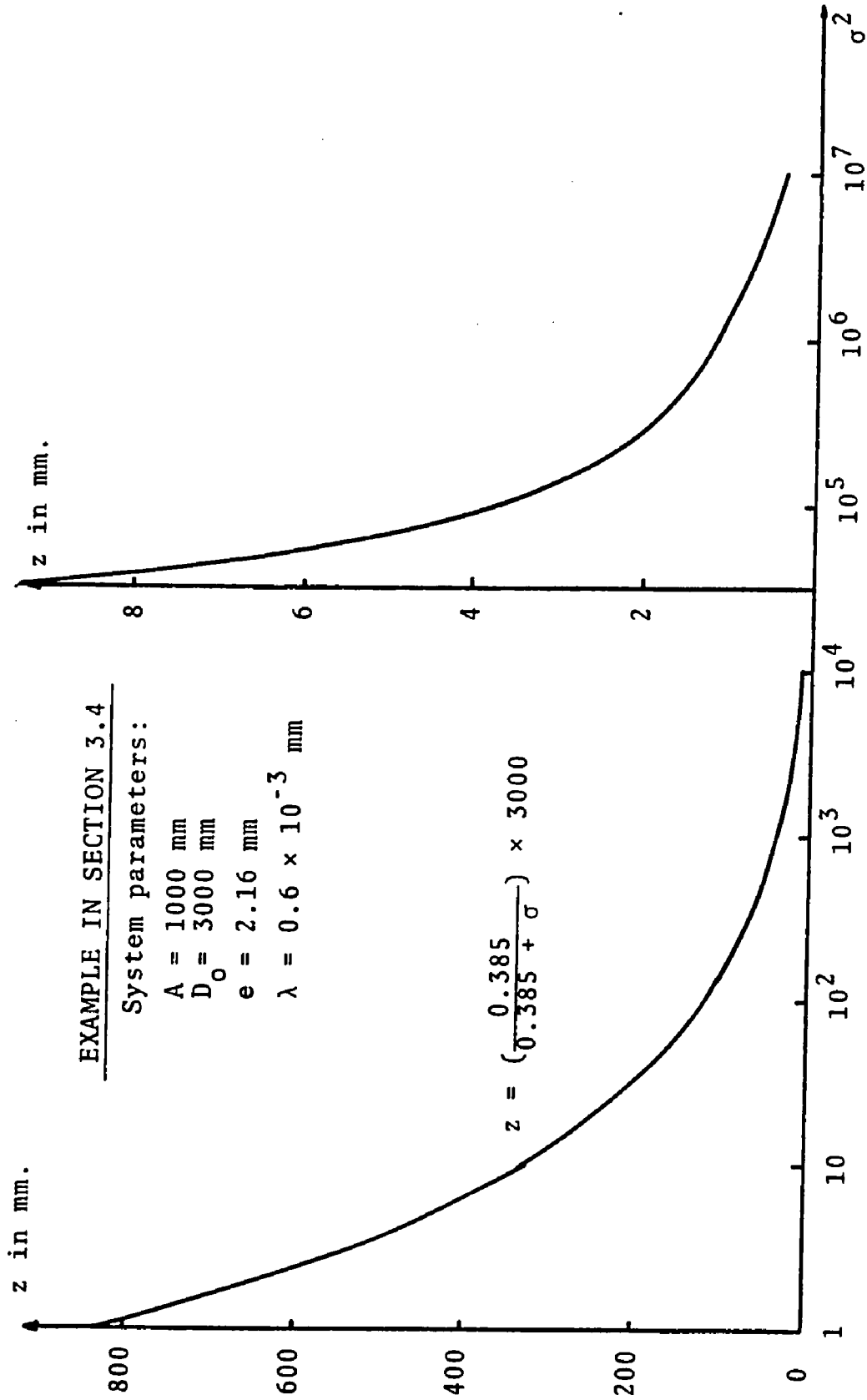


Figure 3-5. The available object space distance z vs the information-reduction factor σ^2 in two dimensions

to search for a scatter plate which would yield the best performance. The scatter-plate method can be roughly described as follows: A transparent scatter plate is placed at the location of the hologram (see Fig. 3-3). The scatter plate scatters the object wavefront passing through it in all directions. A small hologram is placed in the far field of the scattered light to record part of the scattered light. In the reconstruction process, the small hologram is put back in place. The wavefront reconstructed by the small hologram is traced back through the scatter plate and forms a real image at its original place. The information content recorded by the small hologram can be much less than what is required if a big hologram is used at the scatter plate's location. However, theoretical and experimental studies of the scatter-plate method have encountered several difficulties.

First, the mathematical formulation of the actual system is very complicated. Experimental results have not provided enough information to obtain an adequate mathematical model for the system. Information about the basic requirements for the system performance is also lacking. Second, even if a simple mathematical model is formulated for the system based on several simplification assumptions and a performance criterion is chosen, we find the problem is still too difficult to solve without the use of a digital computer. Third, the search for an optimum scatter plate is quite similar to the search for a good imaging lens in the early days of optics. Improvements and perfection of a scatter-plate method will occur through the design of the scatter plate, the innovative use of the scatter plate in experiments, and the development of new techniques for fabricating a high-quality scatter plate. Only a limited number of scatter plates were available to the author for an experimental study. One of the tested scatter plates, the fly's-eye lens array, has certain features which justify the further investigation presented in Chapter 4. In this section, a mathematical formulation of the system will be presented. In Subsection 5.3.1., a mathematical model for the system is carefully derived, assumptions for the model are clearly stated, and the optimization problem of the scatter plate in terms of mathematical operations is defined by four questions. Although the answers to these questions are complicated and unattainable for most cases, possible answers are suggested and difficulties which exist in reaching them are

discussed in detail. In Subsection 3.5.2., the special case of a periodic scatter plate is considered and treated with some simplifying assumptions. The input signals for the system are also assumed to be periodic. Depending on the type of input signals, the answers to the questions mentioned above can be obtained using digital computers.

3.5.1. THE OPTIMIZATION OF THE SCATTER-PLATE METHOD

Figure 3-6 illustrates a scatter-plate system. In the recording process, a small hologram is placed in the far field behind the scatter plate to record part of the scattered light. In the reconstruction process, the small hologram is placed back at its original position. The wavefront reconstructed by the small hologram then propagates back through the scatter plate and forms a real image at the object position. The image is viewed within the viewing aperture A by an observer. For the purpose of this analysis, it is assumed that the object space is limited between the scatter plate and the viewing aperture. The analysis is carried out in one dimension throughout this presentation. The mathematical model for the system can be derived as follows:

Consider a wavefront $U(y)$ of the object space just in front of the scatter plate. The transmitting function of the scatter plate is assumed to be complex and is denoted by $p(y)$; $p(y)$ is limited within the scatter-plate aperture H so that the wavefront emerging to the right of the scatter plate is given by $U_1(y) = U(y)p(y) \text{ Rect}(y/H)$. The E-M field disturbance at the hologram plane K , caused by the wavefront $U_1(y)$ is given by

$$U_2(y_1) = \alpha_1 \int_{-H/2}^{H/2} U(y)p(y) e^{j\frac{\pi}{\lambda Z}(y_1-y)^2} dy \quad (3-8)$$

where α_1 is a complex constant and the Fresnel approximation has been used. This wavefront $U_2(y)$ will be recorded by the hologram aperture K and will propagate back to the scatter plate in the reconstruction process. The reconstructed wavefront to the right of the scatter plate is obtained by again using the Fresnel approximation,

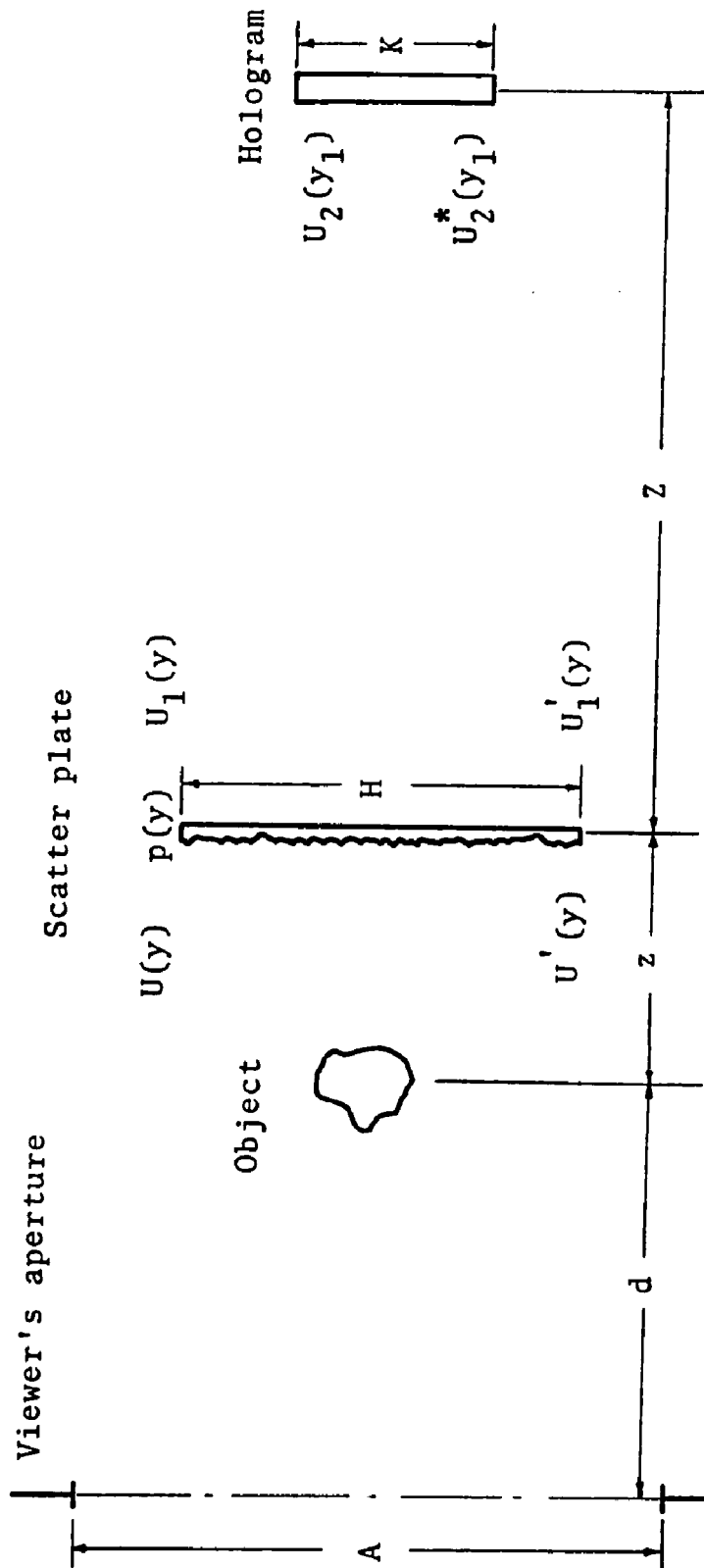


Figure 3-6 The schematic diagram of a scatter-plate system

$$U_1'(y) = \alpha_2 \int_{-K/2}^{K/2} U_2^*(y_1) e^{j\frac{\pi}{\lambda Z}(y-y_1)^2} dy_1 \quad (3-9)$$

where α_2 is a complex constant and where * denotes the conjugate of the wavefront $U_2(y_1)$. Using Eq. (3-8) and Eq. (3-9), $U_1'(y)$ can be expressed as

$$U_1'(y) = e^{j\frac{\pi}{\lambda Z}y^2} \alpha_1 \alpha_2 \int_{-H/2}^{H/2} U^*(t) p^*(t) e^{-j\frac{K}{\lambda Z}t^2} \text{sinc}\left[\frac{K}{\lambda Z}(t - y)\right] dt \quad (3-10)$$

$U_1'(y)$ is then passed through the scatter plate again. The reconstructed wavefront $U'(y)$ emerging to the left of the scatter plate is

$$\begin{aligned} U'(y) &= U_1'(y) p(y) \text{Rect}(y/H) \\ &= p(y) \text{Rect}(y/H) e^{j\frac{\pi}{\lambda Z}y^2} \alpha_1 \alpha_2 \int_{-H/2}^{H/2} U^*(t) p^*(t) e^{-j\frac{K}{\lambda Z}t^2} \\ &\quad \text{sinc}\left[\frac{K}{\lambda Z}(t - y)\right] dt \end{aligned} \quad (3-11)$$

Equation (3-11) is the mathematical formulation of the system where $U'(y)$ is the output of the system and $U^*(y)$ is the input of the system. The conjugate of the input signal indicates that the direction of propagation of the signal wavefront has been reversed. The performance of the system is now described and four questions are posed. The answers to these questions are the fundamental goals of this study.

Question 1: For a given input signal $U^*(y)$, is it possible to choose a scatter-plate function $p(y)$ so that, except for a multiplicative constant, the reconstructed signal $U'(y)$ will be exactly equal to $U^*(y)$?

Question 2: For a given input signal $U^*(y)$, if the answer to Question 1 is no, then is there a scatter-plate function $p(y)$ so that the reconstructed signal $U'(u)$ is the best approximation to the original signal $U^*(y)$?

Question 3: For a given set of signals $\{U^*(y)\}$, is it possible to find a scatter-plate function $p(y)$ so that each of the reconstructed signals will be exactly equal to its original signal except for a multiplicative constant?

Question 4: If the answer to Question 3 is no for a given set of input signals $\{U^*(y)\}$, then is there a scatter-plate function $p(y)$ so that the system performance is optimum?

The answers to these questions depend on the definitions of the phrases, "the reconstructed signal is the best approximation to the original signal," and "the system performance is optimum," and on the particular input signal or the set of input signals chosen. In this section, Question 1 is studied in detail and some comments about Question 2 are made. Questions 3 and 4 will be treated in Section 3.4.2.

For a given input signal $U^*(y)$, is there a function $p(y)$ such that

$$\lambda U'(y) = U^*(y)$$

where λ is a constant? Using Eq. (3-11), the above equation can be written as

$$U^*(y) e^{-j\frac{\pi}{\lambda Z}y^2} = \lambda p(y) \int_{-H/2}^{H/2} p^*(t) U^*(t) e^{-j\frac{\pi}{\lambda Z}t^2} \text{sinc} \left[\frac{K}{\lambda Z}(t - y) \right] dt \quad (3-12)$$

for $|y| \leq H/2$. Let $V(y) \triangleq U^*(y) e^{-j\frac{\pi}{\lambda Z}y^2}$ and substitute into Eq. (3-12). Equation (3-12) becomes

$$V(y) = \lambda p(y) \int_{-H/2}^{H/2} p^*(t) V(t) \text{sinc} \left[\frac{K}{\lambda Z}(t - y) \right] dt \quad (3-13)$$

for $|y| \leq H/2$.

Since $U^*(y)$ is a given function, $V(y)$ can also be regarded as a given function. If Eq. (3-13) is used, Question 1 becomes: "For a given $V(y)$, is there a solution for the integral Eq. (3-13)?" A problem arises because Eq. (3-13) is not a linear integral equation. To solve Eq. (3-13), let us first assume $p(y)$ is a pure phase function denoted by

$$p(y) = e^{j\phi(y)}$$

where $\phi(y)$ is a real function. Equation (3-13) then becomes

$$V(y)e^{-j\phi(y)} = \lambda \int_{-H/2}^{H/2} V(t)e^{-j\phi(t)} \operatorname{sinc}\left[\frac{K}{\lambda Z}(t - y)\right] dt \quad (3-14)$$

for $|y| \leq H/2$.

Defining $V(y)e^{-j\phi(y)} \triangleq S(y)$ in Eq. (3-14), Eq. (3-14) becomes

$$S(y) = \lambda \int_{-H/2}^{H/2} S(t) \operatorname{sinc}\left[\frac{K}{\lambda Z}(t - y)\right] dt \quad (3-15)$$

for $|y| \leq H/2$.

Equation (3-15) is a well known integral equation, the solutions of which are prolate spheroidal wave functions. A prolate spheroidal wave function is denoted by $\psi_i(y)$ which has a corresponding eigen value λ_i . Moreover, the set of functions $\{\psi_i(y)\}$ are the only non-zero solutions of integral Eq. (3-15). A linear combination of the members of the set $\{\psi_i(y)\}$ is not necessarily a solution of Eq. (3-15), although in some special cases, it can be a solution if the eigen values $\{\lambda_i\}$ are equal for some functions of $\{\psi_i(y)\}$. Since $\psi_i(y)$ is the only solution of Eq. (3-15)

$$S(y) \triangleq V(y) e^{-j\phi(y)} = \psi_i(y) \quad \text{for } i = 0, 1, 2, \dots$$

or

$$V(y) = \psi_i(y) e^{j\phi(y)} \quad \text{for } i = 0, 1, 2, \dots \quad (3-16)$$

Equation (3-16) indicates that the amplitude of $V(y)$ should be equal to $\psi_i(y)$ while the phase can be any real function since $\psi_i(y)$ is a real function for all i . Unfortunately, for general viewing system parameters, $\psi_i(y)$ is a very rapidly varying function, while the practical wavefront of an ordinary object has an almost uniform amplitude when its wavefront reaches the scatter plate. This means that the mathematical solution of Eq. (3-15) does not represent a general wavefront of an ordinary object which is some distance away from the scatter plate. The $\psi_i(y)$ function does represent some special signal when the object is planar and is close to the scatter plate so that the signal wavefront generated by the object has an amplitude approximately equal to one of the $\{\psi_i(y)\}$ functions.

When a general scatter-plate function $p(y)$ is considered, the complex valued function $p(y)$ is denoted by

$$p(y) = A(y) e^{-j\phi(y)}$$

where $A(y)$ and $\phi(y)$ are real functions. If $p(y)$ is substituted into Eq. (3-13) and the $e^{j\phi(y)}$ term moved to the right side of the equation, Eq. (3-13) becomes

$$V(y) e^{-j\phi(y)} = \lambda A(y) \int_{-H/2}^{H/2} V(t) e^{-j\phi(t)} A(t) \operatorname{sinc}\left[\frac{K}{\lambda Z}(t - y)\right] dt \quad (3-17)$$

for $|y| \leq H/2$.

If $\phi(y)$ is set equal to the phase of $V(y)$, then

$$V(y) e^{-j\phi(y)} = |V(y)| \triangleq R(y).$$

Equation (3-17) thus becomes

$$R(y) = \lambda A(y) \int_{-H/2}^{H/2} R(t)A(t) \operatorname{sinc}\left[\frac{K}{\lambda Z}(t - y)\right] dt \quad (3-18)$$

for $|y| \leq H/2$,

where $R(y)$ and $A(y)$ are real functions. As stated before, Eq. (3-18) is a nonlinear integral equation. One method of solving Eq. (3-18) is to linearize the equation by assuming a solution containing a given dominant function $G(y)$ plus a small unknown variance function $\delta(y)$. The integral equation is then linearized in terms of the unknown variance function $\delta(y)$. The unknown variance function $\delta(y)$ can be solved by a general method of solving a linear integral equation with a standard form. The solution of the variance function $\delta(y)$ is compared to the given function $G(y)$ to make sure that it is indeed much smaller than that function. If it is not, the given function has to be reselected and the whole procedure repeated. The success of this approach depends on the type of input signal $R(y)$ and the choice of $G(y)$. It is also dependent on experience in solving nonlinear integral equations of this type. A computer may be needed to solve the linearized equation. The procedure is complicated and success is not guaranteed for all possible input signals. Since this approach does not yield a satisfactory solution, Question 2 must be considered. Refer to Question 2 on page 80.

In the phrase "... the reconstructed signal $U'(y)$ is the best approximation to the original signal $U^*(y)$ ", "best approximation" indicates that a measure of the difference between signals $U'(y)$ and $U^*(y)$ must be established. However, information from previous experimental studies is insufficient to define a suitable measure consistent with practical system performance. Moreover, even with the mean-square error criterion, the solution to such a problem still remains difficult. Even if the answers to Questions 1 and 2 can be obtained, they do not help very much for a practical viewing system since only one signal can be treated and it is of no interest to reconstruct only one signal in a holographic movie system. This analysis, however, can be applied

to other problems, such as credit card identification. The next section contains a discussion of a more realistic approach in which a periodic scatter plate is considered. The questions proposed earlier are answered by means of this approach.

3.5.2. THE PERIODIC SCATTER-PLATE APPROACH

In the previous section, it was pointed out that the treatment of a single signal does not characterize a practical movie system. In a practical movie system, the input signal could contain a wide range of possible signals. Formulating a mathematical description of the possible input signals for a 3-D object scene is a difficult problem. There is little information on a suitable mathematical description of a typical movie scene. In this section, the set of all possible input signals is assumed to have a certain mathematical simplicity so that the problems are tractable. One of the difficulties encountered in the previous section is the treatment of a continuous function for which case a computer has not been found to be very useful. Moreover, the optimization of a continuous function in an integral expression involves using calculus of variations theory which is hard to implement in a computer program. To avoid these difficulties, a periodic scatter-plate function is assumed. The possible input signals are also assumed periodic. Both the scatter-plate function and the input signal are assumed to be bandlimited so that they can be represented by a finite set of frequency coefficients. To eliminate difficulties with an integral equation, the aperture of the scatter plate is assumed to be much larger than the period of the scatter-plate function. The effect of the scatter-plate aperture on each of the frequency components of the input signal can therefore be neglected. With these assumptions, Eq. (3-11) can be written as

$$V'(y) = p(y) \left[V(y)p(y) \star \text{sinc}(Wy) \right]^* \quad (3-19)$$

where

$$V(y) \triangleq U(y) e^{j(\pi/\lambda Z)y^2}, \quad V'(y) \triangleq U'(y) e^{-j(\pi/\lambda Z)y^2}, \quad W = \frac{k}{\lambda Z}.$$

Functions $V(y)$ and $p(y)$ are assumed periodic with the same period y_0 , i.e.,

$$V(y) = \sum_{\ell=-L}^L a_{\ell} \exp \left(j \frac{2\pi \ell y}{y_0} \right)$$

$$p(y) = \sum_{p=-P}^P b_p \exp \left(j \frac{2\pi p y}{y_0} \right)$$

where $\pm L$ and $\pm P$ are the highest frequency components of functions $V(y)$ and $p(y)$ respectively. They correspond to the spectral bandwidth in the continuous function case; a_{ℓ} and b_p are the frequency coefficients of $V(y)$ and $p(y)$ respectively and are given by

$$a_{\ell} = \frac{1}{y_0} \int_0^{y_0} V(y) \exp \left(-j \frac{2\pi \ell y}{y_0} \right) dy$$

$$b_p = \frac{1}{y_0} \int_0^{y_0} p(y) \exp \left(-j \frac{2\pi p y}{y_0} \right) dy$$

Using Eq. (3-19), the coefficient of $V'(y)$ can be found in terms of a_{ℓ} and b_p .

Letting $C(y) \triangleq V(y)p(y)$, we have

$$\begin{aligned} C(y) = V(y)p(y) &= \sum_{\ell=-L}^L \sum_{p=-P}^P a_{\ell} b_p \exp \left[j \frac{2\pi}{y_0} (\ell + p)y \right] \\ &= \sum_{m=-(L+P)}^{L+P} c_m \exp \left(j \frac{2\pi m y}{y_0} \right) \end{aligned}$$

where

$$c_m = \sum_{\ell=-L}^L a_{\ell} b_{m-\ell}.$$

The operation $C(y) \star \text{sinc}(Wy)$ in Eq. (3-19) corresponds to a frequency cut-off of the frequency spectrum of function $C(y)$ if the frequency bandwidth of function $\text{sinc}(Wy)$ is smaller than the frequency bandwidth of $V(y)$. Let's assume that the maximum frequency after this operation is $\pm M$. Let $D(y) \triangleq C(y) \star \text{sinc}(Wy)$. Then

$$D(y) = C(y) \star \text{sinc}(Wy)$$

$$= \sum_{m=-M}^M c_m \exp\left(j \frac{2\pi m y}{y_0}\right)$$

where, in general, $M < L$.

Equation (3-19) can then be written as

$$\begin{aligned} v'(y) = D^*(y)p(y) &= \sum_{m=-M}^M c_m^* \exp\left(-j \frac{2\pi m y}{y_0}\right) \cdot \sum_{p=-P}^P b_p \exp\left(j \frac{2\pi p y}{y_0}\right) \\ &= \sum_{m=-M}^M \sum_{p=-P}^P c_m^* b_p \exp\left[j \frac{2\pi(p-m)}{y_0} y\right] \\ &= \sum_{n=-(M+P)}^{M+P} e_n \exp\left(j \frac{2\pi n y}{y_0}\right) \end{aligned}$$

where

$$e_n = \sum_{m=-M}^M c_m^* b_{n+m}.$$

e_n can also be written in terms of a_ℓ and b_p .

$$e_n = \sum_{m=-M}^M \sum_{\ell=-L}^L a_\ell^* b_{m-\ell}^* b_{n+m} \quad (3-20)$$

Equation (3-20) represents the basic system operation in terms of the frequency coefficients of the input signal $V(y)$, the output signal $V'(y)$, and the system function $p(y)$. The goal now is to determine the choice of the frequency coefficients b_p for the system function $p(y)$ which yields the best system performance. In other words, the frequency coefficients e_n of the output signal are compared with the frequency coefficients a_ℓ of the input signal to determine the best choice of b_p . Several assumptions are required before this can be done. First, the frequency bandwidth of $p(y)$ or the number P has to be chosen. For this study, P is set equal to L . It should be recognized that a different choice of P will affect the final results. Second, the output signal $V'(y)$ contains higher spatial frequencies than the input signal $V(y)$. In order to compare these two signals, $V'(y)$ is defined as follows:

$$V'(y) \triangleq V_1(y) + V_2(y)$$

where

$$V_1(y) \triangleq \sum_{n=-L}^L e_n \exp \left(j \frac{2\pi n y}{y_0} \right)$$

$$V_2(y) \triangleq \left(\sum_{n=-(M+P)}^{-L} + \sum_{n=L}^{M+P} \right) \left[e_n \exp \left(j \frac{2\pi n y}{y_0} \right) \right]$$

Function $V_1(y)$ contains the same frequency orders as $V(y)$ and is compared with $V(y)$ to obtain a measure of the system resolution. Function $V_2(y)$ contains all the frequency components which are higher than those

in the input signal $V(y)$ and, thus, serves as a measure of the system noise background. The above assumptions can be justified if the frequency bandwidth of the system is limited by the scatter-plate aperture. No reconstructed signal can have better resolution obtained with a diffraction-limited aperture. The higher-order frequencies in the output signal have no relevance to the system resolution at all. They only contribute to the noise of the system.

Comparison of $V_1^*(y)$ with $V(y)$, using a mean-square error criterion as a measure of the difference between $V_1^*(y)$ and $V(y)$, yields the following expression:

$$\rho[V(y), V_1^*(y)] = ||V(y) - V_1^*(y)||^2 = \left[\sum_{n=-L}^L |a_n - e_{-n}^*|^2 \right]^{1/2} \quad (3-21)$$

where $\rho[V(y), V_1^*(y)]$ is the mean-square error of two functions $V(y)$ and $V_1^*(y)$. The mean-square error criterion for two periodic functions, in terms of their frequency coefficients, is defined as a metric in the Hilbert space H , where $V_1^*(y)$ and $V(y)$ are the function elements. A detailed derivation of this approach is given in Appendix B. Using Eq. (3-21) as a measure for the system performance, the four questions listed in the previous section can be answered by making use of the periodic structure of the system.

For Question 1, a set of system parameters b_p must be found so that $\rho = 0$ for a given set a_ℓ characterizes the input signal. In Eq. (3-21), ρ is zero iff each term $|a_n - e_{-n}^*|^2$ is zero. There are $2L + 1$ terms in Eq. (3-21). Examine Eq. (3-20), which is the expression of e_{-n}^* , and compare it with a_ℓ . There are $2L + 1$ equations and $2L + 1$ unknowns b_p . Hence, in general, there exists a solution for the $2L + 1$ (b_p) since there exists $2L + 1$ independent equations. This answers Question 1.

If there is no solution to the above $2L + 1$ equations, Question 2 must be investigated. To answer Question 2, an optimum b_p must be found, so that Eq. (3-21) is minimum. To find the minimum

value of ρ , a partial differentiation of ρ with respect to each unknown b_p is performed with Eq. (3-21). Then, the resulting equation is set to zero. Since each term in the expression for e_{-n}^* is in a form of $b_i b_j$, the above operation results in $2L + 1$ simultaneous linear equations for $2L + 1$ unknowns. In general, there is a unique nonzero solution for such a set of simultaneous linear equations. Question 2 is, therefore, answered in this case. A periodic approach allows one to deal with a set of possible input signals and, hence, to easily answer Questions 3 and 4.

To answer Question 3, a set of parameters b_p must be found so that, for each input signal in a given set of possible input signals, ρ is always equal to zero. This requires the determination of $2L + 1$ unknown b_p from $N(2L + 1)$ equations, where N is the number of possible input signals. Because of the basic properties of simultaneous equations, this is usually impossible to do, except for special cases.

Question 4 can be treated in several ways, depending on whether the input signal is a set containing a finite number of possible signals or a random function with an infinite number of components. If the input signal is a set containing a finite number of possible signals, then for each possible input signal, a corresponding ρ is obtained from Eq. (3-21). In this case, there are several possible criteria for describing the performance of the system, two of which are commonly used. One requires that the sum of all the ρ 's be a minimum; the other requires that each ρ is less than a minimum value ρ_0 . When the input signal is a random function with an infinite number of components, the expected value of the mean-square error is used as the measure of the system performance. The calculation of the expected value of ρ is strongly dependent on the character of the random signal. The intention here is not to treat all the possible input signals (only two of them have been mentioned above). To demonstrate the feasibility of this periodic approach, a simple random input signal case is presented below.

Consider a random input signal such that each frequency coefficient a_ℓ is a random variable with the following properties:

- 1) $E(a_\ell) = 0$ for all ℓ
- 2) $E(a_\ell^2) = \alpha^2$ for all ℓ
- 3) a_i and a_j are independent for all i, j .

Hence,

$$\begin{aligned} E(a_i a_j) &= E(a_i) E(a_j) = 0 & \text{if } i \neq j \\ &= \alpha^2 & \text{if } i = j \end{aligned}$$

For this signal, the expected value of ρ is

$$\begin{aligned} \rho_e &\triangleq E(\rho) = E\left\{\rho[V(y), V_1^*(y)]\right\} \\ &= E\left[\left(\sum_{n=-L}^L |a_n - e_{-n}^*|^2\right)^{1/2}\right] \\ &= \left[\sum_{n=-L}^L E(|a_n - e_{-n}^*|^2)\right]^{1/2} \end{aligned} \tag{3-22}$$

where the expected value of $|a_n - e_{-n}^*|^2$ for each n is given by

$$E(|a_n - e_{-n}^*|^2) = E(|a_n|^2) + E(|e_{-n}^*|^2) - E(a_n^* e_{-n}^*) - E(a_n e_{-n})$$

$$E(|a_n|^2) = \alpha^2$$

$$E(|e_{-n}^*|^2) = \alpha^2 \left| \sum_{i,j}^{\Omega_n} b_i b_j^* \right|^2$$

$$E(a_n^* e_{-n}^*) = \alpha^2 \sum_{i,j}^{\Omega_n} b_i b_j^*$$

$$E(a_n e_{-n}) = \alpha^2 \sum_{i,j}^{\Omega_n} b_i^* b_j$$

where $\sum_{i,j}^{\Omega_n} b_i^* b_j$ and $\sum_{i,j}^{\Omega_n} b_i b_j^*$ represents a sum of certain terms of a

form $b_i b_j$ with i and j belonging to a space Ω_n which is a function of n . In a practical calculation, the above two expressions can be written directly from the given data.

Equation (3-22) now becomes

$$\rho_e = \alpha^2 \left(2L + 1 + \sum_{n=-L}^L \left| \sum_{i,j}^{\Omega_n} b_i b_j^* \right|^2 - \sum_{n=-L}^L \sum_{i,j}^{\Omega_n} b_i b_j^* - \sum_{n=-L}^L \sum_{i,j}^{\Omega_n} b_i^* b_j \right) \quad (3-23)$$

In Eq. (3-23), ρ_e is a function of the $2L + 1$ frequency coefficients (b_p). To find the minimum value of ρ_e in terms of $2L + 1$ frequency coefficients, the same procedure used for Question 2 should be followed. There, a partial differentiation and the solution of simultaneous equations were involved. The above operation is simple if a digital computer is used and the number L or P is small.

3.6. THE FLY'S-EYE LENS ARRAY

In Section 3.4, a model for information reduction of a holographic viewing system has been achieved by a sampling of the recording hologram. The model considers the redundancy not required by the human eye. In Section 3.5, the scatter-plate method for information reduction is analysed mathematically. In the experimental study, the fly's-eye lens array is superior to the other available scatter plates. Further study of the fly's-eye lens array reveals that there is a direct analogy between the fly's-eye lens array system and the sampling method system proposed in Section 3.4. The fly's-eye lens array system has features inherent in both the sampling method and the scatter-plate method. Therefore, a detailed study of the fly's-eye lens array system is highly desirable. In this section, the study is carried out by geometric optics. More refined treatment of the system presented in Chapter 4 is based on the diffraction theory and other considerations.

Consider the fly's-eye lens array system in Fig. 3-7. The system parameters and various assumptions are listed below:

System Parameters:

1. m - Aperture of individual lenslet
2. l - Resolution aperture of object point at lenslet
3. α - Viewing angle
4. A - Observer's aperture
5. K - Hologram size
6. H - Fly's-eye lens array aperture
7. f - Focal length of each lenslet
8. X - Object size
9. d - Distance between the object and the viewer's aperture
10. z - Distance between the object and the fly's-eye lens array
11. Z - Distance between the image and the hologram
12. s - Distance between the fly's-eye lens array and the image
13. D_o - Distance between the fly's-eye lens array and the viewer's aperture

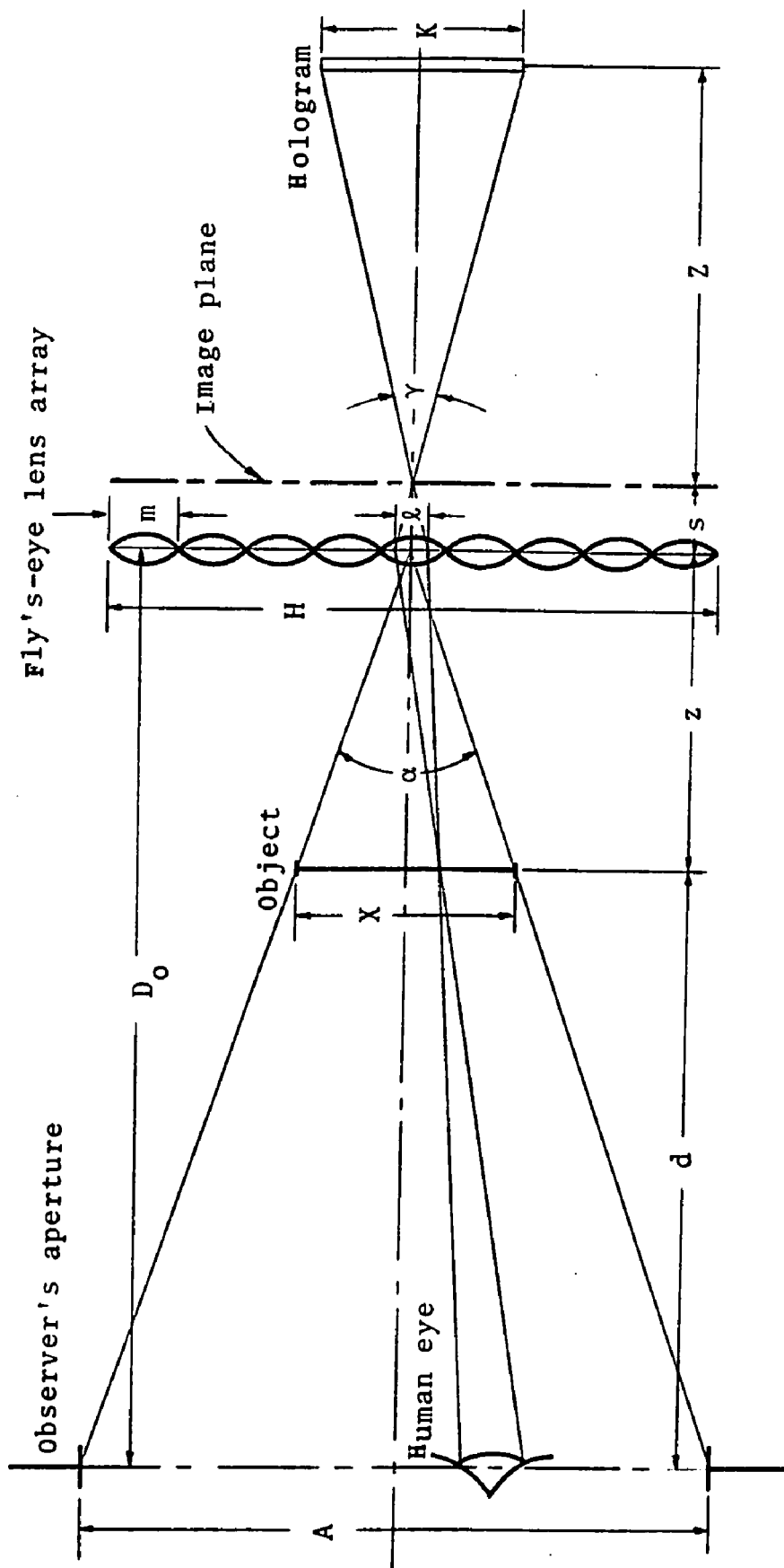


Figure 3-7. The fly's-eye lens array system

Basic Assumptions

1. The analysis is carried out by geometric optics. Each lenslet is assumed to be an ideal, single, thin lens. The inactive area between each lenslet is assumed negligible and the full aperture of each individual lenslet m satisfies the criterion

$$m \geq \frac{D_o}{3600} = \frac{d + z}{3600} . \quad (3-24)$$

The resolution aperture ℓ of an image point reconstructed by the hologram K at each lenslet satisfies

$$\ell \geq 3600\lambda \frac{z}{d} . \quad (3-25)$$

2. The distances D_o and Z are assumed to be much larger than the apertures A , H , and K . Hence, each lenslet is assumed to have the same viewing angle α . The images of the viewer's aperture A formed by each lenslet are assumed to be the same behind each lenslet. They are centered on the axis of each lenslet. The reconstructed resolution aperture ℓ for a fixed hologram size K is assumed to be the same for each lenslet and it is given by

$$\ell = \frac{K}{Z} s . \quad (3-26)$$

Now, for the first assumption, Eqs. (3-24) and (3-25) were given in Section 3.4, where m is the interval between two adjacent recording elements and ℓ is the size of the recording elements on the hologram. In the sampling method system, all the object points are reconstructed by a fixed array of equally spaced apertures ℓ in the hologram plane. In the fly's-eye lens array system, however, the aperture ℓ has to be at a different location within the lenslet aperture m to reconstruct a different object point. This is the major difference between the system performance of the fly's-eye lens array system and the sampling method system. Since the lenslet aperture m is less than $D_o/3600$, the reconstructed object will not appear granular because of the separation of reconstructed rays coming from the adjacent lenslets.

When the second basic assumption is used, the study of the fly's-eye lens array system can be carried out by consideration of a small, single lenslet imaging system, shown in Fig. 3-8; y is the image of the viewer's aperture A formed by the lenslet; β is an angle within which any size hologram is capable of reconstructing each point of the image y within the lenslet aperture m . Given a hologram K , within the angle β , each point of the image y can be reconstructed by the hologram to provide a resolution aperture ℓ within the lenslet aperture. At any particular point inside the viewer's aperture, he can see one object point against each lenslet aperture m . Since m and ℓ satisfy Eqs. (3-24) and (3-25), the viewer sees no degradation of the reconstructed object. The number of lenslets which can be illuminated by the reconstructed wavefront during the reconstruction process depends on the size of β as well as the distance between the hologram and the fly's-eye lens array. For a given hologram K and angle β , the effective fly's-eye lens array aperture H is proportional to the distance Z . Z is approximately equal to the distance between hologram and the fly's-eye lens array since $Z \gg s$. On the other hand, the resolution aperture is inversely proportional to the distance, as given by Eq. (3-26). Hence, there is a compromise between the image resolution and the effective fly's-eye lens array aperture. In the following, the relation between H and ℓ is therefore explored.

β is a function of s , m , and α and is given by

$$\beta = \frac{m}{s} - \alpha \quad (3-27)$$

s is related to the focal length f of the small lenslet and the distance D_0 by the lens formula

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{D_0}$$

The value of f is dependent on the practical, obtainable power of the small lenslet. Hence, in practice, f is fixed for a given fly's-eye lens array which, in turn, determines the value of s for a given D_0 .

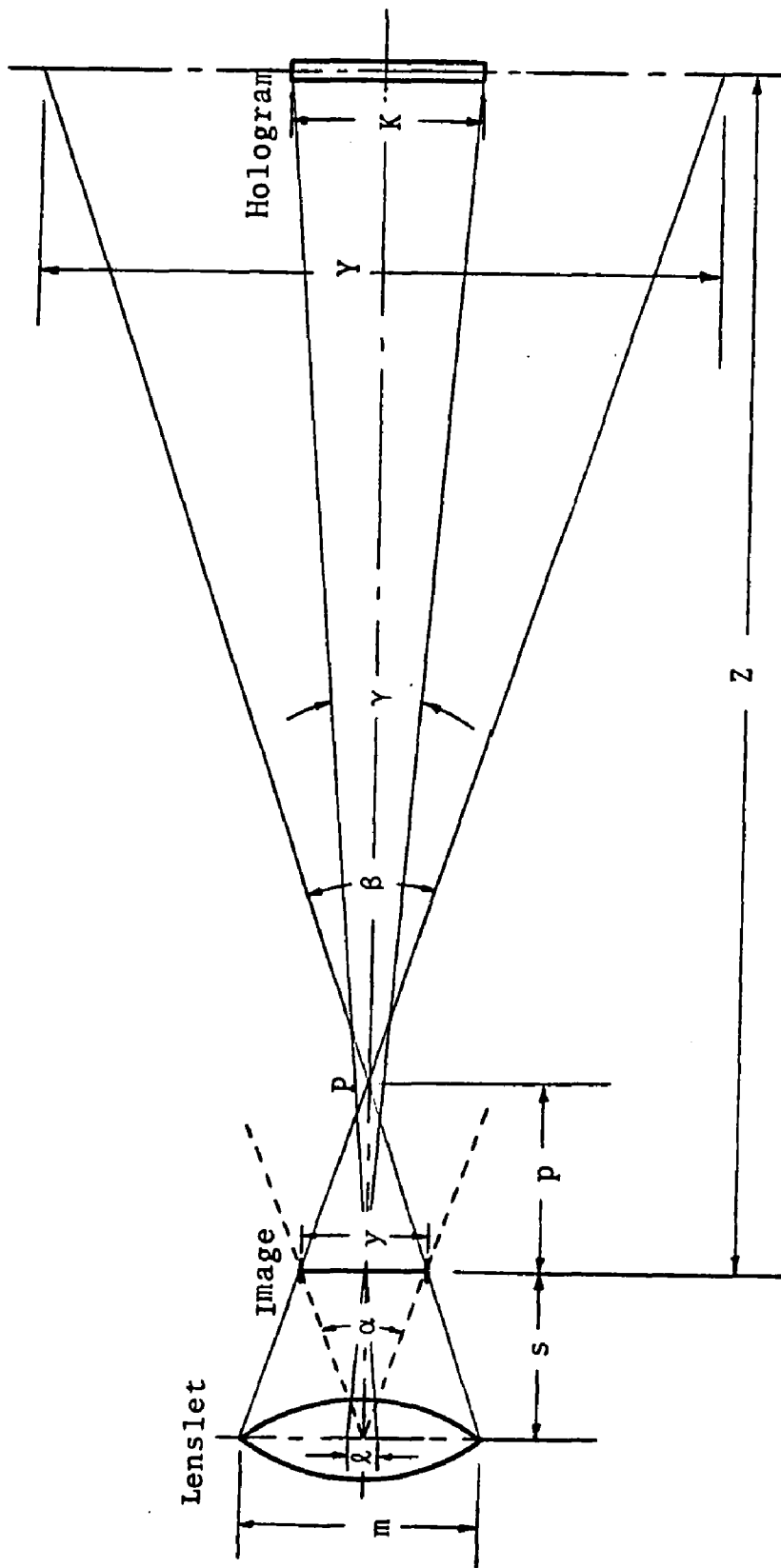


Figure 3-8. The angle β in the study of fly's-eye lens array

Let Y be the aperture at the hologram plane which subtends the angle β at the point P in Fig. 3-8. With the simple geometric relations and Eq. (3-27), Y can be written as

$$Y = m \frac{Z(m - y)\alpha - y^2}{s(m - y)\alpha + y^2} \quad (3-28)$$

where Z is the distance between the image y and the hologram K . For a given fly's-eye lens array, s is fixed and Y can be related to H and ℓ .

The number of effective lenslets is given by $(Y - K)/m$. This is obtained by assuming that a hologram anywhere within the angle β can reconstruct the whole image y within the lenslet aperture. The effective fly's-eye lens array aperture H is therefore

$$H = m \frac{Y - K}{m} = Y - K$$

On the other hand, Y is linearly proportional to Z which is inversely proportional to ℓ by Eq. (3-26). If Eq. (3-26) is substituted into Eq. (3-28), H can be related to ℓ by

$$H = Y - K = K \left(\frac{m - \ell}{\ell} \right) - \left(\frac{K}{\ell} \alpha + \alpha \right) s$$

or

$$H = \frac{K(m - y)}{\ell} - (K + y) \quad (3-29)$$

Equation (3-29) is an important function for the system which relates the resolution aperture to the effective fly's-eye lens array aperture. In case $y \ll K$, Eq. (3-29) becomes

$$\ell = \frac{K(m - y)}{H + K} \quad (3-30)$$

This is the resolution aperture of the fly's-eye lens array system.

For a given fly's-eye lens array system, y is usually determined by the lenslet focal power and the distance D_0 ; D_0 should be such that

Eq. (3-24) is satisfied for the lenslet aperture m . The resolution aperture ℓ of the system is then determined by Eq. (3-29) for a given fly's-eye lens array aperture and the hologram size. The distance Z can then be determined by Eq. (3-26).

In conclusion, the information-reduction factor of this system is given by

$$R = \frac{INF_1}{INF_2} = \frac{AH/\lambda D_o}{HK/\lambda Z} = \frac{A/D_o}{K/Z} = \frac{ys}{s\ell} = \frac{y}{\ell} \div \frac{\alpha}{\gamma} \quad (3-31)$$

where $INF_1 = AH/\lambda D_o$ is the information content which has to be recorded if a hologram with size H is used to replace the fly's-eye lens array and $INF_2 = HK/\lambda Z$ is the information content of the fly's-eye lens array system. The system resolution aperture ℓ is inversely proportional to the information-reduction factor R . In other words, the information-reduction factor R can be increased only by sacrificing other features of the system--the resolution or the viewing angle. Their interaction relationships are given by Eqs. (3-26), (3-29), and (3-31).

CHAPTER 4

THE STUDY OF THE FLY'S-EYE LENS ARRAY SYSTEM

It was pointed out in the previous chapter that the fly's-eye lens array system is the best experimentally demonstrated scatter-plate system and that the performance of the fly's-eye lens array system is similar to the sampling-method system. Since the performance of the sampling-method system is optimum or close to the optimum for a 3-D viewing system, it is worthwhile to investigate the fly's-eye lens array system further. In Sections 4.1 and 4.2, the fly's-eye lens array system's resolution and the signal-to-noise ratio based on the diffraction theory analysis are studied. The aberrations of the small lenslet in this system affect its final resolution, a problem studied in Section 4.3. One of the practical engineering problems, the change from the recording wavelength to the readout wavelength, is considered in Section 4.4. In Section 4.5, the fly's-eye lens array is used to implement a 3-D air-traffic control display system. This application demonstrates one of the interesting properties of the fly's-eye lens array in providing a 3-D system.

4.1. THE RESOLUTION OF THE FLY'S-EYE LENS ARRAY SYSTEM

The performance of a fly's-eye lens array system has been studied in Section 3.6 from the viewpoint of geometric optics. In this section, the system is studied from the viewpoint of the diffraction theory in order to predict the resolution of the system. The signal-to-noise ratio of the system will be studied in the next section. The system model for this study is quite similar to the one shown in Fig. 3-8 and is plotted again in Fig. 4-1.

In geometric optics analysis, a point object is assumed some distance x_1 from the lens array. The rays from the object point form an image point behind each lenslet. These image-forming rays continue to the hologram plane where only some of the rays are recorded. In the

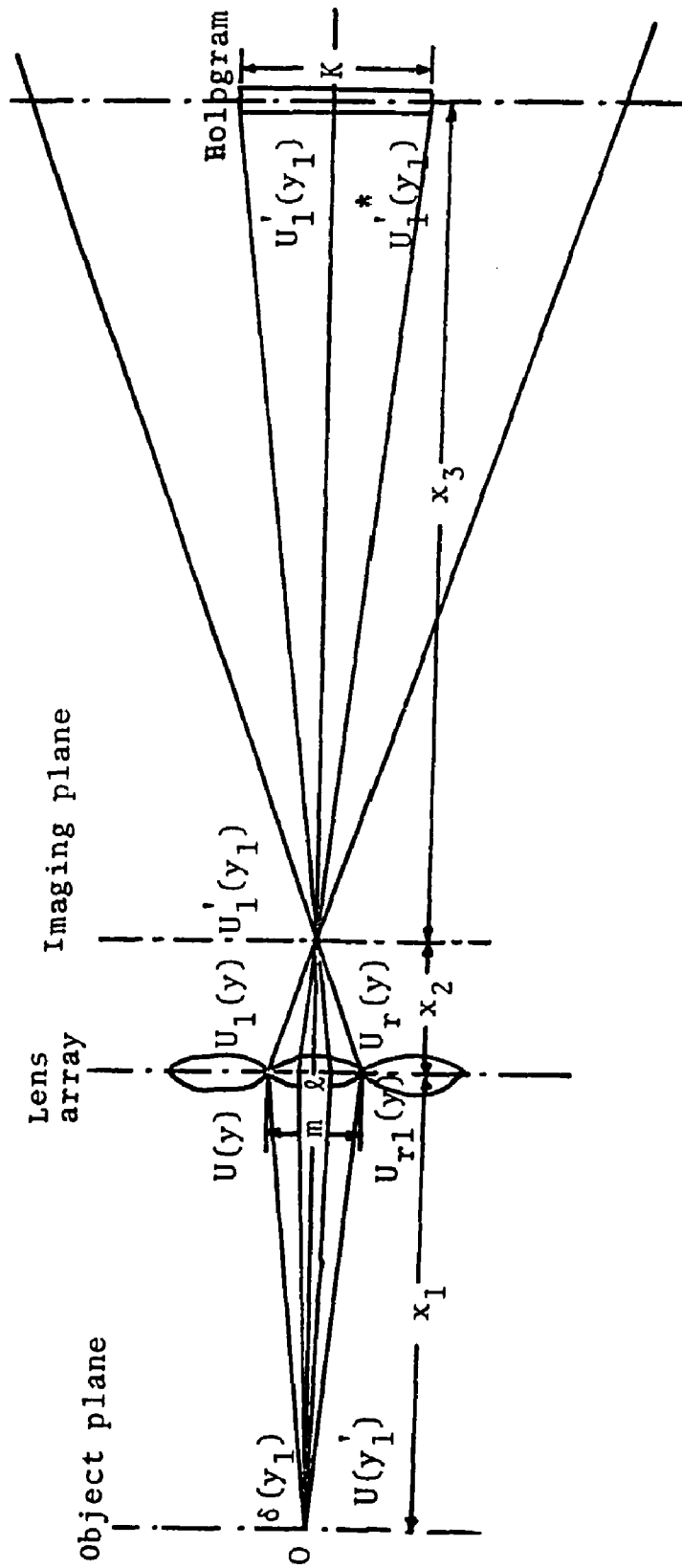


Figure 4-1. System model for the analysis of the fly's-eye lens array

reconstruction process, only those rays which are intercepted by the hologram aperture will be reconstructed to form an aperture ℓ in each small lenslet. These rays then continue on to reconstruct the original object point. The resolution of the constructed point is thus calculated to be $\lambda x_1 / \ell$ because of the effective ray aperture ℓ of each small lenslet. The many reconstruction rays from each lenslet only change the details within the envelope of the intensity pattern given by one single lenslet. Hence, the resolution is determined solely by a single lenslet. Therefore, in the analysis which follows, only a single lenslet is considered.

Starting from the object point, the diffraction effect of the lenslet aperture and the hologram aperture is considered in each aperture plane. The wavefront at each plane has been labeled in Fig. 4-1. The analysis of the system is based on the assumption that both the object point and the hologram are on-axis. If points and the hologram shifted from the central position, the results would be the same.

Consider the point object O in Fig. 4-1. If the Fresnel integral approximation is used, the E-M field in front of the lens aperture m is given by

$$U(y) = C_1 \exp \left(j \frac{k}{2x_1} y^2 \right)$$

where C_1 is a constant and k is the wave number. The transmission function of the lenslet is assumed to be

$$\exp \left(-j \frac{k}{2f} y^2 \right)$$

where f is the focus length of the lenslet. The field emerging to the left of the lenslet is thus

$$\begin{aligned} U_1(y) &= U(y) \exp \left(-j \frac{k}{2f} y^2 \right) = C_1 \exp \left[jk \left(\frac{1}{2x_1} - \frac{1}{2f} \right) y^2 \right] \\ &= C_1 \exp \left(-j \frac{k}{2x_2} y^2 \right) \end{aligned}$$

where x_2 is the distance of the image point from the lens and is given by a single len's image forming relation

$$\frac{1}{f} = \frac{1}{x_1} + \frac{1}{x_2} .$$

If the effect of the lens aperture m is considered and the Fresnel approximation used, the field at any receiving plane x distance away to the left of the lens is

$$\begin{aligned} U'_1(y_1) &= \int_{-m/2}^{m/2} U_1(y) \exp\left[j \frac{k}{2x} (y_1 - y)^2\right] dy \\ &= C_1 \exp\left(j \frac{k}{2x} y_1^2\right) \int_{-m/2}^{m/2} \exp\left[j \frac{k}{2} \left(\frac{1}{x} - \frac{1}{x_2}\right) y^2\right] \exp\left(-j \frac{ky}{x} y\right) dy \end{aligned} \quad (4-1)$$

In Eq. (4-1), the first quadratic phase term, $\exp\left(j \frac{k}{2x} y_1^2\right)$, is just the phase retardation in any receiving plane in reference to the origin of the lens aperture. It is, however, the integral part of Eq. (4-1) which is of major concern. If, for example, $x = x_2$ where the image of the object point is sharply focused, we have

$$U'_1(y_1) = C_1 \exp\left(j \frac{k}{2x_2} y_1^2\right) m \operatorname{sinc}\left(m \frac{y_1}{\lambda x_2}\right)$$

The integral part of Eq. (4-1) yields a real since function which defines the image resolution of a lens system as $\lambda x_2/m$. For planes other than $x = x_2$, the integral part of Eq. (4-1) does not yield a real sinc function, and the intensity pattern is usually wider. The exact shape of the intensity pattern given by the integral part of Eq. (4-1) can be obtained from a table of Fresnel integrals and a point-by-point plot of the curve.

In the following, the problem is approached by using one of the possible approximations [11] to yield a closed form for the integral

$$\int_{-m/2}^{m/2} \exp \left[j \frac{k}{2} \left(\frac{1}{x} - \frac{1}{x_2} \right) y^2 \right] \exp \left(-jk \frac{y_1}{x} y \right) dy \triangleq G(F)$$

$$\triangleq \int_{-m/2}^{m/2} \exp(j\pi B y^2) \exp(-j2\pi F y) dy \quad (4-2)$$

where

$$B \triangleq \frac{1}{\lambda} \left(\frac{1}{x} - \frac{1}{x_2} \right) \quad \text{and} \quad F \triangleq \frac{y_1}{\lambda x} .$$

Equation (4-2) is a function of F and is the Fourier transform of the product of a rectangular window and $\exp(j\pi B y^2)$. Both transforms are known. In order to obtain $G(F)$, we have to convolve the two transforms by approximating one of the transforms by a Gaussian curve. The Fourier transform z_1 of the rectangular window is $z_1 = [\sin(\pi m F)] / \pi m F$. The Fourier transform z_2 of $\exp(j\pi B y^2)$ is $z_2 = C \exp(-j\pi F^2 / B)$. To make the convolution of z_1 with z_2 , we approximate $z_1 = \exp(-2.04 m^2 F^2)$. This approximation, shown with the exact function in Fig. 4-2 is quite good in the main lobe. For the convolution, we then obtain $z_1 * z_2$ and Eq. (4-2) becomes

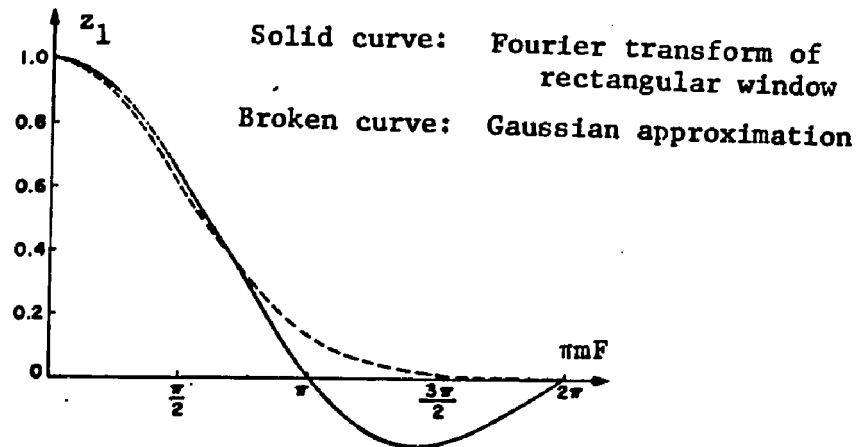


Figure 4-2. The Gaussian approximation of a since function

$$\begin{aligned}
 G(F) &= \int_{-m/2}^{m/2} \exp(j\pi B y^2) \exp(-j2\pi F y) dy \\
 &\approx K \exp\left(-\pi F^2 \frac{0.65m^2 + j 0.42m^4 B}{1 + 0.42m^4 B^2}\right) \\
 &= K \exp\left(-\pi F^2 \frac{0.65m^2}{1 - j 0.65m^2 B}\right)
 \end{aligned} \tag{4-3}$$

where

$$K = m \left(\frac{1}{1 - j 0.65m^2 B} \right)^{1/2}$$

is just a constant which does not interest us at this time. The exact derivation of Eq. (4-3) and some discussion on its validity will be given in Appendix C. The amplitude of Eq. (4-3) is a Gaussian function of the form $\exp(-at^2)$. If this function is approximated by a sinc function like that in Fig. 4-2, we have the following relation:

$$\exp(-at^2) \triangleq \exp(-2.04m^2 t^2) \approx \frac{\sin(\pi m t)}{\pi m t} \tag{4-4}$$

The width of the main lobe of sinc function $\frac{1}{m}$ is generally used as a measure of the resolution of an imaging system. If the same criterion is applied to the Gaussian function and Eq. (4-4) is used, the corresponding resolution of a Gaussian function $\exp(-at^2)$ is

$$R = -\frac{1}{m} = \left(\frac{2.04}{a} \right)^{1/2}$$

Applying this equation to the amplitude of Eq. (4-4), we have the resolution of $G(F)$

$$R = \left(\frac{1 + 0.42m^4 B^2}{m^2} \right)^{1/2}$$

If $F \triangleq \frac{y_1}{\lambda x}$ is used and the resolution of $G(F) = U'_1(y_1)$ in terms of the y_1 variable is desired, we have

$$R = \left[\frac{1 + 0.42 \frac{m^4 B^2}{2}}{m^2} (\lambda x)^2 \right]^{1/2} = \left\{ \left(\frac{\lambda x}{m} \right)^2 + 0.42 \left[\frac{m(x - x_2)}{x_2} \right]^2 \right\}^{1/2} \quad (4-5)$$

Equation (4-5) is the resolution of the field $U_1'(y_1)$ at any receiving plane x distance away from the lens aperture m . As we predicted before, if $x = x_2$

$$R = \frac{\lambda x_2}{m}$$

Let $x = x_2 + x_3$ whereby Eq. (4-5) becomes

$$R = \left\{ \left[\frac{\lambda(x_2 + x_3)}{m} \right]^2 + \left(0.65 \frac{m x_3}{x_2} \right)^2 \right\}^{1/2} \quad (4-6)$$

In Eq. (4-6), the first term is the resolution caused by diffraction effects, while the second term resembles the width of the beam predicted by geometric optics--except for the constant 0.65. The existence of a constant 0.65 in the second term is not surprising since the measure of the resolution is chosen arbitrarily and the first term cannot be neglected completely. Hence, Eq. (4-6) serves as a reasonable measure for the beamwidth in any receiving plane.

The field at the hologram plane is also given by Eq. (4-2), except that we now have the case of $x = x_2 + x_3$ with $x_3 \gg x_2$. Since $x_3 \gg x_2$, the Gaussian amplitude wavefront must be very broad. Hence, for a hologram with very small aperture, the amplitude of $U_1'(y_1)$ at the hologram aperture can be regarded as constant. This leaves $U_1'(y_1)$, given by Eq. (4-2), rather complicated. Since the hologram aperture K is very small, we can assume the resolution of the image point at $x = x_2$, given by lens aperture m , is much smaller than the resolution which will be reconstructed by hologram aperture K at image point $x = x_2$. Under this assumption, the field at the hologram aperture can be written as

$$\begin{aligned} U_1'(y_1) &= C \exp\left(j \frac{k}{2x_3} y_1^2\right) \\ &= C \exp\left(j \frac{\pi}{\lambda x_3} y_1^2\right) \end{aligned}$$

which is just the field caused by a point object. Mathematically, we can assume $0.65m^2B \gg 1$ in Eq. (4-3); then, Eq. (4-3) becomes

$$\begin{aligned} U_1'(y_1) &= C \exp\left(\frac{\pi F^2}{jB}\right) \\ &= C \exp\left(j \frac{\pi}{\lambda x_3} y_1^2\right) \end{aligned}$$

which leads to the same phase function.

This phase function is recorded by the hologram and is then reconstructed back to the fly's-eye lens. The field reconstructed by the hologram, at any receiving plane to the right of the hologram, is denoted by $U_r(y)$ and is given by

$$U_r(y) = \int_{-K/2}^{K/2} \exp\left(-j \frac{\pi}{\lambda x_3} y_1^2\right) \exp\left[j \frac{\pi}{\lambda x} (y - y_1)^2\right] dy_1 \quad (4-7)$$

Equation (4-7) is essentially the same expression as Eq. (4-1), except that the lens aperture m has been changed to the hologram aperture K and the role of x_2 becomes x_3 . Therefore, the results we discussed previously for Eq. (4-4) can be applied to Eq. (4-7) as well. According to Eq. (4-6), the resolution for $U_r(y)$ at $x = x_3$ is given by

$$R = \frac{\lambda x_3}{K}$$

while the resolution of the wavefront in the lenslet aperture is given by

$$R = \left\{ \left[\frac{\lambda(x_3 + x_2)}{K} \right]^2 + \left(0.65 \frac{Kx_2}{x_3} \right)^2 \right\}^{1/2} \quad (4-8)$$

It is very interesting to consider the optimum K which will yield minimum resolution in front of lenslet aperture m . This can be found by taking the derivative of R in Eq. (4-8) with respect to K and solving for K from $dR/dK = 0$. The optimum K is thus found to be

$$K_{\text{opt}} = \left[\frac{\lambda(x_2 + x_3)x_3}{0.65x_2} \right]^{1/2}$$

With $K = K_{\text{opt}}$, the minimum R is given by

$$\begin{aligned} R = R_{\text{min}} &= \left[\frac{0.65\lambda(x_3 + x_2)x_2}{x_3} + \frac{0.65\lambda(x_2 + x_3)x_2}{x_3} \right]^{1/2} \\ &= \left[\frac{1.3\lambda(x_2 + x_3)x_2}{x_3} \right]^{1/2} \end{aligned}$$

The minimum resolution occurs when the first term equals the second term in Eq. (4-8), i.e., when the resolution caused by diffraction effect equals the resolution given by geometric optics.

To proceed further, we write down the expression of field $U_r(y)$ in front of the lenslet aperture m by using Eq. (4-3) in Eq. (4-7):

$$\begin{aligned} U_r(y) &= \exp \left[\frac{-2.04K^2}{(\lambda x_4)^2 (1 + 0.42K^4 B)} y^2 \right] \\ &\quad \exp \left\{ j \left[\frac{\pi}{\lambda x_4} + \frac{0.42K^4 B}{(\lambda x_4)^2 (1 + 0.42K^4 B)} \right] y^2 \right\} \end{aligned}$$

where

$$x_4 \triangleq x_2 + x_3, \quad B \triangleq \frac{x_2}{\lambda x_4 x_3}.$$

If we further define

$$a \triangleq \frac{2.04K^2}{(\lambda x_4)^2 (1 + 0.42K^4 B^2)}$$

$$b \triangleq 1 + \frac{0.42K^4 B}{(\lambda x_4) (1 + 0.42K^4 B^2)},$$

then the above expression for $U_r(y)$ can be written as

$$U_r(y) = \exp(-ay^2) \exp\left(j \frac{\pi}{\lambda x_4} by^2\right) \quad (4-9)$$

To find the reconstructed electrical magnetic field at the object point after $U_r(y)$ pass through the lenslet function again, we have

$$\begin{aligned} U(y'_1) &= \int_{-m/2}^{m/2} U_r(y) \exp\left(-j \frac{\pi}{\lambda f} y^2\right) \exp\left[j \frac{\pi}{\lambda x_1} (y - y'_1)^2\right] dy \\ &= \exp\left(j \frac{\pi}{\lambda x_1} y'^2_1\right) \int_{-m/2}^{m/2} \exp\left\{j\pi \left[\left(\frac{1}{\lambda x_1} - \frac{1}{\lambda f} + \frac{b}{\lambda x_4}\right) + j \frac{a}{\pi}\right] y^2\right\} \\ &\quad \exp\left[-j2\pi\left(\frac{y'_1}{\lambda x_1}\right) y\right] dy \end{aligned} \quad (4-10)$$

The integral in Eq. (4-10) has the same form as Eq. (4-3) and thus can be simplified to give a closed form as follows:

$$U(y'_1) = C \exp\left\{-\frac{2.04m^2 \left(1 + 0.65m^2 \frac{a}{\pi}\right)}{(\lambda x_1)^2 \left[\left(1 + 0.65m^2 \frac{a}{\pi}\right)^2 + \left(0.65m^2 d\right)^2\right]} y'^2_1\right\} \times$$

$\exp(j \text{ phase term})$

where

$$d \triangleq \frac{1}{\lambda x_1} - \frac{1}{\lambda f} + \frac{b}{\lambda x_4} \triangleq \frac{b}{\lambda x_4} - \frac{1}{\lambda x_2}$$

The resolution of $U(y_1')$ is thus given by

$$R = \left\{ \frac{(\lambda x_1)^2 \left[\left(1 + 0.65m^2 \frac{a}{\pi} \right)^2 + (0.65m^2 d)^2 \right]}{m^2 \left(1 + 0.65m^2 \frac{a}{\pi} \right)} \right\}^{1/2}$$

$$= \left[\left(\frac{\lambda x_1}{m} \right)^2 + (\lambda x_1)^2 0.65 \frac{a}{\pi} + \frac{(\lambda x_1)^2 0.42m^2 d^2}{1 + 0.65m^2 \frac{a}{\pi}} \right]^{1/2} \quad (4-11)$$

Equation (4-11) is the final resolution of the reconstructed object point.

However, each term within Eq. (4-11) has its physical meaning. The first term $(\lambda x_1/m)^2$ results from the lenslet aperture m predicted by diffraction theory. The second term $(\lambda x_1)^2 0.65 \frac{a}{\pi}$ can be written in a different form. This term results from the amplitude of $U_r(y)$ in Eq. (4-9) where the resolution of the amplitude $\exp(-ay^2)$ is $R = (2.04/a)^{1/2}$. Hence,

$$a = 2.04/R_r^2$$

Substituting the above equation into $(\lambda x_1)^2 0.65 \frac{a}{\pi}$, we have

$$(\lambda x_1)^2 0.65 \frac{a}{\pi} = (0.65\lambda x_1/R_r)^2 \quad (4-12)$$

Equation (4-12) has the same form of resolution predicted by diffraction theory for a aperture R_r , except for a constant 0.65. Hence, this term can be said to result from the effect of the resolution of the reconstructed wavefront in front of the lenslet which, in turn, is determined by the hologram aperture; R_r is a function of K .

The third term

$$\frac{(\lambda x_1)^2 0.42m^2 d^2}{1 + 0.65m^2 \frac{a}{\pi}}$$

represents the out-of-focus resolution. It can be written as

$$\frac{(\lambda x_1)^2 0.42m^2 d^2}{1 + \left(\frac{0.65m}{R_r}\right)^2}$$

Both R_r and d are functions of K . When K is very large, $R_r \rightarrow \infty$ and $d \rightarrow 0$, so this term can be omitted in Eq. (4-11). When K is very small, $R_r \rightarrow \infty$ and $d \rightarrow -\frac{1}{\lambda x_2}$. The above expression then becomes

$$\left(\frac{0.65mx_1}{x^2}\right)^2. \text{ The term } \left(\frac{0.65mx_1}{x^2}\right)^2 \text{ represents an out-of-focus resolution.}$$

Hence, Eq. (4-11) can be written as

$$R = \begin{cases} \left[\left(\frac{\lambda x_1}{m}\right)^2 + \left(\frac{0.65\lambda x_1}{R_r}\right)^2 + \left(\frac{0.65mx_1}{x_2}\right)^2 \right]^{1/2} & \text{for small } K \\ \left[\left(\frac{\lambda x_1}{m}\right)^2 + \left(\frac{0.65\lambda x_1}{R_r}\right)^2 + \frac{(\lambda x_1)^2 0.42m^2 d^2}{1 + \left(\frac{0.65m}{R_r}\right)^2} \right]^{1/2} & \text{for medium } K \quad (4-13) \\ \left[\left(\frac{\lambda x_1}{m}\right)^2 + \left(\frac{0.65\lambda x_1}{R_r}\right)^2 \right]^{1/2} & \text{for large } K \end{cases}$$

where

$$R_r = \left[\left(\frac{\lambda x_4}{K}\right)^2 + \left(0.65K \frac{x_2}{x_3}\right)^2 \right]^{1/2}$$

Therefore, the final resolution is thus determined by m , K ,

x_1, x_2, x_3, x_4 in a closed form. Curves can be plotted by using a computer to compare its values with those predicted by geometric optics. The above formula, like Eq. (4-13), is only valid in a Fresnel region, i.e., $x_1 > m$ and $x_1 > K$. Hence, in Eq. (4-13), you can't let $x_1 = 0$ and expect $R = R_r$. Anytime the Fresnel approximation integral is used, it is always assumed that the distance between the two planes considered is larger than the aperture of the fields.

4.2. THE SIGNAL-TO-NOISE RATIO OF THE FLY'S-EYE LENS ARRAY SYSTEM

Consider the noise caused by the diffraction effect of a small hologram aperture. More precisely, a point object O in Fig. 4-1 forms a point image behind a small lenslet aperture m . This point image is then reconstructed by a small hologram aperture K . As a result of the diffraction effect of this limited hologram aperture, the reconstructed wavefront $U_r(y)$ in front of the lenslet aperture m has energy outside that aperture. Those rays which are within m will come to a focus to form a reconstructed object point, even though the focusing may not be perfect. However, there is no chance that those rays outside the lenslet aperture m can help form the image point. They are scattered as a noise background and are defined as the noise of the system. We define those rays within m as signals, even though not all the rays will give the real signal brightness by focusing at the same point.

With the above definitions, we can proceed to derive the expression of the signal-to-noise ratio for the system. Refer to Eq. (4-9), in which the $U_r(y)$ in front of the lenslet aperture m is given. The intensity of the wavefront is thus

$$\begin{aligned} |U_r(y)|^2 &= U_r(y)U_r(y)^* = \exp(-2ay^2) \\ &= \exp\left[-\frac{4.08K^2}{(\lambda x_4)^2(1 + 0.42K^2B^2)} y^2\right] \end{aligned}$$

The power which is incident on m is

$$S \triangleq (2a/\pi)^{1/2} \int_{-m/2}^{m/2} \exp(-2ay^2) dy \quad (4-14)$$

where

$$(\pi/2a)^{1/2} = \int_{-\infty}^{\infty} \exp(-2ay^2) dy$$

is the factor normalizing the total power of the wavefront. The signal-to-noise ratio is given by

$$\frac{S}{N} \triangleq \frac{(2a/\pi)^{1/2} \int_{-m/2}^{m/2} \exp(-2ay^2) dy}{1 - (2a/\pi)^{1/2} \int_{-m/2}^{m/2} \exp(-2ay^2) dy} \quad (4-15)$$

Equation (4-14) can be simplified as follows:

$$\begin{aligned} S &\triangleq (2a/\pi)^{1/2} \int_{-m/2}^{m/2} \exp(-2ay^2) dy = \frac{m(2a)^{1/2}}{2} \int_0^1 \exp(-t^2) dt \\ &\triangleq \operatorname{erf} \left\{ \frac{m(2a)^{1/2}}{2} \right\} \end{aligned}$$

and Eq. (4-15) thus becomes

$$\frac{S}{N} = \frac{\operatorname{erf} [m(2a)^{1/2}/2]}{1 - \operatorname{erf} [m(2a)^{1/2}/2]} \quad (4-16)$$

where

$$a \triangleq \frac{2.04K^2}{(\lambda x_4)^2 (1 + 0.42K^4 B^2)}$$

Equation (4-16) is a monotonic increasing function of $m(2a)^{1/2}/2$. This means the signal-to-noise ratio will increase as either m or $\sqrt{2a}$ increases, $\sqrt{2a}$ is a function of K given by

$$(2a)^{1/2} = \frac{2.02K}{(\lambda x_4) \left(1 + 0.42K^4 B\right)^{1/2}}$$

It can be seen that

$$(2a)^{1/2} \rightarrow 0, \text{ as } K \rightarrow 0$$

$$(2a)^{1/2} \rightarrow 0, \text{ as } K \rightarrow \infty$$

Hence, there must be a value of K where $\sqrt{2a}$ is maximum which, in turn, makes the S/N maximum. To find this K , we set $d\sqrt{2a}/dK = 0$ and solve for K . The optimum K is

$$K = K_{\text{opt}} = \left(\frac{\lambda x_4 x_3}{0.65 x_2} \right)^{1/2} \quad (4-17)$$

Equation (4-17) is exactly the same as the optimum K , where the resolution of $U_r(y)$ in front of m is minimum. This is not surprising, since $\sqrt{2a}$ is inversely related to the resolution and the conclusion satisfies our intuitive feeling; i.e., for a fixed lenslet aperture m , the maximum signal energy is obtained when the resolution is a minimum.

Equation (4-16) can also be written in terms of m and R_r

$$\frac{S}{N} = \frac{\text{erf} \left[m(1.02)^{1/2} / R_r \right]}{1 - \text{erf} \left[m(1.02)^{1/2} / R_r \right]} \approx \frac{\text{erf}(m/R_r)}{1 - \text{erf}(m/R_r)} \quad (4-18)$$

Comparing Eq. (4-18) and Eq. (4-13), we note that we cannot increase both the system resolution and its signal-to-noise ratio by changing

R_r , but we can do this by increasing the lenslet aperture m . The limitation of lenslet aperture is imposed by other factors.

4.3. THE EFFECT OF THE THIRD-ORDER ABERRATIONS ON THE FLY'S-EYE LENS ARRAY SYSTEM

In the previous two sections, the study of the performance of the fly's-eye lens array system assumed that each small, individual lenslet is a diffraction-limited single thin lens. In practice, however, the small lenslet will not be aberration free because of fabrication difficulties. Therefore, severe aberrations which affect the resolution of the final image always exist on each lenslet.

In Section 4.1, the resolution of the final image is represented by a single, circular aperture ℓ in the lenslet aperture m for an aberration-free lenslet. When third-order aberrations exist in the lenslet and when the hologram only captures part of the ray for a single object point through the lenslet aperture, the reconstructed aperture for a point object will no longer be a solid circular aperture. The exact shape of the reconstructed aperture depends on the lenslet aberration and the hologram location. In this section, the effects of each third-order aberration term on the image resolution are studied. A similar system model as shown in Fig. 4-1 is used, but the notations are different for presentation convenience. The system model is plotted in Fig. 4-3.

Using the notation given by O'Neil in his book [1], and depicting his Fig. 4-1 as Fig. 4-4 here, we can proceed to attack our problem. The aberration function $\Delta(u, v, h)$ of a symmetrical optical system (a lens) is given by

$$\begin{aligned} \Delta(\rho, \phi, h) = & a_0 + b_0 h^2 + b_1 \rho^2 + b_2 h \rho \cos \phi + c_0 h^4 + c_1 \rho^4 \\ & + c_2 h^2 \rho^2 \cos^2 \phi + c_3 h^2 \rho^2 + c_4 h^3 \rho \cos \phi + c_5 h \rho^3 \cos \phi \\ & + d_0 h^6 + d_1 \rho^6 + \dots \end{aligned} \quad (4-19)$$

where

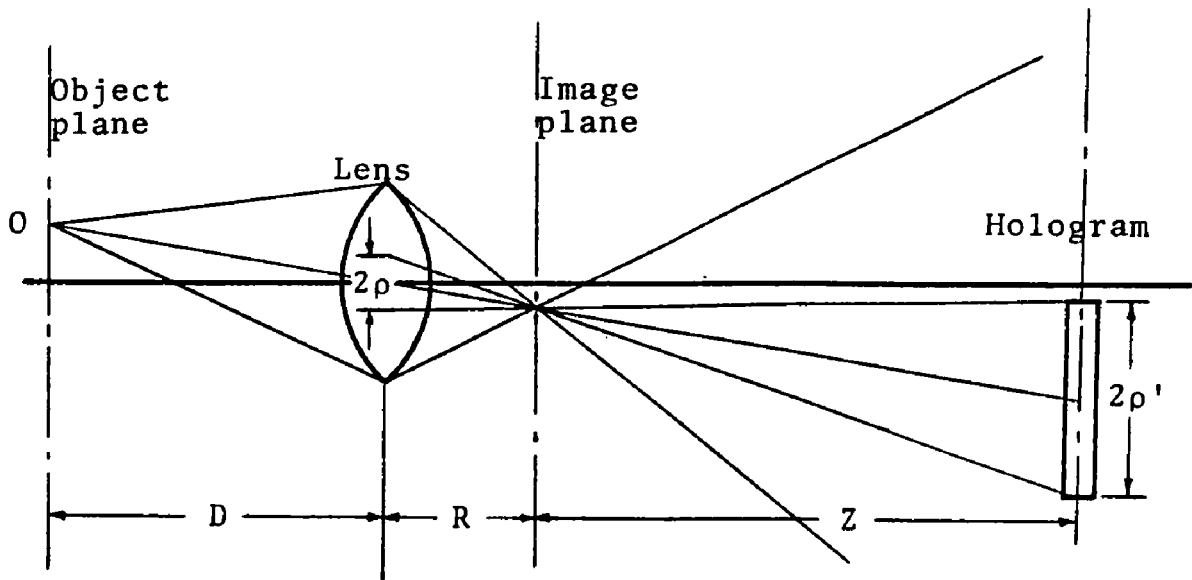


Figure 4-3 The single-lens optical system model

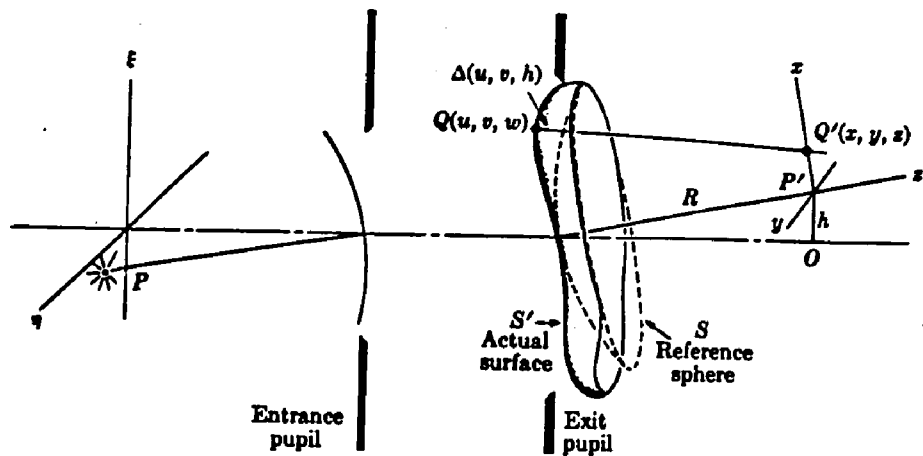


Figure 4-4 The optical system for aberration study

$$u = \rho a \cos \phi, \quad \rho = \sqrt{(u^2 + v^2)/a^2} \quad 0 \leq \rho \leq 1$$

$$v = \rho a \sin \phi, \quad \phi = \tan^{-1} v/u \quad a = \text{radius of the exit pupil} \\ (a = m \text{ in our case})$$

The equation of a ray leaving the exit pupil at Q, normal to the wave-front and intersecting the receiving plane at Q', is given by

$$\frac{x - u}{u - R \left(\frac{\partial \Delta}{\partial u} \right)} = \frac{y - v}{v - R \left(\frac{\partial \Delta}{\partial v} \right)} = \frac{z - \omega}{\omega} \quad (4-20)$$

The Gaussian image point P' is the origin of our coordinate system (x, y, z). Hence, the point (x = 0, y = 0, z = 0) refers, in general, to an off-axis image point P', and not to the axial image point 0. With the approximation $\omega \doteq R$, Eq. (4-20) can be written as

$$x \doteq - \frac{z}{R}(u) + (R + z) \frac{\partial \Delta}{\partial u} \quad (4-21a)$$

$$y \doteq - \frac{z}{R}(v) + (R + z) \frac{\partial \Delta}{\partial v} \quad (4-21b)$$

Equation (4-21) is the fundamental equation relating the coordinates (u, v) in the lens aperture to the coordinates (x, y) at any receiving plane z. When z = 0, Eq. (4-21) reduces to

$$x_0 \doteq R \left(\frac{\partial \Delta}{\partial u} \right) \quad (4-22a)$$

$$y_0 \doteq R \left(\frac{\partial \Delta}{\partial v} \right) \quad (4-22b)$$

Using Eq. (4-22) and considering a ring of rays $\rho = \text{const}$ in the lens aperture, we can plot the corresponding curves traced out in the Gaussian receiving plane. These curves are the so-called aberration curves and, for different terms in $\Delta(u, v, h)$, they represent different

imperfect images. The third-order Seidel aberrations result from the following terms in $\Delta(u, v, h)$:

- | | | |
|----|---|---------------------------------|
| 1) | $c_1 \rho^4$ | spherical aberration |
| 2) | $c_2 h^2 \rho^2 \cos^2 \phi + c_3 h^2 \rho^2$ | astigmatism and field curvature |
| 3) | $c_4 h^3 \rho \cos \phi$ | distortion |
| 4) | $c_5 h \rho^3 \cos \phi$ | coma |

For each of the above Seidel aberrations, the corresponding aberration curves have been studied extensively [8, 25]. Here we reverse the order.

Assuming a given size and shape of a hologram in any receiving plane away from the Gaussian image plane, the corresponding curve in the lens aperture is traced out for each Seidel aberration term. Direct investigation can be accomplished with Eq. (4-21), in which the relation between (x, y, z) and (u, v) has been uniquely established for each term in Δ . Given a well-defined curve function, $f(x, y) = 0$, in a receiving plane, the corresponding curve in the lens aperture can be obtained by substituting Eq. (4-21) into $f(x, y) = 0$. Moreover, since the circular aperture of a hologram is commonly used in discussing the resolution performance of the system, we can also define a polar coordinate (ρ', ϕ') in the (x, y) plane; i.e.:

$$\begin{aligned} x &= \rho' \cos \phi' \\ y &= \rho' \sin \phi' \end{aligned} \tag{4-23}$$

With Eq. (4-23) above, the corresponding relation between ρ' and (u, v) or (ρ, ϕ) is found to be

$$\begin{aligned} \rho'^2 &= \left(\frac{z}{R}\right)^2 (u^2 + v^2) + (R + z)^2 \left[\left(\frac{\partial \Delta}{\partial u}\right)^2 + \left(\frac{\partial \Delta}{\partial v}\right)^2 \right] \\ &\quad - 2\left(\frac{z}{R}\right)(R + z) \left[u \left(\frac{\partial \Delta}{\partial u}\right) + v \left(\frac{\partial \Delta}{\partial v}\right) \right] \end{aligned} \tag{4-24}$$

or

$$\rho'^2 = \left(\frac{z}{R}\right)^2 a^2 \rho^2 + (R+z)^2 \left[\left(\frac{\partial \Delta}{a \partial \rho}\right)^2 + \left(\frac{\partial \Delta}{a \rho \partial \phi}\right)^2 \right] - 2 \left(\frac{z}{R}\right) (z+R) \rho \left(\frac{\partial \Delta}{\partial \rho}\right) \quad (4-25)$$

Equations (4-24) and (4-25) are very convenient if we have a circular hologram aperture centered at the origin, i.e., $\rho' = \text{constant}$

A discussion of the effect of each third-order aberration on the system resolution follows:

1) Spherical Aberration: $\Delta = c_1 \rho^4$

For this aberration, consider a circular hologram centered at the origin of a given receiving plane z . Equation (4-24) becomes

$$\rho'^2 = \left[\left(\frac{z}{R}\right) a \rho - (z+R) \left(\frac{\partial \Delta}{a \partial \rho}\right) \right]^2$$

or

$$\begin{aligned} \rho' &= \left(\frac{z}{R}\right) a \rho - (z+R) \left(\frac{\partial \Delta}{a \partial \rho}\right) \\ &= a \rho \left[\frac{z}{R} - (z+R) 4c_1 \left(\frac{\rho}{a}\right)^2 \right] \end{aligned} \quad (4-26)$$

where ρ' is the radius of the hologram aperture and $a \rho$ is the radius in the lens aperture. A close look at Eq. (4-26) shows that the first term on the right is the term predicted by Gaussian optics without aberrations; hence, the effect of the spherical aberration on the system resolution is well represented by the second term. In the second term, c_1 is the aberration coefficient which can be represented by

$$c_1 = \frac{a^2}{4R^2} \delta L \quad (4-27)$$

where δL is the maximum deviation for the marginal rays $\rho = 1$ at its Gaussian image point. With Eq. (4-27) in (4-26), we get

$$\rho' = a\rho \left[\frac{z}{R} - \frac{(z+R)}{R^2} \delta L \rho^2 \right] \quad (4-28)$$

Since $z \gg R$, Eq. (4-28) becomes

$$\begin{aligned} \rho' &= a\rho \left(\frac{z}{R} - \frac{z}{R} \frac{\delta L}{R} \rho^2 \right) \\ &= a\rho \frac{z}{R} \left(1 - \frac{\delta L}{R} \rho^2 \right) \end{aligned} \quad (4-29)$$

It can be seen from Eq. (4-29) that unless $\frac{\delta L}{R} \ll 1$, i.e., c_1 is very small, the effect of spherical aberration on the system resolution cannot be neglected. Moreover, depending on where δL is positive or negative, the effect on resolution can be either good or bad. Figure 4-5 shows the plot of ρ' versus ρ for $\frac{\delta L}{R} = \pm 0.1, \pm 0.2$, and ± 0.5 .

2) Astigmatism and Field Curvature: $c_2 h^2 \rho^2 \cos^2 \phi + c_3 h^2 \rho^2$

For this aberration, if we consider the same hologram considered in the case of spherical aberrations, Eq. (4-25) becomes

$$\rho'^2 = u^2 \left[\frac{z}{R} - (R+z)(c_2 + c_3) \left(\frac{2h^2}{a^2} \right) \right]^2 + v^2 \left[\frac{z}{R} - (R+z) c_3 \left(\frac{2h^2}{a^2} \right) \right]^2$$

or

$$\frac{\frac{u^2}{\left[\frac{z}{R} - (c_2 + c_3)(R+z) \left(\frac{2h^2}{a^2} \right) \right]^2}}{\rho'} + \frac{\frac{v^2}{\left[\frac{z}{R} - c_3(R+z) \left(\frac{2h^2}{a^2} \right) \right]^2}}{\rho'} = 1 \quad (4-30)$$

This equation represents an ellipse on the lens aperture. The semi-major axis is

$$\rho_{\text{major}} = \frac{\rho'}{\frac{z}{R} - (c_2 + c_3)(R+z) \left(\frac{2h^2}{a^2} \right)} \quad (4-31a)$$

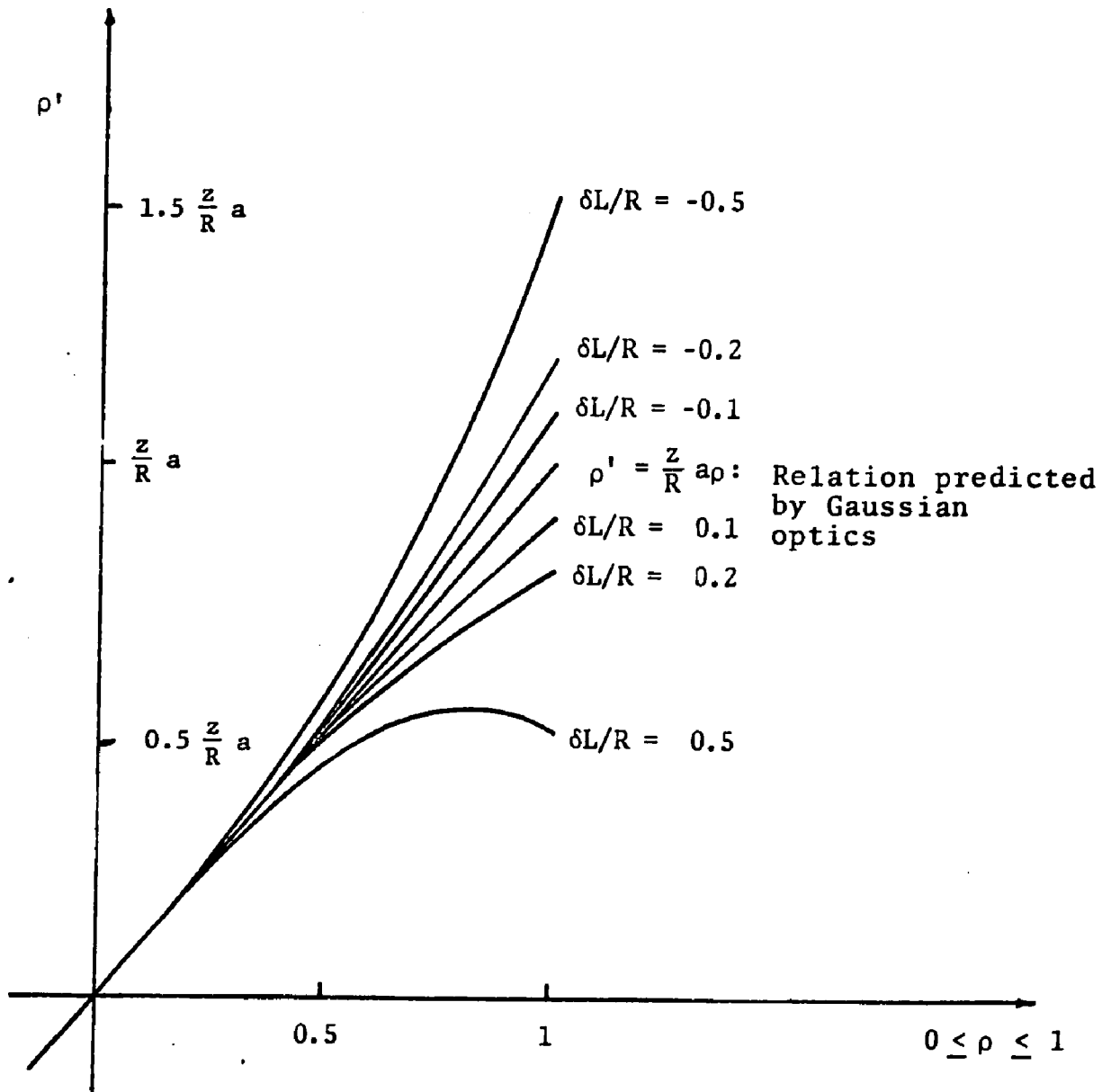


Figure 4-5. Relation between hologram aperture ρ' and reconstructed beam aperture ρ at the lens aperture for different spherical aberrations deviation $\delta L/R$

and a semi-minor axis is

$$\rho_{\text{minor}} = \frac{\rho'}{\frac{z}{R} - c_3(R + z) \left(\frac{2h^2}{a^2} \right)} \quad (4-31b)$$

A circular hologram with radius ρ' will form an elliptical beam pattern in the lens aperture. Thus, for the discussion on the resolution, we must choose the semi-minor axis as the effective aperture for the reconstructed object resolution. Equation (4-31b) shows the relation between hologram aperture and the corresponding resolution aperture in the lens aperture. Equation (4-31b) can be written in the form below:

$$\rho' = \rho_{\text{minor}} \left[\frac{z}{R} - c_3(R + z) \left(\frac{2h^2}{a^2} \right) \right] \quad (4-32)$$

The first term in Eq. (4-32) is just the relation predicted by Gaussian optics without aberrations. Hence, the second term can well represent the effect of this aberration on the system resolution performance. The choice of Eq. (4-31b) instead of Eq. (4-31a) as a measure of system resolution is based on the assumption that both c_2 and c_3 are positive. In this case, the minor axis of the ellipse is given by Eq. (4-31b). Since c_2 and c_3 can be both positive and negative independently, the practical choice of the minor axis should be either Eq. (4-31a) or (4-31b), whichever is smaller. The effect of this aberration on the system resolution depends on the sign of the second term. A comparison of the second term between Eqs. (4-29) and (4-32) reveals that the deviation of the system resolution from its Gaussian value is nonlinear because of the ρ^3 term for the spherical aberration, while it is linear for the astigmatism and field curvature aberration.

3) Distortion: $c_4 h^3 \cos \phi$

For this term, since each object point focuses to a point image in the Gaussian image plane, the aberration does not change the resolution aperture for each lenslet predicted by Gaussian optics. The

relative position of the image point behind the lens array is affected by this term however. The relocation of each image point in the image plane can be compared to the use of a different coding system and the final resolution of the system will not be affected in general.

4) Coma: $c_5 h \rho^3 \cos \phi$

The results for this term are complicated because of the complexity of the expression for the beam pattern in the lens aperture corresponding to an on-axis, circular hologram. For such a hologram, Eq. (4-25) becomes

$$\rho'^2 = \left(\frac{z}{R}\right)^2 (u^2 + v^2) + c^2 (R + z)^2 (9u^4 + 8u^2 v^2 + v^4) \\ - 2c \left(\frac{R}{z}\right) (R + z) (3u^3 + 3uv^2)$$

where c is a constant. Because this expression represents unfamiliar curves, we will consider another approach.

Consider a straight line $x = K$ in a receiving plane; Eq. (4-21) becomes

$$x = K = -\frac{z}{R} u + (R + z) \left(\frac{\partial \Delta}{\partial u}\right) \\ = -\frac{z}{R} u + (R + z) c_5 h (3u^2 + v^2) \\ = -\frac{z}{R} + \beta (3u^2 + v^2) \quad (4-33)$$

where $\beta = c_5 h (R + z)$. Rearranging Eq. (4-33), we get

$$\frac{\left(u - \frac{z}{6R\beta}\right)^2}{\frac{1}{3\beta} \left[K + 3\beta \left(\frac{z}{6R\beta}\right)^2\right]} + \frac{v^2}{\frac{1}{\beta} \left[K + 3\beta \left(\frac{z}{6R\beta}\right)^2\right]} = 1 \quad (4-34)$$

This is an ellipse in the lens aperture with the center shifted to

$$\left(u = \frac{z}{6R\beta}, v = 0\right).$$

The semi-major axis is given by

$$\left\{\frac{1}{\beta}\left[K + 3\beta\left(\frac{z}{6R\beta}\right)^2\right]\right\}^{1/2} = \left[\frac{1}{\beta}\left(K + \frac{z^2}{12R^2\beta}\right)\right]^{1/2} \quad (4-35a)$$

while the semi-minor axis is given by

$$\left[\frac{1}{3\beta}\left(K + \frac{z^2}{12R^2\beta}\right)\right]^{1/2} \quad (4-35b)$$

Since this curve corresponds to a straight line ($z = K$) in the receiving plane and K only appears in both semi-axes, for a set of straight lines ($x = K$) in the receiving plane, the corresponding curve in the lens aperture is a set of ellipses centered at $\left(u = \frac{z}{6R\beta}, v = 0\right)$. In the special case, when $K = 0$, Eq. (4-34) becomes

$$\frac{\left(u - \frac{z}{6R\beta}\right)^2}{\left(\frac{z}{6R\beta}\right)^2} + \frac{v^2}{3\left(\frac{z}{6R\beta}\right)^2} = 1$$

Further investigation reveals that when $K = -\frac{z^2}{12R^2\beta}$, the corresponding ellipse shrinks to the center point $\left(u = \frac{z}{6R\beta}, v = 0\right)$. Figure 4-6 shows the set of ellipses corresponding to different positive and negative K .

Figure 4-6 shows that with a rectangular hologram with boundary $x = K_1$, $x = K_2$, and the boundaries in y coordinates are large enough to cover all the rays in the y direction, the corresponding aperture pattern in the lens aperture would be either: (1) a ring between two ellipses; or (2) a solid ellipse centered at $(u = z/6R\beta, v = 0)$. The pattern depends on whether or not the line $x = -z^2/12R^2\beta$ is covered inside the hologram aperture. Since the shape of the hologram and the corresponding shape of the beam pattern in the lens aperture are not

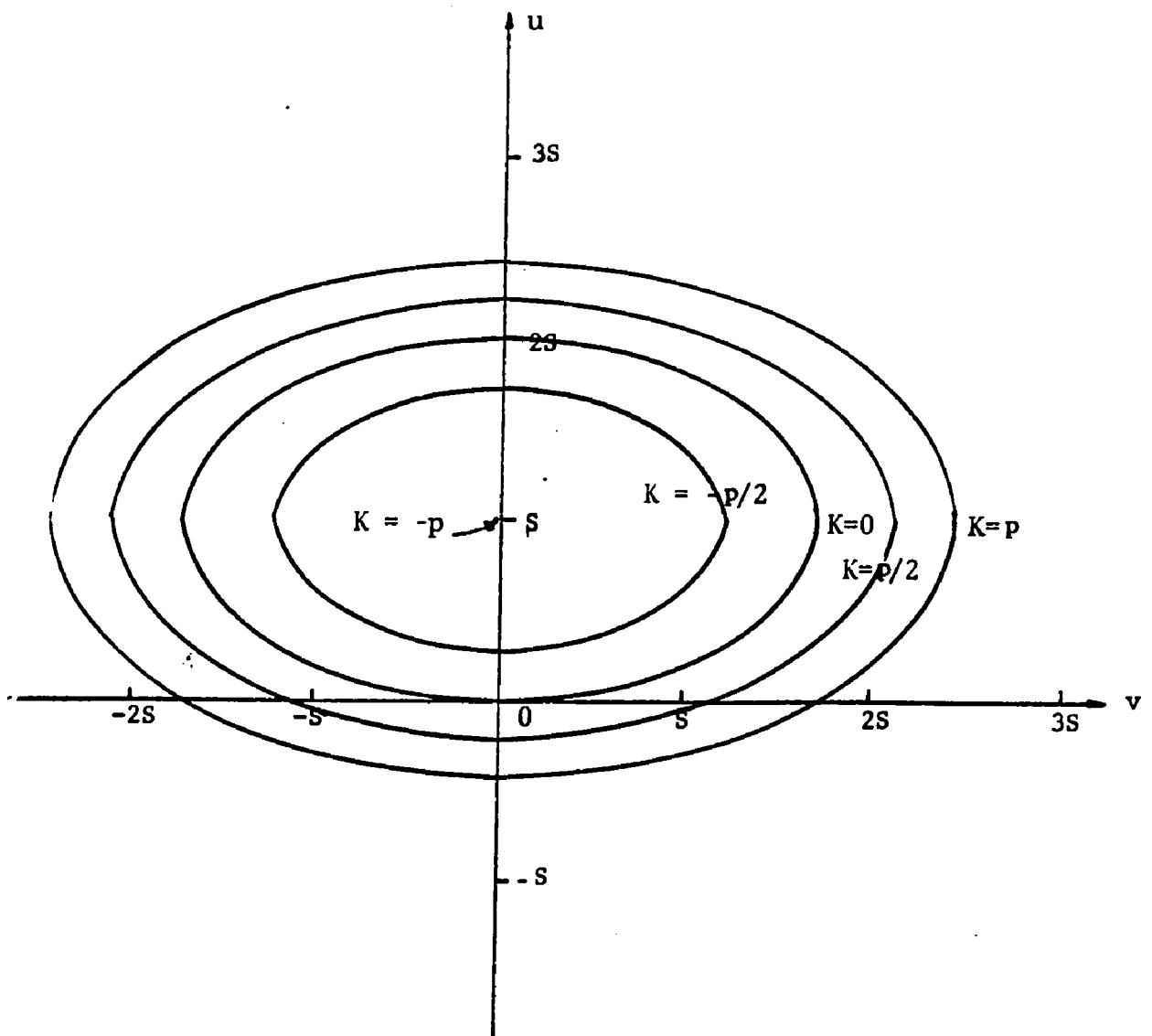


Figure 4-6. The set of ellipse curves in the lens aperture corresponding to different K where

$$s \triangleq \frac{z}{6R\beta}, \quad p \triangleq \frac{z^2}{12R^2\beta}$$

relevant for this aberration, the discussion on the effect of resolution is limited. Only in one, simple situation can we give a reasonable discussion. We assume a rectangular hologram with $x = K_1 = -z^2/12R^2\beta$; let $K_2 > K_1$. In this case, the width of the hologram in the x direction is given by $K_2 + z^2/12R^2\beta$, and the corresponding beam pattern in the lens aperture is a solid ellipse with semi-minor axis (given by Eq. 4-35b) as a resolution aperture. We thus have

$$\rho = 2\left(\frac{1}{3\beta} t\right)^{1/2} \quad (4-36)$$

where ρ is the length of the minor axis of the solid ellipse and $t = K_2 + z^2/12R^2\beta$ is the hologram aperture. Equation (4-36) is plotted in Fig. 4-7; the relation between ρ and t , predicted by Gaussian optics, is also plotted for comparison. At the point where $K_2 = 0$, or $t = z^2/12R^2\beta$, $\rho = z/3R\beta$. This aperture is almost four times that predicted by Gaussian optics. The maximum deviation occurs at the point $t = z^2/3R^2\beta$, where the aperture difference between the two ρ 's is equal to $z/3R\beta$.

In conclusion, the effect of the Seidel aberrations on the system resolution can be studied by tracing the exact shape of the reconstructed beam pattern at the lens aperture for a given hologram. However, the reconstructed beam pattern at the lens aperture yields simple expressions only for very simple hologram shapes at several special locations. These simple expressions can serve as the basis for discussion. The curves of the system resolution have been plotted to show the effect of different aperrations against the Gaussian predictions.

4.4. THE CHANGE OF WAVELENGTH, HOLOGRAM SCALE FACTOR, AND MAGNIFICATION

There are several practical problems in the fly's-eye lens array system: changes of wavelength in the recording and reconstruction process; misalignment of the hologram and the readout beam in the reconstruction process; and the effects of the displacement of the hologram and the readout beam. In this section, the effects caused by wavelength change will be studied.

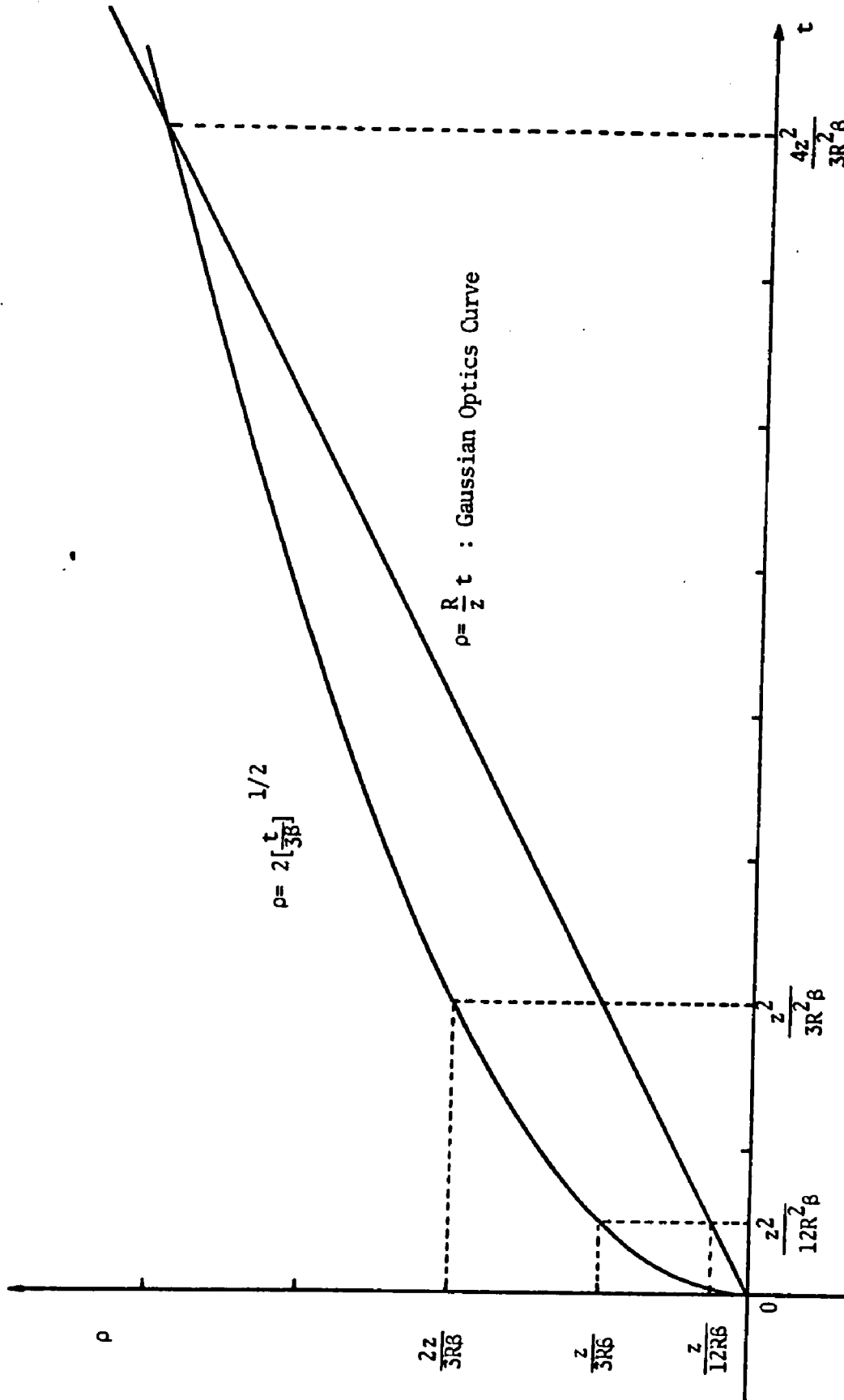


Figure 4-7. Relation between hologram size t and the reconstructed resolution aperture ρ at the lens aperture for coma aberration. The Gaussian optics curve is also included

Wavelength changes in the reconstruction process of the fly's-eye lens array system has two main results: (1) the magnification of the reconstructed image in the image plane of the fly's-eye lens array is changed; and (2) the aberrations of the reconstructed image appears. To cope with these results, the scale factor of the hologram size and the fly's-eye lens array can be changed in relation to the change of wavelength to avoid severe distortion in the final image. Also, the reference beam and the readout beam can be arranged to eliminate some of the aberration terms.

The problem is approached with Meier's [20] equations on image coordinates, magnification, and third-order aberrations.

$$\frac{1}{z_i} = \frac{1}{z_c} - \frac{u}{s} \left(\frac{1}{z_o} - \frac{1}{z_r} \right) \quad (4-37)$$

$$\frac{x_i}{z_i} = \frac{x_c}{z_c} - \frac{u}{s} \left(\frac{x_o}{z_o} - \frac{x_r}{z_r} \right) \quad (4-38)$$

$$\frac{y_i}{z_i} = \frac{y_c}{z_c} - \frac{u}{s} \left(\frac{y_o}{z_o} - \frac{y_r}{z_r} \right) \quad (4-39)$$

$$M = \frac{s}{\left(1 - \frac{s^2}{u} \frac{z_o}{z_c} - \frac{z_o}{z_r} \right)} \quad (4-40)$$

Equations (4-37), (4-38), and (4-39) express the relationship between four point-source coordinates. The subscripts i, o, r, and c correspond to image, object, reference, and construction point source, respectively. In Eq. (4-40), M is the lateral magnification for the reconstructed image, $u = \frac{\lambda_c}{\lambda_r}$ is the wavelength ratio, and s is the scale factor of the hologram. The coefficients of the third-order Seidel aberration terms are

given by spherical aberration

$$S = \frac{1}{z_c^3} - \frac{u}{s^4 z_o^3} + \frac{u}{s^4 z_r^3} - \frac{1}{z_i^3} \quad (4-41)$$

coma

$$C_x = \frac{x_c}{z_c^3} - \frac{ux_o}{s^3 z_o^3} + \frac{ux_r}{s^3 z_r^3} - \frac{x_i}{z_i^3} \quad (4-42)$$

astigmatism

$$A_x = \frac{x_c^2}{z_c^3} - \frac{ux_o^2}{s^2 z_o^3} + \frac{ux_r^2}{s^2 z_r^3} - \frac{x_i^2}{z_i^3} \quad (4-43)$$

field curvature

$$F = \frac{x_c^2 + y_c^2}{z_c^3} - \frac{u(x_o^2 + y_o^2)}{s^2 z_o^3} + \frac{u(x_r^2 + y_r^2)}{s^2 z_r^3} - \frac{x_i^2 + y_i^2}{z_i^3} \quad (4-44)$$

and distortion

$$D_x = \frac{x_c^3 + x_c y_c^2}{z_c^3} - \frac{u(x_o^3 + x_o y_o^2)}{s z_o^3} + \frac{u(x_r^3 + x_r y_r^2)}{s z_r^3} - \frac{x_i^3 + x_i y_i^2}{z_i^3} \quad (4-45)$$

Throughout this section, $u \neq 1$. The two parameters s and M provide four possible cases for coping with the change of wavelength $u \neq 1$:

$$\begin{cases} M \neq 1 \\ s \neq 1 \end{cases} \quad \begin{cases} M \neq 1 \\ s = 1 \end{cases} \quad \begin{cases} M = 1 \\ s \neq 1 \end{cases} \quad \begin{cases} M = 1 \\ s = 1 \end{cases}$$

In the first two cases, $M \neq 1$ means that there is a magnification of the reconstructed image. Therefore, the fly's-eye lens array must be enlarged equally so that it matches the reconstructed image and introduces no distortion in the final image. Discussions on each of the above four cases follows. Studies are made on the choice and arrangement of the reference and reconstruction beam that will reduce the aberration as much as possible.

Case 1:

$$\begin{cases} M \neq 1 \\ s \neq 1 \end{cases}$$

This is the least restricted case. Since the image magnification and the hologram scale factor can be chosen at any specific value, all the aberrations can be eliminated simultaneously with the following two arrangements:

$$A \begin{cases} z_c = z_r = \infty \\ \frac{x_c}{z_c} = -\frac{u}{s} \frac{x_r}{z_r} \\ \frac{u}{s} = 1 \end{cases} \quad B \begin{cases} z_r = z_o \\ z_c = \pm s z_o \\ \frac{x_c}{z_c} = -\frac{u x_r}{s z_r} \\ \frac{u}{s} = 1 \end{cases}$$

In the arrangement A, a collimated reference beam is used, both in recording and in reconstruction. With $z_c = z_r = \infty$, Eq. (4-40) becomes $M = s$. The hologram and the fly's-eye lens array should be magnified or demagnified by the same amount. The conditions $u/s = 1$ and $x_c/z_c = -(u/s)(x_r/z_r)$ are sufficient to eliminate all the aberrations listed in Eqs. (4-41) to (4-45). As a result, $M = u = s$.

In arrangement B, the point-source reference beam is put at the same plane where the object point is located. In the reconstruction process, a convergent wavefront, which comes to a focus at the point $s z_o$,

should be used. To eliminate coma and astigmatism simultaneously, $u/s = 1$ is necessary. From Eq. (4-40), $M = s = u$ in this arrangement.

Case 2:

$$\begin{cases} M \neq 1 \\ s = 1 \end{cases}$$

The required enlargement of the hologram in Case 1 is not desirable because it involves a conventional imaging step. Case 2 is intended to avoid this step. Unfortunately, if $s = 1$, all the aberrations cannot be simultaneously eliminated. Arrangements can be achieved to eliminate or reduce some of the aberrations listed in Eq. (4-41) to Eq. (4-45). For example:

$$\begin{array}{l} \text{A} \begin{cases} z_c = z_r = \infty \\ \frac{x_c}{z_c} = -\frac{u}{M} \frac{x_r}{z_r} \end{cases} \quad \text{B} \begin{cases} z_r = z_o \\ z_c = -Mz_o \\ \frac{x_c}{z_c} = -\frac{u}{m} \frac{x_r}{z_r} \end{cases} \end{array}$$

Arrangement A can eliminate coma completely and reduce astigmatism, as well as eliminating some of the terms in field curvature and distortion. Since $u \neq s$, some of the aberration terms cannot be eliminated. With this arrangement, however, $M = s = 1$ from Eq. (4-40), so that neither the hologram nor the fly's-eye lens array have to be scaled. Arrangement B can eliminate spherical aberration and coma completely. Astigmatism, field curvature, and distortion can be reduced by $x_c/z_c = -(u/s)(x_r/z_r)$. The astigmatism can be eliminated completely by using $z_c = -uz_o$ instead of $z_c = -sz_o$ in this arrangement, but the coma cannot be eliminated. From Eq. (4-46), M becomes $M = -uz_c/z_o$. If $z_c = -uz_o$ is used, $M = u^2$; if $z_c = -sz_o$ is used, $M = u$.

Case 3:

$$\begin{cases} M = 1 \\ s \neq 1 \end{cases}$$

This is the case in which the fly's-eye lens array remains unchanged. The aberrations of Eq. (4-41) to Eq. (4-45) cannot be eliminated simultaneously. However, some of the aberrations can be eliminated. Examples are:

$$\begin{array}{ll} \text{A} & \begin{cases} z_c = z_r = \infty \\ \frac{x_c}{z_c} = -\frac{u}{s} \frac{x_r}{z_r} \end{cases} \\ \text{B} & \begin{cases} z_r = z_o \\ z_c = -\frac{sz_o}{u} = -uz_o, \quad s = u^2 \\ \frac{x_c}{z_c} = -\frac{u}{s} \frac{x_r}{z_r} = -\frac{1}{u} \frac{x_r}{z_r} \end{cases} \end{array}$$

Arrangement A is exactly the same as arrangement A in Case 2; s is forced to be equal to 1 by the requirement of $M = 1$. Arrangement B eliminates spherical aberration and astigmatism completely. Part of the aberration terms in field curvature and distortion can also be eliminated by $x_c/z_c = -(u/s)(x_r/z_r)$. The coma cannot be eliminated in this case by $z_c = -sz_o$, since it makes $M \neq 1$ and $s = u^2$ in this arrangement.

Case 4:

$$\begin{cases} s = 1 \\ k = 1 \end{cases}$$

This is the case that is practically desirable if all the aberrations can be eliminated or reduced. From Eq. (4-40), Case 4 requires $z_c = -(1/u)(z_r)$. Arrangement A from both Cases 2 and 3 can be used here. For a second example, consider the following arrangement:

$$\begin{cases} z_r = z_o \\ z_c = -\frac{1}{u} z_o \\ \frac{x_c}{z_c} = -\frac{u}{s} \frac{x_r}{z_r} \end{cases}$$

Here, $z_c = -\frac{1}{u} z_o$ is obtained by $z_r = z_o$ and $z_c = -\frac{1}{u} z_r$. Only the spherical aberration can be completely eliminated. Part of the aberration terms in astigmatism, field curvature, and distortion can be eliminated by

$$\frac{x_c}{z_r} = -\frac{u}{s} \frac{x_r}{z_r}.$$

Except for the first general case, the aberrations can only be reduced in the remaining three cases. When the system requirements are precise, as in Case 4, more aberration will be present. The aberration-balancing technique used by Latta [22] can be applied in such cases. Since the technique deals with only one point, the possibility of any practical improvement by such a technique is still questionable. The presence of aberration may not be too severe if only one object point is to be reconstructed by the system. This is the case discussed in the next section.

4.5. THE APPLICATION OF THE FLY'S-EYE LENS ARRAY TO A SIMPLE THREE-DIMENSIONAL AIR-TRAFFIC CONTROL DISPLAY SYSTEM

As the last section of this chapter, a fly's-eye lens array is applied to a simple three-dimensional air-traffic control display device, first proposed by J. Upatnieks [24]. In this device, a fly's-eye lens array is used as a viewing screen, while the position of each airplane is represented by a dot of light behind the screen. The dot of light is actually recorded by a small hologram behind the fly's-eye lens array. The position of each dot is continuously adjusted according

to the actual motion of the aircraft. The position of each aircraft is supplied by ground-based information radar or other sources, such as on-board altimeters. This information is then processed by a computer in order to provide the appropriate commands for adjustments of the display. The adjustment of the position of each light dot in the device is accomplished by adjustment of the reconstruction illuminating beam as well as the hologram. The relation between the light-dot position and the required adjustments to produce the change in its position has not been studied analytically. In this section, this relationship is investigated and the basic equations between various system parameters are derived.

The system model for this study is sketched in Fig. 4-8. The real image of a point object is reconstructed to the left of the fly's-eye lens array for the convenience of the analysis although, in practice, the virtual image behind the fly's-eye lens array may be used. A description of the basic mechanism of the system device follows.

A point source O is recorded through the fly's-eye lens array. After reconstruction, this point-source image is supposed to be able to move to any place within a given terrain (denoted by airport terrain) simply by adjustment of both the reconstruction illuminating beam and the hologram. The goal here is to find out if this can be done theoretically and further derive the basic parameter relationships for the adjustment. First, consider the movement of a point source along the optical axis. Figure 4-9 shows a section of Fig. 4-8 in which two points O_1 and O_2 , with different distances z_1 and z_2 from the fly's-eye lens array, are considered. Suppose point O_1 forms a set of image dots behind the fly's-eye lens array--one image dot for each lenslet. If this set of image dots is magnified by a magnification M , can the new set of image dots correspond to a new point object, say O_2 , at different distances from the fly's-eye lens array? What is the relationship between the magnification and the positions of two points?

Let $I_1(n)$ denote the coordinate of the point image behind the n 'th lenslet of the fly's-eye lens array for the object O_1 . Then, from the simple geometry of Fig. 4-9, $I_1(n)$ is given by

$$I_1(n) = nm + n \frac{z_1'}{z_1} \quad (4-46)$$

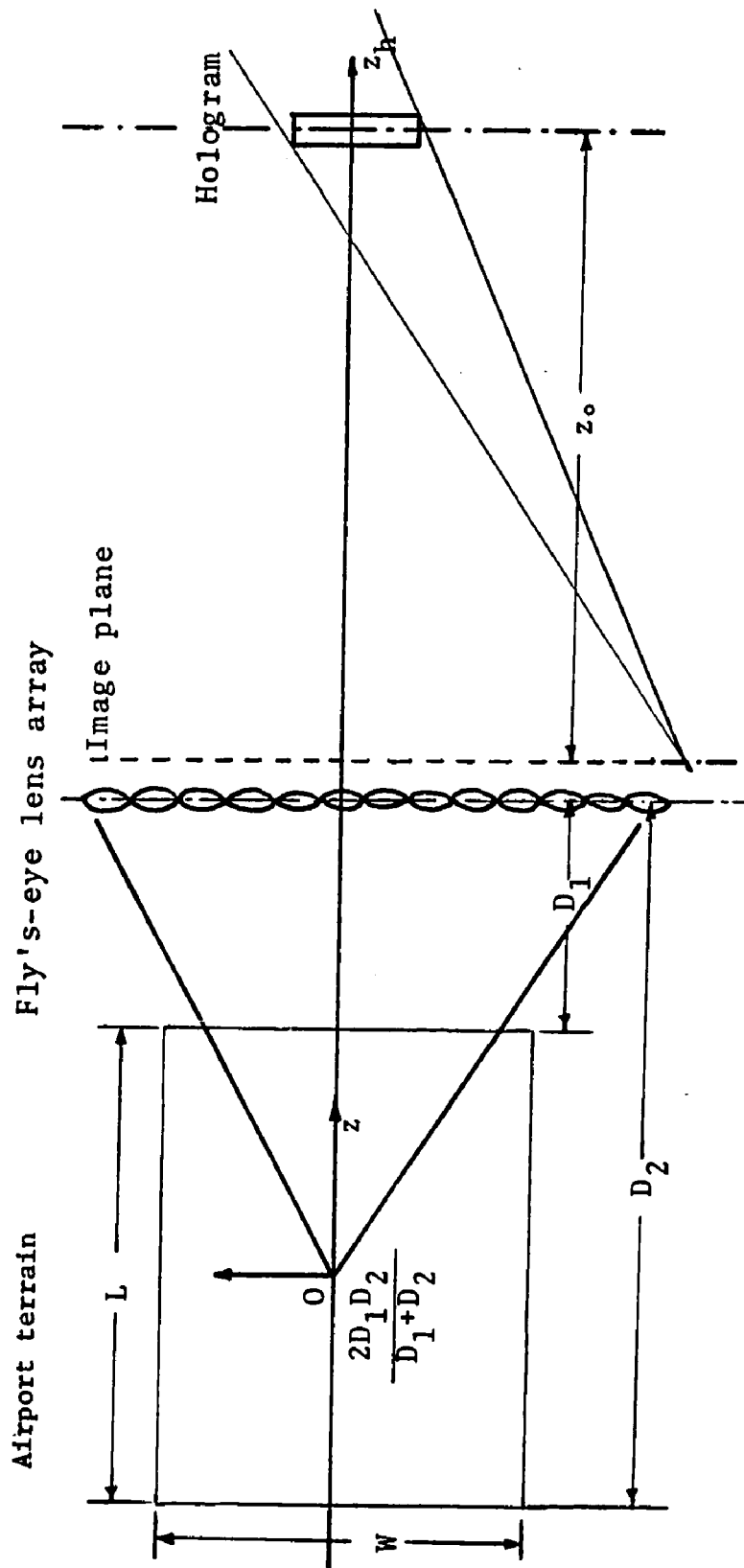


Figure 4-8. The model of a simple 3-D air-traffic control display system.

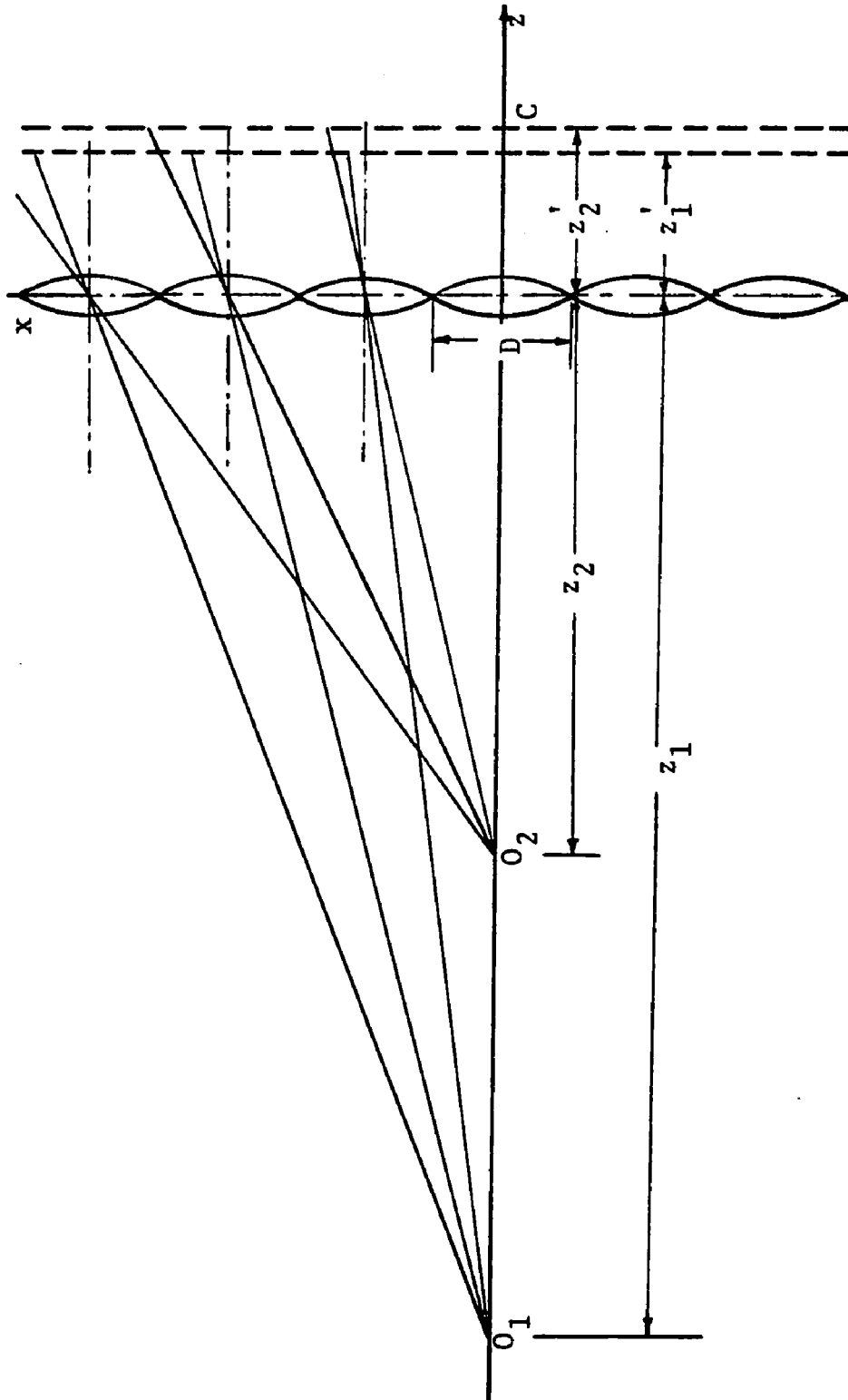


Figure 4-9. The model of the fly's-eye lens array for the study of the axial movement of a point object.

where m is the aperture of each lenslet. These image dots are equally spaced with spacing distance $m + (z'_1/z_1)m$. The coordinates for the other set of image dots of object point O_2 are

$$I_2(n) = nm + n \frac{z'_2}{z_2} m \quad (4-47)$$

Now, if Eq. (4-46) is magnified by a factor M and is set equal to Eq. (4-47):

$$MI_1(n) = I_2(n) \quad (4-48)$$

Equation (4-48) must be true for all n ; n is the integer corresponding to each lenslet. We further let $M = 1 + \Delta M$, where ΔM is the increment of the magnification. We thus have

$$(1 + \Delta M)nm + (1 + \Delta M) n \frac{z'_1}{z_1} m = nm + n \frac{z'_2}{z_2} m$$

Since the above equation has to be true for all n , it immediately yields

$$\Delta M = \left(\frac{z'_2}{z_2} - \frac{z'_1}{z_1} \right) \left(\frac{z_1}{z_1 + z'_1} \right) \quad (4-49)$$

Equation (4-49) is the basic equation which relates the magnification and the object point position. Clearly, if $z_2 < z_1$, then $(z'_2/z_2) - (z'_1/z_1)$ is positive. This means ΔM is positive. We thus need a magnification to move point O_1 to point O_2 . The above consideration did not mention the required z -axis adjustment of the point-image set since $z'_1 \neq z'_2$. We assume that the required adjustment has been made. Moreover, since z'_1 and z'_2 are related to z_1 and z_2 by the single-lens function,

$$\frac{1}{f} = \frac{1}{z_1} + \frac{1}{z'_1}$$

$$\frac{1}{f} = \frac{1}{z_2} + \frac{1}{z'_2}$$

Equation (4-49) can be expressed in terms of z_1 , z_2 , and f

$$\Delta M = \frac{f}{z_1} \left(\frac{z_1 - z_2}{z_2 - f} \right) \div \frac{f}{z_1} \left(\frac{z_1 - z_2}{z_2} \right) \quad (4-50)$$

where f is the focal length of the small lenslet.

In Eq. (4-50), the approximation is valid when $z_2 \gg f$ (which is usually the case). This equation is also a very useful expression of the object point position and the required magnification.

We now use Eq. (4-50) to derive the optimum position for the point-object source O_1 that must be recorded in our system (Fig. 4-8). The criterion for the optimum is as follows: once a point source, say O_1 , is recorded at some position O within the airport terrain (i.e., between D_1 and D_2 in Fig. 4-8) the required incrementation of the magnification ΔM to move the reconstructed point to the two extreme boundaries D_1 and D_2 should be the same in their absolute value. Let z_o denote the optimum position for the object point. Then ΔM at D_1 and D_2 is given by

$$\frac{f}{z_o} \left(\frac{z_o - D_1}{D_1} \right) \text{ and } \frac{f}{z_o} \left(\frac{z_o - D_2}{D_2} \right)$$

respectively. Equating the absolute value of these two expressions for ΔM and solving for z_o , we get

$$z_o = \frac{2D_1 D_2}{D_1 + D_2}$$

Surprisingly, the optimum position does not happen at the center of the terrain.

So far, we have considered the case in which the magnification of the set of the image dots formed by the fly's-eye lens array for a point object O_1 , takes place when the center of magnification (denoted by point C in Fig. 4-9) is on the axis of the object point O_1 . The center of magnification is defined as the point which does not change its position during the magnification process. If the center of magnification does not occur at the point which has the same (x, y)

coordinates as the object point during the magnification process, then the new (x, y) coordinates of the object point after the magnification will not be the same as the old one. In other words, the object point will be shifted laterally or vertically after the magnification. The new coordinates can be found easily, however, since the relation between the object point x, y coordinates and the corresponding image-point coordinates behind each lenslet is very simple.

In the above analysis, we have assumed that the object point is on the axis of one single lenslet, which is not always the case. However, analysis of other cases can be carried out similarly. In practice, however, the size of each lenslet is rather small compared to the size of the fly's-eye lens array aperture, the airport terrain and its distance from the lens array. This makes the above analysis rather accurate, even if we only consider discrete positions in the object space and its image space. As pointed out before, the extension of the analysis to the continuous case yields no further complications.

To illustrate this, consider a small lenslet A and its image plane in Fig. 4-10, where a point O is chosen as the center of magnification and point x is the coordinate of an image point formed by this lenslet for an object point O_1 . If the coordinates axis is also chosen to be at O, then a magnification M makes the x coordinate $M_x = (1 + \Delta M)x$. The displacement ΔM of the image dot x during the magnification process yields a corresponding displacement

$$\Delta M \times \left(\frac{z_2}{z_1} \right)$$

in the object space from a point O_2 to O_3 . The object point should be located at O_2 if the center of magnification is chosen at the image-dot position x. The vertical location of the object point after a magnification process depends solely on where the center of magnification is chosen while the axial location of the object point is still governed by Eq. (4-50). If the center of magnification is chosen at one of the image points which is within the same aperture of the small lenslet, as the object point does, then we can say that the object point has been moved axially by a change of magnification. This situation is shown in Fig. 4-10 by points O_1 , O_2 , and C, where C is chosen as the center of magnification.

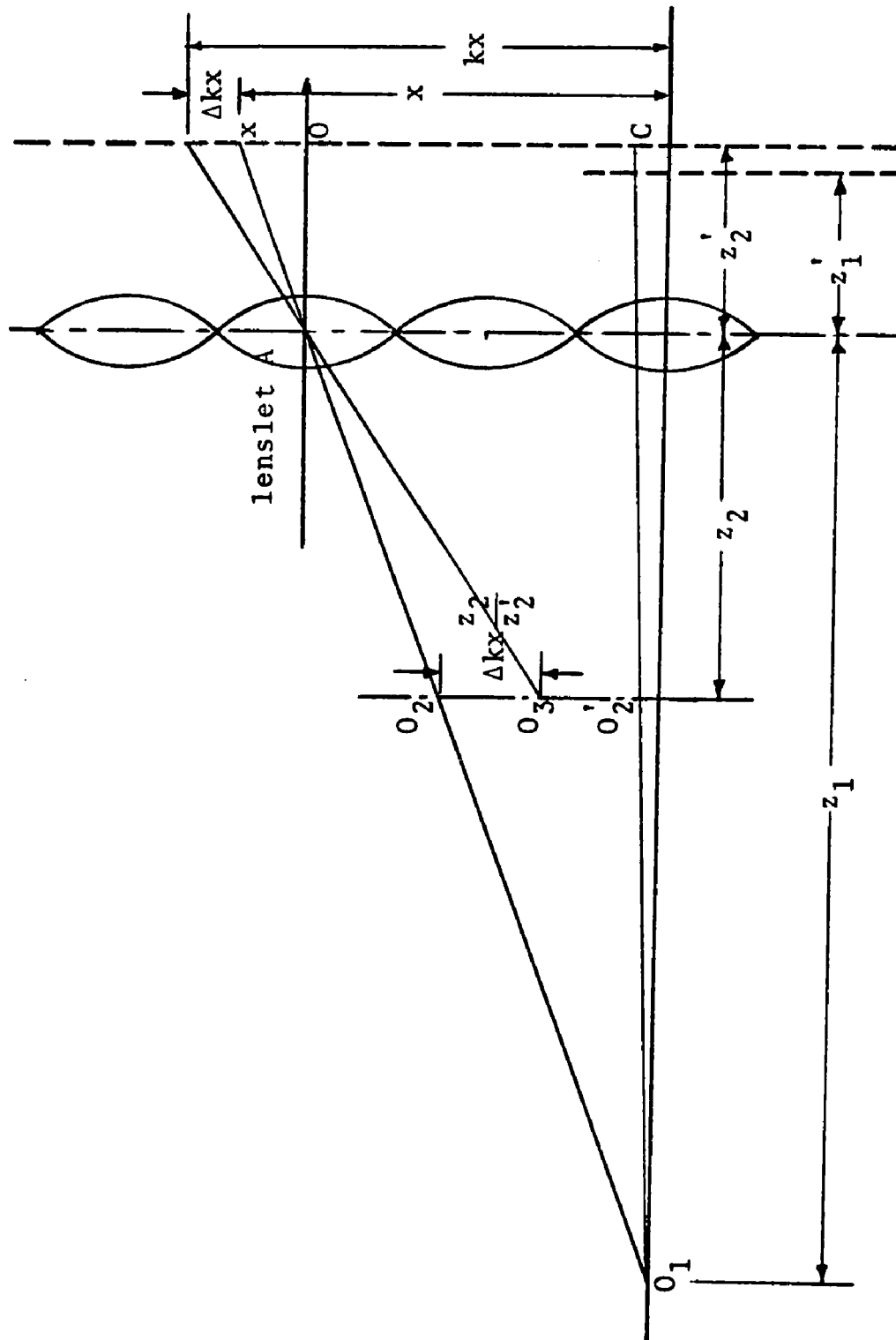


Figure 4-10. The model of the fly's-eye lens array for the study of the vertical movement of a point object.

So far, we have studied the basic property of the fly's-eye lens array of providing the axial movement of a point object source by magnifying its image dots behind the fly's-eye lens array. Now we want to further investigate how we can provide such a magnification by the movement of the hologram and the readout beam behind the lens array. The system setup is sketched in Fig. 4-11 (which is the right part of Fig. 4-8). We first consider the case where the hologram is parallel to the fly's-eye lens array and is located far enough away from the fly's-eye lens array so that the analysis can be carried out by using Meier's [20] equations on magnification and object-image relations. (See Section 4.4, Eqs. 4-37 to 4-40.) Coordinators of image, object, reference, and readout points are, respectively; (x_i, y_i, z_i) , (x_o, y_o, z_o) , (x_r, y_r, z_r) and (x_c, y_c, z_c) . M is the lateral magnification. In our case, $u = s = 1$. Equations (4-37) through (4-40) can then be simplified to

$$\frac{1}{z_i} = \frac{1}{z_c} - \left(\frac{1}{z_o} - \frac{1}{z_r} \right) \quad (4-51)$$

$$\frac{x_i}{z_i} = \frac{x_c}{z_c} - \left(\frac{x_o}{z_o} - \frac{x_r}{z_r} \right) \quad (4-52)$$

$$\frac{y_i}{z_i} = \frac{y_c}{z_c} - \left(\frac{y_o}{z_o} - \frac{y_r}{z_r} \right) \quad (4-53)$$

$$M = \frac{1}{1 - \frac{z_o}{z_c} - \frac{z_o}{z_r}} \quad (4-54)$$

Refer to Fig. 4-11. The object for the hologram is the set of image dots a distance z_o away from the hologram. Our immediate goal now is to find a configuration for the reference and readout beams so that the reconstructed image points have a magnification M which can be controlled by the movement of the hologram as well as by the readout beam. From Eq. (4-54), we note that we cannot let $z_r = \infty$, since if $z_r = \infty$, z_c

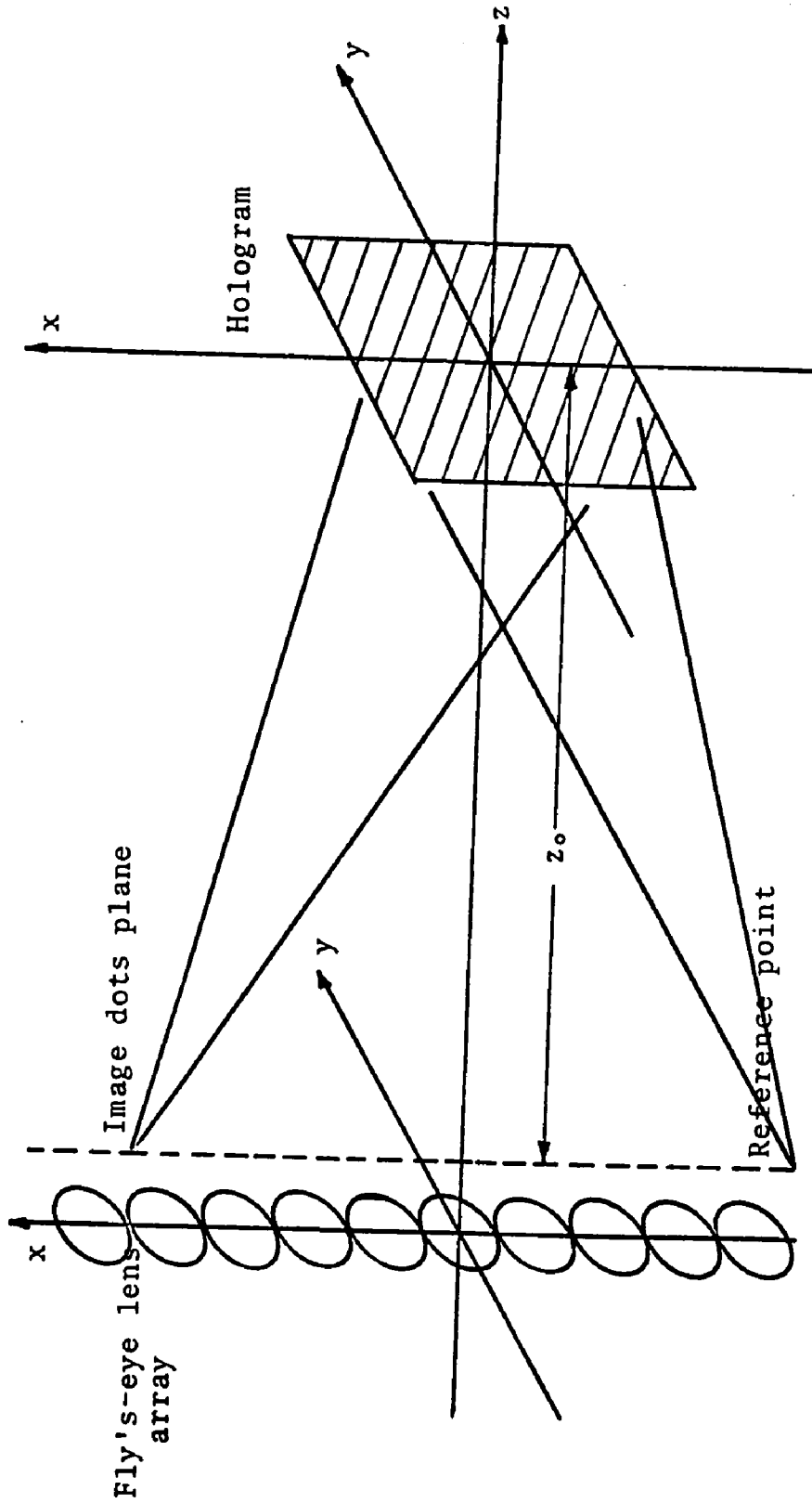


Figure 4-11. The model of the fly's-eye lens array and the hologram for the study of the air-traffic control system

has to approach ∞ in order to have $M \rightarrow 1$. This is undesirable in practical applications. Further investigation reveals that if $z_r = z_o$, then, from Eq. (4-54)

$$M = -\frac{z_c}{z_o} \quad \text{or} \quad z_c = -Mz_o \quad (4-55)$$

Thus, we can control the magnification by controlling z_c . Moreover, from Eq. (4-51), $z_i = z_c$, which indicates that the reconstructed image dots will be in the same plane where the readout point is located. Hence, in practice, we can put the readout point at the plane where the reconstructed image dots are desired and move the hologram along the axial axis to provide the necessary magnification.

Our next concern is the center of magnification in the magnification process. To provide that information, we must refer to Eq. (4-52) and consider the choice of x_r . For practical reasons, x_r is chosen to be $-\frac{H}{2}$, where H is the aperture of the fly's-eye lens array. With this choice of x_r , Eq. (4-52) becomes

$$\frac{x_i}{-Mz_o} + \frac{x_o}{z_o} = \frac{x_c}{-Mz_o} - \frac{H}{2z_o}$$

or

$$\frac{x_i}{-M} + x_o = \frac{x_c}{-M} - \frac{H}{2} \quad (4-56)$$

Equation (4-55) is used to determine the center of magnification. From Eq. (4-56), for a fixed x_c , the object-point coordinate x_o is related to the image coordinate x_i . The center of magnification is determined when the two coordinates x_o and x_i are equal in Eq. (4-56):

$$\left(1 - \frac{1}{M}\right)x_m = \frac{x_c}{-M} - \frac{H}{2} \quad (4-57)$$

where x_m is the center of magnification. Equation (4-57) shows the relationship between the center of magnification x_m and the readout coordinate x_c . The same approach can be applied to the y direction.

With $y_r = 0$, the corresponding equation of Eq. (4-57) is

$$\left(1 - \frac{1}{M}\right) y_m = \frac{y_c}{-M}$$

The above analysis has provided for moving a point object along the axis. Now we consider the means to move the object point vertically. Clearly, a vertical movement of the image dots behind the fly's-eye lens array provides a corresponding vertical movement of the object point as long as the movement of each image dot is within the aperture of each lenslet. The relation between the movement of the object point and the corresponding image point is

$$\frac{x}{z_1} = \frac{x'}{z'_1} \quad (4-58)$$

where (x, z_1) are the object coordinates and (x', z'_1) are the corresponding image-dot coordinates behind one lenslet in which the center of magnification is assumed to be located. In order to move the object point vertically, we only need to move the whole set of image dots vertically, according to Eq. (4-58), in which x' is the required movement of the center of magnification.

To move the center of magnification vertically by the hologram set-up, we have to refer to Eq. (4-56) again, where x_c has to be chosen such that the difference between x_1 and x_0 is equal to x' , and x_0 is the center of magnification. For the purpose of illustration, refer to Fig. 4-9, where an object point, O_1 , is on the axis of one small lenslet which is also chosen as the coordinate axis. If the center of magnification is also chosen to be the point on axis, then $x_m = 0$ in Eq. (4-57), and we have $x_c = -\frac{H}{2} M$. This means that if we want to move the object point O_1 along the axis, we need to move the hologram by the amount of Eq. (4-55) and, at the same time, the position of x_c has to be moved by the amount of $-\frac{H}{2} M - \frac{H}{2} = -\frac{H}{2} \Delta M$ from its reference position $-\frac{H}{2}$. To move the object point vertically, refer to Eq. (4-56) again where $x_0 = x_m = 0$, $x_1 = x_m + x' = x'$; we have

$$\frac{x'}{-M} = \frac{x_c}{-M} - \frac{H}{2} \quad \text{or} \quad x' = x_c + \frac{H}{2} M \quad (4-59)$$

With Eq. (4-58) substituted into Eq. (4-59), we have

$$x = \frac{z}{z'} \left(x_c + \frac{H}{2} M \right) \quad (4-60)$$

Equation (4-60) is the final basic relation between the object position x and the readout point source location, x_c . The same analysis applied to the y direction yields

$$y = \frac{z}{z'} y_c$$

The coordinates of the object point are (x,y) and (x_c, y_c) are the coordinates of the readout focus point.

We can derive a direct relationship between the movement of the hologram and the corresponding movement of the object point by using Eq. (4-50) and Eq. (4-55). Assume that the point object is first recorded at its optimum position, i.e.,

$$z_1 = \frac{2D_1 D_2}{D_1 + D_2} \quad \text{in Eq. (4-50).}$$

We then define a new variable z from this optimum position:

$$z \triangleq \frac{2D_1 D_2}{D_1 + D_2} - z_2 \quad \text{in Eq. (4-50).}$$

Equation (4-50) then becomes

$$\Delta M = \frac{f(D_1 + D_2)^2 z}{2D_1 D_2 [2D_1 D_2 - z(D_1 + D_2)]} \quad (4-61)$$

$$\text{Let } D_1 + D_2 \triangleq p$$

$$2D_1 D_2 \triangleq q .$$

Equation (4-61) becomes

$$\Delta M = \frac{fp^2 z}{q(q - pz)} = \frac{fp^2 z}{q^2 - pqz}$$

From Eq. (4-55), we have

$$z_c = -Mz_o = -(1 + \Delta M)z_o = -z_o - \Delta Mz_o$$

The movement of the hologram from its center location, z_o , is ΔMz_o and is defined by a new variable z_h . Thus

$$\Delta Mz_o = z_h = \frac{z_o fp^2 z}{q^2 - pqz} \quad (4-62)$$

Equation (4-62) is the final basic relation between object coordinate z and hologram coordinate z_h . Note that z and z_h are referred to an origin different from that defined before. The small distance $z'_1 - z'_2$ between two different planes of image points for two different object points, O_1 and O_2 , has been neglected in the above consideration, since, in practice, a ground glass will be put at the plane of image points.

One alternative configuration of the hologram setup for this system is to have the hologram not parallel to the fly's-eye lens array. This configuration is sketched in Fig. 4-12 where the hologram's normal bisects the angle formed by the fly's-eye lens array axis and the ray from a point reference source. The analysis of this setup can be approached by using Champagne's [21] equation for the image-forming parameters of the hologram. These equations are written below:

$$\frac{1}{R_i} = \frac{1}{R_c} \pm \left(\frac{1}{R_o} - \frac{1}{R_r} \right) \quad (4-63)$$

$$\sin(\alpha_i) = \sin(\alpha_c) \pm [\sin(\alpha_o) - \sin(\alpha_r)] \quad (4-64)$$

$$\cos(\alpha_i) \sin(\beta_i) = \cos(\alpha_c) \sin(\beta_c) \pm [\cos(\alpha_o) \sin(\beta_o) - \cos(\alpha_r) \sin(\beta_r)] \quad (4-65)$$

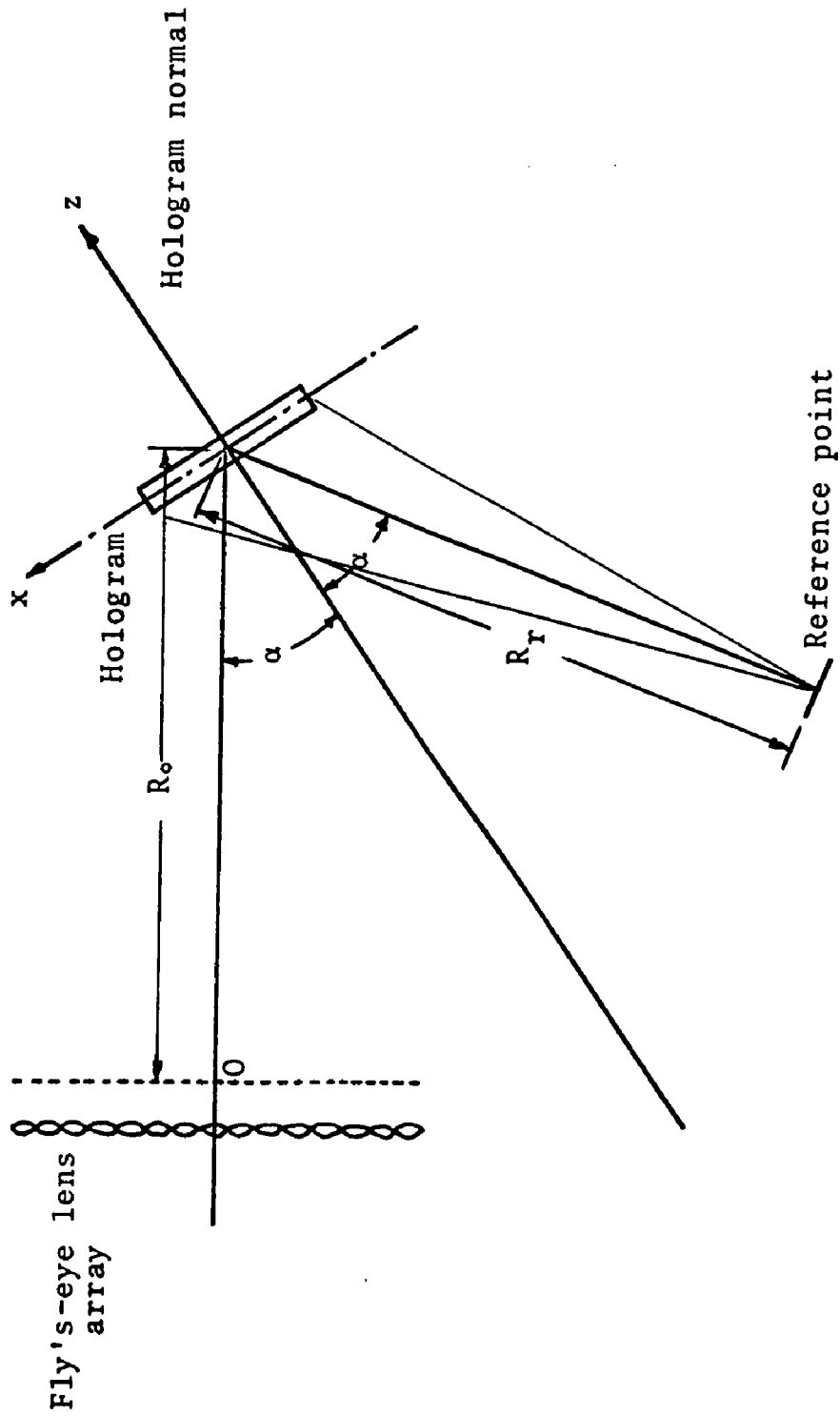


Figure 4-12. An alternative configuration of the air-traffic control system

$$M_{i\alpha} = \left[\frac{\cos(\alpha_o)}{\cos(\alpha_i)} \right] \left(\frac{1}{1 - \frac{R_o}{R_c} - \frac{R_o}{R_r}} \right) \quad (4-66)$$

where we assume $u = s = 1$.

Since these equations only deal with point source objects and images, an extension to the plane object and image is necessary. We thus define the optical axis of the object plane and the image plane to be the central normal axis of the fly's-eye lens array and its image, as shown in Fig. 4-12. We further assume that the size of the fly's-eye lens array and the hologram are small compared to the distance between them so that the magnification and the displacement derived for the center point 0 of the fly's-eye lens array can be applied to the whole array. With this assumption, we now can proceed to find the relations between different system parameters. We note that in our setup for point 0, we have

$$R_o = R_r$$

$$\alpha_o = -\alpha_r = \alpha.$$

From Eq. (4-63), we get $R_i = R_c$ and from Eq. (4-66)

$$M_{i\alpha} = \left[\frac{\cos(\alpha_o)}{\cos(\alpha_i)} \right] \left(\frac{R_c}{R_o} \right) \quad (4-67)$$

If we choose $\alpha_c = -\alpha_r = \alpha$, Eq. (4-64) yields

$$\begin{aligned} \sin(\alpha_i) &= \sin(\alpha) - 2 \sin(\alpha) \\ &= -\sin(\alpha) \end{aligned}$$

or

$$\alpha_i = -\alpha = -\alpha_o \quad (4-68)$$

The negative value of α_i indicates the image is real. From Eq. (4-67) with Eq. (4-68), we have

$$M_{i\alpha} = \frac{R_c}{R_o}$$

or

$$R_c = M_{i\alpha} R_o = R_i \quad (4-69)$$

Equation (4-69) is exactly the same as Eq. (4-55), indicating that the same movement of the hologram can be made to provide the necessary magnification for the fly's-eye lens array as long as we keep $\alpha_c = -\alpha_r = \alpha_o = \alpha$. The center of magnification in this case is 0 which can be proved with the formula for the angular magnification in the α direction for point 0:

$$\text{angular magnification} = \frac{\Delta \alpha_i}{\Delta \alpha_o} = \pm \left[\frac{\cos(\alpha_o)}{\cos(\alpha_i)} \right]$$

Only when $\alpha_o = \alpha$, $\alpha_i = -\alpha_o = \alpha$, then $\frac{\Delta \alpha_i}{\Delta \alpha_o} = 1$. This means point 0 is fixed during the magnification process. To move the object point vertically, the center of magnification has to be moved vertically. We refer to Eq. (4-64) in which we set $\alpha_c = \alpha + \Delta \alpha_c$ and $\alpha_i = -(\alpha + \Delta \alpha)$. Equation (4-64) then becomes

$$2 \sin(\alpha) = \sin(\alpha + \Delta \alpha_c) + \sin(\alpha + \Delta \alpha_i) \quad (4-70)$$

Equation (4-70) is the basic relation between the angular movement of the readout beam and the angular movement of the image plane. Probably the only advantage of this configuration is that we can move the object point vertically by just tilting the reconstructing mirror. The disadvantage of this configuration is the complexity of the hologram movement, more aberration, and more sensitivity.

CHAPTER 5

THE BEAMWIDTH REDUCTION METHOD OF A THREE-DIMENSIONAL HOLOGRAPHIC VIEWING SYSTEM

As we have stated in the first chapter, an alternative method of solving the three-dimensional display problem is to reduce the beamwidth of the object scene. It has been experimentally demonstrated that this method has more potential than other three-dimensional viewing methods. Information-reduction methods for three-dimensional viewing systems have inherent basic limitations. However, for the beamwidth-reduction method, the recording materials required to store a large amount of information are commercially available at a reasonable price. In this chapter, several aspects of the beamwidth reduction method are discussed. Most of the problems in using this method involve engineering considerations. A large amount of experimental work was performed in perfecting this type of system.

5.1. THE SINGLE-LENS MOVIE SYSTEM AND ITS SB CHARACTERISTICS

The objective of the beamwidth-reduction method for a 3-D viewing system is to use a lens to focus a large object scene onto a small recording hologram so that the information can be recorded on small size movie film (i.e., 35-mm movie film). The simplest form of such a system is the single-lens movie system, discussed in Ref. [10] by E. N. Leith. The system is briefly discussed here. Figure 5-1 shows a schematic diagram of the system. A single lens is situated between the observer's aperture A and the hologram aperture K. The focal length f of the lens is such that the aperture A is imaged to the hologram aperture K. In this way, an observer positioned at any point in aperture A can see the image field against the background of the lens aperture H. A suitable object region is indicated by the shaded area in Fig. 5-1. In this system, the available object region contains the volume behind the lens H. In the sampling method, as shown in Fig. 3-4, only the region in front of aperture H is useful.

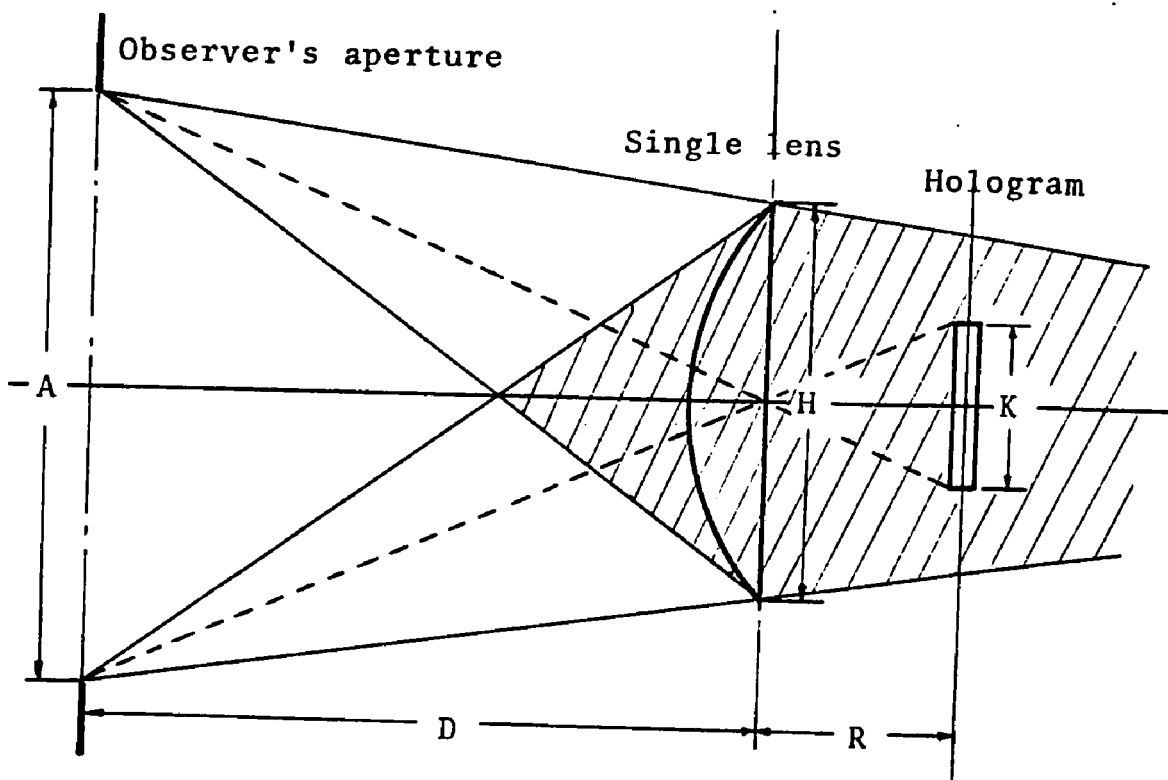


Figure 5-1 The Single-lens system

The quality of the single lens need not be very high since the aberrations introduced by a low-quality lens can be compensated for by the holographic recording and reconstruction process. Severe aberrations introduced by the low-quality lens affects the system performance when the viewing aperture A gets larger and when various system parameters change their values.

For a given distance D and a hologram size K , the observer's aperture A is related to the focal length f by

$$f = \frac{DK}{A + K}$$

A large A requires a large f -number lens, and more aberrations will be present for large f -number lenses. This is the subject of the following two sections.

For the present, consider the amount of data to be recorded by the system. The SB plot concept is used for this study. Refer to Fig. 5-1 where the aperture K is the image of aperture A. We have

$$\frac{A}{D} = \frac{K}{R} \quad \text{or} \quad M = \frac{A}{K} = \frac{D}{R}$$

where M is the magnification.

Assume an electromagnetic field $f_A(x, y)$ exists at aperture A. The spectrum of $f_A(x, y)$ is $F_A(f_x, f_y)$. Since there is an image of $f_A(x, y)$ at K, the electromagnetic (E-M) field at K should be the same as $f_A(x, y)$, except for a magnification factor M. This is true if the diffraction effect is neglected. The E-M field at K is denoted by $f_K(x, y)$. The SB plot of the field $f_A(x, y)$ at the left side of lens aperture H is plotted in Fig. 5-2. It is designated by branch A. The SB plot of the field $f_K^*(x, y)$ at the right side of the aperture H is also plotted. Since the field $f_K(x, y)$ is the conjugate of $f_K^*(x, y)$ and it is not propagating toward the lens aperture H, its SB plot represents the conjugate of the field $f_K^*(x, y)$. If $F_K(f_x, f_y)$ is the spectrum of $f_K^*(x, y)$, $F_K^*(-f_x, -f_y)$ is the spectrum of $f_K(x, y)$. The SB plot of $f_K(x, y)$, plotted in Fig. 5-2, is designated by branch B. Branch C is the SB plot for $f_K^*(x, y)$. The space bandwidth product of branch A should equal the space bandwidth product of branch B.

From Fig. 5-2, we have

$$(SW)_A = \frac{AH}{\lambda D} \quad \text{and} \quad (SW)_B = \frac{KH}{\lambda R}.$$

Since $\frac{A}{D} = \frac{H}{R}$, $(SW)_A = (SW)_B$. This number, $(SW)_A = (SW)_B$, represents the amount of information recorded by the system. It can be seen that the information content is preserved in the system.

We now investigate further how the SB plot of branch A is transferred to the SB plot of branch B by the lens in Fig. 5-2. The function representing a lens is given by

$$e^{-j\left(\frac{k}{2f}\right)(x^2+y^2)}$$

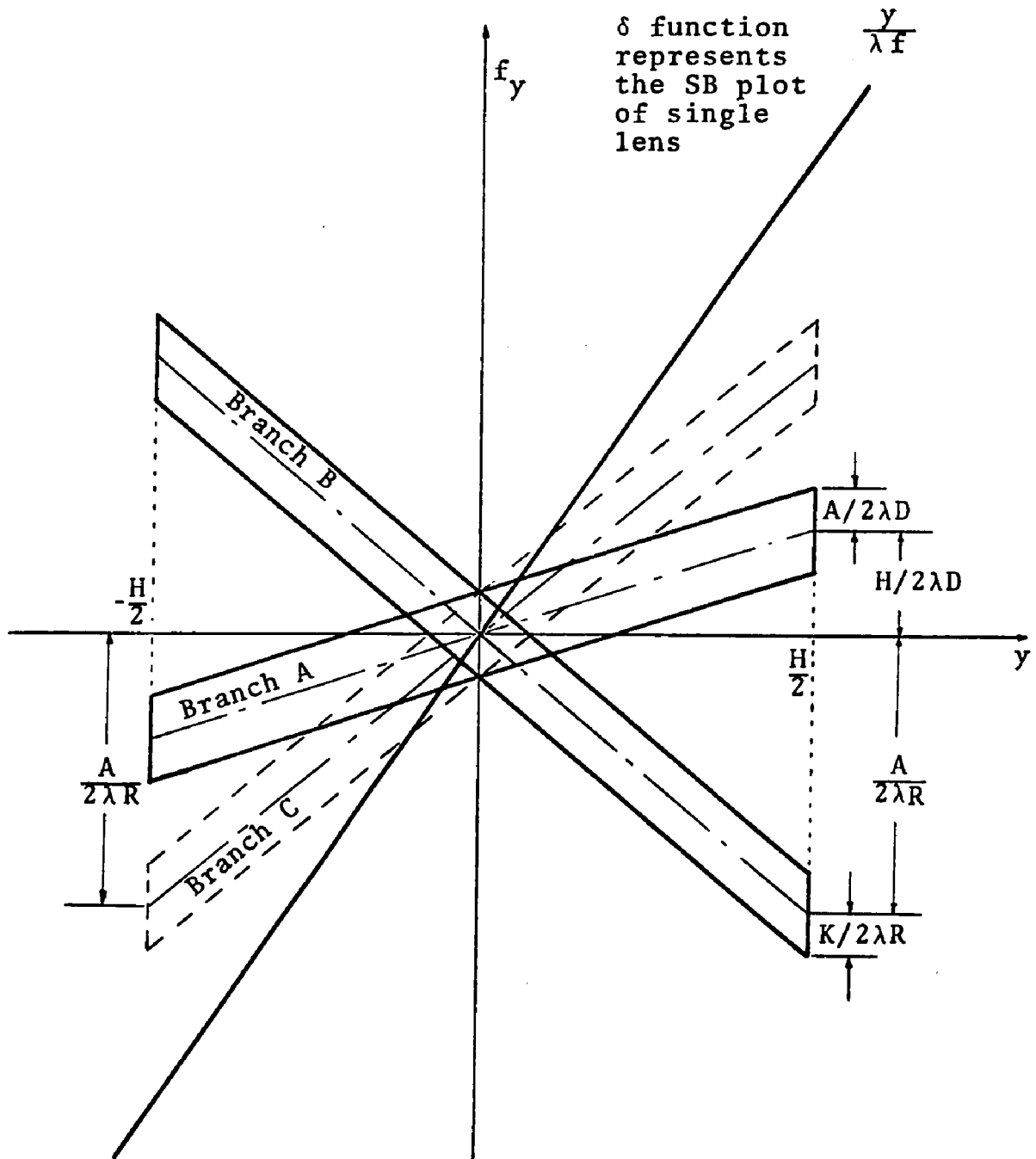


Figure 5-2. The characteristic of the SB plot of a single-lens system

Since the E-M field behind the lens equals the E-M field in front of the lens times the lens function, the corresponding relation in the spatial frequency domain involves a correlation. The SB plot of branch B is the correlation of the SB plot of branch A with the SB plot for the lens function. The lens function has the same form as the field due to a point source a distance f away from the lens.

From Section 2.2, we note that the SB plot for a point source is a δ -function with slope $\frac{1}{\lambda f}$ in the SB plot coordinate system. This δ -function is also shown in Fig. 5-2. At each point y along the lens aperture, the spectral bandwidth for branch B should equal the correlation of the spectral bandwidth A with the δ -function. At any point y in Fig. 5-2, the correlation yields

$$\frac{y}{2\lambda f} - \frac{y}{2\lambda D} = \frac{y}{2\lambda R}$$

or

$$\frac{1}{f} = \frac{1}{D} + \frac{1}{R}$$

This is exactly the lens imaging relation. Thus, we have proven the validity of the SB plot representation for the system. We have also demonstrated the usefulness of the SB plot in obtaining a better understanding of the system.

The role of a lens in the SB plot is identical to that of a δ -function with a slope $\frac{1}{\lambda f}$. It is interesting to consider the case where a point object at A is imaged to a point at K by a single lens H. Since the SB plot for a point source is also a δ -function with a given slope, the SB plots for the object point, the image point, and the lens are three straight δ -function lines in the SB plot having the relation

$$\frac{y}{2\lambda f} = \frac{y}{2\lambda D} + \frac{y}{2\lambda R}$$

for each point y .

5.2. THE ABERRATIONS OF THE SINGLE-LENS MOVIE SYSTEM CAUSED BY THE HOLOGRAM APERTURE

In the previous section, it was shown that the aberrations in the single-lens movie system will not affect the reconstructed image since

they are eliminated because of the recording and reconstruction process. However, if the viewer's aperture gets larger, a high f-number lens is required. The severe aberrations existing in the high f-number lens will affect the final image if the hologram does not capture all of the diffracted rays by the single lens. This situation is most noticeable for those points near the edge of the viewing aperture. Here, some of the rays are likely to miss the hologram because of aberrations. In this case, a viewer close to the edge of the viewing aperture sees very annoying background patterns on the lens aperture. A dark ellipse is the most commonly seen pattern. In this section, studies are made to identify various patterns resulting from various aberration terms. The hologram shape affects the pattern but in this study the shape is assumed to be a circle or a rectangle. The analysis is quite similar to Section 4.3, where a single-lens model is used to study the effect of the lens aberration on system resolution. In this section, the notation of Section 4.3 is adopted but the emphasis is on the shape of the reconstructed beam pattern in the big lens aperture.

Equation (4-21) is the fundamental equation which relates the coordinates (u, v) on the lens aperture to the coordinates (x, y) at any receiving plane z . If $z = 0$ or if z is very small compared to R , then Eq. (4-22) might be useful. The presentation here is similar to that of Section 4.3. The shape of the reconstructed beam pattern in the lens aperture caused by each third-order aberration term, is studied for both the circular hologram and the rectangular hologram.

1) Spherical Aberration: $\Delta = c_1 \rho^4$

For this aberration, a circular hologram centered at the origin of a given receiving plane z gives a circle that is centered in the lens aperture. The radii of the two circles are related to Eq. (4-26). For $z = 0$, their relation is

$$\rho' = - \frac{4c_1 R}{a} \rho^3$$

If a rectangular hologram centered at the origin is considered, the corresponding curves in the lens aperture are

$$\pm K = -\frac{z}{R} u + \frac{(R+z)}{a} (u^2 + v^2) u$$

(5-1)

$$\pm K = -\frac{z}{R} v + \frac{(R+z)}{a} (u^2 + v^2) v$$

where $\pm K$ is the coordinate of the hologram boundary in both x and y . The beam pattern of Eq. (5-1) looks like Fig. 5-3.

2) Astigmatism and Field Curvature: $c_1 h^2 \rho^2 \cos \phi + c_3 h^2 \rho^2$

For this aberration, a centered circular hologram yields a centered ellipse on the lens aperture, a relationship described by Eq. (4-30). The semi-major axis and the semi-minor axis for the ellipse are also given in Section 4-3. For a rectangular hologram, Eq. (4-21) for this aberration becomes

$$x = \left[-\frac{z}{R} + 2(c_2 + c_3) \frac{(R+z)h^2}{a^2} \right] u$$

$$y = \left[\frac{z}{R} + 2c_3(R+z) \frac{h^2}{a^2} \right] v$$

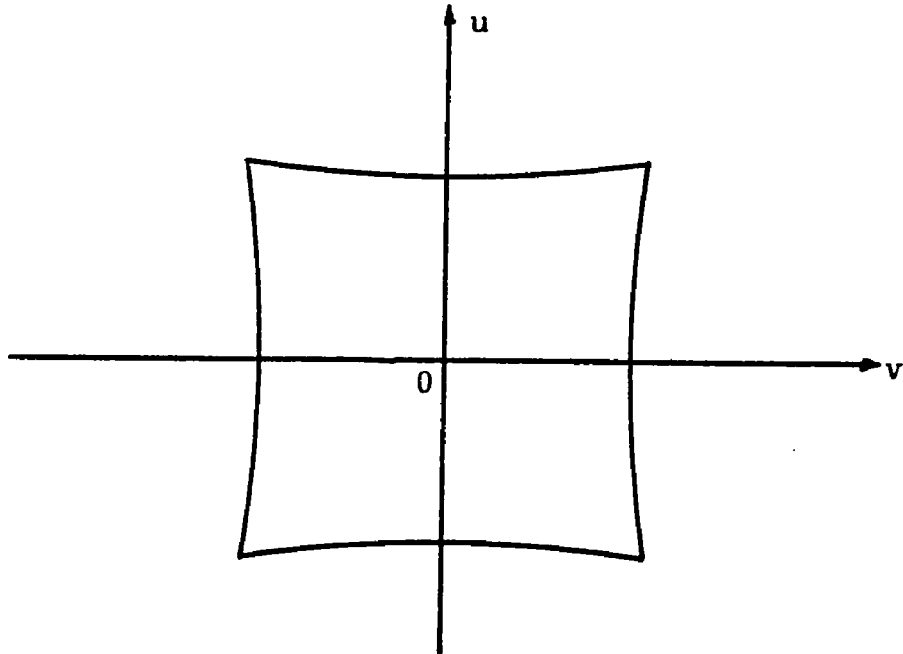


Figure 5-3. The beam pattern of Eq. (5-1)

This equation indicates very clearly that a straight line in the receiving plane will match a straight line in the lens aperture. Hence, in general, a rectangular hologram in any receiving plane will match a rectangular beam pattern in the lens aperture. There are, however, two special receiving planes where the rays coming from the lens aperture focus to a straight line. In one case, where $z = (2R^2h^2/a^2)(c_2 + c_3)$, the rays focus to a line $x = 0$. In the other case, where $z = (2R^2h^2/a^2)(c_3)$, the rays focus to a line $y = 0$. In these two planes, only the length of the focused line intercepted by the hologram aperture determines the shape of the beam pattern in the lens aperture. The hologram shape does not effect the beam pattern. The approximate beam pattern for these two cases is shown in Fig. 5-4.

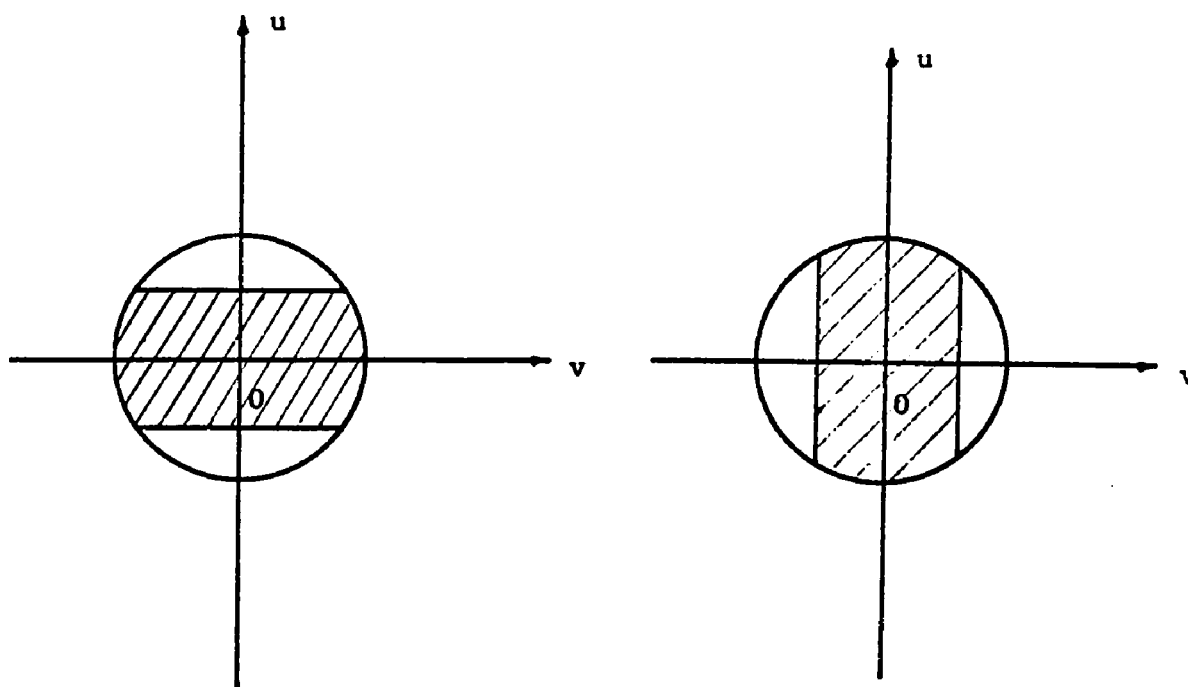


Figure 5-4. The beam pattern for the astigmatism and field curvature when the rays focus to a line $x=0$ or $y=0$

3) Distortion: $c_4 h^3 \rho \cos \phi$

For this term, since each object point will focus to a point image in the Gaussian image plane, the beam shape in the lens aperture will not be affected, as predicted from Gaussian optics. However, the shape of the viewer's aperture is deformed.

4) Coma: $c_5 h \rho^3 \cos \phi$

As was pointed out in Section 4.3, the beam pattern in the lens aperture corresponding to an on-axis circular hologram is

$$\rho'^2 = \left(\frac{z}{R}\right)^2 (u^2 + v^2) + c^2 (R + z)^2 (9u^4 + 10u^2 v^2 + v^4) - 2c \left(\frac{z}{R}\right) (R + z) (3u^3 + 3uv^2)$$

This expression is complex and difficult to visualize. The curve can be plotted by a digital computer, but we will not investigate its shape here.

Attention is now given to a hologram having a rectangular shape. In Section 4.3, Eq. (4-34) represents curves in the lens aperture corresponding to straight lines $x = K$ in a receiving hologram plane. For different K , Eq. (4-34) represents a set of ellipses in the lens aperture. When $K = -(z^2/12R^2\beta)$, Eq. (4-34) becomes a point ($u = z/6R\beta$, $v = 0$). Figure 4-6 shows the set of ellipses. For a rectangular hologram, we have to investigate the corresponding curves in the lens aperture for a straight line $y = \ell$ at any receiving plane. Using Eq. (4-21b) we have

$$y = \ell = -\frac{z}{R} v + 2\beta uv$$

This equation can be seen more clearly by rotating the (u, v) axis 45° to a new axis (u', v') . We have

$$\frac{\left(u' - \frac{z}{R} \frac{1}{2\sqrt{2}\beta}\right)^2}{\frac{\ell}{\beta}} - \frac{\left(v' + \frac{z}{R} \frac{1}{2\sqrt{2}\beta}\right)^2}{\frac{\ell}{\beta}} = 1 \quad (5-2)$$

Equation (5-2) is a hyperbolic curve with a major axis rotated 45° in the (u, v) coordinate system and centered at $(\frac{1z}{4R\beta}, 0)$. As in the case where $x = K$, Eq. (5-2) represents a set of hyperbolic curves corresponding to a set of straight lines $y = \ell$ in the receiving plane. Figure 5-5 shows the set of hyperbolic curves and its corresponding ℓ values. For a given rectangular hologram, with boundaries $x = K_1$, $x = K_2$, $y = \ell_1$, $y = \ell_2$, in the receiving plane, the corresponding beam pattern in the lens aperture can be found by identifying the corresponding curves of Eq. (4-34) for $x = K$, and Eq. (5-2) for $y = \ell$. Figure 5-6 is the superposition of two sets of curves of Figs. 4-6 and 5-5. From Fig. 5-6, it can be seen that for a given rectangular hologram near the origin of a receiving plane, the corresponding beam pattern in the lens aperture can be different. This depends on where the hologram is located. Figure 5-7 shows several different positions of holograms and their corresponding beam patterns. In the special case where $z = 0$, Eq. (4-34) reduces to

$$\frac{\frac{u^2}{K}}{\frac{3\beta}{3\beta}} + \frac{\frac{v^2}{K}}{\frac{3\beta}{3\beta}} = 1$$

and Eq. (5-2) reduces to

$$\frac{\frac{u'^2}{\ell}}{\frac{\rho}{\rho}} - \frac{\frac{v^2}{\ell}}{\frac{\rho}{\rho}} = 1$$

The corresponding sets of curves are plotted in Fig. 5-8. For a single-lens movie system, it can be seen that this aberration introduces the most significant deformation for the background of the lens aperture.

To conclude this section, we note that the beam pattern in the lens aperture, caused by the Seidel aberrations, can be studied by using Eq. (4-21) and tracing out the exact shape of the pattern for different hologram shapes. Studies included in this section are simple examples which are easily analyzed. For more complicated cases, a digital computer is required in order to plot out the exact beam pattern.

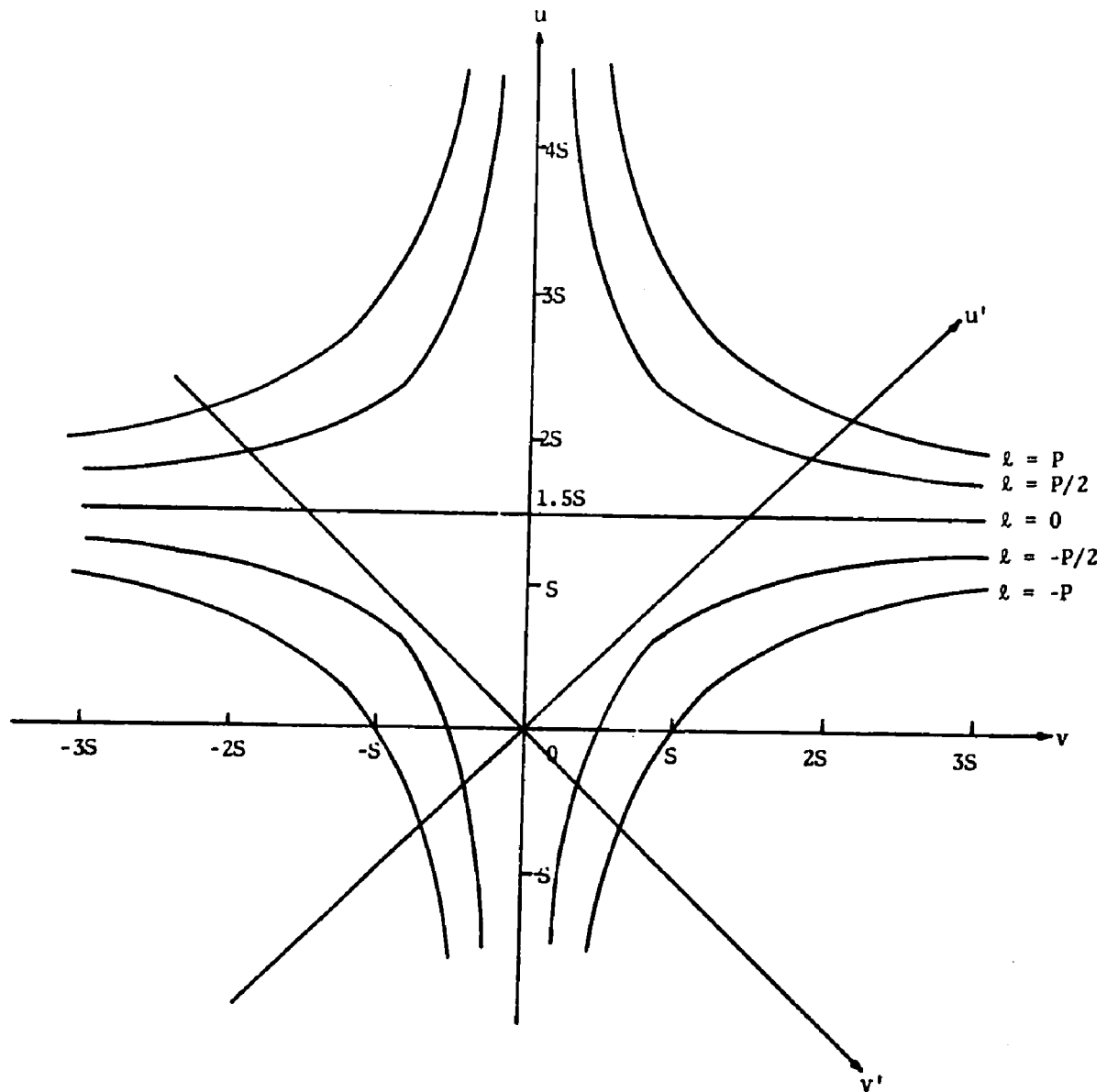


Figure 5-5. The set of hyperbolic curves in the lens aperture corresponding to different l where

$$S \triangleq z/6R\beta \quad , \quad P \triangleq z^2/12R^2\beta$$

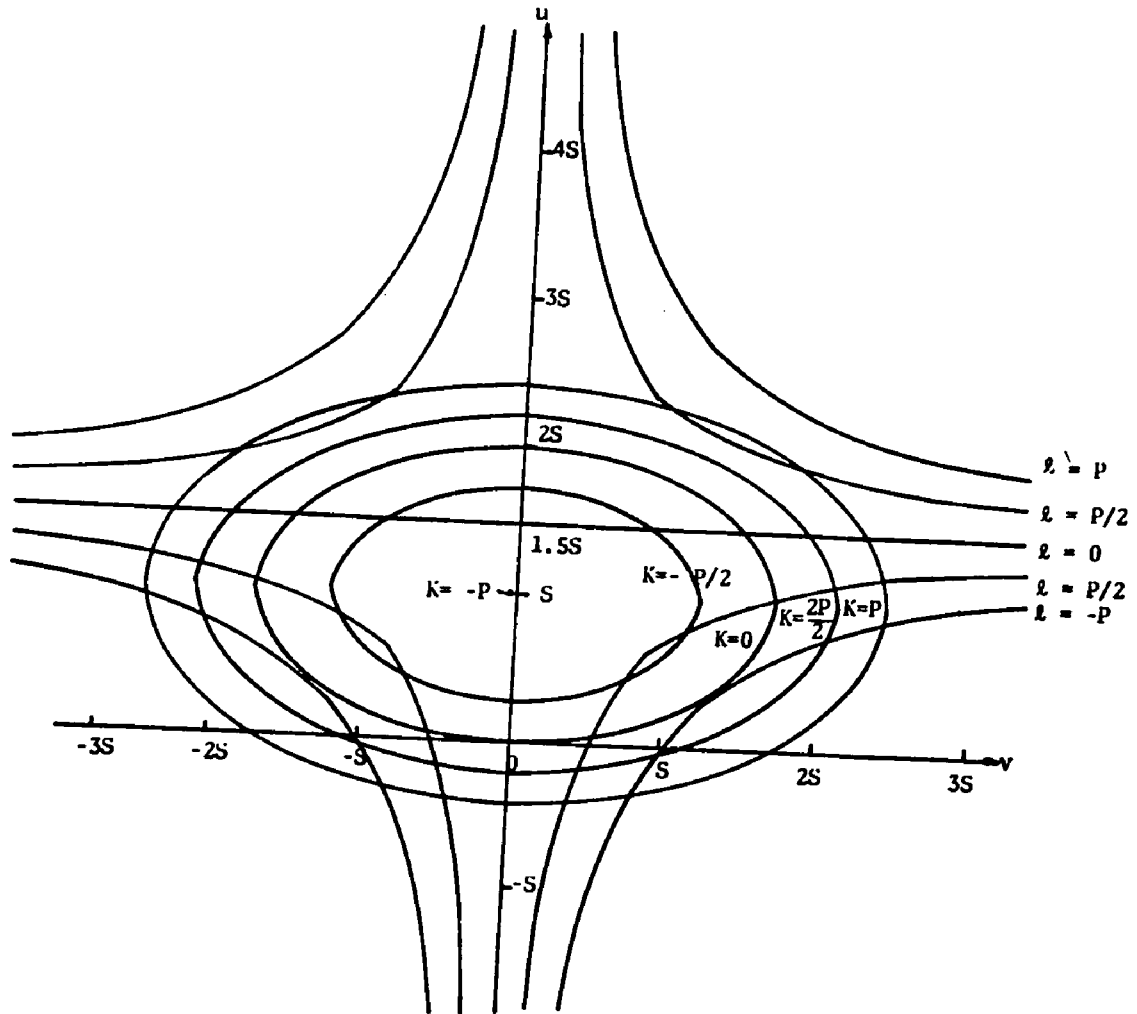


Figure 5-6. The set of curves in the lens aperture corresponding to $x = K$, $y = l$ in the aberration study of coma

$$S \triangleq z/6R\beta \quad , \quad P \triangleq z^2/12R^2\beta$$

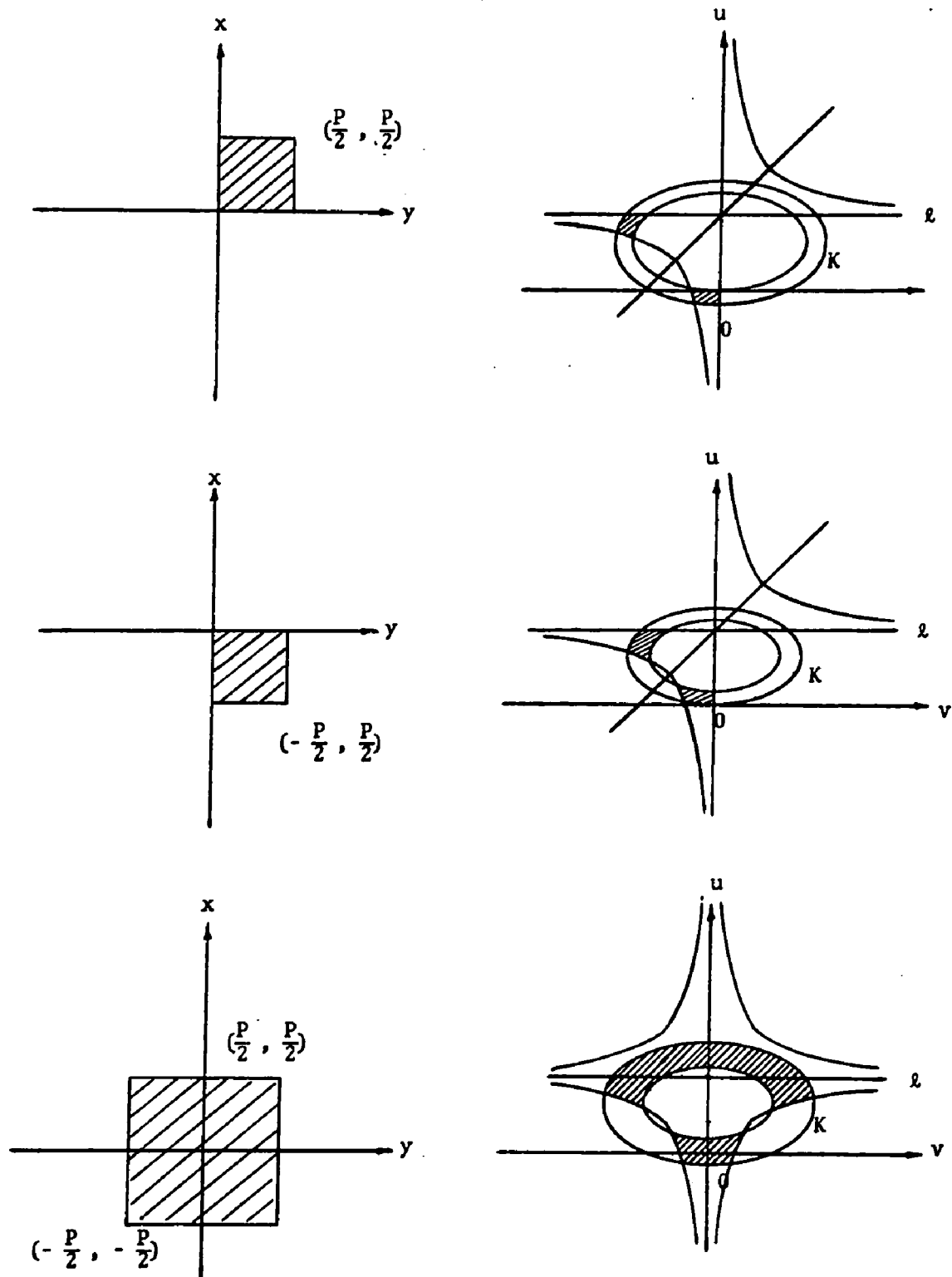


Figure 5-7. Different hologram configurations and their corresponding beam pattern in the lens aperture (The hatch area)

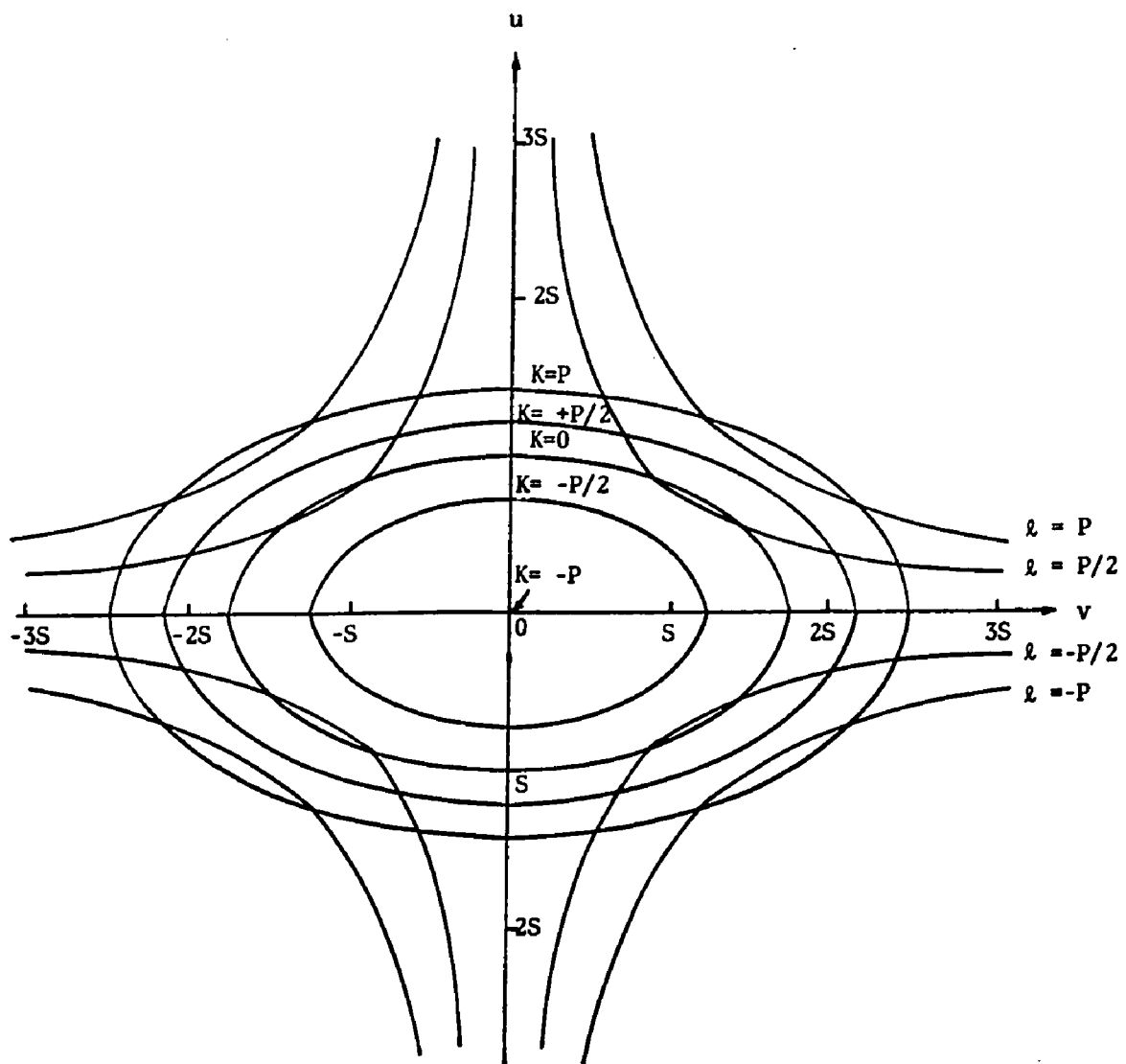


Figure 5-8. Special case of Figure 5-6 with
 $z = 0$

$$S \triangleq z/6R\beta \quad , \quad P \triangleq z^2/12R^2\beta$$

5.3. THE ABERRATIONS OF THE SINGLE-LENS MOVIE SYSTEM CAUSED BY OTHER SYSTEM PARAMETERS

In a single-lens movie system, we note that, ideally, no aberrations will appear in the reconstructed image since the aberrations inherent in the single-lens system cancel out during the reconstruction process. We do, however, have aberrations if the reconstruction process is not perfect. Various errors in the reconstruction process can cause degradation in the final image. One obvious error is a relative shift between the hologram and the single-lens system. Other causes can be a change of wavelength or a misalignment between hologram and readout beam. The aberrations introduced by the readout beam itself may also cause aberrations in the final reconstructed image. In this section, the aberrations caused by a lateral ΔX shift of the hologram relative to the lens system are investigated, as well as aberrations caused by other factors.

5.3.1. ABERRATIONS CAUSED BY ΔX SHIFT OF THE HOLOGRAM

Referring to Fig. 5-9, let us consider an object point (X'_0, Y'_0) a distance D to the right of the single lens H . In the recording process, this point will generate an ideal spherical wavefront G' (usually called a Gaussian reference sphere) in front of H . When passed through the lens, the wavefront emerging from the right of the lens is no longer a Gaussian reference sphere since the aberrations of the lens have been introduced. Let the wave aberration at G , caused by the single lens, be denoted by $\phi_1(X'_0, Y'_0, o, X, Y)$. An aberrated wavefront is recorded by the hologram and reconstructed during the readout process. If the reconstruction is perfect during the readout process, then the same aberration function $\phi_2(X'_0, Y'_0, o, X, Y) = -\phi_1(X'_0, Y'_0, o, X, Y)$ is retained at the wavefront G , but the propagation direction is reversed. When this aberrated wavefront passes through the lens the two aberrations are exactly cancelled and result in a perfect Gaussian reference sphere wavefront G' which emerges from the left side of the lens to form a sharp point image at (X'_0, Y'_0) again. Let $\phi_1(X'_0, Y'_0, o, X, Y)$ denote the aberration of the lens and $\phi_2(X'_0, Y'_0, o, X, Y) = -\phi_1(X'_0, Y'_0, o, X, Y)$ be the wave

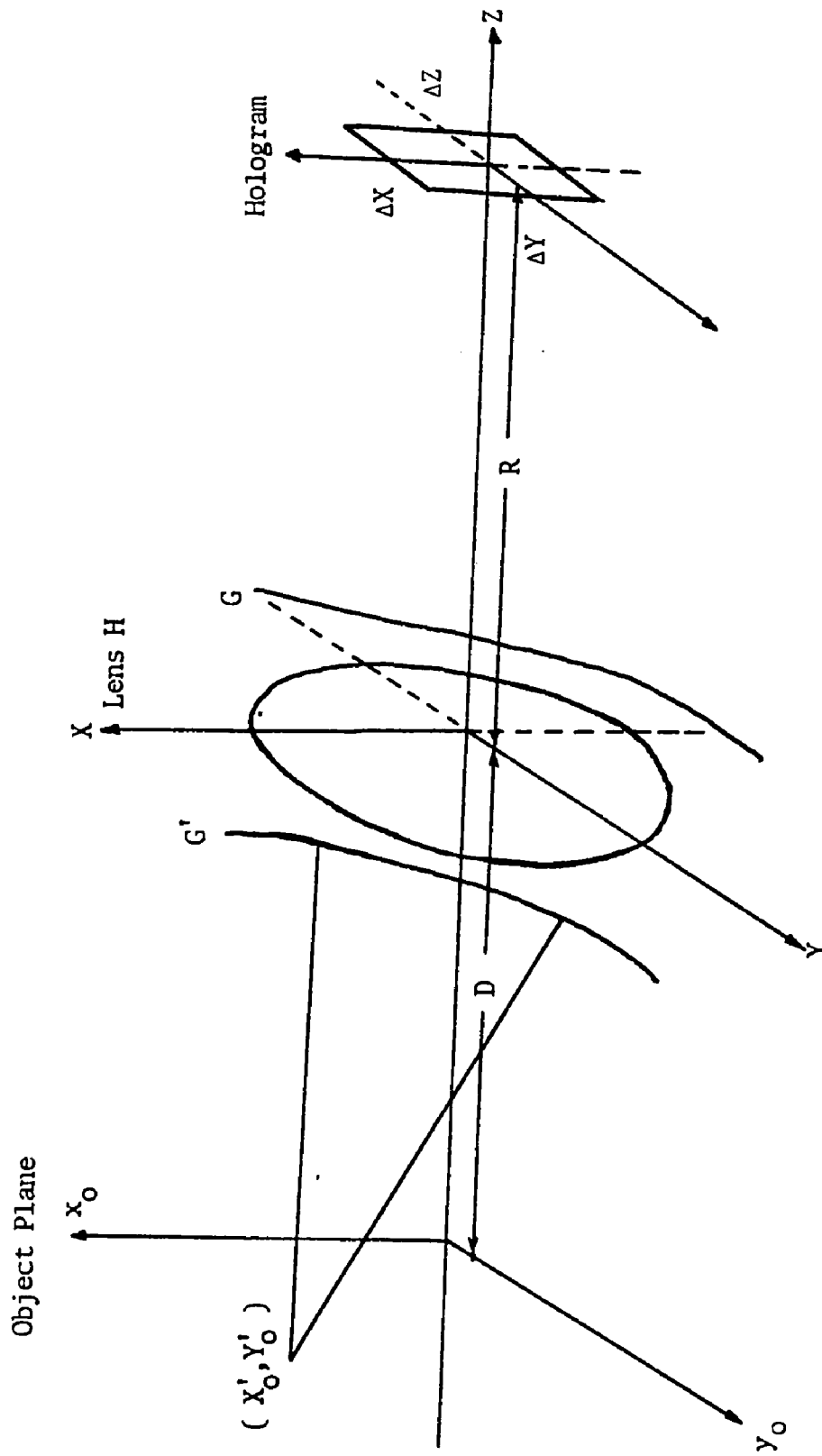


Figure 5-9. The schematic diagram of a single-lens movie system

aberration at G reconstructed by the hologram. The total wave aberration at G' is $\phi(X'_0, Y'_0, o, X, Y) = \phi_1 + \phi_2 = \phi_1 - \phi_1 = 0$. This is the case for a perfect reconstruction. Suppose during the readout process the hologram is moved by a distance ΔX in the lateral direction relative to the lens system; then the total wave aberration at G' becomes

$$\begin{aligned}\phi(X'_0, Y'_0, o, X, Y) &= \phi_1 + \phi_2 \\ &= \phi_1(X'_0, Y'_0, o, X, Y) - \phi_1(X'_0, Y'_0, o, X - \Delta X, Y)\end{aligned}\quad (5-3)$$

where $\phi_2 = -\phi_1(X'_0, Y'_0, o, X - \Delta X, Y)$ is the aberration caused by a shift ΔX of the hologram. It can be seen that ϕ_1 and ϕ_2 no longer cancel each other.

Now, consider the kind of aberration which is introduced because of the ΔX shift of the hologram. The notation for this analysis closely follows Section 5.2 and 5.3 of Born and Wolf's book [8]. We assume a single symmetrical lens system and consider only the Seidel aberrations. When we change coordinates from X'_0, Y'_0, X, Y in Eq. (5-3) to Seidel variables x_0, y_0, ξ_1, η_1 , and consider only the third-order aberrations (the Seidel aberrations), Eq. (5-3) becomes

$$\begin{aligned}\phi_{\text{total}}^{(4)}(x_0, y_0, \xi_1, \eta_1) &= \phi^{(4)}(x_0, y_0, \xi_1, \eta_1) - \\ &\quad - \phi^{(4)}(x_0, y_0, \xi_1 - \Delta\xi_1, \eta_1)\end{aligned}\quad (5-4)$$

where $\phi^{(4)}(x_0, y_0, \xi_1, \eta_1)$ is the third-order aberration term for which a symmetrical system can be expressed in a form

$$\phi^{(4)} = -\frac{1}{4} B\rho^4 - Ck^4 - \frac{1}{2} D r^2 \rho^2 + E r^2 k^2 + F \rho^2 k^2 \quad (5-5)$$

where

$$r^2 = x_0^2 + y_0^2, \rho^2 = \xi_1^2 + \eta_1^2, k^2 = x_0 \xi_1 + y_0 \eta_1$$

Equation (5-5) is a general form of the Seidel aberrations where each term represents a particular type of departure of the wavefront from the ideal spherical form. By substituting Eq. (5-5) into Eq. (5-4), we can immediately see what type of aberration will occur because of the original lens aberrations. It is advantageous to study Eq. (5-4) by substituting Eq. (5-5) term by term. In Eq. (5-5), term $-\frac{1}{4} B\rho^4$ represents a spherical aberration in the original lens system. Substituting this term into Eq. (5-4), we have

$$\phi^{(4)} = B\rho^2 \xi_1 \Delta \xi_1 + \dots (\Delta \xi_1)^2 \quad (5-6)$$

Since $\Delta \xi_1 \ll 1$, we can only consider the first-order term $\Delta \xi_1$ and neglect the effect of the higher-order $(\Delta \xi_1)^2$. Moreover, polar coordinates are also introduced so that $\xi_1 = \rho \cos \theta$, $\eta_1 = \rho \sin \theta$. Equation (5-6) then becomes

$$\phi^{(4)} = B\rho^3 \cos \theta \Delta \xi_1$$

In the coefficient of $\Delta \xi_1$, $\rho^3 \cos \theta$ represents coma. This coma is independent of r^2 (or x_o, y_o). For a fixed ρ , the circle has radius $|B\rho^2|$ and is centered at $-2B\rho^2$. We may conclude that if the original lens system has spherical aberration, then a ΔX shift of the hologram will introduce a coma aberration into the final reconstructed image. With the same kind of manipulation, we get the following coefficient of $\Delta \xi_1$ for the other types of aberration.

<u>Aberration Terms in the Original Lens System</u>	<u>Coefficient of $\Delta \xi_1$ in Equation (5-4)¹</u>
$-Ck^4$	$2C(x_o^2 \rho \cos \theta + x_o y_o \rho \sin \theta)$
$-\frac{1}{2} Dr^2 \rho^2$	$Dr^2 \rho \cos \theta$
Er_k^2	$-Er^2 x_o$
$F\rho_k^2$	$-2F(x_o^2 \rho^2 \cos^2 \theta + y_o^2 \rho^2 \cos \theta \sin \theta)$
	$-Fx_o \rho^2$

The physical interpretation of the above $\Delta\xi_1$ terms is as follows:

The expression $-2Fx_o\rho^2 \cos^2 \theta - Fx_o\rho^2$ is the astigmatism and curvature term. From

$$\begin{aligned}\Delta &= -2Fx_o\rho^2 \cos^2 \theta - Fx_o\rho^2 \\ &= -2Fx_o\xi_1^2 - Fx_o(\xi_1^2 + \eta_1^2) \\ &= -3Fx_o\xi_1^2 - Fx_o\eta_1^2\end{aligned}$$

it can be seen that in the sagittal ($\xi_1 = 0$) and tangential ($\eta_1 = 0$) planes, astigmatism and curvature appear as longitudinal focal errors; the wavefront possesses a different curvature in these two planes. The total separation between two focal lines is $4n Fx_o$, where n is the refractive index of the image space.

$$\text{Term } -2Fy_o\rho^2 \cos \theta \sin \theta = -2Fy_o\xi_1\eta_1:$$

For this term, the ray-aberration components are

$$\begin{aligned}\Delta^{(3)}_x &= x - x_o = \frac{\partial \phi}{\partial \xi_1} = -2Fy_o\eta_1 = -2Fy_o\rho \sin \theta \\ \Delta^{(3)}_y &= y - y_o = \frac{\partial \phi}{\partial \eta_1} = -2Fy_o\xi_1 = -2Fy_o\rho \cos \theta\end{aligned}$$

following Eq. (5) in Section 5.3 of Ref. [8]. This term gives a spherical aberration with the radius increasing with the first power of ρ , as well as the coordinate y_o . Both of the two terms above are derived from the original coma term $F\rho^2k^2$. For a general object point (x_o, y_o) , both aberration terms occur because of a ΔX shift of the hologram. If we consider an object point with $y_o = 0$, then only the astigmatism and curvature terms occur while for an object point with $x_o = 0$, only the spherical aberration term occurs.

$$\text{Term } \rho \cos \theta (2Cx_o^2 + Dr^2) = (2Cx_o^2 + Dr^2) \xi_1$$

For this term, the ray-aberration components are

$$\Delta^{(3)}_x = \frac{\partial \phi}{\partial \xi_1} = 2Cx_o^2 + Dr^2$$

$$\Delta^{(3)}_y = \frac{\partial \phi}{\partial \eta_1} = 0$$

This is just a distortion term.

$$\text{Term } 2Cx_o y_o \rho \cos \theta = 2Cx_o y_o \eta_1$$

For this term, the ray-aberration components are

$$\Delta^{(3)}_x = \frac{\partial \phi}{\partial \xi_1} = 0$$

$$\Delta^{(3)}_y = \frac{\partial \phi}{\partial \eta_1} = 2Cx_o y_o .$$

This, also, is a distortion term. Hence, the original astigmatism and curvature type of aberration contributes to the distortion of the final reconstructed image. The above analysis reveals some of the details of the aberration problems when lateral shift ΔX of hologram in our system occurs during the reconstruction process. The aberrations caused by other means are discussed briefly in the next section.

5.3.2. DISCUSSION OF ABERRATIONS CAUSED BY OTHER MEANS DURING THE RECONSTRUCTION PROCESS

Referring to Fig. 5-9 again, we have considered the problem of aberrations caused by a ΔX hologram shift. Disregarding aberration, a shift in the Z direction only moves the final image in the same direction. Since the hologram can be regarded as a recording of the object, the final image is just the image of the object recorded by hologram through the single-lens-imaging system. If the third-order Seidel aberration terms are considered, however, a ΔZ shift causes aberration in the final reconstructed image. The analysis in

this case is not as easy as the case of ΔX shift. The fact is that the two aberration terms ϕ_1 and ϕ_2 in Eq. (5-3) are no longer derived from the same aberration function. Nevertheless, ϕ_1 and ϕ_2 will still be in a form of Eq. (5-5). This means that whatever aberration terms exist in the original lens system also appear in the final image because of the ΔZ shift.

Having considered the effect of translation along the z axis, let us examine the effect of a rotation of the hologram relative to the lens system. A rotation of the hologram about any axis can be decomposed into a translation plus a rotation around the center of the lens point O , as shown in Fig. 5-10. Since D is much larger than the displacement of the object point (X_o, Y_o) because of the rotation, this displacement can be considered as a shift $\Delta X_o = D \tan \theta$ in the ϕ expression. We can study the total aberration caused by a rotation of angle θ using the same kind of derivation as in the case of ΔX shift of the hologram. In this case we have

$$\begin{aligned} \phi_{\text{total}}^{(4)}(x_o, y_o, \xi_1, \eta_1) &= \phi_1^{(4)}(x_o - x_o, y_o, \xi_1, \eta_1) \\ &\quad - \phi_1^{(4)}(x_o, y_o, \xi_1, \eta_1) \end{aligned}$$

The details of the derivation are neglected here.

Finally, consider the effect of changing the wavelength of the readout beam in the readout process. A straightforward analysis is very complicated, since a change in wavelength causes both aberrations and a change in magnification of the reconstructed wavefront in the lateral and the longitudinal directions. If an arrangement is made so that the sole effect of changing the wavelength is to change the size of the reconstructed wavefront just in front of the lens, then the analysis is much simpler. In this case, the variables x_o, y_o, ξ_1, η_1 in Eq. (5-3) change to $mx_o, my_o, m\xi_1, m\eta_1$, where m is the magnification factor. From this equation, the total aberration of Eq. (5-4) is

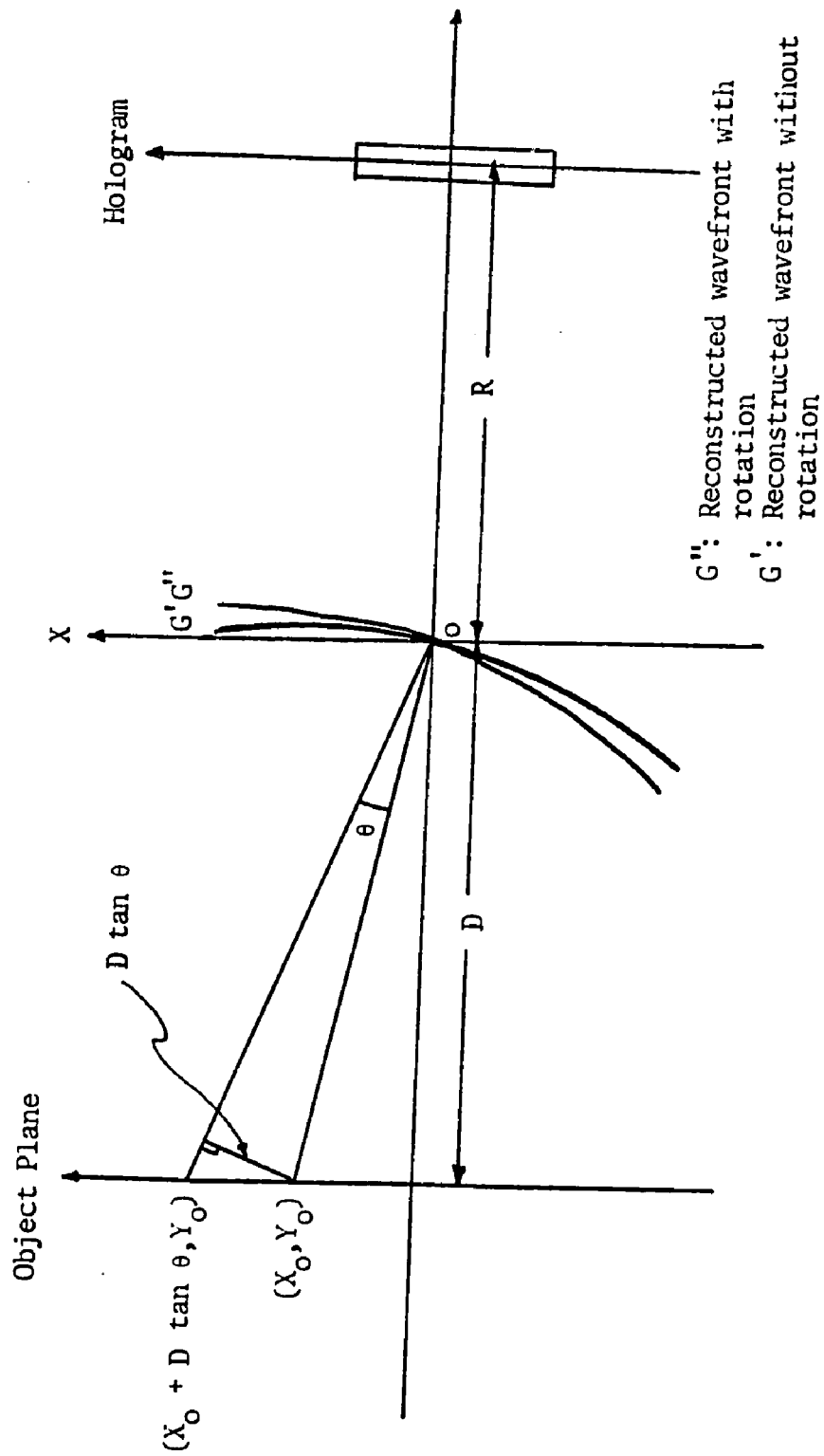


Figure 5-10. The schematic diagram of a single-lens movie system for the study of a rotation aberration

$$\begin{aligned} \phi_{\text{total}}^{(4)} = \phi_m^{(4)} - \phi^{(4)} = (m^4 - 1) \left(-\frac{1}{4}B\rho^4 - Ck^4 - \frac{1}{2}Dr^2\rho^2 \right. \\ \left. + Er_k^2 + F\rho_k^2 \right) \end{aligned}$$

This means a change of wavelength would cause the same aberration as the original lens, but with a larger factor $(m^4 - 1)$. Since m is the magnification factor caused by the wavelength change, a minor change in wavelength causes severe aberration in the final reconstructed image.

5.4. OTHER ENGINEERING CONSIDERATIONS

Most of the work for this system involves engineering considerations and experiments to improve the system performance. In this section, some of the engineering considerations are discussed. The results of the experiments are reported in Chapter 6.

5.4.1. THE PSEUDOSCOPIC IMAGE INVERSION PROBLEMS (INCLUDING THE USE OF THE FLY'S-EYE LENS ARRAY)

The images which can be seen in a single-lens system are pseudoscopic. In order to obtain an orthoscopic image, two methods are considered. One method is use of the single-lens system twice. In the first lens system, a real image is formed. In the second lens system, an orthoscopic image is obtained by use of the real image formed by the first lens system. This latter method has been discussed in Ref. [10] by E. N. Leith. In Ref. [10], four basic types of systems have been considered and the special advantages and disadvantages have been discussed for each type. In our experiment, the VV type of system has been tried with unsatisfactory results. This is caused by too many (16) reflecting lens surfaces throughout the whole recording and reconstruction process. The minimum number of lens surfaces for the whole process is eight. Hence, the surface quality of the lens for this method should be very good. A second method for pseudoscopic image inversion is the use of a fly's-eye lens array and a reflecting mirror behind it. This method has been studied theoretically and experimentally

by several researchers [26-29]. The arrangement is shown in Fig. 5-11, in which the pseudoscopic image is obtained from a real object. The basic principle of this setup has been described by C. B. Burckhardt, et al. [26]. Because of the availability of high-quality fly's-eye lens arrays, we did some experiments on the image quality. The results are reported in Section 6.4. The arrangement in Fig. 5-11 can be applied to any imaging system to accommodate the image inversion. An alternative arrangement, in which two fly's-eye lens arrays with ground glass between them is used has been proposed by A. H. J. DE.L. Saint Genies [29]. The use of a lenticular sheet manufactured by the 3M Company has also been tried without good results.

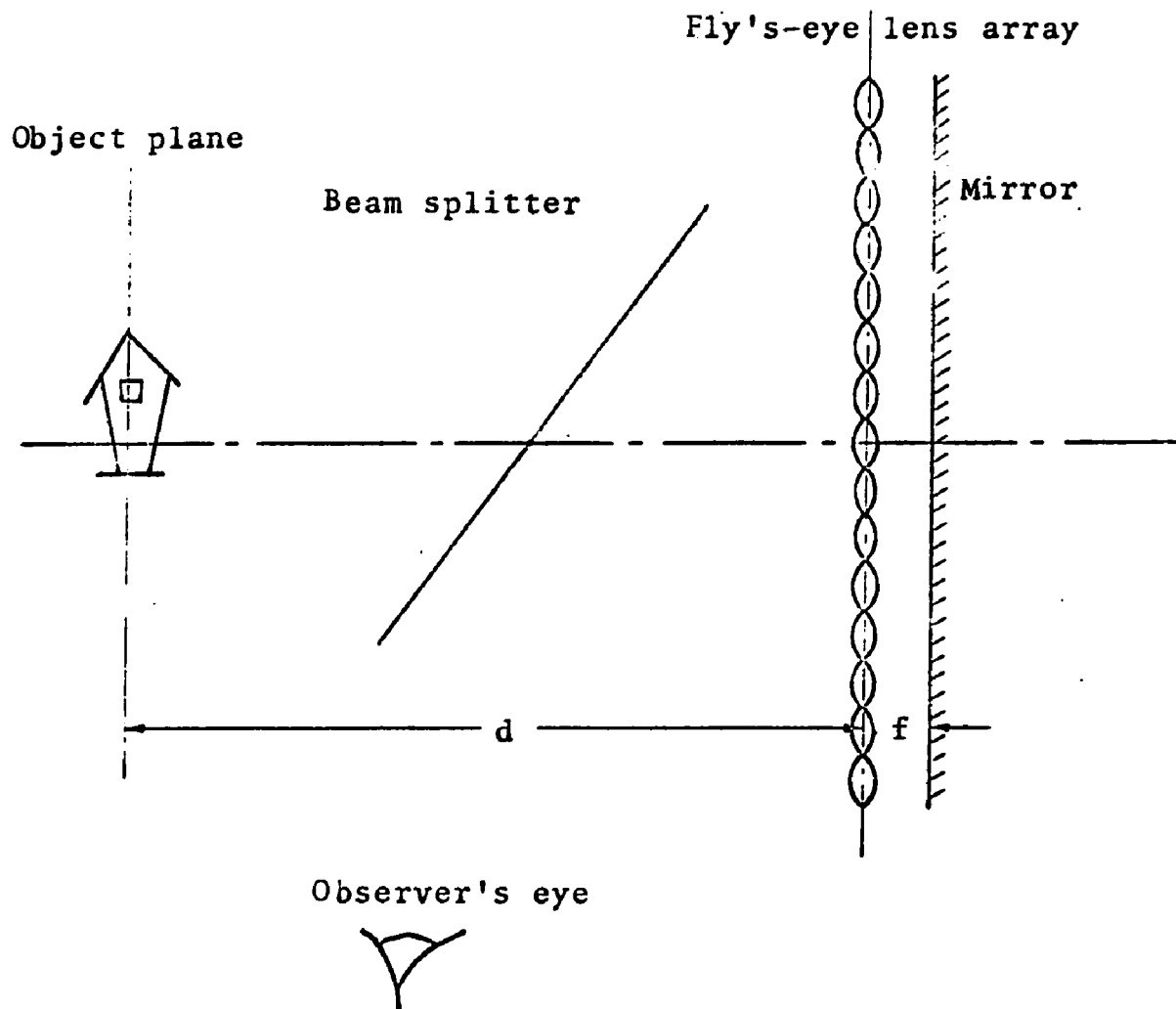


Figure 5-11. The system setup for pseudoscopic image inversion

5.4.2. THE NOISE CONSIDERATION OF THE SINGLE-LENS MOVIE SYSTEM

Noise is one of the fundamental problems in any 3-D display system. In the fly's-eye lens array system, we have discussed the signal-to-noise ratio based on diffraction theory in Section 4.2. The noise of the holographic film has not yet been considered. In the single-lens movie system, the noise of the hologram film is quite sensitive to the system performance. This is because the resolution of the system is usually good and the signal-to-noise ratio thus becomes a dominant factor of the system performance. There are two major sources of noise in the hologram recording and reconstruction process. One is the inter-modulation product noise resulting from the nonlinearity of the film. Second is the film-grain noise caused by the random fluctuation of the film grains. These two kinds of noise have been studied by several researchers [30-32]. The methods to reduce these two kinds of noise depend on the choice of recording media and the techniques of film processing. Efforts to reduce them have been made by J. Upatnieks and C. Leonard during this study. Another two sources which cause the noise background of the reconstructed image are the film-relief image [33] and the variation of film-emulsion thickness [37]. The method to reduce these two noises is to liquid gate the movie film. The liquid is chosen so that the index of refraction of the liquid matches the index of refraction of the film emulsion. The liquid should also be somewhat volatile so that the movie film can be dried quickly after it runs through the liquid gate. The film-noise improvement resulting from a liquid gate is shown in Section 6.5.

5.4.3. THE PRODUCTION OF HOLOGRAPHIC MOVIE FILM AND THE FILM-DRIVE PROBLEM

During the course of this study, a sequence of about 15 frames of movie strips was produced by a quite laborious method. A continuous-wave argon laser was used. Each frame requires about five minutes, including the time for exposing the film and moving the object scene by hand. A period of about three minutes waiting time is necessary before each exposure to let the object scene become stable after the disturbance of

the rearrangement. A pulse laser would have helped in this experiment. The stability of the laser power to provide equal exposure for each frame is also important. Our experience indicates that even with rather stable laser power, one or two frames will sometimes become dim. The stability problem of the laser power, the optical bench, the object scene, and the whole system is an important area which requires more improvement for further advancement.

Another problem for the system is the film-drive problem. As mentioned before, a liquid gate is necessary for the film to reduce the film noise. A film transport with liquid gate was therefore designed and built for the system. The film moves through the gate in steps rather than continuously. The supply of the liquid is fed by a liquid pump. For this film transport, the film can be driven at a rate of up to five frames per second with a satisfactory reconstruction image. When the frame rate increases, the air gets into the gate and degrades part of the image. More liquid is also drained out of the gate when the drive speed gets higher. This limits the frame speed. We use xylene in our liquid gate. This xylene is volatile enough, but its vapor is harmful to human health. A good ventilation system is necessary to operate this film transport. The ultimate goal of 24 frames per second can be obtained by a different design of the film transport with much more effort. The liquid in use should also be chosen so that it will not be harmful to one's health.

CHAPTER 6

THE EXPERIMENTAL RESULTS AND CONCLUSIONS

A large number of experiments have been done throughout the course of this study. In some cases, an experiment preceeded a theoretical interpretation. In other cases, a theory was first derived and an experiment was then performed to verify the theory. In this chapter, the experimental results of the fly's-eye lens array system, the scatter-plate system, the single-lens movie system and the pseudoscopic image inversion techniques are reported. Comments and interpretations are included.

6.1. THE EXPERIMENTAL RESULTS OF THE SCATTER-PLATE METHODS

To perform the experiment for the information-reduction method, a fly's-eye lens array was used as a scatter plate. The set up for the experiment was similar to Fig. 3-7, and a point source at the edge of the fly's-eye lens array was used as reference. USAF resolution test target 1951 was used as the object. The system parameters for which the notation of Fig. 3-7 were used are:

$$m = 1\text{mm}$$

$$s = 3\text{mm}$$

$$Z = 247\text{mm}$$

$$z = 5\text{cm}, 10.5\text{cm}, 15\text{cm}, 20\text{cm}, 25\text{cm}, 30\text{cm}, 40\text{cm}$$

$$K = 10\text{mm}, 5\text{mm}, 2.5\text{mm}$$

where three hologram apertures are considered. Figure 6-1 shows the experimental results of the image resolution against its theoretical predictions. In Fig. 6-1, each solid line represents the resolution

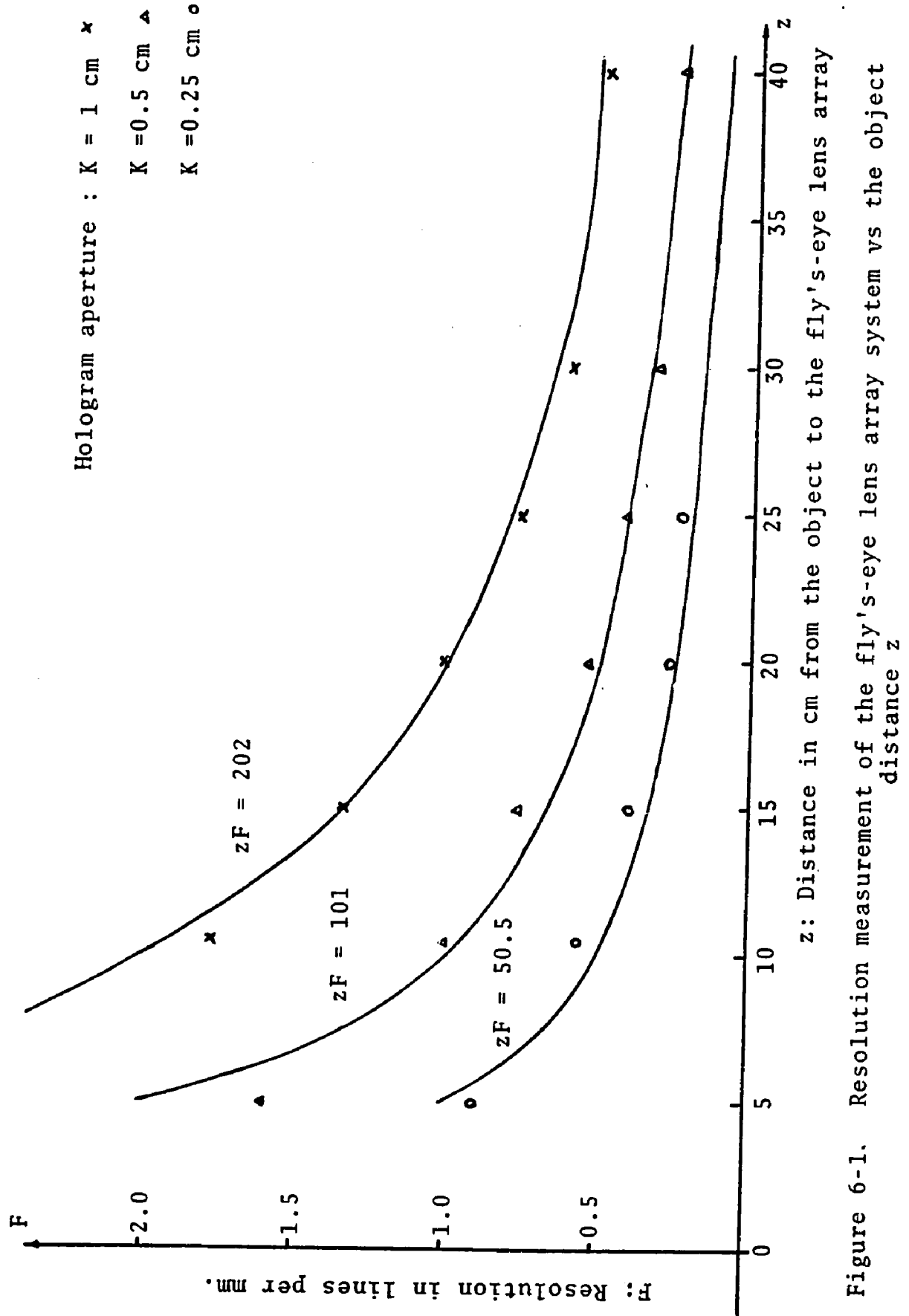


Figure 6-1. Resolution measurement of the fly's-eye lens array system vs the object distance z

of the object target versus the distance between the object and the fly's-eye lens array predicted by geometric optics. The experimental results for different hologram apertures $K = 1\text{cm}$, 0.5cm and 0.25cm at different object distances are represented by symbols x , Δ , o , respectively. The experimental results are consistent with the theory when the object distance is sufficiently large. The theoretical curve for $K = 1\text{cm}$ is calculated as an example. If the system parameters given previously are used, the resolution aperture ℓ at the single lenslet can be calculated to be

$$\ell = \frac{Ks}{Z} = 0.121\text{mm} .$$

The spatial frequency F of the reconstructed image is the reciprocal of the image resolution R :

$$\frac{1}{F} = R = \frac{\lambda z}{\ell}$$

or

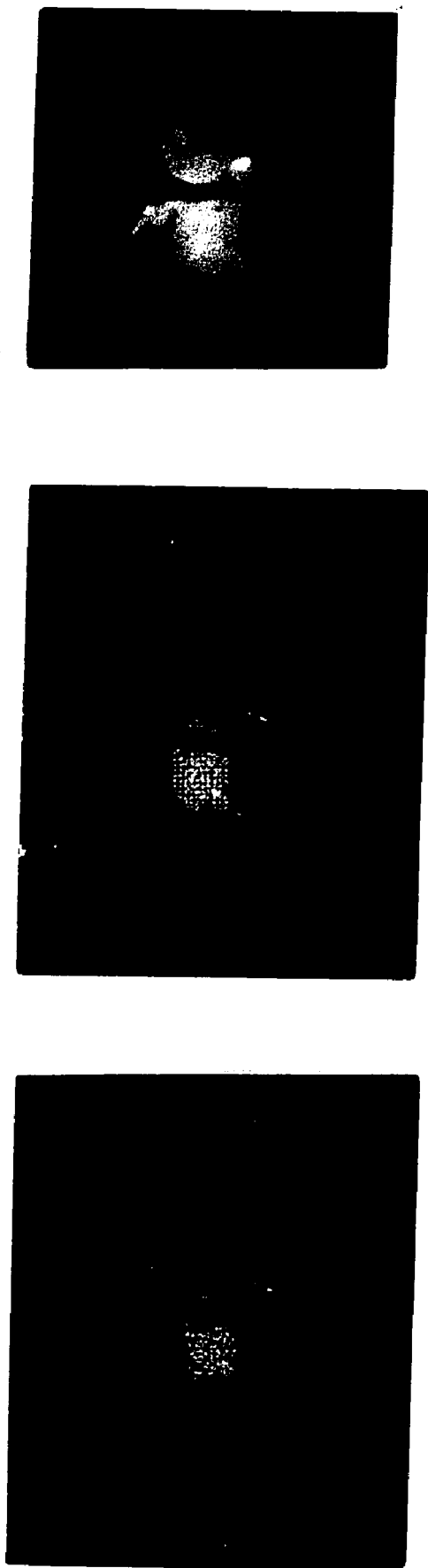
$$zF = \frac{\ell}{\lambda} = \frac{0.121}{\lambda}$$

where z is the distance between the object and the fly's-eye lens array. When $z = 20\text{cm}$, F is calculated to be 1.01 lines per mm. The curve

$$zF = \frac{0.121}{\lambda} = \frac{0.121}{0.6 \times 10^{-3}} = 202$$

is plotted in Fig. 6-1. The experimental results are obtained by examining the resolution group in the USAF resolution chart.

One practical problem of the fly's-eye lens array system is the presence of laser speckle in the reconstructed image. The laser speckle is caused by the interference of the coherent light. Hence it will be present in any coherently illuminated imaging system. Figure 6-2 (a) shows a 3-D object with laser speckle. The



(a)

(b)

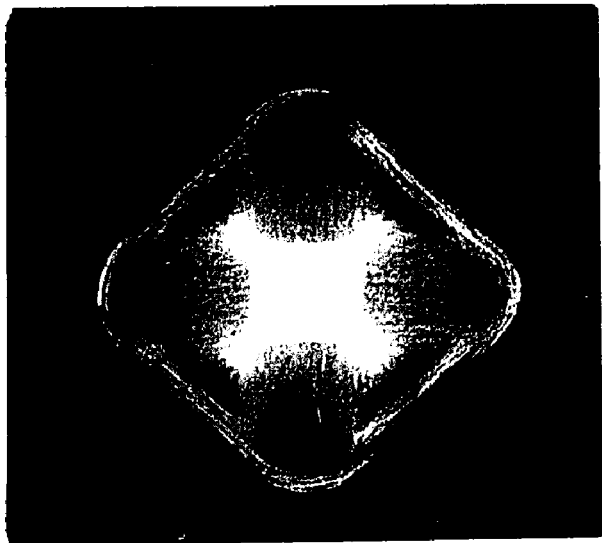
(c)

Figure 6-2. The reconstructed 3-D images when a fly's-eye lens array system is used

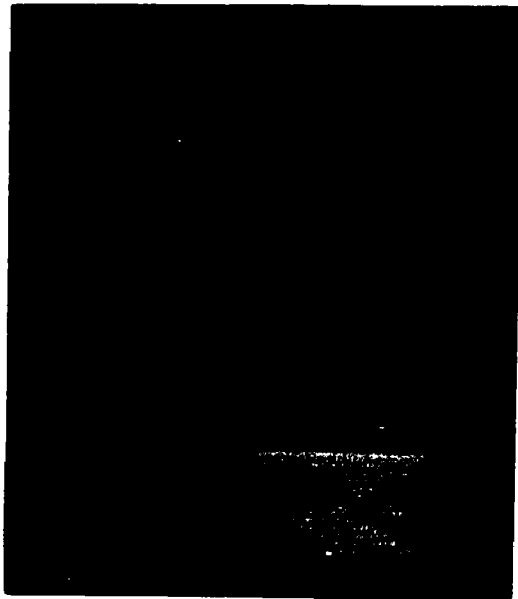
- (a) This picture shows the presence of laser speckle pattern in the reconstructed image
- (b) This picture shows the speckle pattern is reduced by placing a rotating etched glass in the illuminating beam
- (c) This picture shows the periodic structure of the fly's-eye lens array has been smoothed out by putting a diffuser at the central plane of the reconstructed image

laser speckle can be reduced by using incoherently or partially incoherent light in the reconstruction process. This can be achieved by putting a rotating ground glass at the focal point of the reconstruction beam. The rotating ground glass must be put at the place where it will not affect the wavefront of the reconstruction beam. For diverging and converging types of reconstruction beam, the focal point is the only place to put the rotating ground glass without affecting the reconstruction wavefront. In a collimated reconstruction beam, the ground glass can be located anywhere. Figure 6-2 (b) was obtained by putting a rotating ground glass at the focal point of the reconstruction beam. The laser speckle has been greatly reduced by this method. The picture in Fig. 6-2 (b) has some fine structures because of the periodic nature of the fly's-eye lens array. This structure can be smoothed out by putting a diffuser at the center of the reconstructed image. The depth of field for the image will be reduced but the fine structure can be eliminated. Figure 6-2 (c) shows the result obtained from using this method which can be used only when the required depth of field is small.

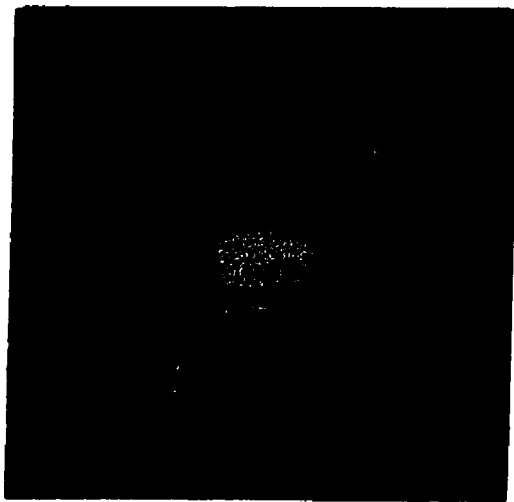
For purposes of comparison, another type of scatter plate, the ground glass, was tried by D. B. Brumm [16]. For this type of scatter plate, the resolution is not degraded very much, but the noise increases when the hologram aperture decreases. In the fly's-eye lens array system, the resolution degrades substantially when the hologram aperture decreases. Several other scatter plates have been tried and found to yield poor resolution and signal-to-noise ratio. One scatter plate supplied by Holotron has the spatial spectrum shown in Fig. 6-3 (a). Figure 6-3 (b) shows a reconstructed image using this scatter plate. The object was diffusely illuminated with laser light. The information reduction factor for this picture is 62. Figure 6-3 (c) shows a reconstructed image of another type of object. The object was illuminated by a ground glass in the background. The picture was taken by focusing the camera on one of the two poles. A random type of scatter plate was also tried in the study. Figure 6-4 (a) shows the



(a)

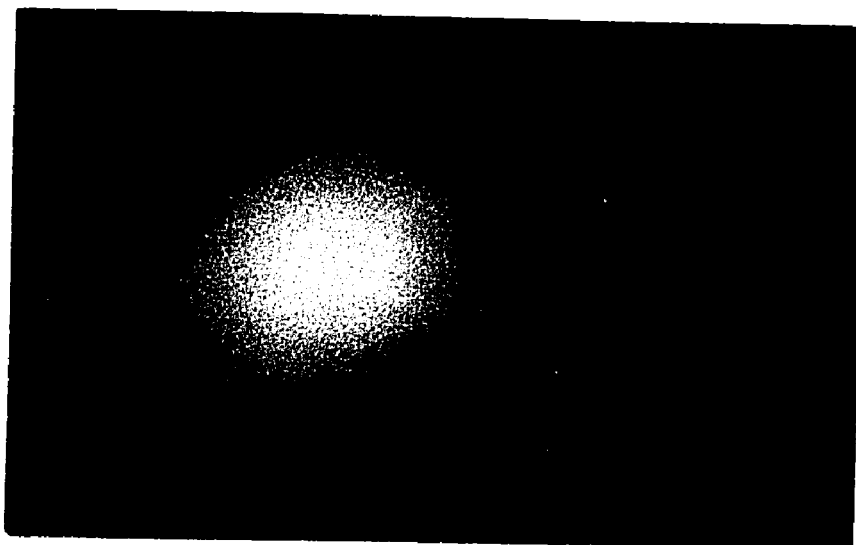


(b)

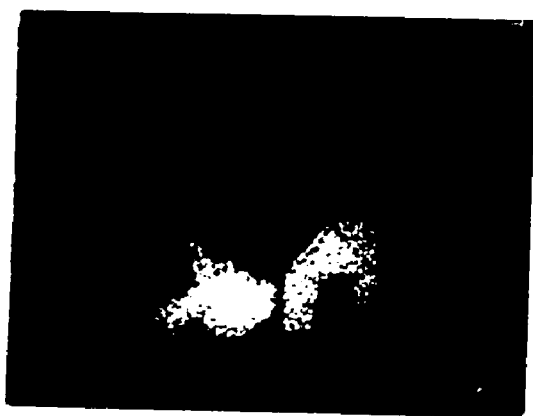


(c)

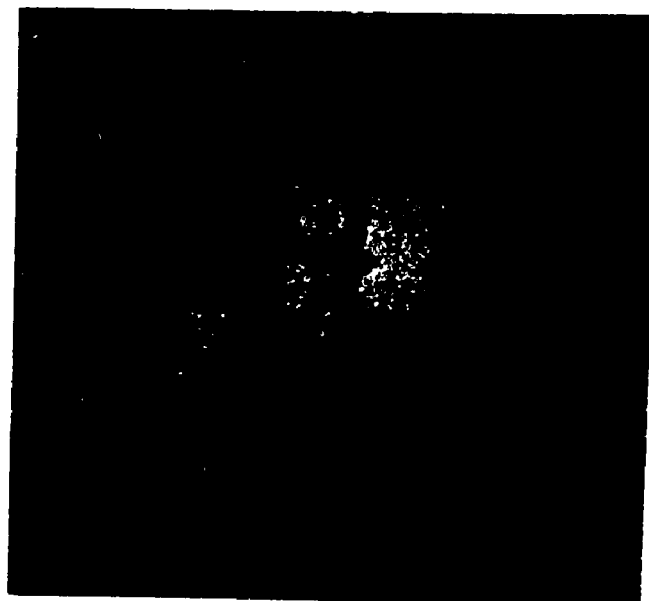
Figure 6-3 (a) The spatial spectrum of a scatter plate supplied by Holotron.
(b) The reconstructed 3-D image using the above scatter plate. The object is diffusely illuminated.
(c) The reconstructed image of a silhouette object.



(a)



(b)



(c)

Figure 6-4. The reconstructed images of a random scatter-plate system

(a) A transparent type of object

(b) A diffusely illuminated 3-D object

(c) A silhouette object

picture of a transparent type of object. Figure 6-4 (b) and (c) shows the reconstructed images of the same type of objects as Fig. 6-3 (b) and (c) respectively.

6.2. THE ABERRATION IN A SINGLE-LENS MOVIE SYSTEM

In this section, we present the experimental results for the aberrations in a single-lens movie system to verify the analysis made in Section 5.3. The set up of the experiment was exactly the same as shown in Fig. 5-7, where $D = 34$ inch, $R = 18$ inch. A single condensing lens 10 inches in diameter was used. This lens has severe spherical aberration as will become evident later. A point object source was placed at the center of the object plane (i.e., $X_0 = 0$, $Y_0 = 0$) and the rays from the point source were recorded on the hologram with a collimated reference beam. To reconstruct this point source, a reversed collimated beam was used to illuminate the hologram and a real image was formed at the original location of the point source. When we examined the rays which converged to the point source during the reconstruction, we found that the rays were not coming uniformly from the whole aperture of the lens. Instead, there was a black circle within the aperture of the lens where the reconstruction rays were formed. This is shown in Fig. 6-5, where the picture was taken somewhere between the lens and the real image point. As shown in Fig. 6-6, the lens being used has severe spherical aberration. Hence, during the recording process, the rays from the outer portion of the lens were focused closer to the lens, while the rays from the inner portion of the lens were focused farther away from the lens. Since the hologram is placed somewhere near the image point, these rays form a circle at the hologram plane where the edge of the circle corresponds to the rays coming from a ring in the lens aperture somewhere between the edge of the lens and the center of the lens. In the reconstruction process, the edge of the circle in the hologram plane cannot be reconstructed because of the interference between two continuous wavefronts which intersect at the edge of the circle. This observation indicates

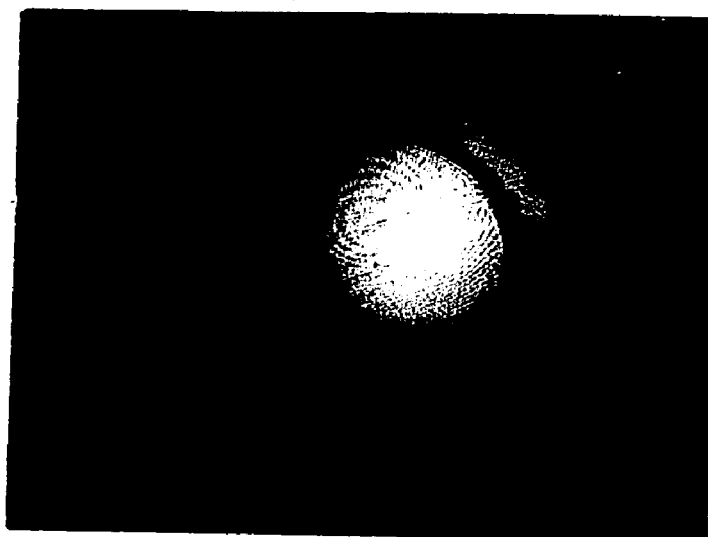


Figure 6-5. Picture of the black circle in the reconstruction beam

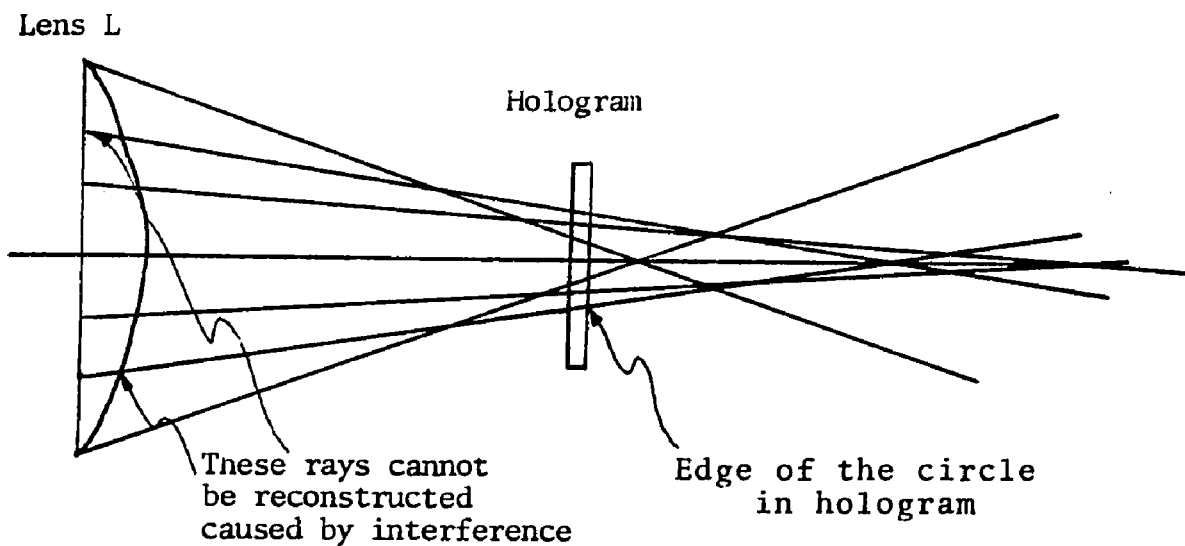


Figure 6-6. Spherical aberration causes an interference ring in the reconstruction beam

one disadvantage of the single lens we used, since a black ring in the reconstructed image degrades the final image. We now proceed to investigate the effect of the displacement of the hologram during the reconstruction process.

First, the hologram is moved $1/8$ inch perpendicular to the optical axis. Figure 6-7 shows the resulting images. The right picture shows the effect when the hologram was moved to the left and the left picture shows the effect when the hologram was moved to the right. The picture in the center is the original perfectly reconstructed point source. It can be seen very clearly that the aberrated image has a severe coma aberration. This demonstrates the theoretical prediction that a displacement of the hologram perpendicular to the optical axis will cause a spherical aberration from the original lens system to appear as a coma aberration in the final image. Experimental verification of the other aberrations in the analysis is not easy to carry out since we cannot isolate any single aberration terms in the experiment.

Second, the effect of a displacement of the hologram along the optical axis is shown in Fig. 6-8, where the center picture is the undistorted point source; the pictures on the right and left correspond to the image when the hologram was moved $1/8$ inch forward and back along the axis. These pictures verified the statement that a shift of the hologram along the axis will introduce the same aberration terms that the original lens system has.

The third and last experiment was concerned with a change of wavelength in the read-out process. Figure 6-9 shows the results. In Fig. 6-9 (a), a perfect point source image was formed with $\lambda = 5120\text{\AA}$, both in recording and reconstruction. Without changing the wavelength used for the recording (i.e., $\lambda = 5120\text{\AA}$), Fig. 6-9 (b) shows the reconstructed image when the wavelength of the read-out beam is changed to $\lambda = 5000\text{\AA}$; Figs. 6-9 (c), (d) and (e) correspond to the wavelengths $\lambda = 4940\text{\AA}$, $\lambda = 4850\text{\AA}$, $\lambda = 4730\text{\AA}$, respectively, in the read-out beam. The more the wavelength changes, the larger the

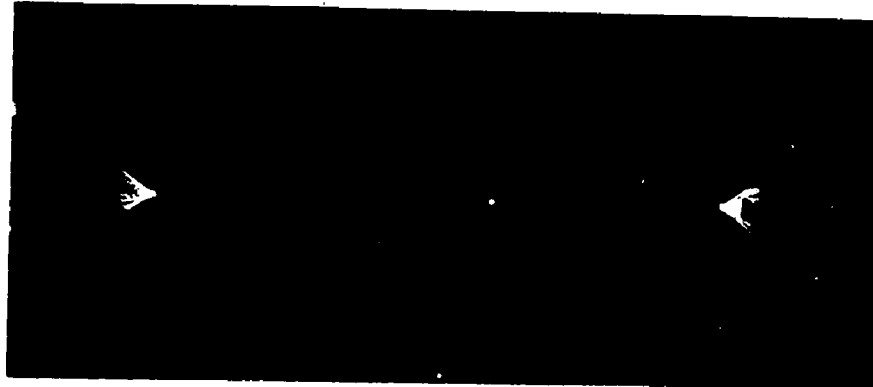


Figure 6-7. Aberrations caused by ΔX shift of the hologram in a single-lens movie system

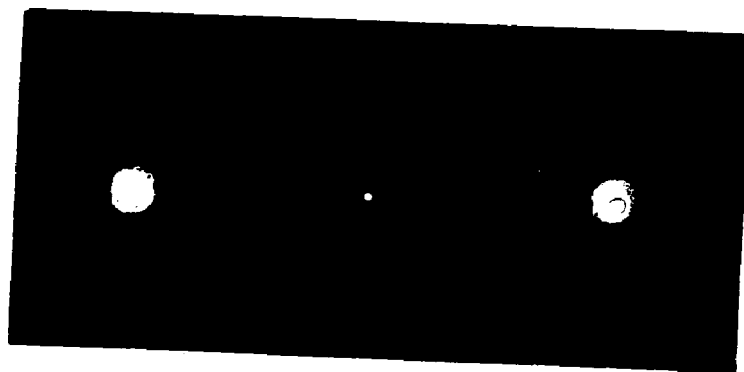
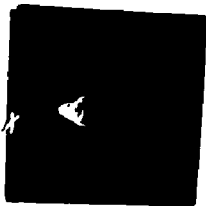


Figure 6-8. Aberrations caused by ΔZ shift of the hologram in a single-lens movie system

(a) $\lambda = 5120\text{\AA}$



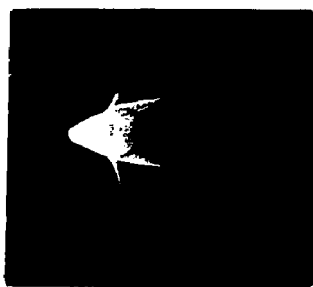
(b) $\lambda = 5000\text{\AA}$



(c) $\lambda = 4940\text{\AA}$



(d) $\lambda = 4850\text{\AA}$



(e) $\lambda = 4730\text{\AA}$

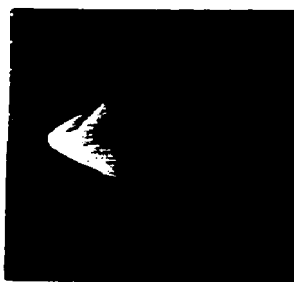


Figure 6-9. Aberrations caused by wavelength change
in a single-lens movie system

aberration pattern becomes. It should also be noted that the aberration patterns in Fig. 6-8 and 6-9 are not similar. This indicates that the details in the analysis of the two cases are completely different. The complexity of the pattern itself indicates that the analysis is difficult.

6.3. THE RESULTS OF THE PSEUDOSCOPIC IMAGE INVERSION

In this section, the experimental results of the pseudoscopic image inversion using a fly's-eye lens array and a mirror are reported. The experimental set up and the system dimensions are shown in Fig. 6-10, where a real object is inverted to a pseudoscopic image. Figure 6-11 (a) is a picture of the real object. Figure 6-11 (b) is the pseudoscopic image with distance $d = 3$ inches between the object and the fly's-eye lens array. Figures 6-11 (c) and (d) correspond to the pictures at distances $d = 8$ " and $d = 13$ " respectively. All pictures were taken with a 4.5/f camera. Figure 6-11 (e) is the same picture as Fig. 6-11 (b) except that the camera was focused on the fly's-eye lens array with 32/f aperture. This indicates that the pseudoscopic image, when viewed with human eye, has a square structure as a background. In order to eliminate this structure, a diffuser can be placed at the center of the reconstructed image, as we did in Fig. 6-2 (c). The image with a diffuser is shown in Fig. 6-11 (f). A retro-reflective screen manufactured by the 3M Company was also tried as an autocollimator to perform the image inversion. The result was much poorer than the fly's-eye lens array.

6.4. THE RESULTS OF THE SINGLE-LENS MOVIE SYSTEM

With a single-lens system, a very high quality image can be reconstructed. Figure 6-12 (a) shows the resolution target reconstructed obtained by use of a Fresnel lens. The Fresnel lens is made of plastic and has a poor surface quality. This makes the reconstructed image somewhat noisy, especially when viewed with the human eye. However, the resolution is still very good because of the cancellation of the lens aberrations. Figure 6-12 (b) is the reconstructed image of the same target when a condensing lens is used.

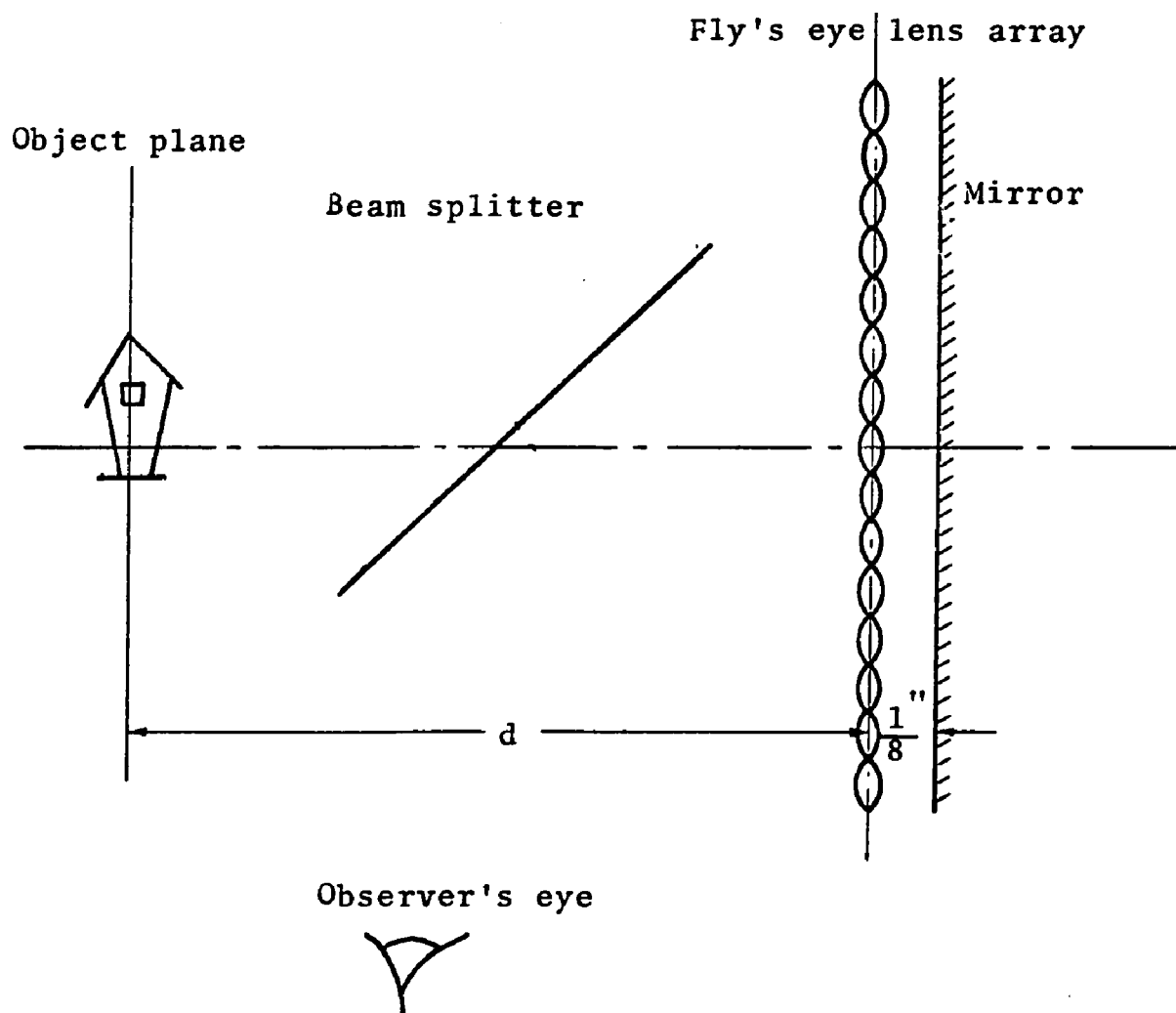


Figure 6-10. The experimental setup for pseudoscopic image inversion

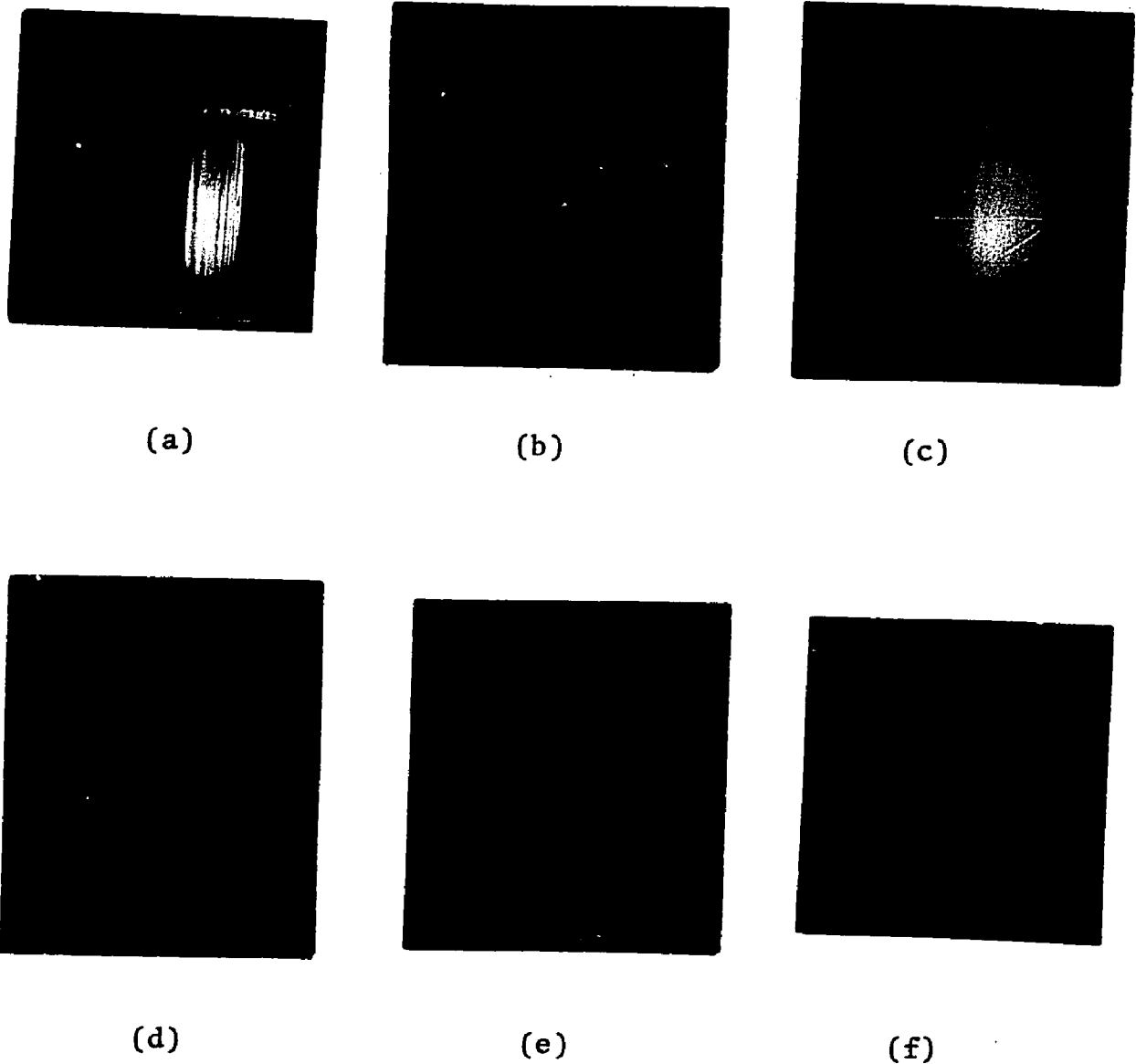


Figure 6-11. The experimental results of the pseudoscopic image inversion

- (a) The original object
- (b) The pseudoscopic image at $d = 3''$
- (c) The pseudoscopic image at $d = 8''$
- (d) The pseudoscopic image at $d = 13''$
- (e) Same picture as (b) except the camera is focused at the fly's-eye lens array
- (f) Same picture as (b) except a diffuser is put at the central plane of the image

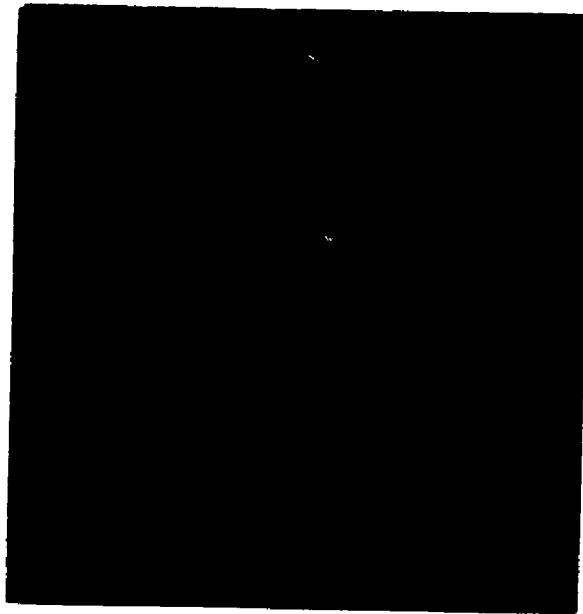


Figure 6-12 (a). The resolution chart reconstructed using a Fresnel lens
Hologram aperture $K = 3.5$ cm
Focal length : 8.5"

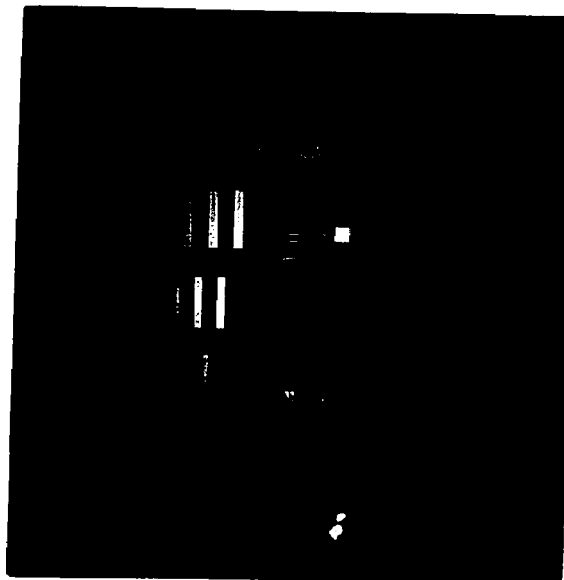


Figure 6-12 (b). The resolution chart reconstructed using a condensing lens
Hologram aperture $K = 3.5$ cm
Lens aperture: 20 cm
F-number: 1.5

The resolution is somewhat better than with the Fresnel lens, but the noise background has been reduced substantially because of better surface quality.

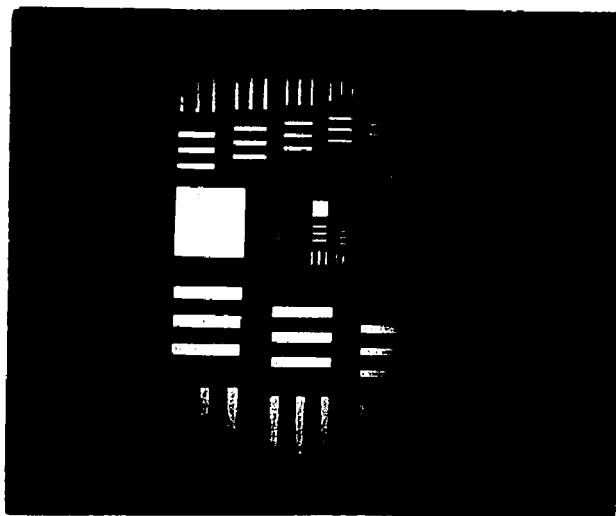
One of the difficulties with the single-lens movie system is the need of a liquid gate to reduce the noise background. Because of the liquid gate, the film cannot be driven fast enough to accommodate the required movie rate of 24 frames per second. The difference made by using a liquid gate can be seen in Figs. 6-13 (a), (b) and (c). Figure 6-13 (a) shows the image target reconstructed without use of a liquid gate in recording or reconstruction. Figure 6-13 (b) shows the case when a liquid gate is used in the reconstruction process; and Fig. 6-13 (c) shows the picture of the same image, when a liquid gate is used in both recording and reconstruction. It can be seen that the signal-to-noise ratio can be increased substantially by use of a liquid gate.

The single-lens movie system can provide a three-dimensional object scene with a very long depth of field. Figures 6-14 (a) and (b) are pictures of two tanks separated in depth by about one foot. The pictures were taken with a camera focused on one of the tanks. The resolution is excellent for the human eye. A condensing lens was used in the above system. Figure 6-14 (c) is a picture of a 3-D object reconstructed from a hologram made with a Fresnel lens. The circular structure of the Fresnel lens background can be seen in the picture.

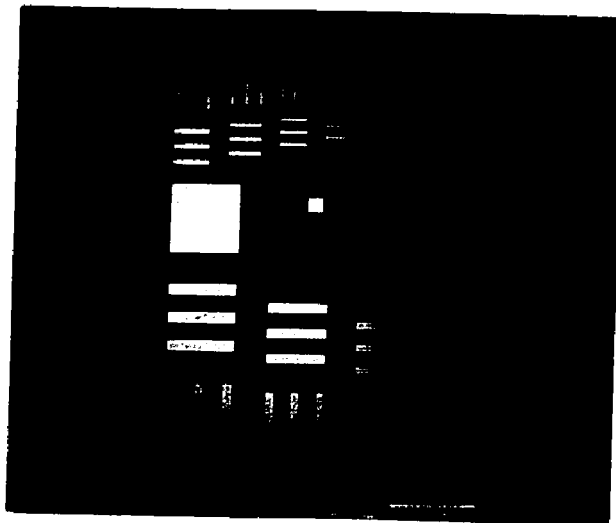
For the single-lens movie system, a movie strip with 15 frames was made in a laboratory set up. The object scene consisted of two tanks (shown in Fig. 6-14) moving parallel to each other. With the movie strip running at 5 frames per second, we can see a 3-D movie three seconds long. The movie has an appearance similar to movie cartoons, except that it is three-dimensional.

6.5. CONCLUSIONS

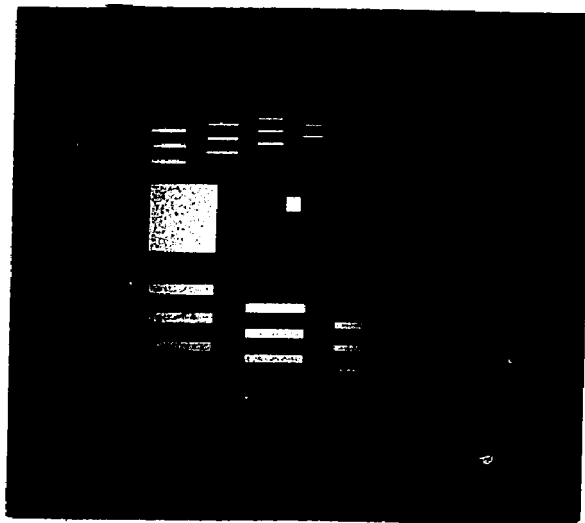
The information capacity and the resolution distribution of a general hologram system is presented. The use of a hologram as a 3-D display system has been studied for different system task



(a) The image reconstructed without liquid gate both in recording and in reconstruction



(b) The image reconstructed using liquid gate in reconstruction but not in recording

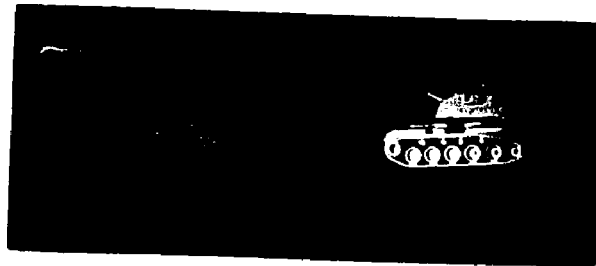


(c) The image reconstructed using liquid gate both in recording and in reconstruction

Figure 6-13. The experimental results of the reconstructed image using liquid gate in a single-lens movie system



(a)



(b)



(c)

Figure 6-14. The reconstructed 3-D images of the single-lens movie system
(a) and (b) are by condensing lens
(c) is by Fresnel lens

considerations. For an information reduction type of a system, it is found that the amount by which information can be reduced depends on the required depth of field for a general 3-D image space. The example in Section 3.4 is a good indication of the optimum limitations on the information reduction type of system. For example, for an information reduction factor of the order of 10^5 , the depth of field is only about 4 mm (Fig. 3-5). This is barely a 3-D image. Hence, the limits on information reduction depend on the required depth of field. A mathematical formulation is employed for the study of the scatter-plate information reduction method. With the assumption that all functions in the system are periodic, the system performance can be calculated by a computer. Because of the periodic assumption and other simplifying assumptions, this approach is limited in its applicability. In the fly's-eye lens array system analysis of the diffraction effects of the small hologram size, lenslet aberrations and wavelength changes are considered.

An alternate method for a 3-D display system is the beamwidth reduction method. This method is more practical and promising for a holographic movie system. The system tolerance is not very critical, so that much more economical components can be used. With the single-lens movie system we built for this study, a displacement of 1/8 inch of the lens relative to the hologram did not degrade the image very much. The signal-to-noise ratio for the system can be reduced to a certain extent by use of a liquid gate. The pseudoscopic image inversion technique using two lens systems is not satisfactory. The fly's-eye lens array and a mirror gives better results, but is not perfect. The film transport problem has been improved but not to the required degree.

Judging from both the theoretical and experimental studies in this dissertation, we conclude that the techniques for building a TV type of 3-D display system using the information reduction methods are still not practical. However, for the movie type of 3-D display system, the difficulties can be overcome to a satisfactory level in the near future.

Appendix A

DERIVATION OF THE NUMBER OF RESOLVABLE



Figure A-1 The viewing configuration for the human eye

space with $\Delta z \ll d$ and $\Delta x \ll d$. We say two points B and C are just resolved by the human eye, if the angle they subtend at the eye is equal to one minute of arc or $1/3600$. For the simple geometry in Fig. A-1, we have

$$\frac{\Delta x}{d} = \frac{1}{3600} = \frac{\lambda}{e}$$

The ability of the human eye to resolve two points in depth, however, depends on the existence of two human-eye lenslets. We say the two points A and B are just resolved in depth by the human eye if the difference between the two angles subtended by the two human eye lenslets at point A and B is one minute of arc or $1/3600$. For the simple geometry in Fig. A-1, we have

$$\theta_A = \theta_B + \theta$$

or

$$\theta = \theta_A - \theta_B .$$

It is thus very clear from Fig. A-1, that if points B and C are just resolved by one human eye, points A and B will be just resolved in depth by two human-eye lenslets. We thus have

$$\frac{\Delta x}{\Delta z} = \frac{E}{d}$$

or

$$\Delta z = \frac{d}{E} \cdot \Delta x = \frac{d^2}{3600E} = \frac{d^2 \lambda}{eE}$$

where Δz is the resolution in depth for the human eye. Referring to Fig. 3-1 where the object space has a depth D and using Δz defined above as the resolution in depth for the human eye, we can give the number of resolvable points in depth by

$$\frac{D}{\Delta z} = \frac{DeE}{\lambda d^2}$$

This completes the derivation. From this derivation, we see that the resolution in depth of the human eye is closely related to the resolution of two object points for the human eye.

Appendix B

DEFINITION OF A MEAN-SQUARE ERROR CRITERION FOR THE TWO
BANDLIMITED PERIODIC FUNCTIONS IN A HILBERT SPACE

Let two periodic functions $V(y)$ and $V_1^*(y)$ be the elements of a Hilbert space H where the inner product is defined by

$$\langle V(y), V_1^*(y) \rangle \triangleq \frac{1}{y_0} \int_0^{y_0} V(y) V_1^*(y) dy$$

and y_0 is the period for the two functions $V(y)$ and $V_1^*(y)$. Let

$$\alpha_k \triangleq e^{j \left(\frac{2\pi k y}{y_0} \right)}$$

Then the set $\{\alpha_k\}$ forms an orthonormal set in H . Since $V(y)$ and $V_1^*(y)$ are bandlimited periodic functions with the same bandwidth, we assume the set $\{\alpha_k\}$ contains finite elements for k from $-L$ to $+L$ and $\{\alpha_k\}$ forms a complete orthonormal set in H . Hence, H is separable and for each $V(y)$ or $V_1^*(y)$ in H we have

$$\begin{aligned} V(y) &= \sum_{\ell=-L}^L \langle V, \alpha_\ell \rangle \alpha_\ell \\ &= \sum_{\ell=-L}^L a_\ell \exp \left[j \frac{2\pi \ell y}{y_0} \right] \end{aligned}$$

and

$$\begin{aligned} V_1^*(y) &= \sum_{n=-L}^L \langle V_1^*, \alpha_n \rangle \alpha_n \\ &= \sum_{n=-L}^L e_n \exp \left[j \frac{2\pi n y}{y_0} \right] \end{aligned}$$

The inner product of $V(y)$ and $V_1^*(y)$ is

$$\langle V(y), V_1^*(y) \rangle = \sum_{\ell=-L}^L a_{\ell} e_{\ell}^*$$

The norm induced by this inner product is

$$||V(y)|| = \left[\sum_{\ell=-L}^L a_{\ell}^2 \right]^{\frac{1}{2}}$$

The metric induced by the norm is

$$\begin{aligned} \rho \left[V(y), V_1^*(y) \right] &= ||V(y) - V_1^*(y)|| \\ &= \left[\sum_{n=-L}^L |a_n - e_n|^2 \right]^{\frac{1}{2}} \end{aligned}$$

The above expression is the definition of a mean-square error criterion for the two bandlimited periodic functions in a Hilbert space. For more information about the Hilbert space concept, please refer to Reference [38].

Appendix C
THE DERIVATION OF EQ. (4-3) AND
A DISCUSSION ON ITS VALIDITY

In order to derive Eq. (4-3), we first note that the integral formula

$$\int_{-\infty}^{\infty} e^{j\pi b t^2} e^{-j2\pi c t} dt = \left(j\frac{1}{b}\right)^{\frac{1}{2}} e^{-j\pi \frac{c^2}{b}} \quad (C-1)$$

holds for all complex-valued parameters b and c if $\text{Im}(b) \geq 0$.
Eq. (4-3) then follows:

$$\begin{aligned} G(F) &\triangleq \int_{-m/2}^{m/2} \exp(j\pi B y^2) \exp(-j2\pi F y) dy \\ &= \int_{-\infty}^{\infty} \text{rect}\left(\frac{y}{m}\right) e^{j\pi B y^2} \cdot e^{-j2\pi F y} dy \\ &= \mathcal{F}\left[\text{rect}\left(\frac{y}{m}\right)\right] * \mathcal{F}\left[e^{j\pi B y^2}\right] \\ &= m \cdot \text{sinc}(mf) * \left(j\frac{1}{B}\right)^{\frac{1}{2}} \exp\left(-j\frac{\pi}{B} f^2\right) \\ &\triangleq m \left(j\frac{1}{B}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \text{sinc}[mf] e^{-j\frac{\pi}{B}(F-f)^2} df \end{aligned}$$

Here $\mathcal{F}$ represents Fourier Transform and $*$ represents convolution. If we approximate $\text{sinc}[mf]$ by $\exp[-2.04m^2 f^2]$ in the above convolution, we have

$$\begin{aligned}
 G(F) &= m \left(j \frac{1}{B} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp(-2.04m^2 f^2) \exp[-j \frac{\pi}{B} (F^2 - 2Ff + f^2)] df \\
 &= m \left(j \frac{1}{B} \right)^{\frac{1}{2}} \exp\left(-j \frac{\pi}{B} F^2\right) \int_{-\infty}^{\infty} \exp \left[j \pi \left(\frac{-1}{B} - \frac{2.04m^2}{j \pi} \right) f^2 \right] e^{-j 2 \pi \left(\frac{-F}{B} \right) f} df
 \end{aligned}$$

Comparing the above integral with the integral formula Eq. (C-1), we have

$$b = \frac{-1}{B} - \frac{2.04m^2}{j \pi}$$

and

$$c = \frac{-F}{B}.$$

If Eq. (C-1) is used, $G(F)$ then becomes

$$\begin{aligned}
 G(F) &= m \left(j \frac{1}{B} \right)^{\frac{1}{2}} \exp\left(-j \frac{\pi}{B} F^2\right) \left(j \frac{1}{b} \right)^{\frac{1}{2}} \exp\left(-j \pi \frac{c^2}{b}\right) \\
 &= m \left(j \frac{1}{B} \right)^{\frac{1}{2}} \left(j \frac{1}{b} \right)^{\frac{1}{2}} \exp\left(-j \frac{\pi}{B} F^2\right) \exp\left(-j \frac{\pi F^2}{B^2 b}\right) \\
 &= m \left(\frac{-1}{Bb} \right)^{\frac{1}{2}} \exp \left[-j \pi F^2 \left(\frac{1}{B} + \frac{1}{B^2 b} \right) \right],
 \end{aligned}$$

where

$$\begin{aligned}
 \frac{1}{B} + \frac{1}{B^2 b} &= \frac{1}{B} + \frac{1}{B^2 \left(\frac{-1}{B} + j 0.65m^2 \right)} \\
 &= \frac{1}{B} + \frac{1}{B} \left(\frac{1}{-1 + j 0.65m^2 B} \right) = \frac{1}{B} \left(1 - \frac{1}{1 - j 0.65m^2 B} \right)
 \end{aligned}$$

$$= \frac{-j0.65m^2}{1 - j0.65m^2B}$$

and

$$\frac{-1}{Bb} = \frac{1}{1 - j0.65m^2B}.$$

$G(F)$ can then be written as a function of F , m and B in the following form

$$G(F) = m \left(\frac{1}{1 - j0.65m^2B} \right)^{\frac{1}{2}} \exp \left[-\pi F^2 \frac{0.65m^2}{1 - j0.65m^2B} \right].$$

This completes the derivation of Eq. (4-3). We now turn our attention to the validity of Eq. (4-3). The exact formula for $G(F)$ without approximation can be obtained in terms of Fresnel integral functions. It is given by

$$G(F) = \left(\frac{1}{2B} \right)^{\frac{1}{2}} e^{-j\frac{\pi F^2}{B}} [Z(u_2) - Z(u_1)]$$

where $Z(\cdot)$ is the Fresnel integral function the value of which has been tabulated in the Handbook[39]; u_2 and u_1 are functions of F , B and m :

$$u_2 = -\sqrt{2B} \left(\frac{F}{B} - \frac{m}{2} \right) = -\sqrt{\frac{2}{B}} F + \sqrt{\frac{B}{2}} m$$

$$u_1 = -\sqrt{2B} \left(\frac{F}{B} + \frac{m}{2} \right) = -\sqrt{\frac{2}{B}} F - \sqrt{\frac{B}{2}} m.$$

With a computer, the exact curves of $G(F)$ for different parameters B and m can be plotted. A comparison between those plots and the

plots of Eq. (4-3) should reveal the ranges for different parameters where Eq. (4-3) is valid. For our discussion here, we intend to investigate the problem from the integral itself. The only approximation which is made in the derivation of Eq. (4-3) is to use $\exp[-2.04m^2f^2]$ instead of $\text{sinc}[mf]$. Examining the convolution integral

$$\int_{-\infty}^{\infty} \text{sinc}[mf] \cdot \exp[-j\frac{\pi}{B}(F - f)^2] df,$$

we note that the major contribution of the above integral comes from the interval $-\pi/2$ to $\pi/2$ for the quadratic term $\pi(F - f)^2/B$ if the period of the exponential term is much smaller than the period of the $\text{Sinc}[mf]$ function. This is so because the value of the integral contributed by the higher frequencies of the exponential term will average to zero if the envelope of the higher frequencies is slowly varying. We thus have the relation $\sqrt{2B} < 1/m$, where $\sqrt{2B}$ is the period of $\exp[-j\pi(F - f)^2/B]$ and $1/m$ is the period of $\text{sinc}[mf]$. Under this condition, the above integral can be integrated from $F - \sqrt{B/2}$ to $F + \sqrt{B/2}$ for f instead of from $-\infty$ to $+\infty$. The interval of integration is $\sqrt{2B}$. Now, when the $\text{Sinc}[mf]$ is replaced by $\exp[-2.04m^2f^2]$ in the above integral and when the same interval $\sqrt{2B}$ is the integration interval, we want to know under what condition the above approximation will not effect the results very much. Since $\exp[-2.04m^2f^2]$ is very close to $\text{Sinc}[mf]$ in the main lobe, it is clear that the wider the main lobe is the better the approximation will be. We thus want m to be small and $\sqrt{2B}$ to be small. This requirement is also consistent with the requirement $\sqrt{2B} < 1/m$. Of course, the accuracy of the approximation also depends on the value of F . Fortunately, the smaller the absolute value of F is, the better the approximation. Since Eq. (4-3) is used to evaluate the resolution of the system, we are usually only interested in values of F near zero. We thus conclude that Eq. (4-3) is valid if both m and B are small with $\sqrt{2B} < 1/m$.

REFERENCES

1. E. N. Leith, Appl. Opt., Vol. 10, No. 11, 2419 (1971).
2. See Chapter 3, Sections 3.2, 3.3, 3.4.
3. A. Macovski, J.O.S.A., Vol. 60, No. 1, 21 (1970).
4. W. H. Carter and A. A. Dougal, J.O.S.A., Vol. 56, No. 12, December 1966.
5. E. B. Champagne and N. G. Massey, Appl. Opt., Vol. 8, No. 9, September 1969.
6. A. Macovski, Ph.D. Dissertation, 1968, Stanford University, University Microfilms, Ann Arbor, Michigan (Order No. 69-258).
7. L. H. Enloe, W. C. Jakes, Jr. and C. B. Rubinstein, Bell System Tech. J., 47, 1875 (1968).
8. M. Born and E. Wolf, Principles of Optics, Pergamon Press, 1959.
9. J. W. Goodman, Introduction to Fourier Optics, McGraw-Hill Book Co., 1968.
10. "Investigation of Holographic Display Techniques," Final Report, 28000-9-F, Willow Run Laboratories, The University of Michigan, Ann Arbor, Michigan, 1971.
11. C. B. Burckhardt, J.O.S.A., Vol. 58, No. 1, January 1968.
12. E. N. Leith, et al., J.O.S.A., Vol. 74, No. 10, October 1965.
13. E. N. Leith and J. Upatnieks, J.O.S.A., Vol. 56, No. 4, April 1966.
14. L. H. Lin, Appl. Opt., Vol. 7, No. 3, March 1968.
15. C. B. Burckhardt, J.O.S.A., Vol. 58, No. 2, February 1968.
16. D. B. Brumm, Ph.D. Dissertation, The University of Michigan, 1970.
17. K. A. Haines and D. B. Brumm, Appl. Opt., Vol. 7, No. 6, June 1968.
18. D. Fritzler and E. Marom, Appl. Opt., Vol. 8, No. 6, June 1969.
19. K. A. Haines and D. B. Brumm, Proceedings of IEEE, p. 1512, August 1967.
20. R. W. Meier, J.O.S.A., Vol. 55, No. 8, August 1965.
21. E. B. Champagne, J.O.S.A., Vol. 57, No. 1, January 1967.
22. J. N. Latta, Appl. Opt., Vol. 10, No. 3, March 1971.
23. J. M. Moran, Appl. Opt., Vol. 10, No. 8, August 1971.
24. "Proposal for a Three-Dimensional Display of Air Traffic," ORA-71-778-KB1, Office of Research Administration, The University of Michigan, 1971.
25. E. L. O'Neil, Introduction to Statistical Optics, Addison-Wesley Publishing Co., Inc., 1963.

26. C. B. Burckhardt, et al., Appl. Opt., Vol. 7, No. 3, April 1968.
27. H. E. Ives, J.O.S.A., 21 March 1931.
28. R. V. Pole, App. Phy. Lett., Vol. 10, No. 1, 1 January 1967.
29. A. H. J. DE. L. Saint Genies, U.S. Patent 2139855.
30. A. Kozma, Ph.D. Dissertation, The University of London, 1968.
31. A. A. Friesem and J. S. Zelenka, Appl. Opt., Vol. 6, No. 10, October 1967.
32. W. H. Lee and M. O. Greer, J.O.S.A., Vol. 61, No. 3, March 1971.
33. H. M. Smith, J.O.S.A., Vol. 58, No. 4, April 1968.
34. A. Kozma, Optica Acta, Vol. 15, No. 6, 1968.
35. K. O. Hill and G. W. Jull, J.O.S.A., Vol. 61, No. 3, March 1971
36. V. S. Srinivasan, Appl. Opt., Vol. 9, No. 9, September 1970.
37. A. D. Gara and F. T. S. Yu, App. Opt., Vol. 10, No. 6, June 1971.
38. W. A. Porter, Modern Foundations of System Engineering, The Macmillan Co., N.Y., 1966.
39. M. Abramowitz and I. A. Stegun, ed., Handbook of Mathematical Functions, National Bureau of Standards, 1964.