MICROCOPY CHART

AFATL–TR–86–11

Production and Analysis of Computer– Generated Holographic Matched Filters

Steven F. Butler

ELECTRO – OPTICAL TERMINAL BRANCH
ADVANCED SEEKER DIVISION
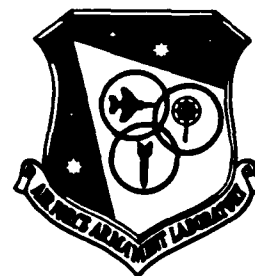
DTIC
SELECTED
APR 2 9 1986
D
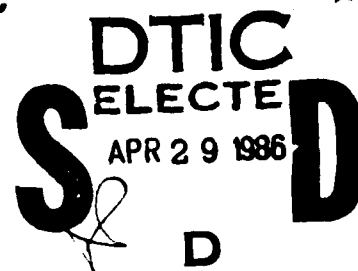
APRIL 1986

FINAL REPORT FOR PERIOD OCTOBER 1982 – SEPTEMBER 1985

APPROVED FOR PUBLIC RELEASE

DTIC FILE COPY

AIR FORCE ARMAMENT LABORATORY

Air Force Systems Command•United States Air Force•Eglin Air Force Base, Florida
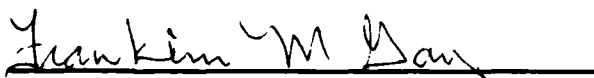
86   4 28   1 52

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

FRANKLIN M. GAY
Technical Director, Advanced Seeker
   Division

Even though this report may contain special release rights held by the controlling office, please do not request copies from the Air Force Armament Laboratory. If you qualify as a recipient, release approval will be obtained from the originating activity by DTIC. Address your request for additional copies to:

> Defense Technical Information Center
> Cameron Station
> Alexandria, Virginia 22314

If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization, please notify AFATL/DLMI, Eglin AFB FL 32542.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION		1b. RESTRICTIVE MARKINGS		
Unclassified		N/A		
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT		
N/A		Approved for public release; distribution unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE				
N/A				

4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)		
AFATL-TR-86-11				

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION		
Advanced Seeker Division	AFATL/DLMI			

6c. ADDRESS (City, State and ZIP Code)	7b. ADDRESS (City, State and ZIP Code)
Air Force Armament Laboratory Eglin AFB, FL 32542-5000	

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
Advanced Seeker Division	DLMI	In-House		

8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS.			
Air Force Armament Laboratory Eglin AFB FL 32542-5000	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
	61601F	2305	E2	01

11. TITLE (Include Security Classification)
Production and Analysis of Computer-Generated Holographic Matched Filters

12. PERSONAL AUTHOR(S)
Steven F. Butler

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM Oct 82 TO Sep 85	March 86	

16. SUPPLEMENTARY NOTATION
Availability of report specified on verso of front cover.

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Lasers, Holography, Matched Filters, Computer-Generated Pattern Recognition.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report presents techniques for the use of computer-generated holograms (CGH) for matched filtering. An overview of the supporting technology is provided. Included are techniques for modifying existing CGH algorithms to serve as matched filters in an optical correlator. It shows that matched filters produced in this fashion can be modified to improve the signal-to-noise and efficiency over that possible with conventional holography. The effect and performance of these modifications are demonstrated. In addition, a correction of film nonlinearity in continuous-tone filter production is developed. Computer simulations provide quantitative and qualitative demonstration of theoretical principles, with specific examples validated in optical hardware. Conventional and synthetic holograms, both bleached and unbleached, are compared.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Steven F. Butler	(904) 882-5252	AFATL/DLMI

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

PREFACE

TABLE OF CONTENTS

v

Accesion For

NTIS CRA&I	☑
DTIC TAB	☐
U announced	☐
Justification	

By

Distribution/

Availability Codes

| Dist | Avail and/or Special |

A-1

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONCLUDED)

LIST OF TABLES

x

SUMMARY

This report presents techniques for the use of computer-generated holograms for matched filtering. Due to the enormity of the information presented, an executive summary is provided. Sections I and II review the appropriate background to introduce the reader to the power of optical matched filtering. This power is particularly useful when a fixed detection algorithm, such as Fourier transformation or correlation, is to be used consistently at an extremely high rate. The parallelism of the optical processor provides high speed processing of large SBP images. Thus, as input arrays provide large images at high speeds, optical correlators will serve as efficient processors to perform specific tasks.

Section III describes the recording of the reference signal on a hologram. The continuous-tone hologram closely matches an actual interferometric hologram. This hologram can be improved by the incorporation of equations (41) and (42). This reduces the dynamic range and yields a recording containing more information. The binary techniques of the B-L, Lee, Haskell, and A-K holograms are presented to indicate the current evolution of binary holograms. The optimum combination appears to be a combination of the GBCGH and the A-K. By increasing the intra-cell sampling to an odd valued N, as in the GBCGH, the hologram can represent large dynamic ranges. In addition, the hologram should be sampled at each pixel, as in the A-K, to reduce or eliminate false images.

Section III also describes the SBP required of the CGH to perform the optical correlation. The SBP drives the size, complexity, and expense of the ГGH. When used as a matched filter, the requirements are more stringent than when used for reconstruction holography. However, when the techniques presented at the end of the Section are incorporated, the SBP requirements can be minimized.

Section IV describes preprocessing techniques useful in optimizing the CGH matched filter. The high-frequency emphasis and phase-only filtering increase the discrimination against false targets but sensitize the filter to scale and rotation. The phase-only filter, although easy to implement, is not unique in its ability to improve signal-to-noise and efficiency. Rather, it is a form of

high frequency emphasis which may or may not be optimum. It does minimize the
dynamic range but may not be the appropriate choice for a specific application.
The rules describing the application of high frequency emphasis should be
applied to each case. The exception is when an on-axis hologram is to be phase
modulated so as to require the reference to contain no amplitude information.

The phase-modulation process applied to the physical hologram should not be
confused with the phase-only filtering step which is applied to the reference
information. That is, a phase-modulated hologram can utilize the normal refer-
ence information, frequency emphasis or phase-only filtering. In general, the
phase modulation provides considerably higher efficiency with an accompanying
loss of signal-to-noise ratio due to the nonlinearity. When using binary
holograms, there is no nonlinearity and phase modulation should always be
incorporated. A binary, phase-modulated, on-axis hologram with the dc term
eliminated, provides the least SBP and highest efficiency. For this reason, it
is recommended for future work in programmable holograms. However, the
bi-phase conversion required for this hologram may not provide sufficient
signal-to-noise ratio for all applications.

Section V describes two approaches for handling scale, rotation and other
deformations. Presently, the Mellin-Fourier approach is hard to implement.
The SDF approach is easy to implement but typically yields low (<10) signal-to-
noise ratios (Reference 29).

Section VI describes the techniques for minimizing the THD of holographic
films. This process is important for optimizing continuous-tone holograms.
The inability to perform this linearization is a severe limitation of conven-
tional matched filter holograms. This Section outlines the steps necessary for
proper modeling of the film. Using a polynomial model and correcting for the
nonlinearity, a 100 times improvement was realized using 8E75 holographic
plates. It is probable that further improvement could be obtained using a
spline model. Another improvement is possible by whitening the modulation
transfer function (MTF). The MTF, like the film linearity, affects the film
response. It can be measured and whitened to prevent undesired frequency
emphasis. The effect on spatially modulated holograms is small due to the
small frequency variations about the carrier and has been omitted from this
report. Future work may include this effect by performing a second frequency

emphasis upon the encoded CGH pattern. The effect would be more pronounced with an on-axis hologram where a wide spread of spatial frequencies is possible. However, on-axis holograms for matched filtering must be phase modulated, complicating the correction to the MTF.

Section VII describes the simulations used to demonstrate the various techniques. The ideal correlation is performed using real valued variables with no hologram encoding. This provides an essentially infinite dynamic range correlation with no CGH encoding noise and serves as a standard by which holographic techniques can be compared.

The continuous-tone simulation provides the best approximation to the performance of a conventional hologram. By eliminating the magnitude terms, the dynamic range is reduced and the hologram is greatly improved. The binary simulation shows that equivalent performance is possible using binary techniques. The example of an SDF matched filter demonstrated a practical example. This example shows that although correlations are obtained, the signal-to-noise and efficiencies are low for SDF filters. For some applications, this may preclude the use of SDF's and require sequential single filter steps. Section VIII demonstrates the implementation of the holograms and shows reconstruction of several examples.

SECTION I
INTRODUCTION

Images and photographs have long been used to identify and locate objects. By photographing an area, perhaps from afar, a scene could be given detailed study. This study might disclose the presence of objects of interest and determine their spatial location. Images from satellites show weather, agriculture, geology and global actions. Special images may contain additional scientific information including object spectral characteristics, velocity, temperature, and the like.

The traditional medium of these images has been photographic film. It is capable of high resolution and is sensitive to visible and near-visible wavelengths. Unfortunately, film based methods are slow due to exposure, processing, and analysis time. This time lag is not a problem for many applications and so film is still the primary medium for reconnaissance. Electronic imagery (TV, radar, etc.) is used for those applications that require faster interpretation. These images can be viewed, like film, by people trained to interpret the particular images. Because of the electronic nature of the images, electronic hardware and computers are used for manipulation of the images.

1. MACHINE VISION

For very high speed retrieval and interpretation, machines must be designed around the specific tasks. Machine interpretation is also necessary when a human is not available. Unmanned robots work in hazardous areas and perform many jobs more efficiently without the encumbrance of human intervention. However, to function and carry out their assigned job, the robots must have information about their surroundings. The ability to interpret imagery from self-contained sensors is necessary for the proper function of a robot. This image interpretation includes guidance, obstacle avoidance, target recognition, tracking, and closed-loop control of robot action. For robot action without human intervention, machine intelligence must have the ability to make decisions based on scene content. Computer image processing and recognition refer to techniques that have evolved in this field in which the computer receives and uses visual information.

1

Image processing techniques prepare or preserve an image for viewing. This includes enhancement, restoration, and reconstruction. Image enhancement techniques are designed to improve image quality for human viewing. For example, correction of a geometrically distorted image produces an obvious improvement in quality to a human observer. Image restoration techniques compensate an image, which has been degraded in some fashion, to restore it as nearly as possible to its undegraded state. For example, an image which is blurred due to camera motion may be improved using motion restoration. To perform the difficult task of image interpretation, extraneous noise must be separated from the desired signals. This may occur in several stages of enhancement where each stage reduces the extraneous noise and preserves the information crucial to object recognition. Image enhancement may include contrast transformation, frame subtraction, and spatial filtering. The goal of image enhancement is to reduce the image complexity so that feature analysis is simplified (Reference 1).

Once the scene has been enhanced, the job of interpretation is simplified. The interpreter must now decide what the remaining features represent. The features present a pattern to the interpreter to be recognized. This pattern *recognition problem may be quite difficult* when a large number of features are necessary to differentiate between two possibilities. Most people have to look closely to see any difference between twins. A computer might have equal difficulty distinguishing a car from a house in low-resolution image.

Recognition involves an interpretation of an image. This includes scene matching and understanding. Scene matching determines which region in an image is similar to a pictorial description of a region of another scene. A reference region or template is provided and systematically compared to each region in a larger image. Here the computer attempts to match models of known objects, such as cars, buildings, or trees, to the scene description and thus determine what is there. The model objects would be described in memory as having certain characteristics, and the program would attempt to match these against various parts of the image. Scene understanding involves a more general recognition problem describing physical objects in a scene based on images. For example, a scene may be divided into regions that match various

objects stored in memory such as a house, tree, and road. Once the scene is divided into known regions, the interrelationship between these regions provides information about the scene as a whole.

When it is necessary to recognize specific objects, correlation techniques are often used (Reference 2). A reference image of the desired object is stored and compared to the test image electronically. When the correlation coefficient is over a specified threshold, the computer interprets the image as containing the object. The correlation procedure may also provide the location of the object in the scene and enable tracking. The correlation coefficient may be used in decision making to determine robot action. Because even a single object may present itself in many ways, correlation procedures are complicated by the immense reference file that must be maintained (Reference 3). Special correlation techniques may provide invariance to specific changes, but a wide range of object conditions (i.e., temperature, color, shape, etc.) make correlation recognition a complicated computer task (Reference 4). The best computer vision systems now available have very primitive capabilities. Vision is difficult for a computer for a number of reasons. The images received by a sensing device do not contain sufficient information to construct an unambiguous description of the scene. Depth information is lost and objects frequently overlap. Vision requires a large amount of memory and many computations. For an image of 1000 X 1000 picture elements, even the simplest operation may require 10^8 operations. The human retina, with 10^8 cells operating at roughly 100 hertz, performs at least 10 billion operations a second. Thus, to recognize objects at a rate even closely resembling human vision, very special processor technologies must be considered. One promising technology has emerged in the form of optical computing.

2. OPTICAL COMPUTERS

Optical computers permit the manipulation of every element of an image at the same time. This parallel processing technique involves many additions and multiplications occurring simultaneously. Most digital processors must perform one operation at a time. Even though the digital processors are very fast, the number of total operations required to recognize patterns in an image is very large. Using optical Fourier transformers, an optical processor can operate on

the image and its Fourier transform simultaneously. This permits many standard image processing techniques, such as spatial filtering and correlation, to be performed at tremendous rates.

The Fourier transform is formed optically by use of a lens. The usual case that is considered in optical computing is the illuminating source located at infinity (by use of an auxiliary collimating lens) and the image transparency located at a distance equal to focal length from the transforming lens. The distribution in the output plane located a focal length behind the transforming lens is the exact Fourier transform of the input distribution. The Fourier transform contains all of the information contained in the original image. However, the information is now arranged according to spatial frequency rather than spatial location. The advantage of such an arrangement is that objects or signals of interest may overlap with noise in the image domain but exist isolated in the frequency domain. This permits the possible separation of signal from noise in the frequency plane when it would have been impossible in the image plane. The image can be transformed into frequency space, frequency filtered and then transformed back into image space with the noise removed. The frequency filter may be low-pass, high-pass, or band-pass, chosen to optimize the filtering of a specific signal. This frequency plane filter is the heart of the analog optical computer.

The frequency plane filter can be constructed in many ways. Low-pass and high-pass filters are accomplished using simple apertures mounted on axis in the frequency plane. More complicated filters are produced optically using holographic techniques. These filters may also be produced using computer-generated holography (CGH). The computer is used to model the desired filter response, mathematically represent the holographic filter, and create a physical filter using a writing device. One of the important advantages of computer-generated holography is that the reference need not exist physically, but only mathematically. This permits mathematical manipulation of the reference prior to creation of the filter for purposes of optimization.

The advantage of an analog optical processor is that it may operate at very high speeds. In addition, the processor typically is smaller, lighter, and

4

consumes considerably less power than an equivalent digital processor (References 5 and 6). When coupled with the ability to manipulate and optimize the frequency plane filter, the optical processor becomes a useful tool. With considerable justification, there is great interest in the robotics community.

3. CONTRIBUTION

This report states that CGH matched filters should be used in an optical correlator to recognize patterns in a complex scene, and describes how to create that filter. The CGH matched filter is superior to interferometric filters due to the ability to preprocess the filter function and control the production of the hologram. The use of optical elements for high speed pattern recognition was first proposed 20 years ago (Reference 7). The concept of using computers to define and generate holograms came only 2 years later (Reference 8). Since that time, considerable effort has been devoted to exploring the potential of these CGH elements for reconstruction holography. Most of this effort was devoted to optimizing the methods for encoding and writing the holograms (References 8 through 13). More recently, interest has grown in the area of efficiency improvement (Reference 14). The efficiency of a hologram for optical correlation must be high in order to utilize low power, light weight diode lasers. In separate but parallel efforts in artificial intelligence, researchers have studied the effects of image enhancement on pattern recognition (Reference 15). Though research in the various fields is proceeding, a unified approach to the interrelation of preprocessing, holographic encoding and physical implementation is lacking. Specifically, the research in CGH to date, has only been for display or reconstruction holography, not matched filtering. This report describes the steps necessary and possible to create practical matched filters using CGH.

The approach presented here ties many areas of research together as they apply to CGH matched filters. Modifications to existing encoding schemes which provide real valued filter patterns for use in an optical correlator are explained in Section III. In addition, Section III defines the space-band-width-product (SBP) required for holographic matched filtering rather than for display holography as is presented in existing literature. This includes procedures for minimizing the SBP required. Preprocessing methods which apply

specifically to matched filtering are presented along with rationale for their use in Section IV. Techniques for the use of CGH matched filters as a pattern recognizer are reviewed in Section V. Linearization methods for writing on film are derived and evaluated in Section VI.

These various considerations are not independent, but are interwoven in the production of CGH matched filters. These interactions can be fully analyzed only with a complete model incorporating all the parameters. Section VII describes such a model created to analyze the preprocessing, encoding, and writing techniques used to produce optimal CGH matched filters. Now that the various methods have been developed and the analytical tools demonstrated, specific examples are presented and analyzed. Section VIII describes approaches for physically producing a transparency including specific examples taken from Section VII. Finally, conclusions based on the analysis are offered in Section IX.

SECTION II
BACKGROUND

For a better understanding of an optical processor operation, the background technology is reviewed here. A number of different types of optical processors are in use today. These include one-dimensional signal processors, two-dimensional image processors and multi-dimensional digital processors. Only two-dimensional image processors used for matched filtering are described here. A matched filter optimizes the signal-to-noise ratio at a specific point when the characteristics of the input are known (Reference 16). Typically, the desired pattern and the nature of the background or noise in the input image are known. Specifically, the input consists of a known signal $s(x,y)$ and an additive noise $n(x,y)$. The system is linear and space invariant with impulse response $h(x,y)$. The criterion of optimization will be that the output signal-to-noise power ratio be a maximum. This optimum system will be called a matched filter for reasons that will become clear as the derivation proceeds.

1. COMMUNICATION THEORY

A system is any unit that converts an input function $I(x,y)$ into an output function $O(x,y)$. The system is described by its impulse response--its output when the input is an impulse or delta function. A linear system is one in which the output depends linearly on the input and superposition holds. That is, if the input doubles, so does the output. More precisely stated, let O_1 be the output when I_1 is the input and O_2 be the output when I_2 is the input. Then the system is linear when, if the input is aI_1+bI_2 the output is aO_1+bO_2. This property of linearity leads to a vast simplification in the mathematical description of phenomena and represents the foundation of a mathematical structure known as linear system theory. When the system is linear, the input and output may be decomposed into a linear combination of elementary components.

Another mathematical tool of great use is the Fourier transform. The Fourier transform is defined by

$$F(u,v) = \int_{-\infty}^{\infty}\!\!\int f(x,y) \exp -j2\pi(ux+vy)\ dx\ dy = F\{f(x,y)\}. \quad (1)$$

The transform is a complex valued function of u and v, the spatial frequencies in the image plane. The Fourier transform provides the continuous coefficients of each frequency component of the image. The Fourier transform is a reversible process, and the inverse Fourier transform is defined by

$$f(x,y) = \int\int_{-\infty}^{\infty} F(u,v) \exp{j2\pi(ux+vy)} \, dx \, dy = F^{-1}\{F(u,v)\}. \quad (2)$$

The transform and inverse transform are very similar differing only in the sign of the exponent appearing in the integrand. The magnitude squared of the Fourier transform is called the power spectral density

$$0_f = \quad |F(u,v)|^2 = F(u,v) \, F^*(u,v). \quad (3)$$

It is noteworthy that the phase information is lost from the Fourier transform when the transform is squared and the image cannot, in general, be reconstructed from the power spectral density. Several useful properties of the Fourier transform are listed here.

Linearity Theorem

$$F\{af_1(x,y) + bf_2(x,y)\} = a F\{f_1(x,y)\} + b F\{f_2(x,y)\}. \quad (4)$$

The transform of the sum of two functions is simply the sum of their individual transforms. The Fourier transform is a linear operator or system.

Similarity Theorem

$$F\{f(ax,by)\} = F(u/a,v/b)/ab \quad \text{where} \quad F(u,v) = F\{f(x,y)\}. \quad (5)$$

Scale changes in the image domain results in an inverse scale change in the frequency domain along with a change in the overall amplitude of the spectrum.

Shift Theorem

$$F\{f(x-a,y-b)\} = F(u,v) \exp[-j(ua+vb)]. \quad (6)$$

8

Translation of patterns in the image merely introduces a linear phase shift in the frequency domain. The magnitude is invariant to translation.

Parseval's Theorem

$$\iint |F(u,v)|^2 \, du \, dv = \iint |f(x,y)|^2 \, dx \, dy \qquad (7)$$

The total energy in the images plane is exactly equal to the energy in the frequency domain.

Convolution Theorem

$$F\{f(x,y) \, g(x,y)\} = \iint F(u,v) F(u_o-u, v_o-v) \, du \, dv \qquad (8)$$

The Fourier transform of the product of two images is the convolution of their associated individual transforms. Also, the Fourier transform of the convolution of two images is the product of the individual transforms.

Correlation Theorem

$$R_{fg}(x,y) = \iint f(x,y) \, f(x-xo, y-yo) \, dxo \, dyo. \qquad (9)$$

The correlation is very similar to the convolution except that neither function is inverted.

Autocorrelation (Wiener-Khintchine) Theorem

$$Off(u,v) = F\{R_{ff}(x,y)\} \qquad (10)$$

This special case of the convolution theorem shows that the autocorrelation and the power spectral density are Fourier transform pairs.

Fourier Integral Theorem

$$f(x,y) = F^{-1}\{F\{f(x,y)\}\} \qquad (11)$$

$$f(-x,-y) = F\{F\{f(x,y)\}\}$$

9

Successive transformation and inverse transformation yield that function again. If the Fourier transform is applied twice successively, the result is the original image inverted and perverted.

It is also useful to define here the impulse function. Also known as the Dirac delta function, it describes a function which is infinite at the origin, zero elsewhere, and contains a volume equal to unity. One definition of the Dirac delta function is

$$\delta(x) = \lim_{a \to 0} (a/\pi) \exp{-a^2 x^2}. \tag{12}$$

The delta function possesses these fundamental properties:

$$\delta(x) = 0 \quad \text{for } x \neq 0 \tag{13}$$

$$\int_{\infty}^{\infty} \delta(x)dx = \int_{\infty}^{\infty} \delta(x)dx = 1 \tag{14}$$

$$\delta(x) = \delta(-x) \tag{15}$$

$$\delta(ax) = (1/a)\,\delta(x) \qquad a \neq 0 \tag{16}$$

$$\int f(x)\,\delta(x-a)dx = f(a). \tag{17}$$

The Fourier transform of the delta function is unity. This property provides a useful tool when studying systems in which an output is dependent on the input to the system. When an impulse is the input to the system, the input spectrum is unity at all frequencies. The spectrum of the output must then correspond to the gain or attenuation of the system. This frequency response of the system is the Fourier transform of the output when an impulse is the input. The output of the system is the impulse response. Thus, the impulse response and the frequency response of the system are Fourier transform pairs. To determine the output of a system for a given input, multiply the Fourier transform of the input by the frequency response of the system and take the inverse Fourier transform of the result. The convolution property shows an equivalent operation is to convolve the input with the impulse response of the system.

$$O(u,v) = I(u,v) \, H(u,v) \qquad (18)$$

$$o(x,y) = F^{-1}\{O(u,v)\} = F^{-1}\{I(u,v) \, H(u,v)\} \qquad (19)$$

$$= \iint i(x_o,y_o) \, h(x-x_o, y-y_o) \, dx_o \, dy_o$$

$$= f(x,y) \circledast h(x,y)$$

where $\circledast$ denotes convolution.

Consider the effect of an additive noise on the input of the system. Although the exact form of the noise $n(x,y)$ may not be known, the noise statistics or power spectral density may be predictable. Thus, the effect of the system on the input is determined by its impulse response or frequency response. That is, when there is knowledge of the input signal and noise, the output signal and noise characteristics can be predicted. The relationship of the input and output are expressed in the following diagram and equations. The letters i and o indicate the input and output terms while the letters s and n indicate the signal and noise portions.

$$s(x,y) + n(x,y) \longrightarrow \boxed{\begin{array}{c} \text{Linear System} \\ h(x,y) \end{array}} \longrightarrow s_o(x,y) + n_o(x,y)$$
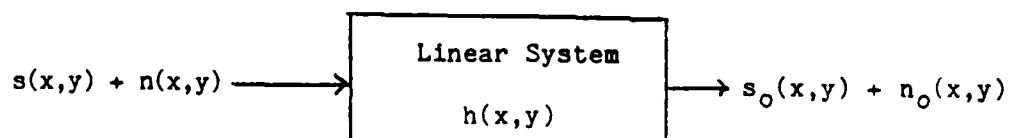
$$i(x,y) = s_i(x,y) + n_i(x,y) \qquad (20)$$

$$o(x,y) = s_o(x,y) + n_o(x,y) \qquad (21)$$

$$O(u,v) = I(u,v) \, H(u,v) \qquad (22)$$

$$S_o(u,v) = S_i(u,v) \, H(u,v) \qquad (23)$$

$$N_o(u,v) = N_i(u,v) \, H(u,v) \qquad (24)$$

Now that the relationships between the input and output of a linear system are known, such a system may be utilized to enhance the input. For example, assume an image has been degraded by some distorting function $d(x,y)$. The original image was convolved with the distorting function, and the spectral contents of the ideal image $F_i(u,v)$ were attenuated by the frequency response $D(u,v)$ of the distorting system. By multiplying the degraded image by the inverse of the $D(u,v)$, the original ideal image is obtained. Any distortion which can be represented as a linear system might theoretically be canceled out using the inverse filter. A photograph produced in a camera with a frequency response which rolls off slowly could be sharpened by Fourier transforming the image, multiplying by the inverse filter, and then inverse transforming. In this case, the inverse filter is one in which the low frequencies are attenuated and the high frequencies are accentuated (high pass filter). Because the high frequencies represent the edges in the image, the edges are accentuated and the photo appears sharper (Reference 17). As indicated in the following diagram, the image is distorted by the function $D(u,v)$ but in some cases can be restored by multiplying by $1/D(u,v)$.

$$f_i(x,y) \Rightarrow \boxed{F_i(u,v) \ X \ D(u,v)} \Rightarrow \quad f_d(x,y) \quad = \quad \text{blurred photograph}$$

$$f_d(x,y) \Rightarrow \boxed{F_d(u,v) \ X \ 1/D(u,v)} \Rightarrow f'_d(x,y) \quad = \quad \text{enhanced photograph}$$

The linear blur of a camera is another classic example. Consider traveling through Europe on a train with your camera. Upon getting home and receiving your trip pictures, you find that all of them are streaked by the motion of the train past the scenes you photographed. Each point in the scene streaked past the camera, causing a line to be formed on the film rather than a sharp point. The impulse response is a line, and the corresponding frequency response of the distorting system is a sinc function (sin u /u). To retrieve the European photo collection, merely multiply the Fourier transform of the pictures by u/sin u and re-image.

In the physical implementation of this process, there are several practical problems. To multiply the image transform by the inverse function, a transparency with the appropriate response is produced. In general, a transparency can only attenuate the light striking it. That is, the transparency can only represent non-negative real values less than one. Herein is the problem. The inverse response required to correct a specific distortion may be complex. In some cases, a combination of two transparencies can be combined to provide complex values. One transparency is used for amplitude or attenuation, and another phase transparency or phase plate is used to provide the appropriate phase shift at each point. A phase transparency can be produced by bleaching film with an appropriate latent image induced in the emulsion. Chu, Fienup, and Goodman (Reference 18) demonstrated a technique in color film which consists of three emulsions. One emulsion was used as an amplitude transparency and another emulsion was used as a phase plate. The appropriate patterns were determined by a computer and the film was given the proper exposure using colored filters.

Even with a two-transparency system, not all distortions are possible to remove. Note that in the linear blur case, the inverse response is u/sin u. The denominator goes to zero for specific values of u, and the response has a pole at those values. The filter cannot represent those values, and the practical filter is merely an approximation to the ideal filter. It is worth noting that when the distorting response reduces a frequency component to zero or below some noise threshold, that component cannot be recovered. That is, information is usually lost during the distorting process and inverse filtering cannot recover it.

It is desirable to remove noise from a corrupted image. Although it is not always possible to remove all of the noise, the relationships between the input and output of a linear system are known. A linear system is optimized when most of the noise is removed. To optimize a system the input must be specified, the system design restrictions known, and a criterion of optimization accepted. The input may be a combination of known and random signals and noises. The characteristics of the input such as the noise spectrum or statistics must be available. The classes of systems are restricted to those which are linear, space-invariant, and physically realizable. The criterion of the

13

optimization is dependent on the application. The optimum filters include the least mean-square-error (Wiener) filter and the matched filter. The Wiener filter minimizes the mean-squared-error between the output of the filter and actual signal input. The Wiener filter predicts the least mean-squared-error estimate of the noise-corrupted input signal. Thus, the output of the Wiener filter is an approximation to the input signal. The output of the matched filter is not an approximation to the input signal but rather a prediction of whether a specific input signal is present. The matched filter does not preserve the input image. This is not the objective. The objective is to distort the input image and filter the noise so that at the sampling location (x_0, y_0), the output signal level will be as large as possible with respect to the output noise.

The signal-to-noise ratio is useful in the evaluation of system performance, particularly in linear systems. In the matched filter, the criterion of optimization is that the output signal-to-noise power be a maximum. The input consists of a known signal $s(x,y)$ and an additive random noise $n(x,y)$. The system is linear and space invariant with impulse response $h(x_0, y_0)$. To optimize the system or filter, maximize the expression

$$R_0 = s_0^2(x_0, y_0) / E\{n_0^2(x,y)\} \tag{25}$$

where $E\{n_0^2(x,y)\} = \iint n_0^2(x,y) \, dx \, dy$

at some point (x_0, y_0). The problem is then to find the system $h(x,y)$ that performs the maximization of the output signal-to-noise ratio. The output signal $s_0(x,y)$ is

$$s_0(x,y) = \iint s_i(x_0, y_0) h(x-x_0, y-y_0) \, dx_0 \, dy_0 \tag{26}$$

and the output noise $n_0(x,y)$ power is

$$\iint |n_0(x,y)|^2 \, dx \, dy = \iint |N_0(u,v)|^2 \, du \, dv$$
$$= \iint |N_i(u,v)|^2 \, |H(u,v)|^2 \, du \, dv. \tag{27}$$

The signal-to-noise output power ratio becomes

$$R_o = \frac{|\int\int s_i(x_o,y_o)h(x-x_o,y-y_o)\ dx_o\ dy_o|^2}{|N_i(u,v)|^2\ |H(u,v)|^2} \quad . \tag{28}$$

Thus, to complete the maximization with respect to $h(x,y)$, the power spectral density or some equivalent specification of the input noise must be known. Once the input noise is specified, the filter function $h(x,y)$ is the only unknown. Equation (28) becomes

$$E\{n_o^2(x_o,y_o) - as_o^2(x_o,y_o)\} > 0 \tag{29}$$

$$N_i^2(u,v)H^2(u,v)\ du\ dv - a|\int\int s_i(x,y)h(x-x_o,y-y_o)\ dx_o\ dy_o|^2 > 0$$

where $R_{o\ max} = 1/a$

and the maximum signal-to-noise ratio at the output is obtained when $H(u,v)$ is chosen such that equality is attained. This occurs when

$$\int\int n_i^2(x,y)\ h(x-x_o,y-y_o)\ dx_o\ dy_o = s_i(x,y). \tag{30}$$

Taking the Fourier transform of both sides and rearranging gives

$$H(u,v) = \frac{S(-u,-v)}{|N_i(u,v)|^2}\ \exp -j(ux_o+vy_o) \quad . \tag{31}$$

Thus in an intuitive sense, the matched filter emphasizes the signal frequencies but with a phase shift and, attenuates the noise frequencies. This becomes clear when the additive noise is white. In this case the noise power is constant at all frequencies and thus has a power spectral density of

$$|N_i(u,v)|^2 = N/2 \qquad \text{where N is a constant} \tag{32}$$

15

From equation 32 the form of the matched filter for the case of white noise is

$$H(u,v) = S_i(-u,-v)\exp -j(ux_o+vy_o) \tag{33}$$

$$= S^*_i(u,v) \exp -j(ux_o+vy_o)$$

or

$$h(x,y) = s(-x,-y).$$

Equation (33) shows that the impulse response of the matched filter (with white noise) is simply the signal image in reverse order (inverted and perverted). Thus, the filter is said to be matched to the signal. Filtering with a matched filter is equivalent to cross-correlating with the expected signal or pattern. That is,

$$O(x,y) = R_{hs}(x,y)$$

$$= \iint s(x_o,y_o)h(x_o-x,y_o-y)\ dx_o\ dy_o \tag{34}$$

Also, it can be seen that the frequency response of the matched filter is equivalent to that of the signal but with the phase negated so that the output of the filter is real. That is, the matched filter removes the phase variations and provides a real valued output (Reference 19).

Matched filters are used extensively in radar signal processing, seismic data processing, and communications. These filters are implemented using electronic circuitry and digital computers. For image processing, the need to process large two-dimensional arrays places a large burden on conventional filtering techniques. For these applications, optical processing techniques provide the highest throughput speeds for matched filtering. One such optical processing technique was proposed by Vander Lugt (Reference 7) in 1969.

2. VANDER LUGT FILTERING

If an image is placed on a transparent sheet and illuminated by a plane wave of coherent light, its Fourier transform is formed using a simple lens (Reference 19). Once the Fourier transform is formed, specific frequency

components in the image can be removed or attenuated. The result may then be inverse Fourier transformed to recreate the modified image. The aperture, which may be replaced by a complicated filter, functions to perform specific filtering operations including Wiener or matched filter. Unfortunately, there are certain limitations to the functions which can be physically implemented. A normal transparency merely attenuates the light passing through it. Its transmission is real and non-negative. Thus, when a transparency film is exposed to a waveform to be recorded, the phase information in the waveform is lost. Two pieces of information, the real and imaginary parts of the waveform, are recorded as only one value, their magnitude. This loss of information can be corrected by taking advantage of the redundancy in the wavefront and the use of additional film space. When using the heterodyning technique proposed by Vander Lugt, the complex waveform can be recorded on photographic film.

Vander Lugt proposed the use of holographic film to store the filter response for a matched filter. A lens is used to Fourier transform the reference and test images. Derivations of the Fourier transforming capabilities of lenses can be found in the literature (Reference 10). The Fourier transform of the reference image is brought to focus on a photographic film. Film is a nonlinear, time-integrating medium and thus only the magnitude of the Fourier transform or power spectral density is recorded. The power spectral density does not contain all of the original image information. Only the autocorrelation of the original image can be obtained upon inverse transformation. Neither the power spectral density nor the autocorrelation uniquely describe the original image. If a plane wave is mixed with the Fourier transform of the reference image at the film plane, the film will record the interference pattern caused by the summation of the two fields. The result on the film then is

$$H(u,v) = 1 + |F(u,v)|^2 + F(u,v)\exp j2\pi av + F^*(u,v)\exp -j2\pi av, \quad (35)$$

which contains a constant, the power spectral density, and two terms due to a spatial carrier fringe formed due to interference with the plane wave. The two spatially modulated terms contain the original image and Fourier transform information. With this Fourier transform recorded on the film, it is placed in the optical filter arrangement and illuminated with the Fourier transform

$G(u,v)$ of the test image $g(x,y)$. The output of the film transparency is the product of its transmittance and the illuminating Fourier transform.

$$O(u,v) = G(u,v) \; H(u,v) \tag{36}$$

$$= G(u,v) + G(u,v) |F(u,v)|^2$$

$$+ \; G(u,v)F(u,v)\exp \; j2\pi av + G(u,v)F^*(u,v)\exp \; -j2\pi av.$$

The product of the transforms from the reference and test images is then Fourier transformed by another lens to obtain the correlation of the two images.

$$o(x,y) = g(x,y) + g(x,y) \circledast h(x,y) \circledast h^*(x,y) \tag{37}$$

$$+ \; g(x,y) \circledast f(x,y) \quad (x,y-a)$$

$$+ \; g(x,y) \circledast f^*(x,y) \star \delta(x,y+a).$$

The first two terms are formed on axis or at the origin of the output plane. The third term is the convolution of the reference and test images and is centered off axis. The last term is the correlation of the reference and test images and is located off-axis opposite the convolution. This optical arrangement provides the entire convolution and correlation images at once while a digital processor must compute one point at a time. In addition to the convolution and correlation processes, additional image plane and frequency plane filtering may be accomplished simultaneously in the same optical arrangement. The convolution, correlation, and any additional linear filtering are accomplished with a single absorbing mask.

When used as a matched filter, the transparency multiplies the expected pattern by its complex conjugate, thereby rendering an entirely real field. This matched transparency exactly cancels all the curvature of the incident wavefront. When an input other than the expected signal is present, the wavefront curvature will not be canceled by the transparency and the transmitted light will not be brought to a bright focus. The expected pattern will be

detected by a bright point of light in the correlation plane. If the pattern occurs in the input plane but is shifted, the bright point of light in the correlation plane will shift accordingly. This provides for the detection of specific patterns in a larger image. The detection and location of specific objects in large complicated images is a job well suited for the high-speed processing capability of the Vander Lugt filter.

SECTION III
COMPUTER-GENERATED HOLOGRAMS

Vander Lugt described a technique by which the holographic matched filter could be produced optically (Reference 7). At that time, no other convenient method existed for the computation and creation of the complicated filter function required. This limitation has faded away with the availability of digital computers with large memories. Using digital computers to determine the filter function and a computer-driven writing device, a transparency with the appropriate filter image can be produced. Using this technique, the computer determines the appropriate value of the matched filter at each point and produces a transparency with that absorption at each point. The resolution required of the writing device depends on the application and, in some cases, may be consistent with optically generated holograms.

Computer-generated holograms (CGH) have found applications in optical information processing, interferometry, synthesis of novel optical elements, laser scanning, and laser machining (References 20 through 23). CGHs can implement computer-optimized pattern-recognition masks (Reference 24). The computer writes the hologram by transferring the transmittance function to an appropriate holographic medium. The computer drives a plotter or scanner and writes the hologram one point at a time. Typically, the primary limitation is writing resolution. A conventional optical hologram may have a resolution of one-quarter of a micron. A system using visible light to write holograms (plotters, flying spot scanners, CRT's, etc.) cannot achieve resolutions much better than several microns. Writing systems utilizing electron beams are currently achieving better than 1-micron resolution. The electron beam systems are typically binary and thus the transmittance function must be quantized in some fashion into two levels, on or off. Binary holograms are attractive because binary computer-graphics output devices are widely available and because problems with nonlinearities in the display and recording medium are circumvented (Reference 12). When photographic emulsions are involved, granularity noise is reduced (Reference 25).

1. CONTINUOUS-TONE HOLOGRAMS

When a hologram is produced optically or interferometrically, a reference wave is superimposed with the wavefront to be recorded. Typically, the reference wave is a tilted plane wave with constant amplitude across the wavefront. The reference wave approaches at an angle 0 relative to the direction of the wavefront to be recorded. The resultant field is

$$E(x,y) = f(x,y) + A\exp(j2\pi ay) \tag{38}$$

where $\qquad a= \dfrac{\sin 0}{\lambda}$

and the amplitude of the reference wave is 1. An interference pattern is produced by the superposition of the waves. The fringe spacing is dependent on the term a, known as the spatial carrier frequency, and the details in the function $f(x,y)$. A photographic film placed into this field records not the field itself but rather the square magnitude of the field. The pattern recorded on the film is then

$$h(x,y) = |f(x,y) + A e^{j2\pi ay}|^2 \tag{39}$$

$$= A^2 + |f(x,y)|^2 + A f(x,y)e^{j2\pi ay} + A f^*(x,y)e^{-j2\pi ay}.$$

The function recorded on the film contains a D.C. bias, A^2, the base band magnitude, $|f(x,y)|^2$, and two terms heterodyned to plus and minus a. These heterodyned terms contain the complex valued information describing the input function $f(x,y)$. If the spatial carrier frequency is sufficiently high, the heterodyned terms are separable and no aliasing exists. The original input function can be retrieved with no distortion by re-illuminating the film with the reference beam and spatially filtering the output to separate the various terms.

To make the hologram of the Fourier transform of an image, the same procedure is applied. That is, the Fourier transform of the image $f(x,y)$ is used as the input to the hologram. Now

$$h(u,v) = A^2 + F(u,v)^2 + A F(u,v)e^{j2\pi au} + A F^*(u,v)e^{-j2\pi au} \tag{40}$$

21

where $F(u,v)$ = Fourier Transform of $f(x,y)$ = $F\{f(x,y)\}$ and

$A\ e^{-j2\pi au}$ = the off-axis reference wave used to provide the spatial carrier for the hologram.

$a = \dfrac{\sin\theta}{\lambda}$ = the filter spatial carrier frequency (θ = off-axis angle)

This filter contains the D.C. bias, A^2; the power spectral density, $|F(u,v)|^2$; and two terms heterodyned to plus and minus a. These heterodyned terms contain the complex valued information describing the Fourier transform of the input $f(x,y)$.

These optically generated holograms are formed interferometrically by combining a plane wave with the wavefront to be recorded. The transmittance of the hologram is a real valued, non-negative function of position on the plate. Recall that the input $F(u,v)$, which was used to create the hologram, is, in general, complex. This conversion from a complex function to a pattern which can be recorded on film is known as coding. The coding performed in optical holography is a natural consequence of the action of the film. Typically, the complex wavefront is coded to a real non-negative function which can be recorded as transmittance values on film. Equation (35) describes a way in which film (a square law detector) would encode the complex input image in an optically generated hologram.

Once produced, the hologram and its interference fringes may be inspected by microscope. The hologram can be copied on another plate by contact print-ing. The hologram consists of real valued positive transmittance varying across the face of the photographic plate. To record the hologram on a com-puter, the transmittance across the surface of the plate is sampled. If the samples are many and the transmittance determined with accuracy, the hologram can be accurately reproduced from the recorded samples. In this way the holo-gram can be represented with some accuracy using digital numbers stored on some electronic media. An electronic device writes the physical hologram. The computer can electronically record, modify an optically produced hologram, and then rewrite the holographic pattern onto another plate. The limitations to such a system include the ability to sample the input hologram sufficiently

often and accurately, the ability to store the large number of sample values, and the ability to rewrite the holographic pattern to film.

If the input wavefront is known, the optical step may be omitted altogether. If the input wavefront can be accurately represented by discrete samples stored electronically, the holographic pattern can be computed. That is, the input is coded to create a function which can be recorded on a transparency. In the case of the matched filter, the Fourier transform of an image is recorded. The image is sampled and stored on the computer, and equation (35) is used to determine the holographic pattern. Note that the continuous variables are replaced by discrete steps. At each sample point the actual value is represented by a finite number. The value may be complex, but the accuracy is limited by the sampling system. In any case the holographic pattern is computed and written to the photographic plate. The writing device is known as continuous-tone when the transmittance of each point in the holographic plate can be controlled over a wide range of transmittance values. The transmittance varies smoothly from clear to opaque, including gray scale values between. These continuous-tone holograms most closely resemble the optically generated holograms when the sampling is dense and many gray scale values are available.

When continuous-tone holograms are written to the photographic plate using equation (35) as the model, they include a D.C. term, a square magnitude term, and the heterodyned terms due to the tilted reference wave. Note that the first two terms are real valued and that the sum of the last two terms is real valued. On the computer, the film process is emulated using equation (35) or other coding schemes for specific applications. The D.C. and square magnitude terms need not be included in the computer-generated hologram as long as the heterodyned terms are scaled and biased to provide results between 0 and 1. The heterodyned terms contain the desired information. Omission of the baseband terms has no adverse effect on the hologram. The square magnitude term typically contains a large dynamic range. Its omission from the coding algorithm helps reduce the dynamic range of the hologram and, in most cases, improves the hologram. Equation (40) can be replaced by the expressions

$$H(u,v) = 2|F(u,v)| + F(u,v)e^{j2\pi au} + F^*(u,v)e^{-j2\pi au} \qquad (41)$$

23

$$H(u,v) = A2 + F(u,v)ej2\pi au + F^*(u,v)e\text{-}j2\pi au \qquad (42)$$

where each of these expressions includes the reference information, is real valued, and is non-negative.

The dynamic range in the hologram, defined as the largest value divided by the smallest value, is limited by the writing device used to create the hologram. Most films have dynamic ranges much less than 10,000. The clearest portions of the film can transmit light no better than 10,000 times better than the darkest portions. If the coding scheme requires a dynamic range of over 10,000, the writing device cannot faithfully reproduce the holographic pattern. Unfortunately, the dynamic range of film is frequently much less than 10,000 and closer to 100. Additionally, the writing device also limits the dynamic range. Most continuous-tone writing devices, which are attached to computers, convert an integer value to an intensity on a cathode-ray tube or flying spot scanner. Due to physical limitations in the writing intensity, the dynamic range is usually much less than 1000. Most commercially available computer-writing devices are designed with a dynamic range of 256 or 8-bit writing accuracy. The resultant transmittance on the film will have one of 256 quantized levels determined by an integer value provided by the computer. Quantization occurs when all values in a specified range are assigned to a quantized value representative of that range. If the quantization steps become large, the quantized level may be a poor estimate of the actual values. The estimate is equivalent to the actual pattern with an additive noise called quantization noise. Quantization noise occurs in computer-generated holograms because the computer-graphic devices have limited gray levels and a limited number of addressable locations in their outputs. Quantizing the holographic pattern into 256 gray scale levels introduces quantizing noise which may be considerable when the dynamic range of the pattern is large. To minimize the quantizing error, the coding scheme must produce a result with a dynamic range compatible with the writing system.

Some writing systems are capable of only two quantized levels. These binary devices are either on or off. Most metal etchers, ink plotters, dot matrix printers, and lithographic printers are binary. The mark they create is either completely on or completely off. To represent the reference pattern on binary media accurately requires specialized coding schemes.

2. BINARY HOLOGRAMS

Binary holograms are attractive because binary computer-graphics output devices are widely available and because problems with nonlinearities in the display and recording medium are circumvented. When photographic emulsions are involved, granularity noise is reduced. Using the normal definition of dynamic range, binary holograms have a dynamic range of 1. The transmittance at each point is completely on or completely off. All gray scale effects must be created by grouping light and dark areas together and averaging over an area large enough to provide the needed dynamic range. In this case the dynamic range is the averaging area. Thus, dynamic range is exchanged for increased area to represent each point. This is similar to Pulse Code Modulation (PCM) in an electronic communication systems (Reference 26). In PCM, each sample value is quantized to M levels. Then each level is represented by a binary code requiring $N = \log_2 M$ bits. Rather than represent each point with a continuous variable with sufficient dynamic range, N binary variables are used. Each variable is either on or off, but N variables are required to provide sufficient dynamic range. This exchanges dynamic range of the variables for the number of variables required. In binary holograms, the variables are not, in general, exponentially weighted as in PCM; thus, M variables are required to represent M levels. It becomes very important to code the hologram such that the number of variables M needed to represent that dynamic range is reasonable.

One of the first practical binary coding schemes was introduced when, in 1966, Brown and Lohmann, (Reference 8), devised a method for complex spatial filtering using binary masks. They coded the Fourier transform of an image $f(x,y)$. When using this method, the complex Fourier transform is sampled and represented at each point by an amplitude and phase. To record a complex filter, both amplitude and phase information are needed on the hologram. However, the hologram transmittance is real-valued, non-negative, and in this case binary. The amplitude can be recorded by opening or closing an appropriate number of binary windows in the hologram, but the phase is not correct. Brown and Lohmann proposed turning the hologram at an angle to the incoming waveform. Thus, along the surface of the hologram, a phase shift occurs. This phase shift is proportional to the position along the hologram. Using this tilted wave technique, a phase shift occurs as the aperture moves up and down the

hologram causing the total path length through that aperture to change. The further the detour through the aperture, the larger the phase shift. Phase shift induced by this technique is known as detour phase. Thus, in the Brown-Lohmann hologram, an aperture is moved up and down to create the appropriate phase shift. The size of the aperture is varied to allow the appropriate amount of light through. To synthesize the complex filter function $F(u,v)$, a continuous function is sampled. The cells of a size Δu by Δv must be sufficiently small that the function F will be effectively constant throughout the cell.

$$F(u,v) = F(n\Delta u, m\Delta v) = F_{nm} = A_{nm} \exp i\theta_{nm} \qquad (43)$$

where n and m are integers.

For each cell in the hologram, the amplitude and phase are determined by the size and position of an aperture as shown in Figure 1. From each cell a complex light amplitude F_{nm} will emerge. The tilted wave must approach at an angle steep enough to allow for a full wavelength of detour phase within one cell. The dynamic range of the amplitude and phase is limited by the number of resolvable points within the cell. If a cell has only 4 by 4 resolvable points, the dynamic range of the amplitude or phase can be no better than 4. The granularity in the amplitude and phase may cause distortion in the reconstructed image. Many points are required to represent a transform with a large dynamic range accurately.

Lee (Reference 9) proposed a method in 1970 which helped relieve some of the phase granularity. The Brown-Lohmann technique represented each cell with an amplitude and phase component. The complex value for each cell may be represented by a magnitude and phase or by the sum of in-phase and out-of-phase terms. The Lee method represents each cell with such a quadrature representation. For each cell the magnitude and phase are converted to real and imaginary components. As in the Brown-Lohmann method, the tilted wave is set to provide a wavelength of delay across the cell. The cell is divided into four components which represent the positive and negative real and imaginary axes.
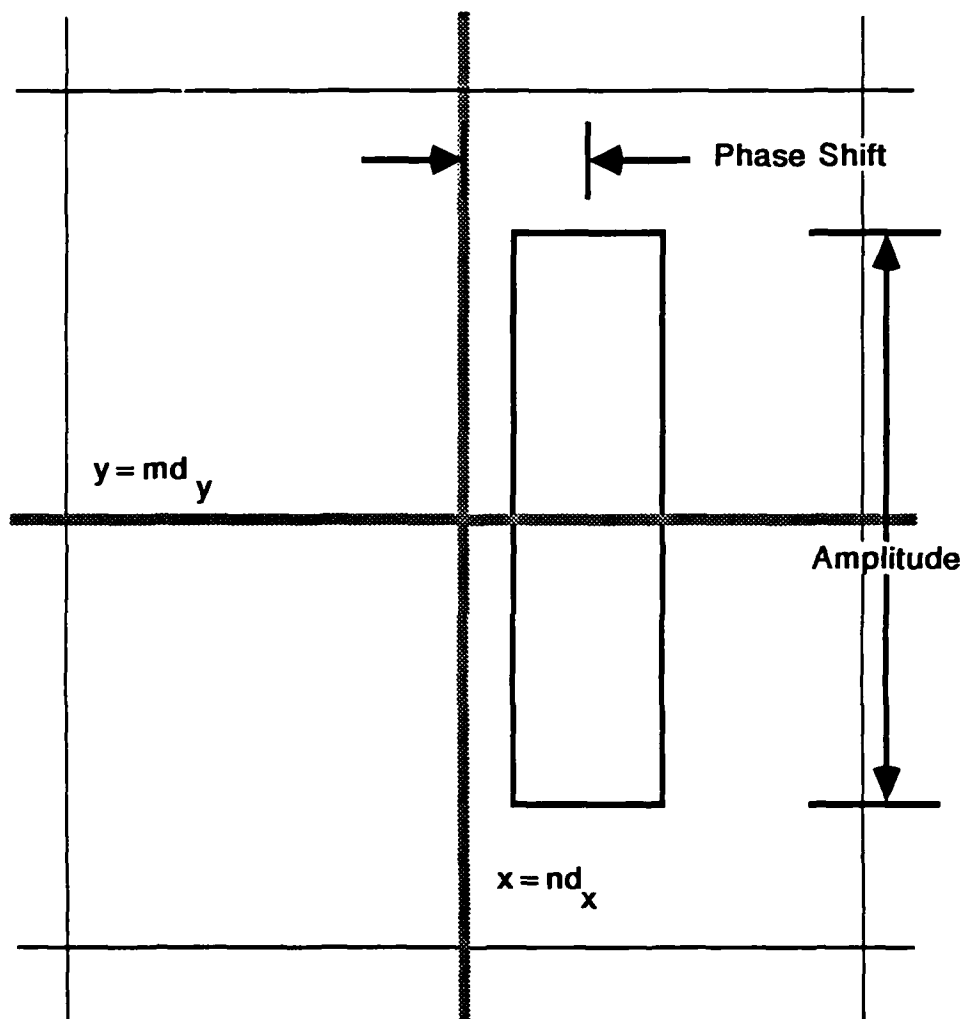
Figure 1. Brown and Lohmann CGH Cell

Lee defined the functions as

$$|F(u,v)|\exp[j\ \phi(u,v)] = F_1(u,v)-F_2(u,v)+jF_3(u,v)-jF_4(u,v) \qquad (44)$$

where

$$
\begin{aligned}
F_1(u,v) &= |F(u,v)|\cos\phi(u,v) \qquad && \text{if } \cos\phi(u,v) > 0 \\
&= 0 \qquad && \text{otherwise,}
\end{aligned}
$$

$$
\begin{aligned}
F_2(u,v) &= |F(u,v)|\sin\phi(u,v) \qquad && \text{if } \sin\phi(u,v) > 0 \\
&= 0 \qquad && \text{otherwise,}
\end{aligned}
$$

$$
\begin{aligned}
F_3(u,v) &= -|F(u,v)|\cos\phi(u,v) \qquad && \text{if } \cos\phi(u,v) > 0 \\
&= 0 \qquad && \text{otherwise,}
\end{aligned}
$$

$$
\begin{aligned}
F_4(u,v) &= -|F(u,v)|\sin\phi(u,v) \qquad && \text{if } \sin\phi(u,v) > 0 \\
&= 0 \qquad && \text{otherwise.}
\end{aligned}
$$

For any given complex value, two of the four components are zero. Each of the components $F_n(u,v)$ is real and non-negative and can be recorded on film. The Lee hologram uses four apertures for each cell shown in Figure 2. Each aperture is positioned to cause a quarter-wave phase shift by increased path length (detour phase). The two non-negative quadrature terms are weighted to vector sum to the appropriate magnitude and phase for each pixel. The two appropriate apertures are opened according to their weight. The Lee method uses continuous-tone variables to represent the two non-zero components. The phase is no longer quantized by the location of the aperture. The phase is determined by the vector addition of the two non-zero components. In a totally binary application of the Lee method, the apertures are rectangles positioned to obtain the quarter-wave shift. The area of each aperture is adjusted to determine the amplitude of each component. Once again, in this binary case, the dynamic range is limited by the number of resolution elements within one cell.

Burckhardt (Reference 10) showed that while the Lee method decomposes the complex-valued $F(u,v)$ into four real and positive components, only three components are required. Each cell can be represented by three components 120 degrees apart. Any point on the complex plane can be represented as a sum of
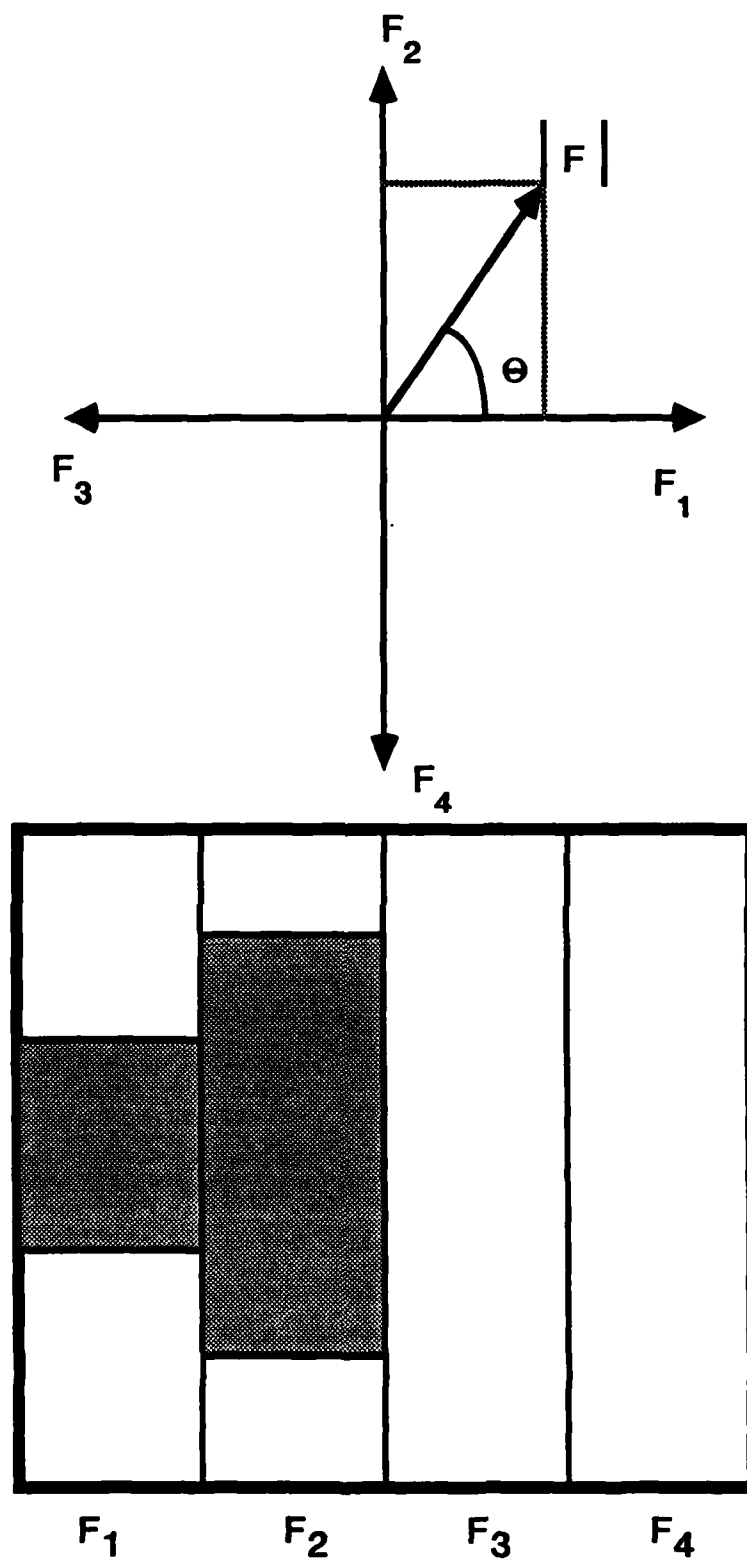
Figure 2. Complex Plane Showing Four Quadrature Components

any two of these three components. As in the Lee method, two non-negative components are chosen to represent each cell. Because only three instead of four components have to be stored, the required memory size and plotter resolution are reduced. Haskell (Reference 11) describes a technique in which the hologram cell is divided into N components equally spaced around the complex plane. It is identical to the binary Lee (N=4) and the Burckhardt (N=3) where N may take larger values. This Generalized Binary Computer-Generated Hologram (GBCGH) uses N columns and K rows of subcells. Each subcell can take a transmittance value of 1 or 0. The phase is delayed by 2/N to provide N unit vectors. The K cells in each component are opened or closed to provide the appropriate weights for each component. The control over the amplitude and phase is not absolute with finite N and K. The result at each cell is the vector sum of components with integer length and fixed direction. Figure 3 shows that various combinations of points turned on or off define an array of specific points addressable in the complex plane. By increasing the number of points N and K, the amplitude and phase can be more accurately matched. When the total number of plotter dots is limited and more subcells used for each cell, fewer cells can exist. Thus, with a limited number of points, the hologram designer must choose between space-bandwidth product (number of cells) and quantization noise.

The GBCGH allows more accurate determination of the amplitude and phase of the cell by using more points. However, the complex sample to be represented was taken at the center of the aperture. If N, the number of points in the cell, is large, the outer pixel may have noticeable error due to the offset in sample location. Allebach (Reference 12) showed that the Lohmann hologram fell into a class of digital holograms which sample the object spectrum at the center of each hologram cell to determine the transmittance of the entire cell. The Lee hologram fell into a class of digital holograms which sample the object spectrum at the center of each aperture to determine its size. He also described a new third class in which the object is sampled at each resolvable spot to determine the transmittance at that spot. Although the function to be recorded should be constant over the entire cell, there is some phase shift across the cell dimensions. By sampling the object spectrum at the center of each aperture rather than at the center of each hologram cell, some of the false images in the reconstruction are removed. By sampling the object spectrum at the center of each resolvable spot in the hologram, the hologram
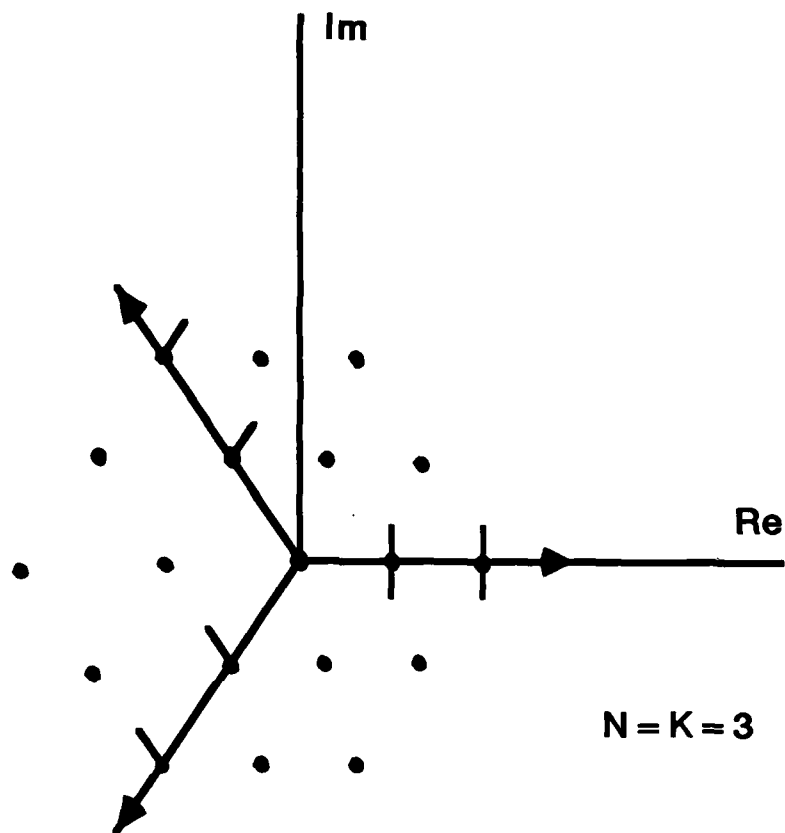
Figure 3. Addressable Amplitude and Phase Locations
Using the GBCGH Method

noise is further reduced. Allebach described an encoding technique in this last category known as the Allebach-Keegan (A-K) hologram (Reference 13). The A-K hologram encodes the complex-valued object spectrum by quadrature components as does the Lee hologram. Unlike the Lee hologram, the A-K hologram compares subsamples within the aperture to an ordered dither to determine whether each pixel is on or off. The input image is padded to provide as many points in the FFT as there are resolvable points. The FFT is decomposed into components spaced a quarter wave apart (or more as in the GBCGH). Each point is then compared to a threshold determined by the threshold matrix. The threshold values are chosen to quantize the amplitude of each component. The threshold values divide the range from zero to the spectrum maximum in steps determined by the Max quantizer (Reference 27). The size of the dither matrix and the corresponding points in the cell can increase as with the GBCGH but the magnitude and phase are sampled at each pixel.

3. SAMPLING AND SPACE-BANDWIDTH REQUIREMENTS

To represent an image on a computer, the image must be sampled and quantized into a set of numbers. To sample a continuous image or function, the value of the function is determined at discrete points. The values of a function $f(x,y)$ are determined at regular intervals separated by Δx and Δy. The continuous independent variables x and y are replaced with discrete sample point denoted by $m\Delta x$ and $n\Delta y$. Here Δx and Δy are the fixed sample intervals and m and n are integers. The sampling rate is $u=1/\Delta x$ in the x direction and $v=1/\Delta y$ in the y direction. To convert the continuous function $f(x,y)$ to a sampled version $f(m\Delta x,n\Delta y)$, multiply $f(x,y)$ with a grid of narrow unit pulses at intervals of Δx and Δy. This grid of narrow unit pulses is defined as

$$s(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta (x-m x, y-n y) \tag{45}$$

and the sampled image is

$$f_s(m\Delta x,n\Delta y) = f(x,y)\ s(x,y). \tag{46}$$

32

The sampled version is the product of the continuous image and the sampling function s(x,y). The spectrum of the sampled version can be determined using the convolution theorem (equation 8).

$$F_s(u,v) = F(u,v) \circledast S(u,v) \qquad (47)$$

where F(u,v) is the Fourier transform of f(x,y)

and   S(u,v) is the Fourier transform of s(x,y)

$$S(u,v) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(u-m\Delta u, v-n\Delta v)$$

where $u = 1/\Delta x$ and $v = 1/\Delta y$

Thus $$F_s(u,v) = \iint F(u-u_o, v-v_o) \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(u_o-m\Delta u, v_o-n\Delta v) \, du_o \, dv_o$$

$$\qquad (48)$$

Upon changing the order of summation and integration and invoking the sampling property of the delta function (equation 17), this becomes

$$F(u,v) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} F(u-m\Delta u, v-n\Delta v). \qquad (49)$$

The spectrum of the sampled image consists of the spectrum of the ideal image repeated over the frequency plane in a grid space $(\Delta u, \Delta v)$. If Δu and Δv are sufficiently large and the ideal function f(x,y) is bandlimited, no overlap occurs in the frequency plane. A continuous image is obtained from the sampled version by spatial filtering to choose only one order m,n of the sum in equation (49). If the image is undersampled and the frequency components overlap, then no filtering can separate the different orders and the image is aliased. To prevent aliasing, the ideal image must be bandlimited and sampled at a rate $\Delta u > 2f_u$ and $\Delta v > 2f_v$. The ideal image is restored perfectly when the sampled

33

version is filtered to pass only the 0,0 order and the sampling period is chosen such that the image cutoff frequencies lie within a rectangular region defined by one-half the sampling frequency. This required sampling rate is known as the Nyquist criterion. In the image, the sampling period must be equal to, or smaller than, one-half the period of the finest detail within the image. This finest detail represents one cycle of the highest spatial frequency contained in the image. Sampling rates above and below this criterion are oversampling and undersampling, respectively. To prevent corruption of the reconstructed image, no overlap of the desired frequency components can occur.

Frequency overlap is also a problem in holography. Recall that in equation (39) the ideal function $f(x,y)$ was heterodyned to a spatial carrier frequency by mixing with an off-axis reference beam, i.e.,

$$h(x,y) = A^2 + |f(x,y)|^2 + A\ f(x,y)e^{j2\pi ay} + A\ f^*(x,y)e^{-j2\pi ay} \qquad (50)$$

and that the spectrum (shown in Figure 4) of this recorded signal is

$$H(u,v) = |A|^2 + F(u,v) \circledast F(u,v) + A\ F(u,v+a) + A\ F(u,v-a) \qquad (51)$$

where $F(u,v)$ is the Fourier transform of $f(x,y)$ and $\circledast$ denotes convolution.

The first term is a delta function at (0,0). The second term is centered on axis (0,0) but has twice the width as the spectrum $F(u,v)$. The third and fourth terms are the Fourier transforms of the $f(x,y)$ but centered off axis at plus and minus a. To prevent frequency overlap, the second term and the heterodyned terms must not overlap. This requires that the spatial carrier frequency, a, used to heterodyne the information, must be sufficiently large. Specifically, this carrier frequency must be larger than three times the one-sided bandwidth of the information spectrum.

In the case of the Vander Lugt filter and the subsequent correlation, the output of the holographic matched filter has the form

$$o(x,y) = g(x,y) + g(x,y) \circledast f(x,y) \circledast f^*(x,y)$$

$$+g(x,y) \circledast f(x,y) \circledast \delta(x,y-a)$$

$$+g(x,y) \circledast f^*(x,y) \circledast \delta(x,y+a). \qquad (52)$$

The output, shown in Figure 5, contains a replica of the test image $g(x,y)$ centered on-axis along with a term consisting of the test image convolved with the autocorrelation of the reference image $f(x,y)$. This term consumes a width of twice the filter size plus the test image size. In addition to the on-axis terms, there are two heterodyned terms centered at plus and minus a. These heterodyned terms have a width equal to the sum of the widths of the test image $g(x,y)$ and reference image $f(x,y)$. Again to prevent overlap of the information terms in the output, a spatial carrier of sufficiently high frequency is required to separate the heterodyned terms from the on-axis terms. Assuming as an example that the test image and the reference image are the same size 2B, the output positions of the various terms can be shown graphically. To prevent the information terms from overlapping with the on-axis terms, the carrier frequency, a, must be chosen to center the heterodyned terms at 5B or more. In the general case, the reference image $f(x,y)$ and $g(x,y)$ may have different sizes. Let $2B_f$ represent the size of the reference image and $2B_g$ represent the size of the test image. Then the requirement on the carrier frequency, a, to prevent aliasing is

$$a = 3B_f + 2B_g. \qquad (53)$$

Sampling and heterodyning cause aliasing when improperly accomplished. The combination of the two in the CGH requires specific attention to detail. To create a CGH from a continuous image $f(x,y)$, it must first be sampled and quantized. According to the Nyquist criteria, there are two samples for the smallest details in the image. The sampling rate is at least twice the highest spatial frequency in the continuous image. If a limited number of sampling points are available, the image should be low pass filtered to limit the highest frequency in the continuous image to half the number of sampling points. This can be accomplished in an electronic sensor by blurring the optics before the detector. When using a television camera to digitize a transparency or film, the camera must be blurred to match the detail in the continuous image to the number of points in the digitizer. The detail required

35
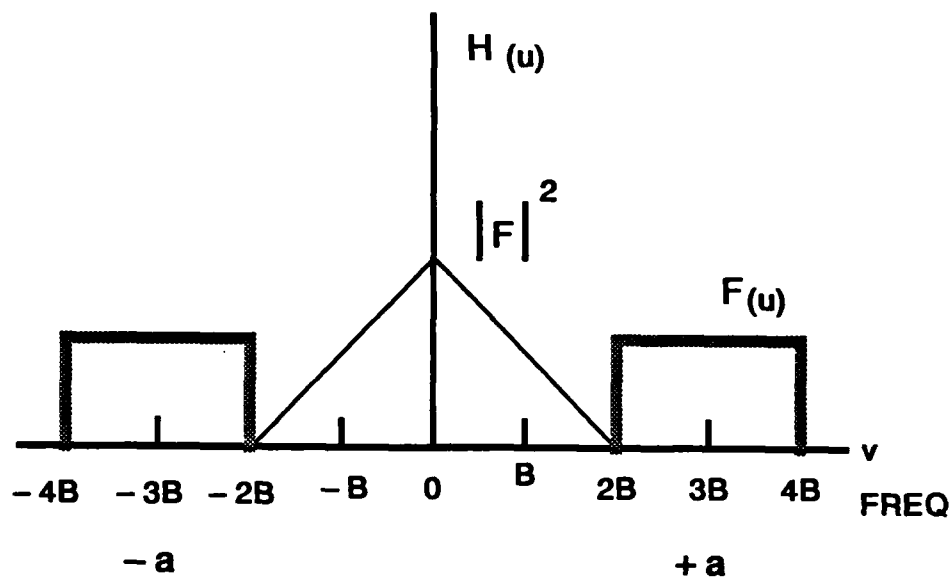
Figure 4.   Spectral Content of an Image Hologram

36
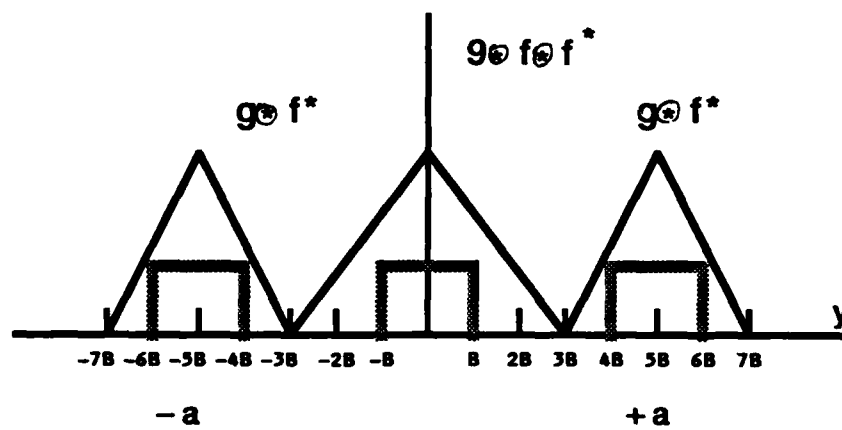
Figure 5.   Spectral Content of a Vander Lugt Filter

in the reference and test images is determined by the pattern or target to be recognized. To detect the presence of a desired target while an unwanted object could appear in the test scene, sufficient detail to discriminate the two must be included. To pick out cars from a scene which contains both cars and trucks, the resolution must be adequate to resolve the differences between the two. This resolution is typically chosen in an ad hoc fashion using the human eye to determine what resolution is required. Computer techniques have been used to quantify the resolution required, but the results are usually not different than what a human would have decided by eye. Although beyond the scope of this dissertation, the bandwidth and specific frequencies best suited to discriminate between targets and clutter can be determined with large computers operating on adequate training sets.

The resolution must be adequate for target recognition. However, over-sampling beyond that resolution required will drive the CGH to impractical limits. The resolution in the test image must match that in the reference image yet the test image usually represents a much larger area and larger total number of points. If the image already exists in digital form, the resolution can be reduced by averaging the image to produce an unaliased image of the appropriate number of points. If an image is blurred or averaged to reduce the highest spatial frequency, the detail above that maximum frequency is lost. That is, all frequency components above the maximum are zero and lost. Sampling the image properly (Nyquist criteria) permits the perfect reconstruction of the averaged image, not the original image.

It is worthwhile to define the concept of space-bandwidth product (SBP) here. The bandwidth of an image is the width of the spatial frequency content to the highest spectral component. The space is the physical length over which the image exists. For example, a piece of film may have a maximum resolution of 100 points/mm with an image which occupies 1 cm along the length of the film. In this case the SBP is 100 points/mm X 10 mm = 1000 points. This is in one dimension. For a square image, the number of points is 1,000,000. The SBP is the number of resolution points in an image. The maximum SBP capability of the film may not be utilized by an image recorded on the film, and the actual SBP of the stored image will depend on the image itself. In general, the bandwidth will be determined by the finest detail in the image and the area of the total image. The area of the smallest detail divided into the total image

area defines the SBP. When a continuous image is sampled at the Nyquist rate, one sample per resolution point in the image is required. Thus, the SBP of the image sampled at the Nyquist rate matches that of the continuous image. The SBP in the sampled image is a very practical detail because each sample must be stored in the computer memory. The number of resolution elements in a 4- by 5-inches holographic film may exceed 108. A computer cannot practically store such a large number of values. With a limited number of memory locations on the computer, the sampling rate and SBP demand careful consideration.

A CGH is created using a digitized image. A continuous film image may be sampled and quantized to create a non-aliased digital image. Some imaging sensors output data in digital format with no further digitizing required. Once the digital image is obtained, the image values may be manipulated on a digital computer. If this digital image is encoded on a continuous-tone CGH using equation (35) as a model, a spatial carrier frequency on the Fourier transform of the image must be induced. The image is encoded as $f(m\Delta x, n\Delta y)$ with a SBP of M x N where M and N are the number of points in the image in each direction. If the Fast Fourier Transform (FFT) is applied to the image, a digital representation of the Fourier transform of the image is obtained. This transformed image $F(m\Delta u, n\Delta v)$ contains the same number of points as the image and obviously the same SBP. If the image contained M points along the x direction, the highest spatial frequency possible in this image would be M/2 cycles/frame. This situation would exist when the pixels alternated between 0 and 1 at every pixel. That is, the image consisted of {0,1,0,1, ...}. The maximum frequency in the transform is M/2 cycles/frame in both the positive and negative directions. The FFT algorithm provides the real and imaginary weights of each frequency component ranging from -M/2+1 cycles/frame to +M/2 cycles/ frame in one cycle/frame steps. This provides M points in the u direction. The same is true for N points in the v direction. Thus, the first point in the FFT matrix is (-M/2+1,-N/2+1), the D.C. term is in the M/2 column and N/2 row, and the last term in the FFT matrix is (M/2,N/2).

It is useful to point out that the FFT describes the frequency components of the image $f(x,y)$. The FFT pattern also contains structure which can also be represented by a Fourier series. That is, the FFT pattern or image has speci- fic frequency components. Because the image and the FFT are Fourier transform pairs, the image describes the frequencies in the FFT pattern. For example, a

spike in the image implies the FFT will be sinusoidal. A spike in the FFT implies the image is sinusoidal. The existence of a non-zero value on the outer edge of the image implies the FFT contains a frequency component at the maximum frequency. A non-zero value on the corner of the image implies the maximum frequency exits in the FFT pattern which is M/2 in the x direction and N/2 in the y direction.

To record the complex Fourier transform as a hologram, the function $F(m\Delta u, n\Delta v)$ must be heterodyned to a spatial carrier frequency so as to create a real non-negative pattern to record on film. To prevent aliasing, the heterodyne frequency must be sufficiently high. The frequency components in the hologram are shown in Figure 6 and consist of the D.C. spike, the power spectral density of the function $F(u,v)$, and the two heterodyned terms. To record the function $F(u,v)$ on film without distortion from aliasing, the spatial carrier frequency must be 3 times the highest frequency component of the FFT pattern. This permits the power spectral density term to exist adjacent to the heterodyned terms with no overlap. The frequencies in the hologram extend to plus and minus 4B. Thus, the hologram has a space-bandwidth product 4 times larger than the original image in the heterodyne direction. When heterodyned in the v direction as implied by equation (35), the resulting hologram matrix must be larger than the original image by 4 times in the v direction and 2 times in the u direction. The spectral content in two dimensions is shown in Figure 7. The space-bandwidth product is very large for this CGH to record the information in $H(u,v)$.

The requirement is even greater when the hologram is to be used as a Vander Lugt filter. When used as a Vander Lugt filter, the CGH must diffract the light sufficiently away from the origin and the additional on-axis terms to prevent aliasing in the correlation plane.

The output of the Vander Lugt filter is shown in equation (37) and the spectral contents are plotted in Figure 5. These spectral components are shown in two dimensions in Figure 8. Here the space-bandwidth product is 7 times larger than the image in the v direction and 3 times larger than the image in the u direction. To produce a correlation image without stretching, the samples in the u and v directions should have the same spacing. Usually for convenience, the hologram contains the same number of points in both directions, giving
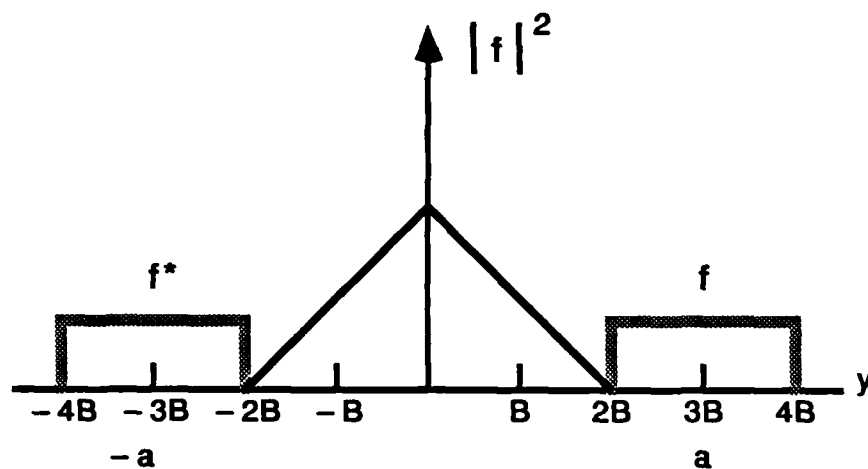
Figure 6. Special Content of a Fourier Transform Hologram
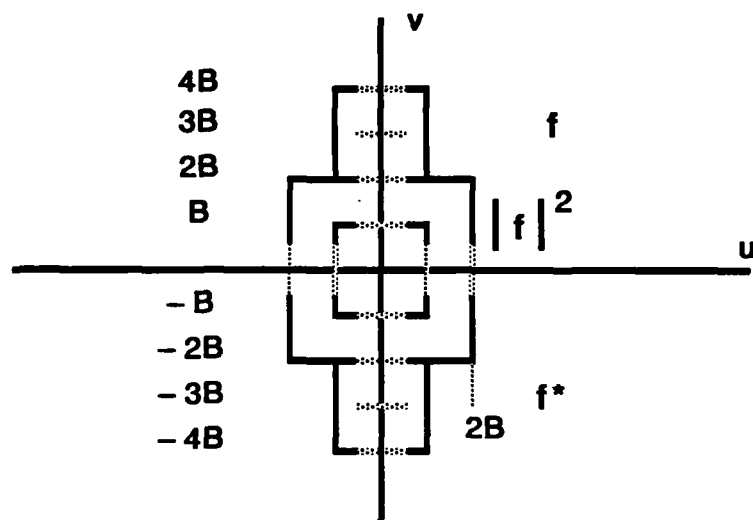
Figure 7.   Two-Dimensional Spectrum of the Fourier Transform Hologram
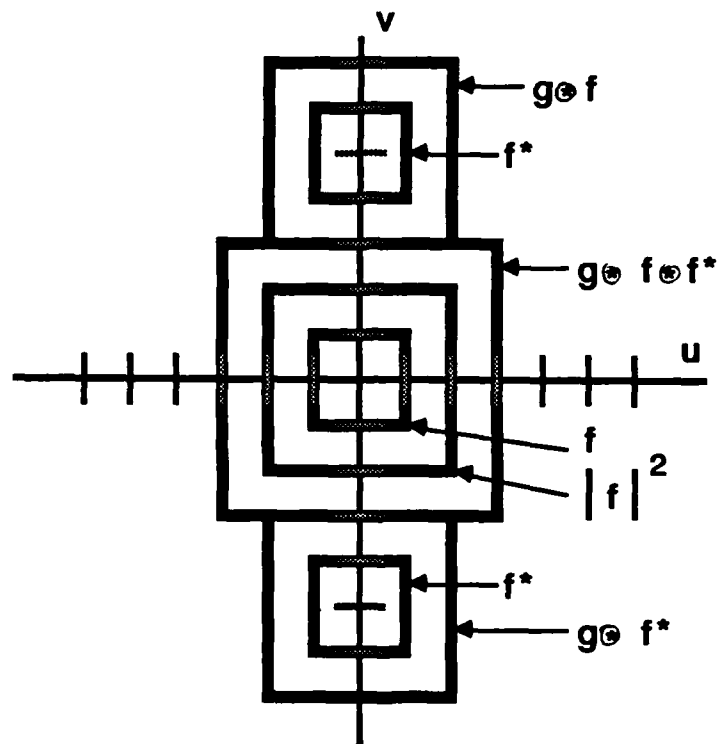
Figure 8. Two-Dimensional Spectrum of the Vander Lugt Filter

a pattern which is 7B by 7B. The FFT algorithm used on most computers requires the number of points to be a power of 2. This requires that the hologram be 8B by 8B. For example, if the original images to be correlated contain 128 by 128 points, the required continuous-tone CGH contains 1024 by 1024 points. In a binary hologram, each continuous tone point or cell may require many binary points to record the entire dynamic range of the image.

This illuminates the key problem with CGH-matched filters. The space-bandwidth product becomes large for even small images. Yet it is the ability of optical processors to handle large images with many points that makes them so attractive. Holograms created with interferometric techniques contain a large amount of information or a large space-bandwidth product. However, these optically-generated holograms lack the flexibility offered by CGH. Holographic filters are produced by either optical or computer prior to their actual use. The filter imparts its required transfer function to the test image without any further computation of the hologram pattern. Even if the task is difficult, production of the filter is a one-time job. The more information stored on the hologram, the greater the potential processing capability of the Vander Lugt filter. To produce powerful yet practical CGH filters, the space-bandwidth product and dynamic range of the hologram must be understood and minimized within design criteria.

One key to reducing the space-bandwidth product of the CGH is to recognize that much of the spectrum is not useful information. The terms in Figure 5 are described as the convolution of f and g, the baseband terms $f \otimes f^* \otimes g$, and the correlation of f and g. Only the correlation term is useful for our purposes in the Vander Lugt filter, but the other terms arrive as a by-product of the square law nature of the film. The two heterodyned terms which result in the convolution and correlation of f and g must come as a pair. That is, when the real part of the heterodyned information is recorded, the plus and minus frequencies exist. The real part, cos θ, can be written as $\exp(j\theta) + \exp(-j\theta)$ using Euler's formula. The plus and minus exponents give rise to the plus and minus frequency terms which become the convolution and correlation terms. The convolution and correlation terms are always present in a spatially modulated hologram.

A more efficient hologram is produced using equation (42). This hologram consists of a D.C. term sufficiently large to produce only non-negative values and the heterodyned terms.

$$H(u,v) = A^2 + F(u,v)e^{j2\pi av} + F^*(u,v)e^{-j2\pi av}. \qquad (54)$$

The output (shown in Figure 9) of the Vander Lugt filter using this hologram is

$$O(u,v) = A^2G(u,v) + F(u,v)G(u,v)e^{j2\pi av} + F^*(u,v)G(u,v)e^{-j2\pi av} \qquad (55)$$

or

$$o(x,y) = A^2g(x,y) + f(x,y) \circledast g(x,y) \circledast \delta(x,y+a) + f(-x,-y)g(x,y) \circledast \delta(x,y-a)$$

$$= A^2g(x,y) + f(x,y) \circledast g(x,y) \circledast \delta(x,y+a) + R_{fg}(x,y) \circledast \delta(x,y-a) \qquad (56)$$

which gives the spectrum shown in Figure 9 assuming $B_f=B_g=B$. Here the spectrum extends to 5B rather than 7B and considerable space saving is possible. However, the 5B is not a power of 2 and most computer systems would still be forced to employ 8B points. The terms in Figure 9 are the convolution term, the image term, and the correlation term. The image term arises from the product of the D.C. term with the test image $g(x,y)$. In a normal absorption hologram, it is not possible to eliminate the D.C. term. The image term takes up the space from -B to B, forcing the spatial carrier frequency to 3B and requiring 5B total space. If the absorption hologram is replaced with a bleached hologram where the phase varies across the hologram, the D.C. term may be eliminated.

As discussed in Section II, film may be bleached to produce a phase modulation. This is accomplished at the expense of the amplitude modulation. However, this phase hologram behaves much like the original amplitude or absorption hologram. One advantage of the bleaching process and the use of phase modulation is the opportunity to eliminate the D.C. term (set it to zero) and reduce the space-bandwidth product. Equation (54) is changed to

$$H(u,v) = F'(u,v)e^{j2\pi av} + F^*{}'(u,v)e^{-j2\pi av} \qquad (57)$$

45
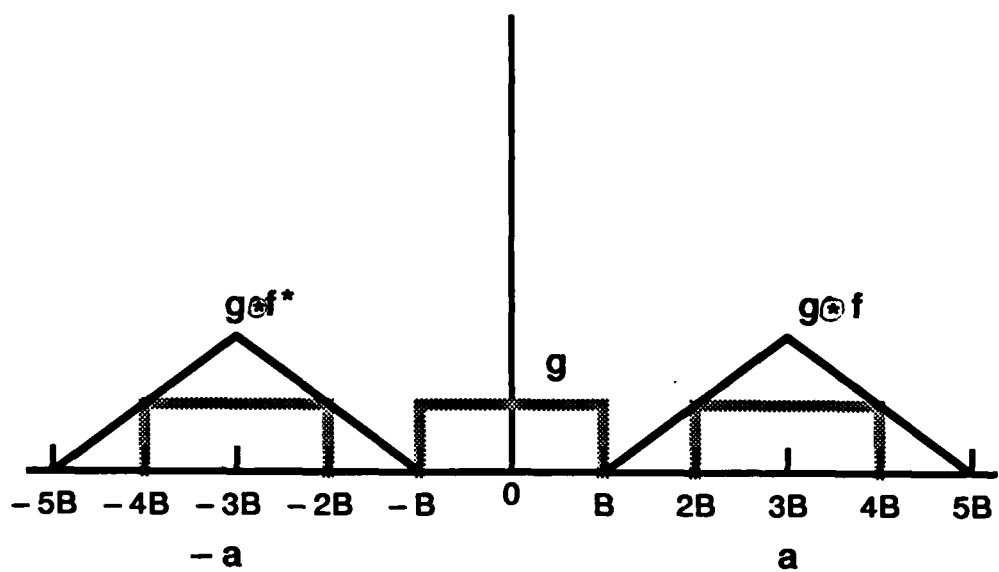
Figure 9. Spectrum of a Modified Vander Lugt Filter

where the prime mark (') indicates the function has been modified by the bleaching process. There is no D.C. term, so the output of the Vander Lugt filter is

$$O(u,v) = F'(u,v)G(u,v)e^{j2\pi av} + F^{*'}(u,v)G(u,v)e^{-j2\pi av} \qquad (58)$$

or

$$o(x,y) = f'(x,y) \otimes g(x,y) \otimes \delta(x,y+a) + f'(-x,-y) \otimes g(x,y) \otimes \delta(x,y-a) \quad (59)$$

$$= f'(x,y) \otimes g(x,y) \otimes \delta(x,y+a) + R_{f'g}(x,y) \otimes \delta(x,y-a)$$

which gives the spectrum shown in Figure 10, assuming $B_f = B_g = B$.

This phase hologram reduces the number of points to 4B, a power of 2. This is the smallest possible size in a spatially modulated hologram. As will be shown later, the phase modulation process may significantly affect the information, and the correlation obtained may be a poor approximation to the ideal correlation.

The Vander Lugt filter is typically used to detect the presence of a small object in a large scene. This implies that B_f may be much smaller than B_g. In any case, the least theoretical hologram size using equation (57) is still twice the size of the reference image and test image combined in the y direction. For example, a large scene consisting of 1024 by 1024 points is to be searched for an object that would occupy 32 by 32 points in that scene. The smallest continuous-tone hologram to perform that correlation would contain 2112 points in the y direction (at least 1088 in the x direction). For most practical applications, the absorption hologram illustrated in Figure 9 would be used. For the same example consisting of a 1024 by 1024 test scene and a 32 by 32 reference image, a square hologram would be at least 2144 by 2144.

Another practical consideration provides some relief in the size of the correlation plane. The correlation of two images creates a correlation image whose size is the sum of the individual image sizes. Non-zero correlation values can exist when any points in the two images overlap. However, the number of points which overlap becomes very small near the outer edge of the

47
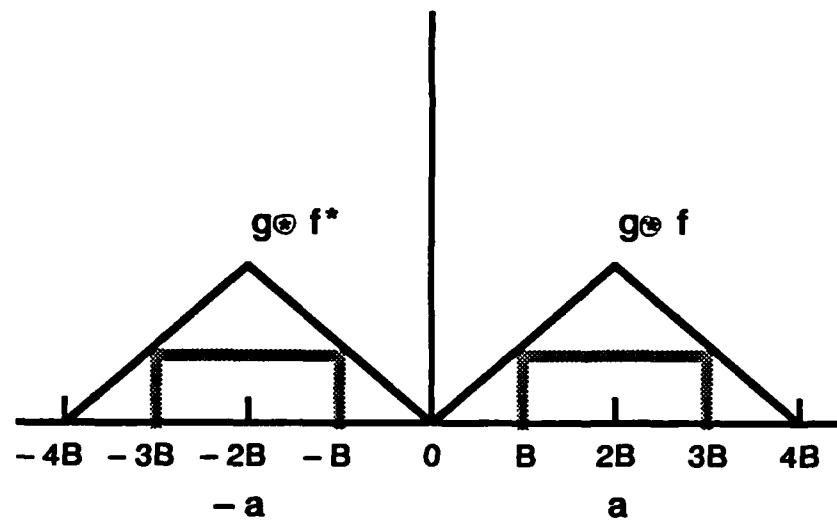
Figure 10. Spectrum of the Zero Mean Vander Lugt Filter

correlation plane. In a practical system, a threshold is set to determine correlations which are targets (above threshold) or background (below background). When the target fills the reference image and is entirely present in the test image, the autocorrelation condition exists and the correlation can be normalized to one. When the target begins to fall off the edge of the test image, correlations will still occur. However, the correlation value will fall from unity by the ratio of the target area present in the test image to the target area in the reference image. A practical rule of thumb might be to ignore the correlations when half of the target falls outside the test image in any direction. This reduces the correlation plane to the size of the test image, offering some relief to the required hologram size. If the outer edge of the correlation plane is ignored, it does not matter if that edge is aliased. This reduces the sampling and heterodyning requirements in the filter hologram especially when the reference contains many points. When using the absorption hologram with 50 percent aliasing (shown in Figure 11), the spatial frequency is

$$a = B_g + B_f. \tag{60}$$

and the number of points in the hologram in the v direction (SBP$_v$) is

$$SBP_v = 2B_g + 3/2 \, B_f. \tag{61}$$

Phase encoding this hologram does not relieve the requirement on the carrier frequency or the total number of points. The edges of the correlation plane will fall into the active correlation region if a or SBP$_v$ is reduced from the values given in equation (60) and (61).

In review, the SBP of the hologram is determined by the following criteria.

a. The required resolution in the reference scene to recognize the desired target.

b. The size of the reference scene. This is not normally a significant factor due to the small size of the reference compared to the size of the test image.

49
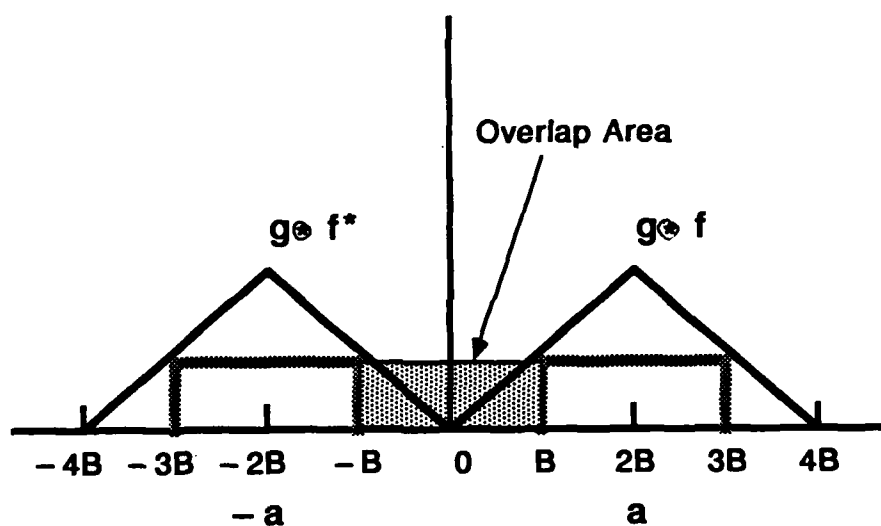
Figure 11. Output of a 50-Percent Aliased Vander Lugt Filter with Absorption Hologram

c. The size of the test scene. The potential advantage of optical processing is to test a large scene for the presence of the reference object. The test image must contain the same resolution as the reference image but includes many times the image area. Thus, the SBP of the test scene is very large and is the driving factor in the size of the CGH-matched filter.

d. Usually, aliasing can be tolerated at the edges. This depends on the threshold and expected intensity of false targets. When 50 percent imposed aliasing can be tolerated, the SBP reduces to an even multiple of two.

e. The dynamic range in the reference scene. The hologram must adequately represent the dynamic range in the reference scene. In the case of binary olograms, many binary points may be required for adequate representation of each hologram cell.

f. Hologram type. The type of CGH produced determines the encoding scheme and number of points required to represent the SBP and dynamic range of the reference while preventing aliasing of the active correlation plane.

g. Incorporate D.C. elimination when possible to minimize on-axis terms.

By following these guidelines it is possible to determine the minimum possible SBP needed in the CGH.

SECTION IV
OPTIMIZATION OF CGH-MATCHED FILTERS

The previous sections describe the basic design techniques employed to create CGH-matched filters. To determine the performance of these filters, specific criterion must be established.

1. PERFORMANCE CRITERIA

Because the matched filter is based on maximizing the signal-to-noise ratio, that criteria is reasonable to apply to the result of the CGH. The matched filter created as a result of a CGH is only an approximation of the ideal filter. The non-linearities of the film, along with the sampling, heterodyning, and quantizing of the CGH process, cause the correlation to be less than ideal. The noise is not just caused by background in the input image but also by artifacts from the hologram. The matched filter was intended to recognize a specific target in a clutter background, yet, in some cases, the target will vary in size and orientation. There is a tradeoff between using high resolution to discriminate against false targets and too much sensitivity for target size and orientation. When modifying the frequency content of the scene to best distinguish target from background, the signal-to-noise ratio may decrease from the ideal. Another important property of the optical matched filter is the efficiency or light throughput. In a practical system, the input image is illuminated by a laser of limited size and power. Typically the laser source could be an IR diode putting out 10 mW.[6] Even if the signal-to-noise ratio is large, the energy reaching the correlation plane may be too small to measure. The efficiency of the hologram, the ratio of the power in the correlation to the power in the input test image, is an important criterion in evaluating a practical CGH-matched filter. Mathematically, it is given as

$$\eta_H = \frac{\int\int |g(x,y) \circledast f^*(x,y)|^2 dx\, dy}{\int\int |g(x,y)|^2 dx dy} \tag{62}$$

where η_H has been coined the Horner efficiency,[28] f is the reference scene, g is the test scene, and $\circledast$ denotes an ideal correlation. The correlation derived

52

from a Vander Lugt-matched filter is not ideal. To determine the Horner efficiency for a CGH-matched filter, equation (62) must include an accurate model of the encoding scheme. This efficiency can be measured experimentally using a known input source and calibrated detectors. Caulfield (Reference 28) estimated that efficiencies for certain matched filters could be as low as 10^{-6}. Butler and Riggins (Reference 29) used models of CGH filters to verify Caulfield's prediction and went on to recommend techniques for improving the efficiency.

The matched filter is used to determine the presence of a target in a large scene. A test scene is correlated with a reference, and the correlation plane is thresholded to indicate the target location. Occasionally, the Vander Lugt filter will generate correlation values above the threshold in areas where no target exists. Accordingly, the correlation of an actual target corrupted by noise may be lower than the threshold. Due to the presence of noise, random and otherwise, the performance of the filter must be measured in terms of the probability of detection and the probability of false alarm. The probability of detection, P_d, is defined as the probability that a target will be indicated when there is a target to be detected. The probability of false alarm, P_{fa}, is defined as the probability that a target will be indicated when there is no target to be detected. These two quantities are correlated by the presence of noise. If the detection threshold at the correlation plane is lowered, the probability of detection is increased, but the probability of false alarm is also increased. As with the efficiency measurements, determining P_d and P_{fa} for CGH-matched filters requires accurate models or optical experiments.

Historically, efficiency was not a concern in laboratory experiments because powerful lasers were available to overcome the hologram loss. When attempts are made to improve the efficiency, the signal to noise ratio may suffer. An efficient hologram is impractical if the signal-to-noise ratio in the correlation plane is so low that P_d goes down and P_{fa} goes up significantly. The performance of matched filters are typically measured in terms of the P_d and P_{fa}, but testing requires modeling the entire system and providing realistic images. All of these measures must be considered for the cases when the test target deviates from the reference. Optimization criteria for optical matched filters depend on the application. To improve the matched filter,

modifications to the filter design have been proposed. These modifications fall into areas of frequency modification, phase filtering, and phase modulation.

2. FREQUENCY EMPHASIS

High frequencies in an image correspond to the small details. Most images contain large continuous areas bounded by sharp edges. The large continuous areas contribute to the D.C. and low frequency content of the image, while the edges contribute to the high frequencies. If the high frequencies are removed from the image through spatial filtering, the sharp edges disappear, the large continuous areas blend together smoothly, and the resultant image appears soft or blurred. A low-pass image may not provide sufficient resolution to discriminate between two similar objects. If the low frequencies are removed from an image, the continuous areas become dark with only the edges remaining. The image appears sharp with well-defined edges and detail. This high-pass image provides, to the human eye, the same or better discrimination of the original image. That is, objects are identified and distinguished at least as well as in the original image. For example, images containing a bright square area and bright circular area are easily distinguished as a square and circle. If the high frequencies are removed, both square and circle appear as blobs with no distinct edges. However, if the low frequencies are removed, the bright area in the center of the square and circle disappears leaving only a bright edge. Yet these bright edges clearly indicate a square and a circle as shown in Figure 4. Even if the square is not filled in, the edge clearly denotes the square shape. The edge of the circular area still defines a circle. The square and circle are easily distinguished in the high-pass images. The information that distinguishes the square from the circle is contained in the high frequencies.

The traditional matched filter, as outlined in Section II, is created from the complex conjugate of the Fourier transform of the reference image. Filtering with such a filter is equivalent to correlating the reference image with a test image. Because most scenes contain large continuous areas with edges, they contain a large D.C. and low frequency component. Most images have spectra where the magnitude tends to drop off drastically with increasing frequency. The energy in the low frequencies may be several orders of
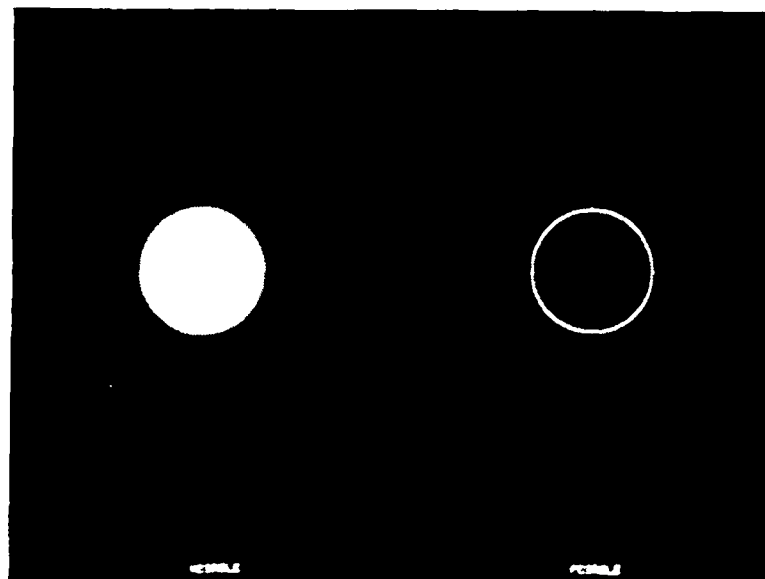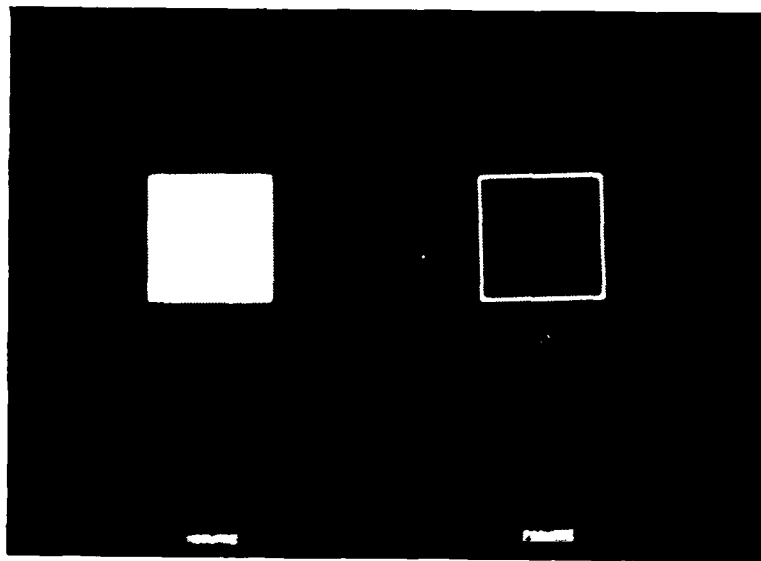
Figure 12. High- Frequency Emphasis of a Square and a Disk

magnitude larger than the high frequencies. However, it is the high frequencies which contain the useful information in separating the desired target from false targets. A practical problem with holography is the dynamic range to be recorded. Film cannot typically induce more than two or three orders of magnitude of dynamic range. To record a hologram of the Fourier transform, the film must accurately record the entire dynamic range of the transformed image. If the dynamic range of the transformed image is too large, the film cannot record the Fourier transform linearly and the correlation is not ideal. The film non-linearity will emphasize some frequencies and attenuate others. The correlation signal-to-noise ratio will suffer if important frequency components are attenuated. To reduce the dynamic range of the transformed image and allow linear recording on the hologram, the useless frequencies in the image should be eliminated. Because the low frequencies contain most of the image energy but little of the information, their omission considerably reduces the dynamic range with little effect on the correlation except to reduce the overall light through the hologram.

To determine which frequencies are important in target discrimination involves considerably more work than can be considered here. In general, a set of target images and a set of non-target images can be compared on a large digital computer to determine which frequencies appear most in the desired target. This requires a large data base of true and false targets. Filtered images are correlated and cross correlated to determine the most discriminating frequencies. In practice, this process is too time consuming. Certain assumptions are reasonable in spatial filtering. It is reasonable to assume that the reference and test images do not have much more detail than is absolutely necessary to distinguish the true target. To reduce the number of points needed in the digital imagery, the original sampling process was accomplished by limiting the spatial frequencies to those required to recognize the target. Thus, the appropriate filter to eliminate unnecessary frequency components will have the form of a high-pass filter. The nature of this high-pass filter is dependent on the application of the matched filter.

The matched filter is created for a specific target. If the target is present, the correlation is larger than for areas of the image where the target is absent. If the target changes slightly from the reference stored on the filter, the correlation drops. In a practical application, small changes in

56

the expected target are the rule rather than the exception. If the target grows in size, rotates, or changes its appearance slightly, the correlation may drop below the threshold. This topic will be discussed further in Section V, but it is necessary to point out that the invariance of the filter to small changes in the target depends heavily on the frequencies used in the correlation. Using the previous example, recall that the high-pass images showing the edges allowed discrimination between the square and circle. If the square were rotated slightly, the results would change. The cross-correlation between a square and a slightly rotated square depends on the frequencies used in the correlation. If only low frequencies are used, considerable rotation can occur with little effect on correlation. If high frequencies are used, the cross-correlation drops quickly with rotation. Thus, a matched filter created from a high-pass image to discriminate against out-of-class targets will not correlate well on in-class targets with small changes. That is, as more high frequency emphasis is applied to the matched filter, the discrimination sensitivity is increased. The probability of false alarm is increased, but the probability of detection drops. The high frequency emphasis is then tied to the P_d and P_{fa} which must be specified for a particular application.

There is another advantage to the frequency emphasis of matched filters. As seen in equation (35), the transmission of the hologram at each point depends on the magnitude of the reference image Fourier transform. Yet the hologram transmission cannot be greater than 1. Depending on the dynamic range of the film, the transmission out at the edge of the hologram corresponding to the high frequencies is very low or zero. As the magnitude drops off for high frequencies, so does the transmission of light through the holographic filter, and hence, filter efficiency is low. However, if the high frequencies are emphasized (boosted), the transmission at those points in a positive hologram is likewise emphasized. This creates an overall increase in the hologram transmission. In an absorption hologram, the light which is not transmitted is absorbed and lost to the system. The throughput or efficiency is highly dependent on the total transmission of the hologram. Thus, by emphasizing the high frequencies, the efficiency of the Vander Lugt filter is increased. Because the maximum transmission is limited to 1 and the dynamic range is limited on the film, the greatest efficiency occurs when most of the frequencies have

equal weighting and the transmission is close to 1 across the entire hologram. This implies that the throughput of the hologram will be largest when the image transform is nearly white.

The following procedures determine the choice of frequency emphasis.

a. Specify the P_d and P_{fa} for the particular application.

b. Choose a high-pass emphasis which satisfies the P_d and P_{fa} requirements. Typical choices include gradient, exponential, and step filters.

c. Because the test image should be filtered in the same fashion as the reference image, the frequency emphasis chosen should be squared before inclusion in the hologram. This permits the pre-emphasis of the test image without a separate stage of spatial filtering. That is, the test image is spatially filtered for pre-emphasis with the same hologram providing the correlation.

d. The test image is typically much larger than the reference image and can thus contain frequencies lower than any contained in the reference. Since those frequency components can never contribute to correlations, all frequencies below the lowest useful frequency in the reference should be truncated to the value of the next smaller term.

e. The frequency emphasis (squared) greatly reduces the dynamic range of most scenes, simplifying the coding of the CGH-matched filter and greatly improving the efficiency. The frequency-emphasized CGH matched filter is created, as shown in Section III, but utilizes a reference image whose frequency content is modified.

$$F'(u,v) = |P(u,v)|^2 F(u,v) \qquad (63)$$

where F' is the modified image transform,

F is the original image transform,

and P(u,v) is the frequency emphasis chosen.

58

3. PHASE-ONLY FILTERS

The preceding section describes techniques in which the high frequencies
are emphasized. This emphasis usually improves the discrimination against
false targets and increases hologram efficiency. Frequency emphasis involves
the multiplication of the image transform by a filter function which attenuates
or amplifies the appropriate frequency components. The filter function adjusts
the spectral magnitude of the image. In the Fourier representation of images,
spectral magnitude and phase tend to play different roles and, in some situa-
tions, many of the important features of a signal are preserved even when only
the phase is retained. Oppenheim (Reference 15) showed that when the magnitude
portion of an image Fourier Transform is set to an arbitrary constant and the
phase left intact, the reconstructed image closely resembles the original.
Features of an image are clearly identifiable in a phase-only image but not in
a magnitude-only image. Statistical arguments by Tescher (Reference 30) and by
Pearlman, and Gray (Reference 31) have been applied to real-part, imaginary-
part, and magnitude-phase encoding of the discrete Fourier transform of random
sequences. They conclude that, for equivalent distortion, the phase angle must
be encoded with 1.37 bits more than the magnitude. Kermisch (Reference 32)
analyzed image reconstructions from kinoforms, a phase-only hologram. He
developed an expansion of the phase-only reconstructed image $I(x,y)$ in the form

$$I(x,y) = A \ [I_0'(x,y) + 1/8 \ I_0'(x,y) \circledast R_0'(x,y)$$

$$+ \ 3/64 \ I_0'(x,y) \circledast R_0'(x,y) \circledast R_0'(x,y) + \ . \ . \ .] \tag{64}$$

where $I_0'(x,y)$ is the normalized irradiance of the original object, $R_0'(x,y)$ is
the two-dimensional autocorrelation function of $I_0'(x,y)$ and $\circledast$ denotes convolu-
tion. The first term represents the desired image, and the higher terms repre-
sent the degradation. Kermisch showed that the first term contributed 78
percent to the total radiance in the image, giving a ratio of 1.8 bits.

The phase-only image typically emphasizes the edges as in the case of the
high-pass filtering as shown in Figure 12. This phase-only filtering is
closely related to the high-pass filter. Most images have spectra where the
magnitude tends to drop off with frequency. In the phase-only image, the

magnitude of each frequency component is set to unity. This implies multi-plying each pixel magnitude by its reciprocal. The Fourier transform tends to fall off at high frequencies for most images, giving a mound-shaped transform. Thus, the phase-only process applied to a mound-shaped Fourier Transform is high-pass filtering. The phase-only image has a high-frequency emphasis which accentuates edges. The processing to obtain the phase-only image is highly non-linear. Although the response $1/|F(u,v)|$ generally emphasizes high frequencies over low frequencies, it will have spectral details associated with it which could affect or obliterate important features in the original. Oppen-heim (Reference 15) proposed that if the Fourier transform is sufficiently smooth, then intelligibility will be retained in the phase-only reconstruction. That is, if the transform magnitude is smooth and falls off at high frequen-cies, then the principal effect of the whitening process is to emphasize the high frequencies and therefore the edges in the image, thereby retaining many of the recognizable features. In Figures 12 and 13 the phase-only filter emphasizes edges more strongly than a gradient filter for the examples shown.

The advantage of using a phase-only image or high-pass image is the increase in optical efficiency of the resultant matched filter. As shown in equation (35), the transmission of each hologram element depends on the magni-tude of the reference image Fourier transform. As the magnitude drops off for high frequencies, so does the transmission of light through the holographic filter, and hence filter efficiency is low. If the magnitude is set to unity (phase-only filter) for all frequencies, the overall efficiency increases dramatically. The image transform is white and thus the throughput of the absorption hologram is highest. Horner (Reference 14) shows that the maximum throughput efficiency of an ideal autocorrelation of a 2-D rect function is only 44 percent, while the autocorrelation using a phase-only filter achieves 100 percent efficiency.

The phase function, $o(u,v)$ of an image Fourier transform is a continuous function. To fabricate a phase-only filter for such an image requires a linear process capable of faithfully reproducing the whole range of values from 0 to 2. If the phase is quantized so as to permit only two values, typically 0 and pi, such a filter is known as a bi-phase filter.
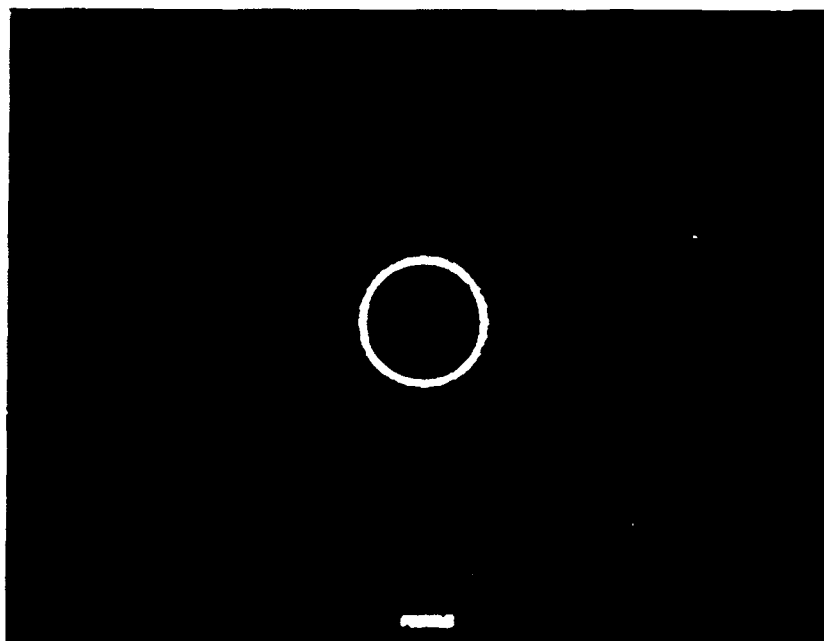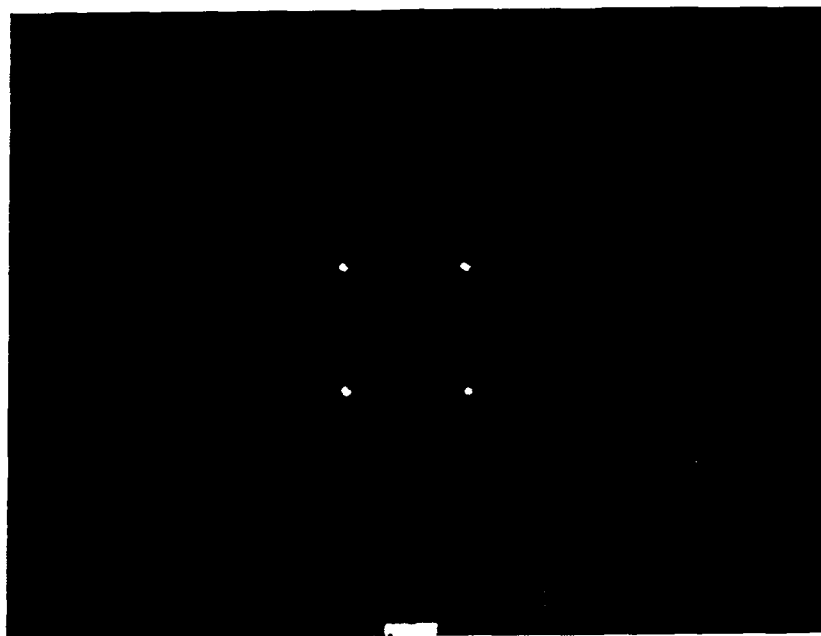
Figure 13.   Phase-Only Filtering of a Square and a Disk

$$H'(u,v) = \text{sgn}\ [\cos\ \phi(u,v)] = +1 \qquad \text{if Re}\ [H(u,v)] > 0 \quad (65)$$
$$= -1 \qquad \text{otherwise}$$

where $H(u,v)$ is the Fourier transform of the filter impulse response $h(x,y)$, the sgn operator gives the sign of the argument, and $H'(u,v)$ is the bi-phase transform. This bi-phase information is an approximation to the phase-only information. In many cases, reconstructions from this bi-phase information contain the same detail as the ideal amplitude and phase information. This would indicate that much of the information in an image is contained in the sign of each pixel or where the zero-crossings occur.

In converting a complex wave, which contains continuous magnitude and phase values, to binary values, much is thrown away. If the reconstructions from the binary image transforms are similar to the original image, then the bi-phase conversion reduces redundancy and eliminates superfluous dynamic range. When this is accomplished in an optical correlator without significant reduction in signal-to-noise ratio, the CGH-matched filter is greatly simplified. Most important is the ability to use binary light modulators. A number of electronic spatial light modulators are commercially available. Of these modulators, several can be used to phase-modulate a light wave. These include deformable paddles, liquid crystals, and magneto-optical modulators. These can be used as bi-polar phase modulators. (Reference 33). If the information in the reference image can be accurately represented using only bi-phase information, binary phase modulators can be used as real-time holographic filters. The ability to adapt the matched filter in real time permits scanning the test image for various targets with varying sizes and orientations. This technique is very efficient because the light is phase shifted and not attenuated.

4. PHASE-MODULATION MATERIALS

Recall that spatially modulated holograms are needed for matched filtering only because film cannot record a complex wavefront. Film can record only real values. Film may be used to record, at baseband, the magnitude of a wavefront, or it may be computer-encoded and phase-modulated (bleached) to record the phase of a wavefront. Thus, without using a spatially modulated hologram, the magnitude or phase may be recorded. If only the phase information of the image is needed to represent the reference image, a baseband hologram which records

the phase portion of the image transform can be used in the optical correlator. This on-axis phase hologram, or kinoform, is recorded as a relief pattern in which appropriate phase delays are induced in the illuminating wavefront. To produce a Fourier transform kinoform, the phase is restricted to a range from - pi to + pi. The arctangent of the ratio of the imaginary and real parts yields such a result. The film is exposed to a pattern, whose intensity is proportional to the desired phase, and bleached to create a relief pattern. (Reference 34) These kinoforms cannot record the amplitude variation of the image transform and thus, the filter formed is a phase-only filter.

Several techniques have been proposed by which the phase could be modified to introduce amplitude variation in the reconstructed wavefront. (References 35 and 36) Chu, Fienup, and Goodman (Reference 18) used multiple layers of film to represent both the phase and amplitude variation. Kodachrome II, for color photography, contains three emulsions. The phase variation was recorded on one emulsion and the amplitude on another. The inventors named this technique Referenceless On-Axis Complex Hologram (ROACH). To introduce amplitude variation to the reconstructed wavefront, light must be removed from the wavefront, resulting in a reduction in efficiency.

The reconstruction from the kinoform is formed on-axis and is a phase-only image. When the phase values are uniformly distributed between - pi and + pi, the D.C. or average term is zero. However, if the phase recording is unbalanced or the phase distribution is not uniform, a D.C. term will exist in the hologram. When used as a matched filter, the kinoform must be carefully phase-balanced to prevent a large D.C. spike from occurring in the correlation plane. Such a spike would be indistinguishable from an actual correlation. If the phase hologram is produced using a real time holographic device, the phase might be controlled using a feedback loop to eliminate the D.C. term prior to correlation. To produce a permanent hologram on film, the exposure and bleaching processes must be carefully controlled.

Bleaching includes several processes which produce a phase hologram from an exposed absorption hologram. The bleached hologram retards the wavefront, causing a phase shift instead of attenuation. The result is generally an increase in diffraction efficiency but often with an accompanying decrease of signal-to-noise ratio. (Reference 37) There are three basic types of bleaches.

The direct or rehalogenizing method converts metallic silver back into silver halide which has a different index than that of the gelatin. Reversal or complementary bleaches dissolve the metallic silver from an unfixed hologram, leaving the undeveloped silver halide which varies the index of refraction. The third process creates a surface relief by shrinking the exposed portions of the hologram by removing the metallic silver. When the emulsion is bleached properly, the attenuation of the transparency can be reduced to the point that phase modulation due to index changes dominates any residual amplitude modulation. Phase modulators prove to be more efficient in terms of the portion of incident illumination that is diffracted to form the desired correlation. A sinusoidal hologram using absorption or amplitude modulation can theoretically diffract only 6.25 percent of the incident energy into an image. Experimentally, the number is about 4 percent (Reference 38). A phase-modulated hologram transmits all of the light (ignoring the emulsion, substrate, and reflection losses). A sinusoidal phase hologram can diffract as much as 33.9 percent of the incident light into the first order.

The bleaching process converts the real function F(u,v), recorded in silver on the film, to a phase delay.

$$H(u,v) = \exp j[\ F(u,v)\ ] \qquad (66)$$

To produce a kinoform, the film is exposed to the phase function $\theta(u,v)$ of the image transform. Upon subsequent bleaching, the film contains the response

$$H(u,v) = \exp j[\ \theta(u,v)\ ]. \qquad (67)$$

The kinoform, produced in this fashion, records the phase-only information of the image transform. The bleaching process is not restricted to phase-only information. Rather, the absorption hologram created from equation (35) can also be bleached.

$$H'(u,v) = \exp j[\ H(u,v)\ ] \qquad (68)$$

$$= \exp j[1 + |F(u,v)|^2 + F(u,v)\exp j2\pi av + F^*(u,v)\exp -j2\pi av]$$

where H'(u,v) is the bleached hologram response. The phase-only information and the phase modulation obtained through bleaching are entirely independent of one another. That is, a phase-modulated hologram can be created from an image whose amplitude and phase are intact or from an image whose amplitude or phase are modified or removed. Considerable confusion continues to exist in the literature in which a phase modulation process seems to imply, by default, phase-only information. Cathey attempted to clarify this confusion in 1970 by defining specific terms for each case. (Reference 39). The holographic process, which is independent of the recorded information, was described as (a) phase holography when only phase modulation was present, (b) amplitude holography when only amplitude modulation was present, and (c) complex holography when both amplitude and phase modulation were present. In an equivalent fashion, the information to be recorded on the hologram can be described as phase-only information or amplitude-only information when either the amplitude or phase portion of the complex waveform are deleted. Thus, an amplitude hologram can be created from phase-only information.

When an amplitude hologram is bleached, the density at each point on the film is mapped to a phase delay. This mapping is linear when the bleaching chemistry is correct. This new phase function on the film is related to the original pattern on the film.

$$H(u,v) = \exp j\{F(u,v)\} \qquad (69)$$

where H(u,v) is the complex function on the film after bleaching and F(u,v) was the original transmission pattern recorded on the film. The exponential expression in equation (66) can be expanded with a series expression.

$$H(u,v) = 1 + jF(u,v) - (1/2)F^2(u,v) - j(1/6)F^3(u,v) + \ldots$$

$$= \sum \frac{[jF(u,v)]^n}{n!} \qquad (70)$$

When reconstructed, this hologram can be expressed as a series of convolutions.

$$h(x,y) = (x,y) + jf(x,y) - (1/2)f(x,y) \circledast f(x,y) -$$

$$-j(1/6)f(x,y) \circledast f(x,y) \circledast f(x,y) + \ldots$$

$$= \sum \frac{j^n f^{(n)}(x,y)}{n!} \qquad (71)$$

where $f^{(n)}(x,y) = f(x,y) \circledast f(x,y) \circledast \ldots \circledast f(x,y)$ n convolutions

and $\qquad f^{(0)}(x,y) = \delta(x,y)$

$\qquad f^{(1)}(x,y) = f(x,y)$

$\qquad f^{(2)}(x,y) = f(x,y) \circledast f(x,y)$

and so on.

Thus, the phase modulation technique is very non-linear and the resultant reconstruction is rich with harmonics. The reconstruction from such a hologram is noisy due to the harmonic content. The higher order correlations are broader, thus contributing less flux into the reconstruction. Phase modulation in the form of bleached and dichromated gelatin holograms have become the rule in display holography due to the bright images. This fact indicates that the noise is acceptable in many cases. In fact, the reconstruction of such display holograms looks very good. Nevertheless, such an example is deceiving because the repeated convolutions and correlations of equation (71) become more detrimental for more complicated objects, especially if the object has low contrast. (Reference 32) The harmonics combine to produce intermodulation terms within the bandpass of the desired information, causing an increase in background noise. When used for matched filtering, the decision to use phase modulation is a balance between hologram efficiency and signal-to-noise ratio.

An interesting case occurs when a binary amplitude hologram is converted to a phase modulation hologram. The bleaching process maps an amplitude of zero and one to a phase shift of plus and minus pi. This equates to an amplitude of plus and minus one. For this binary mapping, the transfer function is 2x-1, which is a linear process. In that sense, the binary hologram is inherently

linear. The binary hologram represents the continuous-tone amplitude hologram by opening more or fewer binary windows. Through the use of many windows, the amplitude can be accurately represented by the appropriate combination of binary values. The subsequent bleaching of the binary hologram is a linear process and thus no additional harmonics are contributed. This provides a means by which high efficiency holograms may be produced without sacrificing signal-to-noise ratio due to non-linearity. A sufficient number of points is necessary in the binary hologram in order to minimize the non-linearity of the binary CGH mapping. When a computer and writing device are available to produce such binary holograms, subsequent bleaching or phase modulation greatly improves the efficiency without any adverse effect on signal-to-noise. This makes digital, phase-modulated holograms very attractive for matched filtering.

SECTION V
PATTERN RECOGNITION TECHNIQUES

Coherent optical correlators have been used as a means of performing 2-D pattern recognition (References 40 through 43). An optical correlator system could scan a large scene for the presence of a specific pattern. The input image is correlated with the impulse response of the matched filter to determine the presence and position of the reference pattern. Because the Fourier transform is shift invariant, equation (6), correlation can occur anywhere in the input image and multiple targets can be recognized simultaneously. However, other changes in the input pattern do affect the correlation function. Rotation, scale changes, and geometrical distortions due to viewing a 3-D scene from various angles can lead to correlation degradation and a corresponding loss in detectability (Reference 44). For example, to recognize a hammer in a box of tools, the reference must be capable of correlating with the hammer when it is laying in any orientation from 0 to 360 degrees. The hammer could lay on either side so that both orientations would need to be included in the reference image. If we were not sure of the distance from the camera to the hammer, we would not be sure of its size in the image.

The fundamental difficulty in achieving a practical recognition system lies in correlation of the reference image with a real-time image which differs in scale, aspect, contrast, and even content when sensed in a different spectral band or at a different time than the reference image. Matched filter pattern recognition systems, both optical and digital, tend to suffer from two types of difficulties. They tend to be too sensitive to differences likely to occur in the desired pattern. These differences are termed within-class variations. Second, they tend to be too insensitive to differences between real and false targets. These are between-class variations. While other deformations in the object condition are possible in specific applications, translation, rotation, and scale are the most common in pattern recognition whether it is accomplished optically or digitally.

1. DEFORMATION INVARIANT OPTICAL PATTERN RECOGNITION

The basic operation performed in an optical processor is a two-dimensional Fourier transform. Matched spatial filters are used to perform correlations

between an input image and a reference pattern. While the reference pattern may exist in the input image, it may be deformed by scale, rotation or geometrical distortion. The Fourier transform is invariant to shift in two dimensions, see equation (6). It is not invariant to scale or rotation, and a dramatic loss in signal-to-noise ratio (3 to 30 dB) occurs for small scale changes (2 percent) or rotation (3.5 degrees) (Reference 44).

In some applications it is desirable to give up translation or shift invariance in substitution for some other deformation invariance. The technique described by Casasent and Psaltis (Reference 45) involves a space variant coordinate transformation to convert the deformation under consideration to a shift in the new coordinate system. Because the optical processor performs two-dimensional transforms, it is insensitive to shifts in two dimensions. Thus, two separate invariances can be accommodated. Scale can be converted to a shift in one direction and the normal shift can be left in the other dimension. This would provide scale invariance, but the resultant correlation would only yield the location of the target in only one dimension (i.e. the x coordinate).

In another example, the scale can be converted to shift in one dimension and rotation converted to shift in another dimension. Such a two-dimensional optical correlator could provide correlations on rotated and scaled objects but would no longer predict the location of the object. The two-dimensional nature of the optical processor allows the correlator to be invariant to both deformations. In order to provide invariance to other deformations two at a time, a coordinate transformation is needed to convert that deformation to a coordinate shift. The Mellin Transform is an excellent example of such a transformation used to provide scale and rotation invariance.

The Fourier transform is invariant to translation shift in two dimensions. To provide invariance to other deformations, a coordinate transformation is needed to convert each deformation to a shift. To provide scale invariance a logarithmic transformation is used. The logarithmic transformation converts a multiplicative scale change to an additive shift. This shifted version will correlate with the logarithmically transformed reference pattern. To provide rotation invariance, a transformation is performed to map the angle to each point in the image to a theta coordinate. If an object rotates in the test

image, it is translated along the theta coordinate. Usually the two transformations are combined into the log r, theta transformation. The test image as well as the reference image is converted to polar form to provide r and theta values for each pixel. The image is transformed into a coordinate system where one axis is log r and the other axis is theta. In this system, scale changes shift the object along the log r axis and rotation shifts the object along the theta axis. Because this transform, known as the Mellin-Fourier transform, is itself not shift invariant, it is normally applied to the Fourier transform of the test image. This provides the shift invariance but loses the location information in the test scene. The cross correlation between the transformed test and reference images no longer can provide the location of the object but does determine the presence of the object, its size, and its rotation relative to the reference pattern.

To perform the Mellin-Fourier transform for shift, scale, and rotation invariance, the input image is first Fourier transformed and the resultant power spectral density recorded. This magnitude array is converted to polar coordinates and the linear radial frequency is converted to a logarithmic radial coordinate. The new coordinate space (log r,theta) is used for cross-correlation of the input image with similarly transformed reference images. A high speed technique is required to convert the image into log r, theta coordinates at a speed compatible with the optical processor. This has been demonstrated using holograms to perform geometrical transformations (References 46 through 50). To do this, the coordinate transforming hologram must influence the light from each point and introduce a specific deflection to the light incorporating such modifications as local stretch, rotation, and translation.

A practical correlator system might incorporate such an optical transforming system or a sensor which collects data in the appropriate format by the nature of its scan pattern. Whether accomplished by the sensor scan or by a coordinate transformation, the logarithmic coordinate transformation is equivalent to resampling an image at logarithmically spaced intervals. An increase in space bandwidth (number of samples) is caused by the oversampling which takes place at small values of the input coordinate. This increased sampling at the input is a cause for concern in a practical correlator design. In such a system, the resolution required at the highest sampling rate fixes the design of the entire system. This may cause the space-bandwidth product required for

70

adequate correlation to exceed the capability of the sensor. However, Anderson and Callary (Reference 51) showed that previous researchers (Reference 52) had overestimated the space-bandwidth requirement and that practical Mellin-Fourier correlators were possible.

2. SYNTHETIC DISCRIMINANT FUNCTIONS

Another technique for recognizing multiple orientations and conditions is to cross-correlate with many different reference images in parallel. The test image can be transformed by many lenses at once, with each Fourier transform falling on an array of reference filters chosen to give reasonable correlation to all conditions. By the proper choice and location of the inverse transform lens, the correlations of all the filters can coincide in one common plane. This parallel setup has been extensively studied by Leib et al (References 53 and 54). They showed that with a memory bank of 23 views of a tank, an optical correlator could obtain a 98 percent probability of detection and 1.4 percent false alarm rate in scenes exhibiting both scale and rotation variations. Unfortunately, this parallel technique is somewhat cumbersome to implement due to alignment of the multiple lenses and filters.

To avoid the need for multiple lenses and filters, it is possible to combine several reference images into one filter. The use of multiple lenses and filters superimposes the outputs of the individual correlators. Because the Fourier transform and correlation are linear, the superposition at the output is equivalent to superimposing the individual filter functions into one filter. Likewise, this is equivalent to superimposing the reference images in the creation of the filter. Rather than create separate filters from many images, a single filter is created from a sum of the images. This simplifies the optical hardware. Caulfield et al (Reference 55) defines a composite matched filter CMF as a single filter which is a linear sum of ordinary matched filters, MF.

$$CMF = \sum_k w_k \, MF_k \qquad (72)$$

These filters can be implemented by either multiple exposure optical holography or computer holography. In the optical hologram, the weights in the linear

combination are obtained by varying the exposure time. The latter approach is to use computers to generate the CMF off-line. In this way, the long-drawn-out creation of the CMF is performed on a digital computer where time is not critical. This takes advantage of the control, dynamic range, and flexibility of digital processors.

Once the CMF function is determined, an optical filter is produced, tested, and optimized. It is then inserted in an optical correlator to take advantage of its real-time processing. To implement the CMF optically, two techniques can be used: transform the digital image to optical image via a high resolution CRT or digitally addressed camera and produce a Vander Lugt Filter in the conventional holographic manner; or, retain the image in a digital format and produce the filter through computer-generated hologram techniques. This latter technique has the advantage of using the dynamic range of the digital computer until the final product is produced. That is, if the CMF function is displayed and transformed optically, the display will limit the dynamic range. By producing a computer-generated holographic filter, the dynamic range is retained till a later stage. In addition, complex filter functions and frequency pre-emphasis can be easily incorporated.

However the CMF is implemented, the weights must be chosen for optimal performance in a specific application. Hester and Casasent (References 56 and 57) developed what is called the Synthetic Discriminant Function (SDF) which is a CMF that gives the same correlation output intensity for each pattern in the training set. The weights required to provide a constant correlation output for each pattern are not unique. Additional constraints can be placed upon the SDF to reduce the response to specific unwanted targets, to reduce dynamic range, or to incorporate other desirable features. Starting with a training set (Figure 14) which adequately describes the conditions in which the desired target could be found, the SDF is formed as a linear combination of all of the training images (Figure 15). The weights are determined using matrix techniques which typically requires considerable time on a large computer (References 58 through 63). The weights are adjusted to provide a correlation with each member of the training set as close as possible to one. That is, the quantity $\Sigma_i |f \cdot g_i - 1|^2$, is minimized where f is the SDF and the g_i's are the images in the training set.
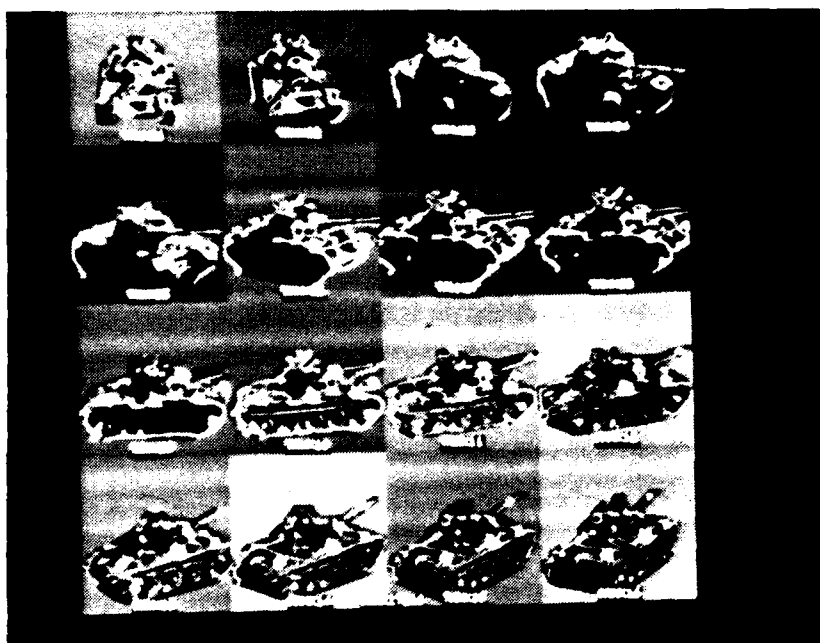
Figure 14. Training Set for the Creation of an SDF
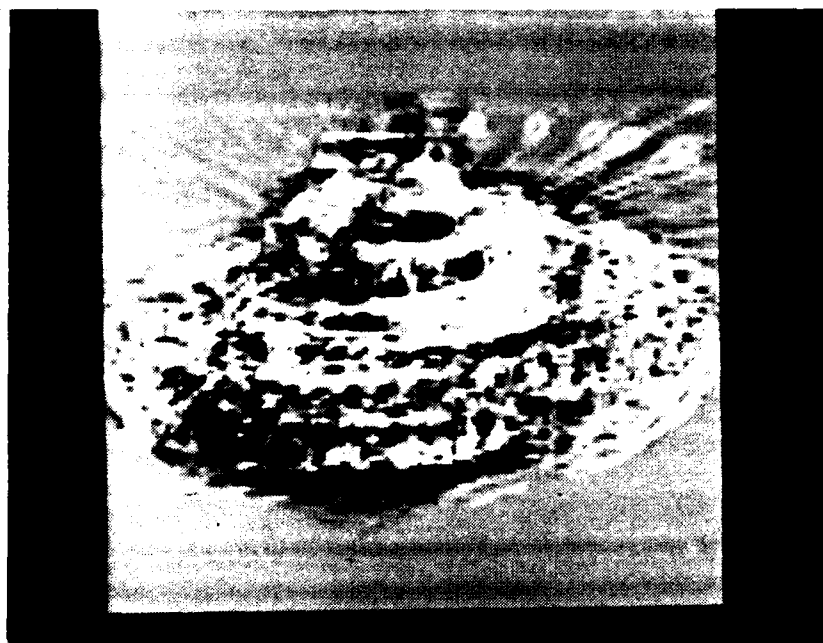
Figure 15. SDF Created from the Images in Figure 14

A constraint on the number of images in the training occurs when the dynamic range is limited. To illustrate this, note that if the medium on which the SDF will be reproduced has limited dynamic range, small values cannot be recorded on the same medium as the large values. The images with larger weights will appear in the SDF of limited dynamic range, while the images with smaller weights will be lost in the noise. As more and more images are combined into the SDF, the sums will become large but the small details in any one image will be too small to appear in the recorded SDF image. This problem is greatly simplified by leaving the SDF in the computer where the dynamic range is not practically limited. That is, to create the hologram pattern on the computer rather than producing the hologram optically. As was shown in Section III, the dynamic range can be reduced by eliminating unnecessary terms from the hologram. However, when producing the hologram optically, the SDF image must be displayed on a device with a definite limit to its dynamic range. This restriction is quite severe and frequently prevents the use of SDFs in optically-generated holographic matched filters. This can be somewhat eliminated by the judicious choice of weights to reduce the dynamic range to a minimum while maintaining adequate performance.

SECTION VI
MATCHED FILTER LINEARITY

Optical matched filters are almost always produced on some type of film or photosensitive surface. The pattern recorded is typically the Fourier transform of the reference scene or some pattern related to the transform. The assumption of the previous chapters was that the film responds as a square-law device. This implies that the transmission of the film responds linearly with the irradiance or exposure on the plate. However, photographic materials are rarely linear, but rather, respond with a typical s curve response. The study of the relationships between the irradiance of the light falling on the film and the resulting blackening produced after development is known as sensitometry. The sensitometry of any photographic material is a crucial link in producing and optimizing a matched filter.

Sensitometry is based upon plotting the density (blackening) of a photographic material as a function of exposure. The blackening of a photographic emulsion is measured in terms of optical density. Light striking a developed photographic negative is partially absorbed by the metallic silver in the emulsion. Opacity is defined as the ratio of the irradiance of light incident on a film to the irradiance of the light passing through the film. This ratio is always greater than unity. The intensity transmittance is defined as the reciprocal of the opacity and thus has a value less than 1. The density of a photographic material is the logarithm (base 10) of the opacity. Densities rarely exceed 3 for normal photographic materials (OD 3 implies only 0.1 percent of the light passes through the film) but certain holographic films are capable of higher densities. At the other extreme, no materials have densities of zero. All materials have some loss in the emulsion, substrate, and surfaces causing the unexposed density or base fog to amount to as much as OD 0.5.

Hurter and Driffield performed the first successful work toward finding a relationship between exposure irradiance and the resulting density of a photographic emulsion. By exposing a photographic film to varying intensities of light, the resultant density can be plotted as a function of exposure. The Hurter and Driffield method of plotting, which has come to be known as the H & D curve, is to plot Optical Density against the logarithm of the exposure (log

It), where I is the irradiance (watts/cm2) and t is the time (seconds). The H & D curve depends on the emulsion type, development, spectral content of the exposing light, age and condition of the emulsion.

Figure 16 shows a typical H & D curve for a photographic emulsion. At low exposures, (point A) there is an attenuation or fog on the film even when the film is unexposed. This occurs due to the losses of the emulsion and sub-strate, and surface reflections. The base fog determines the lowest opacity possible with the film. Typically, the base fog becomes considerably worse as the film ages, especially if it is not stored in a cool place. The silver halide crystals in the emulsion do not respond to light until a certain number of photons strike the crystal. Due to this finite number of photons which must strike the crystal before it responds and the limited size of the crystals, the film is insensitive to light below a certain threshold. At point B, sufficient numbers of photons are available to convert the larger silver halide crystals to metallic silver. This statistical event accounts for the toe of the H & D curve. When the number of photons arriving are larger than the threshold, the conversion of the silver halides to metallic silver depends on the projected area of each crystal. The crystal size is carefully controlled during the production of the film. The distribution of grain sizes provides a linear portion of the H & D curve denoted part C in Figure 16. Because the crystal grain sizes are limited in range, the film saturates at point D on the curve. When all of the silver is converted, additional exposure cannot further reduce the transmission of the developed film. In fact, some films will lose density with additional exposure (part E on the curve). This effect, known as solari-zation is rarely used except by photographic artists for special effects. This curve indicates that the film response over a wide range of exposure is quite non-linear.

1. MEASUREMENT OF FILM CHARACTERISTICS

In order to model the film accurately, it is important to collect an adequate sample of data points. These samples must be gathered in an experi-ment in which the conditions are representative. The exposing energy and wavelength must be in a range consistent with the eventual use of the film. Typically, a sensitometer and densitometer are used to obtain the data points for the response of the film. The densitometer exposes the film to a known
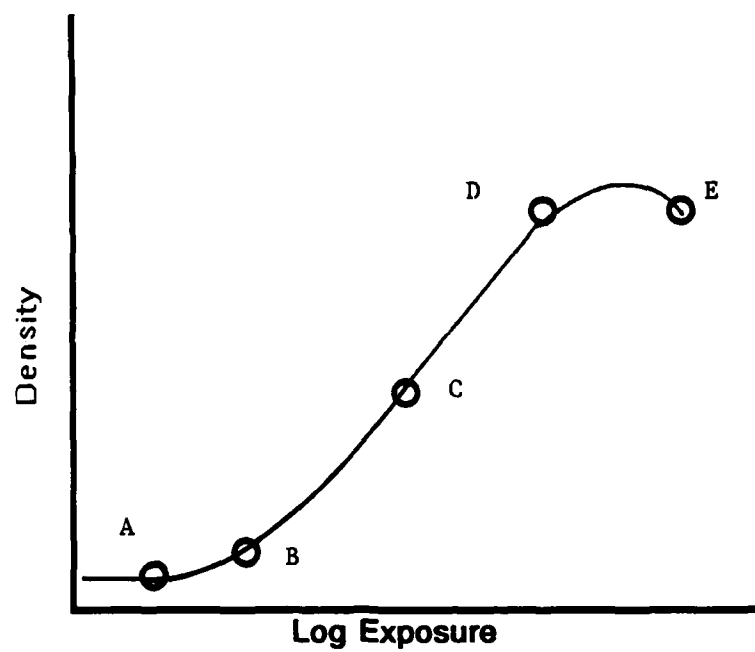
77

Figure 16.   Typical H and D Curve

irradiance for a fixed time period. A Kodak step wedge is placed between the sensitometer light source and the film to provide a wide range of exposures on the film simultaneously.

The EGG sensitometer used in this lab provides the exposure with an xenon flash lamp. The lamp has a consistent output irradiance and the exposure time can be controlled electronically. The light from the flash lamp is diffused to provide a uniform exposure across the entire surface of the step wedge. This provides an accurate and simple means by which the film is exposed to a known energy. This method is used to calibrate most photographic films. Unfortunately, several problems arise when attempting to use this sensitometer for holographic films. The xenon lamp is rich with many spectral lines, giving a brilliant white appearance. The film response to this white light is not necessarily the same as the response to a single laser spectral line. This limitation can be somewhat eliminated by incorporating a spectral filter between the lamp and the film. When the filter is properly chosen, the output of the sensitometer can approximate that of the laser source.

A second limitation of the EG&G sensitometer is the extremely short exposure time. The flash lamp emits light for a brief period much less than a second. Considering that many holographic exposures may last as long as a minute, the time discrepancy between the use of the film and the calibration of the film may be considerable. Reciprocity is a term applied to the consistent response of film to the same energy of light even when the exposure time is changed. When the response does not depend on energy alone, but also exposure time, the film suffers reciprocity failure. Reciprocity failure is common among most films when attempting to use them at very long or very short exposure times. It is important to calibrate the film with an exposure time consistent with the intended use of the film.

Yet another limitation of the EG&G sensitometer is the difficulty in measuring the energy of the exposing light. Although the sensitometer provides a quick and consistent source, its flash is difficult to measure with most radiometers due to their limited ability to integrate the narrow pulse. It is therefore necessary to use film calibrated by another source in order to calibrate the sensitometer. This roundabout calibration lends itself to inaccuracy. These limitations of the EG&G sensitometer suggest that another

79

technique is needed when accurate measurement of the film characteristics is desired. To perform a more accurate measurement a laser, beam expander, and step wedge were set up to expose the film in a controlled fashion. It is interesting to note that even with the limitations of the EG&G sensitometer, the results derived from the sensitometer are quite consistent with those derived from the laser setup. The inaccuracies range from 5 to 20 percent, depending on the films used. Because the EG&G sensitometer is commonly avail-able, easy to use and fast, it is still used in this Laboratory as a check of each new lot of film and to test for development anomalies.

When a new film is received or a new developing technique is tested, a rigorous film calibration is required. As described above, the EG&G sensitom-eter is not always adequate for calibrated exposure of the film. Instead a laser beam is spatially filtered and expanded to provide a broad uniform beam. The irradiance of this beam is measured using an NRC radiometer. This cali-brated beam is then passed through a Kodak step wedge to provide 21 discrete exposure levels. Using the known irradiance of the original beam and the known transmission of each step on the tablet, the irradiance at each point on the photographic plate is determined. The resulting exposure energy density in joules/cm2 at each point is computed as the product of the irradiance (in watts/cm2) and the exposure time (seconds). The exposure time is chosen based on an educated guess to give a transmittance of 0.5 near the middle of the exposure range. It is desirable to have as many of the 21 exposure levels provide points within the dynamic range of the film. It may be necessary to repeat the exposure using different exposure times in order to properly bracket the film response.

A number of holographic film types are used in this Laboratory. These films are mounted on glass substrates and are referred to as photographic plates. These plates include Agfa 10E75, Agfa 8E75, Kodak 120, and Kodak 649F. Full specifications are available from the manufacturer but the general charac-teristics and uses of the plates are outlined here. The Agfa emulsions are on a relatively thick substrate. The 10E75 plate is 50 times more sensitive than the 8E75 plate but with an accompanying loss of resolution. The 10E75 plates have sufficient resolution for two-beam holography, where the fringe spacing is reasonably broad. However, for white-light holography, the resolution of the 8E75 plates is necessary. Each of the Agfa plates has thick emulsions of

approximately 8 microns, permitting volume holography. The Kodak 120 plates utilize a thinner substrate and emulsion. The speed and resolution are between that of the Agfa 10E75 and 8E75 plates. The Kodak 120 plates are useful for two-beam holography but lack the resolution and thickness for white-light holography. The Kodak 649F is one of the first commercially available holographic plates. It has an extremely thick emulsion (25 micron) and is capable of high resolution. Although it is slow, the 649F plate is a popular choice for white light holography. In addition, most phase holograms using bleached halides or dichromated gels use the 649F plates due to its thick emulsion.

Once the various plates have been exposed to a calibrated light source, they are developed using a standardized developing technique. In this case, development consists of 5 minutes in Kodak D-19 at approximately 74^OF with 1-minute agitation intervals. The plate is then placed in Kodak Stop Bath for 1 minute and Kodak Rapid Fix for 4 minutes. After a 30-minute wash, the plate is dipped in Kodak Photo Flo and hung up to air dry. After processing, the intensity transmission of each exposure step is measured. This measurement is performed with a densitometer. A commercial densitometer is readily available and easy to use. Unfortunately, the densitometer gives results in terms of optical density which must be converted to transmittance by $t = 10^{-D/2}$.

The commercial densitometer, as with the sensitometer, is designed for use with photographic products which operate over a broad spectrum of colors. Thus the white light in the densitometer may give results inconsistent with the red laser light used for holographic experiments. To determine whether this was the case or not, an experiment to measure the transmission of the film was devised. The irradiance of the collimated laser beam is measured. The beam is then passed through each step of the developed plate and the existing beam irradiance measured. The power transmission is the ratio of the transmitted irradiance to the incident irradiance. The amplitude transmission is the square root of the power transmission. Using the laser transmission experiment, extensive measurements were performed on each of the photographic plates. Additional measurements were made on the same plates using the densitometer set to the red color position. In the red position, a red filter is placed over the light source. Apparently, the red filter sufficiently matches the red laser light. The results from the commercial densitometer very closely match

81

the results from the laser transmission experiment. The agreement was consistent, with each of the plates indicating that the commercial densitometer results could be trusted. This is significant, due to the ease and speed with which this densitometer is operated. The agreement has been so consistent that the laser setup is used only as an occasional check of the densitometer.

The entire process of exposing the plates to a calibrated light source, developing the plate, and measuring the resultant transmission must be carefully accomplished. It is the film processing which causes many to refer to holography as an art rather than a science. There are many factors which affect the film response. Each of these must be held constant for the calibrations to be meaningful. For example, the base fog is highly dependent on the age of the film. The film speed and contrast depend on the developer temperature and freshness. Film and chemicals must be properly stored and discarded after their expiration date. Unless these conditions are held constant, the film calibrations will be inconsistent and experimental results inconclusive.

2. MODELS FOR FILM NON-LINEARITY

When used in a coherent optical processor, the essential parameter of any optical element is its amplitude transmittance, $T_a(x,y)$. For a photographic emulsion, this in turn is determined by the energy, $E(x,y)$ to which it has been exposed. Thus an emulsion is best characterized experimentally by the transmittance versus exposure (T_a-E) response curve. This is in contrast to the typical H & D curve provided by most film manufacturers. The curves are related. The amplitude transmission (ignoring any phase shift) is the square root of the intensity transmission measured experimentally. The exposure is the same as in the H & D curve but is not plotted logarithmically. These curves also display an s shape due to the saturation at low and high exposures. The T_a-E curves have the opposite slope as the H & D curve because the transmission decreases with exposure while the density increases with exposure.

For the matched filter to be recorded linearly, the recording media must have transmittance directly proportional to exposure, $T_a=cE$, for all values of E. Recall from the previous discussion of optical matched filters that the hologram is created from the interference of the Fourier transform wave and a

reference wave. A lens forms the Fourier transform of the reference image $S(u)=S_O(u)\exp\{i\theta(u)\}$, and a uniform reference beam $R(u)-R_O\exp\{iau\}$ is introduced at an angle o to the optical axis (a=ksin o). A photographic film placed at the focal plane of the Fourier lens receives an exposure

$$E(u) = |S(u) + R(u)|^2 t \qquad (73)$$

where t is the exposure time. Substituting for S(u) and R(u),

$$E(u) = [R_O+S_O(u)]t + tR_OS_O(u)[\exp(i\theta)\exp(-iau) + \exp(-i\theta)\exp(iau)]$$

$$= E_{dc} + E_{ac} \qquad (74)$$

where $E_{dc} = t[R_O + S_O^2]$

and $E_{ac} = tR_OS_O[\exp(i\theta)\exp(-iau) + \exp(-i\theta)\exp(iau)]$

are the local average exposures and the varying components respectively. Note that dc (direct current) and ac (alternating current) subscripts apply here to the slowly varying terms and the high spatial frequency terms, respectively.

The ideal recording medium would have a transmittance directly proportional to exposure, $T_a = cE$, for all values of E. For such a linear material the transmittance would be

$$T_a(u) = c[Edc + Eac]. \qquad (75)$$

When such a transparency is illuminated with a collimated beam, light is diffracted by the film. The first term in equation (75) is real and gives rise to wavefronts propagating near the optical axis. The ac term includes the factor exp(iau), a linear phase shift, which will diffract light waves at angles plus and minus o to the optical axis. Thus if a is chosen appropriately, or the recording angle properly chosen, each of these wavefronts will be separated from the other and from those which propagate on axis. One of the off-axis

83

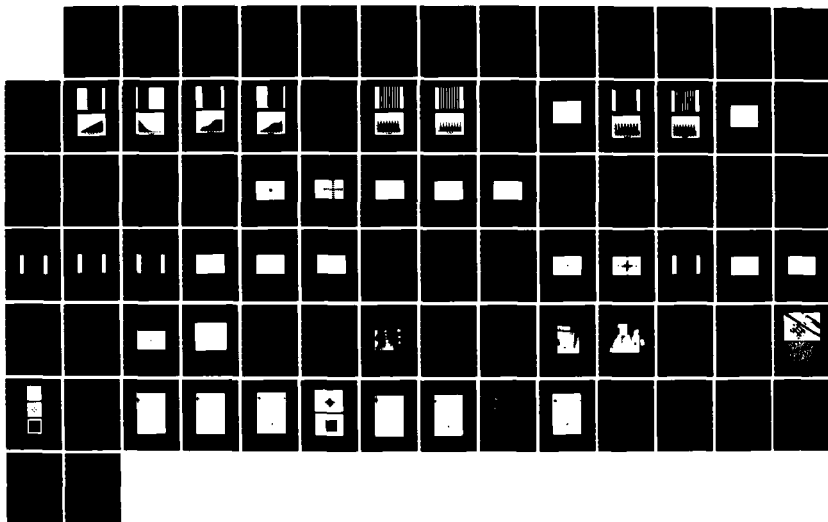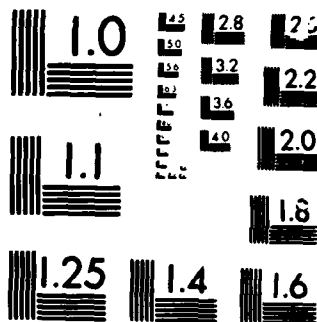MICROCOPY CHART

beams contains the term $S_o(u)\{exp -i\theta(u)\}$ which is the complex conjugate of the reference Fourier transform and thus has the desired optimal filter characteristic.

Unfortunately, photographic emulsions do not exhibit this ideal linear response described above. Rather, they saturate at high and low exposures. To understand the effect of this non-linearity, it is important to describe the actual film response at all exposures and model that response to predict the transmittance of the developed emulsion. This model should be based on actual response curve measurements performed on the film using a calibrated sensitometer and densitometer. This model will assist in the proper choice of average exposure and ratio of reference to signal exposures.

The Hurter-Driffield (H-D) curve has been used extensively to predict photographic response. This plot of photographic density versus log exposure demonstrates the key features of the film saturation. A form of this curve could be described by

$$D = \begin{cases} D_s & \text{when} \quad E > E_s \\ \gamma \log(E) - \log(E_{bf}) + D_{bf} & E_{bf} < E < E_s \\ D_{bf} & E < E_{bf} \end{cases} \quad (76)$$

where s denotes saturation, bf denotes base fog, and γ is the slope of the linear portion of the curve. Modeled in this fashion, the film exhibits a linear transmittance versus exposure only by producing a positive print and developing to net γ of -2. When this is the case, the transmittance becomes

$$T_a = \begin{cases} T_{bf} & \text{when} \quad E > E_s \\ c(E-E_s) + t_{bf} & E_{bf} < E < E_{ss} \\ T_s & E < E_{bf} \end{cases} \quad (77)$$

where the transmittances are those of the positive print and the exposures refer to the original negative. Thus, when the exposure is determined by equation (74) this model predicts linear recording will occur when the signal and reference amplitudes satisfy

$$|R_o - S_{o\ max}|^2 t > E_{bf} \quad \text{and} \quad (78)$$

84

$$|R_o + S_{o\ max}|^2 t < E_s$$

where $S_{o\ max}$ is the maximum signal amplitude in the Fourier plane. This can be expressed in terms of the experimental parameters, E_{dcmax} and K where

$$E_{dcmax} = (R_o^2 + S_{o\ max}^2) \tag{79}$$

and $K = R_o^2/S_{o\ max}^2$.

$E_{dc\ max}$ is the maximum average exposure and K, frequently called the beam balance ratio, is the ratio of the reference to signal beam intensities. In terms of E_s and E_{bs} these become

$$E_{dcmax} = 1/2\ (E_s + E_{bf}) \tag{80}$$

and

$$K = [(E_s^{1/2} + E_{bf}^{1/2})/(E_s^{1/2} - E_{bf}^{1/2})]^2$$

Under these conditions, the maximum amplitude of the diffracted signal is given by $1/4(T_s - T_{bf})$. For all values of $s(p)$ less than s_{max} the diffracted signal will be proportional to $s(p)$, as desired.

The Hurter-Driffield model, although the most common in photographic work, is not directly applicable to holography. Transmittance, the ratio of transmitted amplitude to incident amplitude, is the fundamental parameter in holographic filters. Plotting the transmittance versus exposure is more convenient than inferring the information from the D-log(E) curve. The model as expressed in equation (76) assumes the response is piece-wise linear with two breakpoints. This neglects the smooth non-linearity which exists throughout the film response. Although easier to analyze, the piece-wise linear model does not predict the non-linear effects in the regions near the breakpoints. To predict more accurately this non-linear effect, a polynomial approximation can be made to fit an experimentally measured T_a-E curve.

A polynomial representation provides a more accurate approximation over a wider range of exposure than possible with a linear representation. Such representations have been studied with polynomials of degree three being the most common choice (References 65 through 67). Using a third order polynomial, the transmittance can be expressed as

$$T_a = C_0 + C_1 E_T + C_2 E_T^2 + C_3 E_T^3 \qquad (81)$$

where $E_T = E_{dc} + E_{ac}$. Since E_{dc} is slowly varying, equation (81) can be written

$$T_a = A_0 + A_1 E_{ac} + A_2 E_{ac}^2 + A_3 E_{ac}^3 \qquad (82)$$

where the A_i depends on the average local exposure E_{dc}. Only the ac portion of the exposing light causes fringes which will diffract light in the resultant hologram.

Each power of E_{ac} will diffract light to a unique angle or diffractive order. The terms in equation (82) can be written

$$T_a = T_0 + T_1 + T_2 + T_3 \qquad (83)$$

where T_0 is a real quantity and T_1 contains all those terms which contribute to the first diffraction order. T_2 and T_3 contain only terms which appear in the second and third diffracted orders. Thus T_1 can be separated spatially from all other contributions and contains the filter term which is desired. This term, when expanded, can be written

$$T_a = [(C_1 + 2C_2 E_r + 3C_3 E_r^2) + (2C_2 E_s + 9C_3 E_r E_s + 3C_3 E_s^2)]E_{ac} \qquad (84)$$

where $E_r = R_0^2 t$ and $E_s = So^2 t$ are the reference and signal exposures respectively. Written in this form, the first term in brackets does not depend upon the signal strength. This term produces a linear reproduction as in the ideal case. The second term in brackets contains all contributions depending on the signal strength which would produce a nonlinear result. Thus, in this form, the first term represents the signal coefficient and the second term the nonlinear noise coefficient. This provides a convenient method to determine the signal-to-noise ratio for a given signal and reference exposure.

Several interesting cases are apparent from inspection. When E_r is large, the signal portion is dominant. The coefficient of E_{ac} is maximized when $E_r = E_s$. There is a tradeoff between the need for adequate signal and minimizing signal-to-noise ratio. This choice depends upon factors such as the laser power available and the noise level which can be tolerated. If E_{ac} is maximized by setting $E_r = E_s$ and also set the sum $E_r + E_s = E'$, the exposure value at the inflection point, the values of E_r and E_s are determined uniquely. The result is $E_r = E_s = -C_2/6C_3$. By substituting these values into equation (84), it is seen that the noise term is zero. This choice of values is not very useful for matched filtering because the zero noise condition holds only if E_s does not vary. No information can be stored in this condition. If E_s is allowed to vary, the noise coefficient will increase. The worst case condition is when $E_s = E_r/2$ when the signal-to-noise ratio is 12.4. For many applications, this noise is tolerable and this choice of E_s and E_r will be the preferred choice.

The expression in equation (84) is general for any t-E curve which is described by a cubic polynomial. It is useful to explore analytically the effect of varying reference and signal exposures and to predict the signal and noise in the recorded pattern. However, such analytical methods prove unwieldy when complicated signals are involved. To analyze the more complicated patterns, the pattern and film must be modeled on a computer. The computer provides the brute force to expand the input pattern into the nonlinear terms expected from the film. By modeling the film on the computer, it is possible to plot the experimental film response along with the predicted response. This permits the comparison between the experimental data and various order polynomial fits.

It has been recognized in experiments by this author, that cubic polynomials have difficulty representing accurately the entire range of the film response. In order to use the cubic polynomial to predict the inflection point, only the experimental data points near that point or in the linear region should be considered. Unfortunately, this does not permit accurate modeling in the saturation regions. When modeling an absorption hologram, where only the linear region is utilized, the cubic polynomial will suffice. The cubic polynomial loses accuracy when modeling holograms whose response

extends into the nonlinear regions, such as phase holograms. This problem is greatly relieved by using a 5th order polynomial. This gives a more accurate fit to the t-E curve over the entire useful range of the film.

In order to model accurately the film response, a set of experimental training points must be established. This is accomplished by exposing the film to varying irradiances of light and measuring the density of the film after developing. A Kodak step wedge with densities varying in 21 discrete steps from OD 0.05 to OD 3.05 was placed before the test film. The step wedge and film were then exposed to a known irradiance of laser light for a known duration of time. The irradiance of the laser was attenuated by each step of the wedge to give a wide range of exposures to the film. The exposure for each section of the film is computed based on the laser irradiance, step density, and exposure time. The film is then processed in the darkroom using the standardized technique mentioned previously. This processing plays a major role in the response of the film and thus must be carefully controlled. The developed film is placed on a densitometer to measure the resultant density of each exposed portion. The set of exposure and density points are incorporated into a polynomial fit routine to produce the coefficients.

The output of the polynomial fit routine is shown in Figure 17. All of the important parameters are recorded in order to reproduce the same result in future exposures. The laser irradiance and exposure time are recorded to verify that their combination would not suggest reciprocity failure which occurs when exposure times are very short or very long. It is advised that the exposure time for determining the training set be near that to be used in experiments. This may require that the experiment be run several times to determine what laser irradiance is required. Unfortunately, it may dictate that more laser power is needed than is available. In Figure 17, it is seen that the exposure time was 120 seconds. This is longer than a normal holographic exposure but no additional laser power was available to shorten the exposure time. An independent test verified that a shorter exposure on several points gave results consistent with the 120 second measurement. This indicates that the 120 second exposure did not cause significant reciprocity failure. The film processing procedure is outlined briefly in the printout with

DATE 11 FEB 85   TYPE OF FILM: 10E75 NAH

LOT #595906   DATE FILM RECEIVED 12 DEC 84

LIGHT INTENSITY   2000.00 ERGS/CM**2 .@ .512 micron

EXPOSURE TIME: 120 SEC.

DEVELOPING PROCEDURE
    5 MIN. D-19   75 DEG. F.
    1 MIN. STOP
    4 MIN. FIXER
    30 MIN. WASH
    1 MIN. FOTO-FLO

STEP#	EXPOSURE (ERGS/CM**2)	RESULTANT DENSITY	AMPLITUDE TRANSMISSION
1	1783.	7.87	.000
2	1262.	7.86	.000
3	893.	7.86	.000
4	632.	7.83	.000
5	448.	7.64	.000
6	317.	5.67	.001
7	224.	4.50	.006
8	159.	3.58	.016
9	112.	2.65	.047
10	80.	1.82	.123
11	56.	1.10	.282
12	40.	0.64	.479
13	28.	0.36	.661
14	20.	0.24	.759
15	14.	0.17	.822
16	10.	0.12	.871
17	7.	0.10	.891
18	5.	0.09	.902
19	4.	0.09	.902
20	3.	0.09	.902
21	2.	0.09	.902

THE POINT OF INFLECTION IS: EXPOSURE FOR O.D. =2
EXPOSURE =   32.ERGS/CM**2   IS   89. ERGS/CM**2
AMPLITUDE TRANSMISSION= 0.596

SLOPE OF TANGENT LINE AT INFLECTION POINT IS -.014535
Y INTERCEPT IS 1.062

TAU vs. EXPOSURE IS APPROXIMATED BY A
FIFTH-ORDER LEAST SQUARES FIT

THE POLYNOMIAL IS OF THE FORM: THE COEFFICIENTS FOR
CO + C1*X + C2*X**2 + ... + C5*X**5   EXPOSURE VS. TAU ARE:

CO = 0.903661728E+00   CO = 0.141298828E+03
C1 = 0.132656097E-02   C1 = -.743000000E+03
C2 = -.568270683E-03   C2 = 0.263350000E+04
C3 = 0.838190317E-05   C3 = -.504100000E+04
C4 = -.422005542E-07   C4 = 0.472500000E+04
C5 = 0.553654900E-10   C5 = -.174050000E+04

THE COEFFICIENT OF CORRELATION IS.999964

Figure 17. Computer Output of the Polynomial Fit Routine

89

developing times and temperatures recorded. The data from the measurement of the film includes the step number, the exposure through that step, the density on the developed film, and the associated amplitude transmission.

The computer tabulates several useful points from the data. The exposure and transmission at the point of inflection are provided. The best linear fit is determined by the slope and intercept of a line tangent to the curve through the inflection point. Also, since most phase holograms are exposed to OD 2.0, the appropriate exposure value is provided. Finally, the printout provides the polynomial coefficients for the transmission versus exposure curve and its inverse relationship along with the coefficient of correlation. The H & D plot is found in Figure 18 for the AGFA 10E75 plates. The associated transmission versus exposure is plotted in Figure 19. The plot shows the linear fit through the inflection point and the result of the 5th order polynomial. Figures 20 through 22 are identical to Figures 17 through 19 except the results are for AGFA 8E75 plates.

3. COMPUTER LINEARIZATION OF FILTER RESPONSE

The models presented here include a summation of terms representing various powers or orders. The linear term or first order represents the desired term where the transmission is linear with exposure. If the linear term represents the signal and the higher order terms represent noise, then the recording on the film is optimized for signal-to-noise ratio when the higher order terms are minimized or eliminated. As was shown in the previous section, the choice of beam balance ratio in the conventional hologram determines which part of the t-E curve is occupied by the information. When the beam balance ratio is high, the information resides in the linear portion. The signal-to-noise ratio is high but there is little variation in density on the film and the efficiency is low. When the beam balance ratio is unity, the entire density range of the film is used, giving high efficiency. Since the information extends into the saturation regions, the high-order nonlinear terms are large and the signal-to-noise ratio suffers.

Using the computer models for the film response, the nonlinear effects in conventional holography are predictable. It is impossible to affect the results, except by varying the average irradiance and the beam balance ratio.
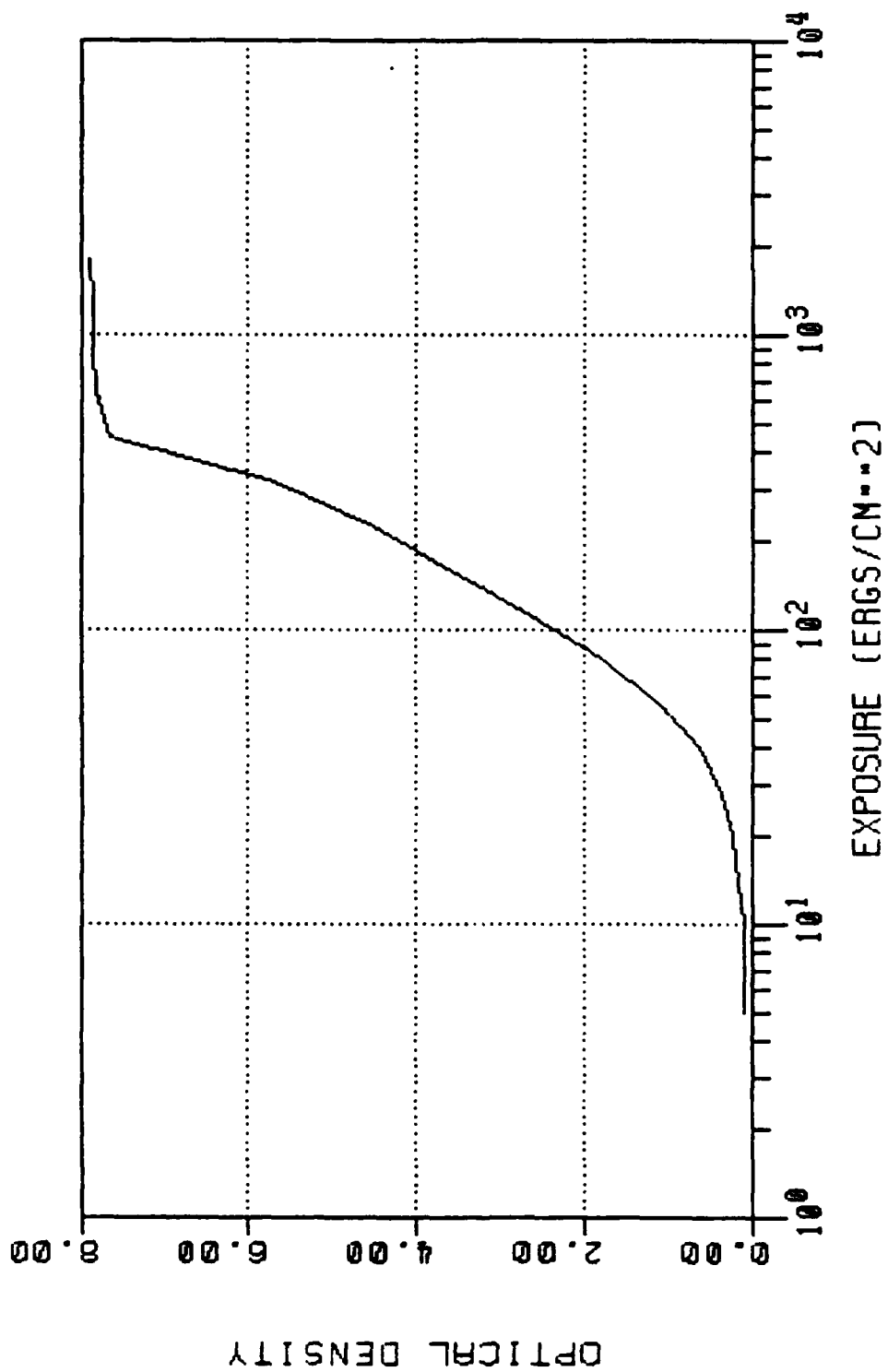
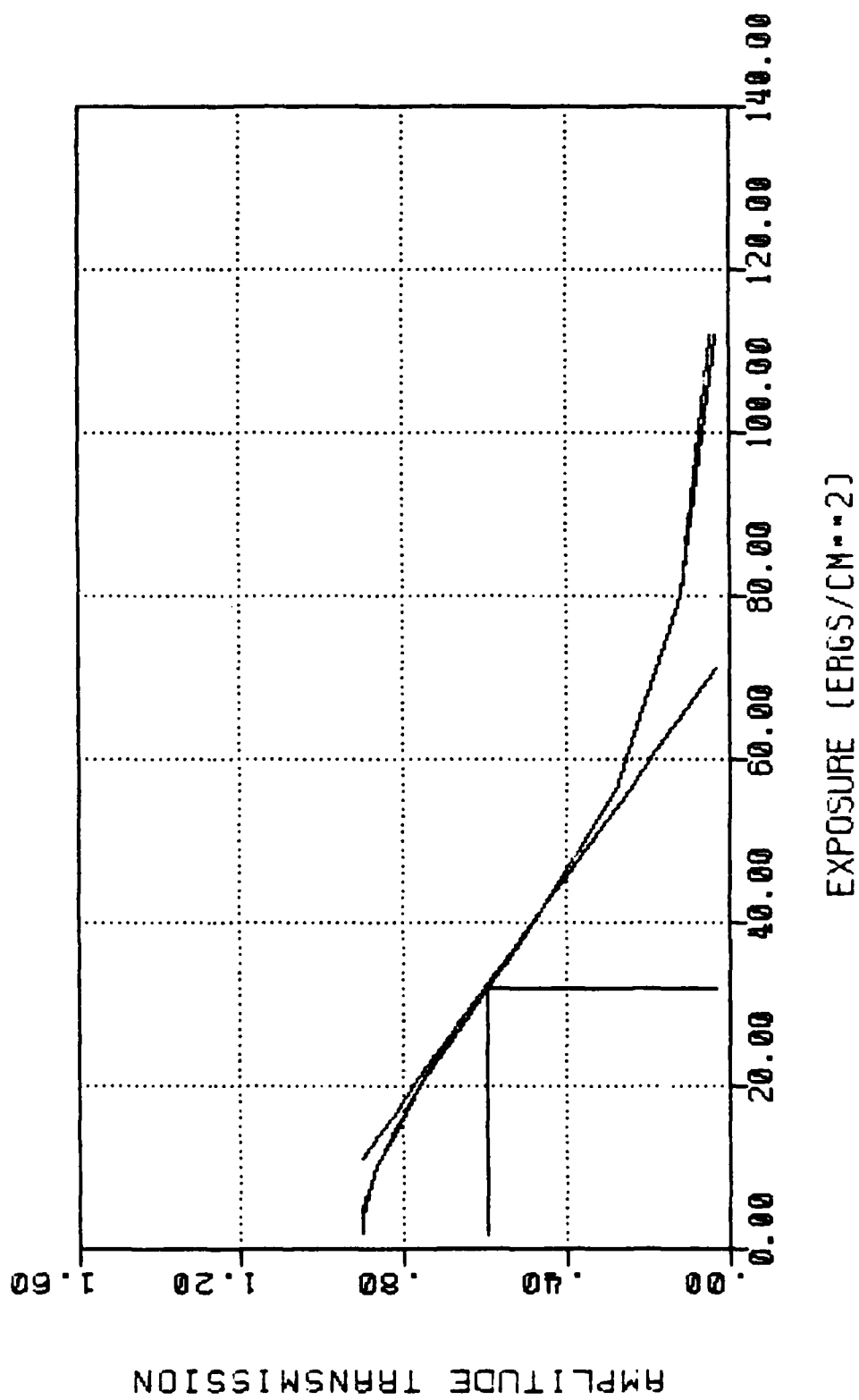Figure 18. H and D Plot for AGFA 10E75 Photographic Plates

Figure 19.   Amplitude Transmission Versus Exposure for AGFA 10E75 Plates

DATE 20 DEC 84   TYPE OF FILM: 8E75 HD

LOT #595707   DATE FILM RECEIVED 9 APR 84

LIGHT INTENSITY 10000.00 ERGS/CM**2 ,@ .512 micron

EXPOSURE TIME: 300 SEC.

DEVELOPING PROCEDURE
    5 MIN. D-19   75 DEG. F.
    1 MIN. STOP
    4 MIN. FIXER
    30 MIN. WASH
    1 MIN. FOTO-FLO

STEP#	EXPOSURE (ERGS/CM**2)	RESULTANT DENSITY	AMPLITUDE TRANSMISSION
1	8913.	4.28	.007
2	6310.	4.27	.007
3	4467.	4.07	.009
4	3162.	3.69	.014
5	2239.	3.12	.028
6	1585.	2.77	.041
7	1122.	2.29	.072
8	794.	1.84	.120
9	562.	1.38	.204
10	398.	0.98	.324
11	282.	0.67	.462
12	200.	0.45	.596
13	141.	0.28	.724
14	100.	0.18	.813
15	71.	0.12	.871
16	50.	0.10	.891
17	35.	0.08	.912
18	25.	0.08	.912
19	18.	0.08	.912
20	13.	0.08	.912
21	9.	0.08	.912

THE POINT OF INFLECTION IS: EXPOSURE FOR O. D. =2
EXPOSURE =   163. ERGS/CM**2   IS   889. ERGS/CM**2
AMPLITUDE TRANSMISSION= 0.684

SLOPE OF TANGENT LINE AT INFLECTION POINT IS -.001774
Y INTERCEPT IS 0.973

TAU vs. EXPOSURE IS APPROXIMATED BY A
FIFTH-ORDER LEAST SQUARES FIT

THE POLYNOMIAL IS OF THE FORM: THE COEFFICIENTS FOR
$CO + C1*X + C2*X**2 + ... + C5*X**5$ EXPOSURE VS. TAU ARE:

CO = 0.944401741E+00   CO = 0.246046875E+04
C1 = -.120447576E-02   C1 = -.256770000E+05
C2 = -.391120557E-05   C2 = 0.125204000E+06
C3 = 0.106440439E-07   C3 = -.284328000E+06
C4 = -.869682104E-11   C4 = 0.295216000E+06
C5 = 0.233255255E-14   C5 = -.113680000E+06

THE COEFFICIENT OF CORRELATION IS 999746

Figure 20.   Computer Output of the Polynominal Fit Routine for
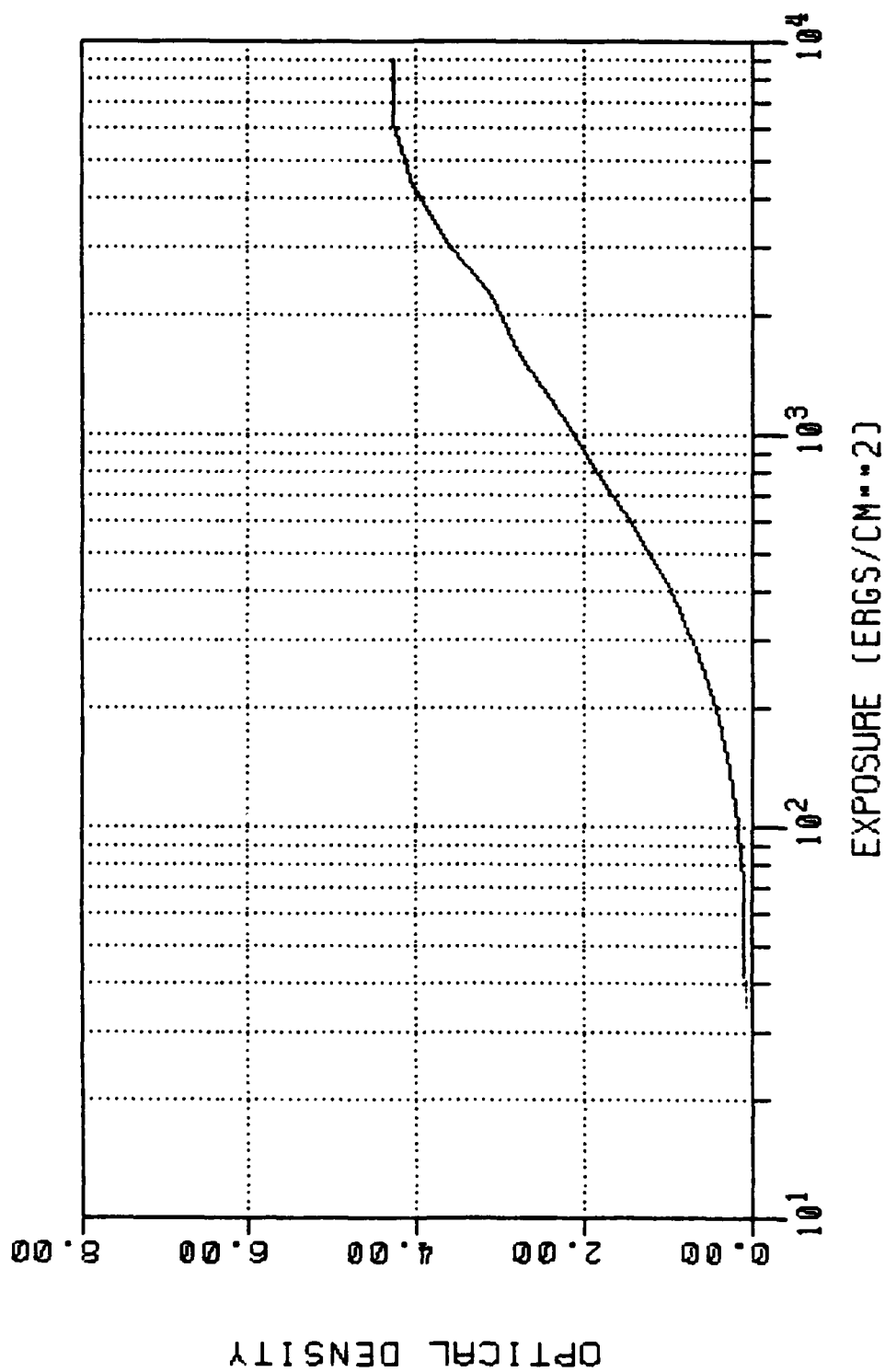   8E75 Plates

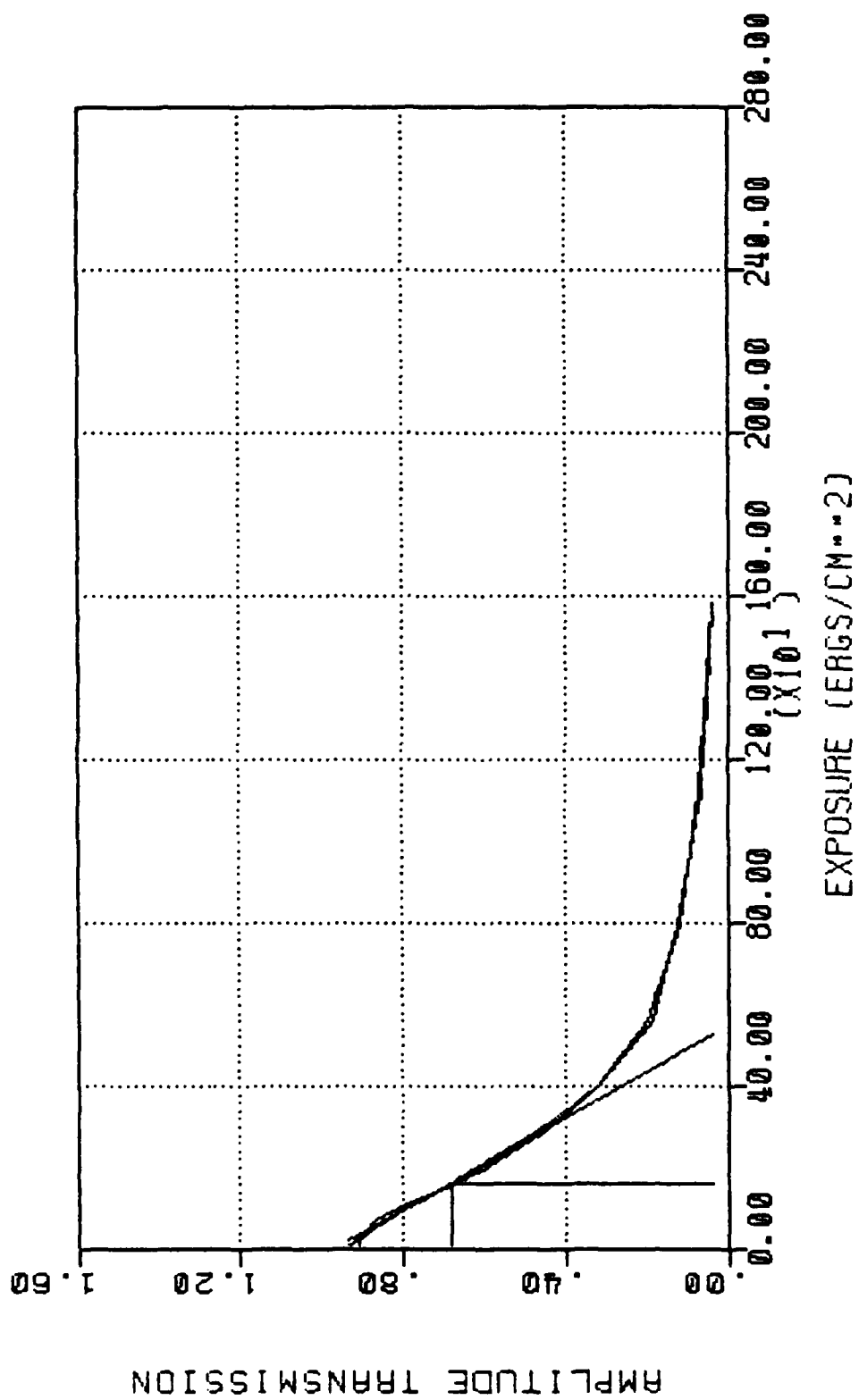Figure 21   H and D Plot for AGFA 8E75 Photographic Plates

94

Figure 22. Amplitude Transmission Versus Exposure for AGFA 8E75 Plates

To linearize the hologram, it is necessary to produce the hologram as a CGH. The techniques discussed in the previous sections are used to compute the holographic filter function. Rather than writing the function directly to film, the pattern is predistorted. Such a predistorted image, when written to the film, will be distorted by the film so as to yield the desired result.

In order to linearize the filter response over the entire dynamic range of the film, it is necessary to model the film over that range. The preceding sections describe the modeling process. Once the film response has been determined experimentally, the computer is used to determine the transfer function of the film. The transfer function is described by the computer expression of the t-E curve. Whether that expression is first order, third order, fifth order, or spline, the inverse relationship can be applied to the desired pattern prior to writing on the film. By predistorting the pattern, the film can be driven linearly to the extremes without saturation or nonlinearity. This permits the film to operate over its entire dynamic range while maintaining a linear response. Thus, the efficiency can be maximized along with the signal-to-noise ratio. When using phase holograms, the mapping from exposure to phase is quite nonlinear. The predistorting process permits accurate control of the phase despite the inherent nonlinearity of the bleaching process. This ability to optimize many aspects of the hologram independently is an overwhelming advantage of computer-generated holograms.

This linearization process was applied to Agfa 8E75 film to test the signal-to-noise improvement. Figure 23 shows a linear gradient provided as an input to an 8E75 photographic plate. Figure 24 shows the resultant transmission on the plate. This figure demonstrates the typical saturation effects noted in film. Note that the once linear gradient is grossly distorted. This is the normal effect incurred in conventional holography. If the pattern in Figure 23 is predistorted, yielding Figure 25 and submitted as the input to the 8E75 plate, the output in Figure 26 is obtained. The predistorted result in Figure 27 is superior to the normal exposure.

To analyze the extent of the distortion and the effect on the signal-to-noise ratio, a suitable test pattern is utilized. The nonlinearity manifests itself in the form of high frequency components due to the powers expressed upon the input pattern. To measure the extent of the nonlinearity, the
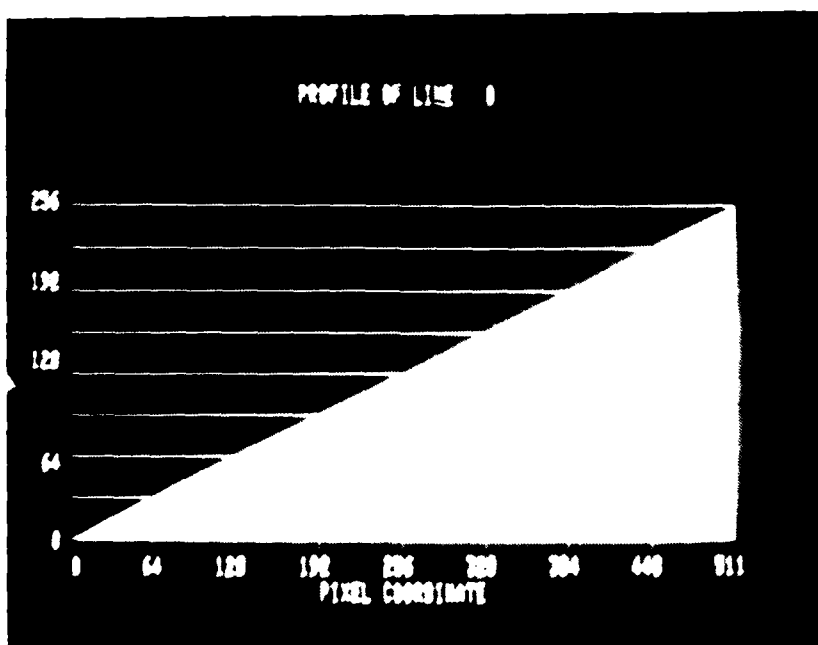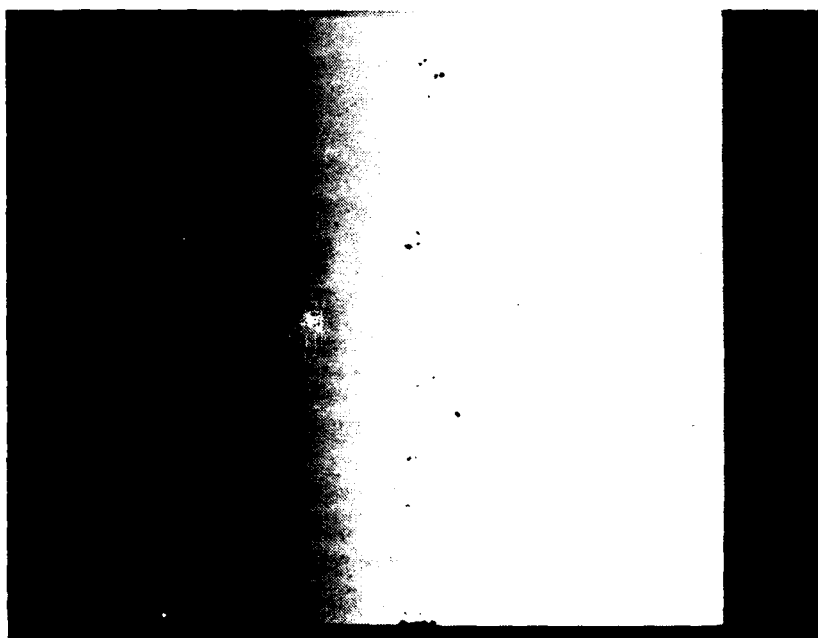
Figure 23. Image and Plot of a Linear Gradient Used for a Test Input
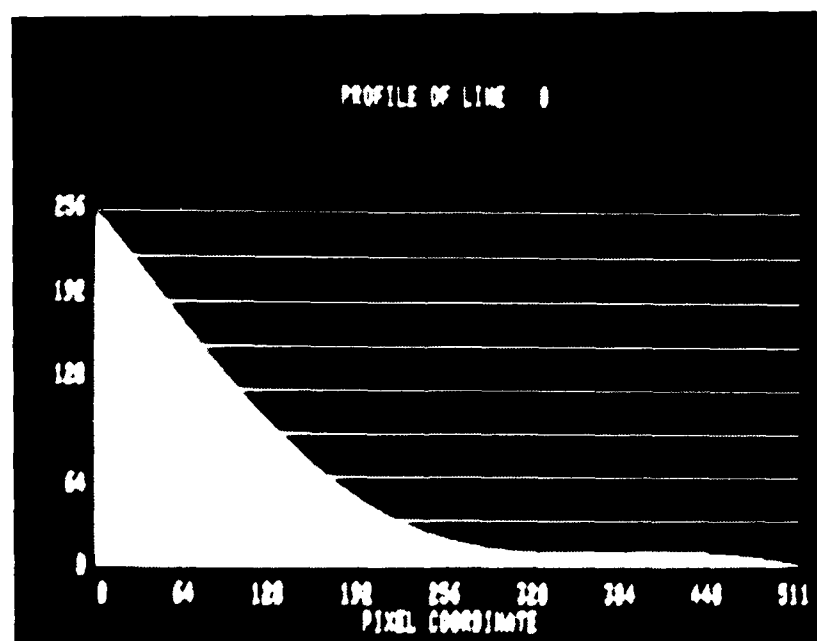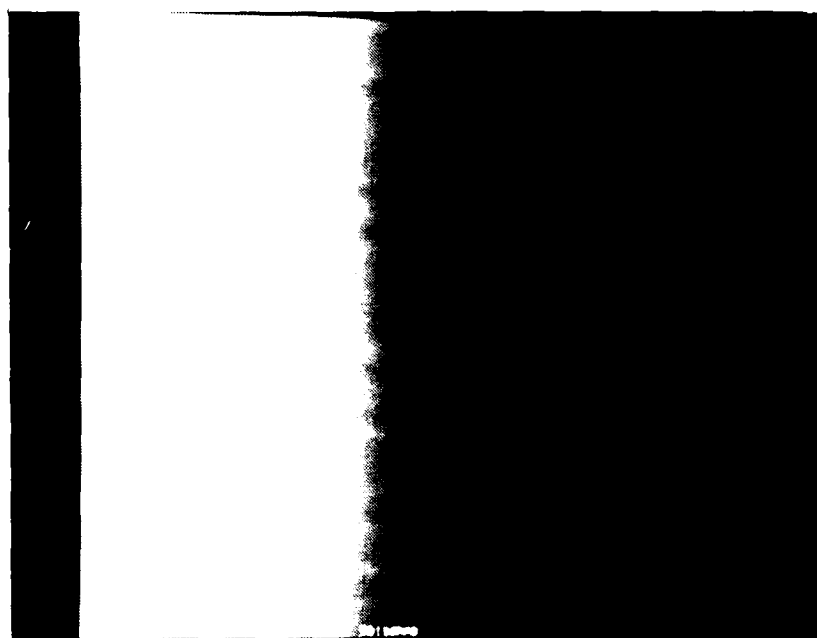
97

Figure 24.   Image and Plot of the Output Transmission on Film from
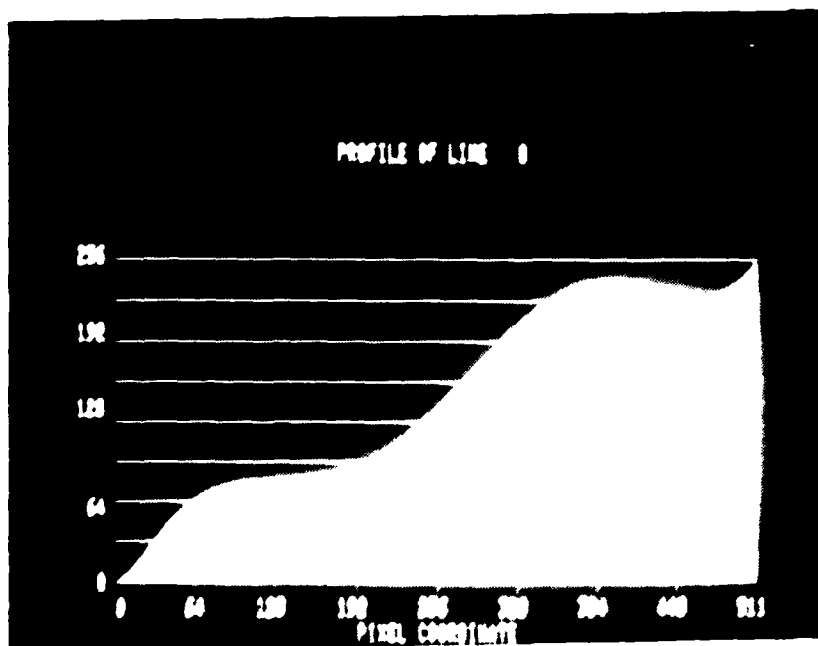a Gradient Input
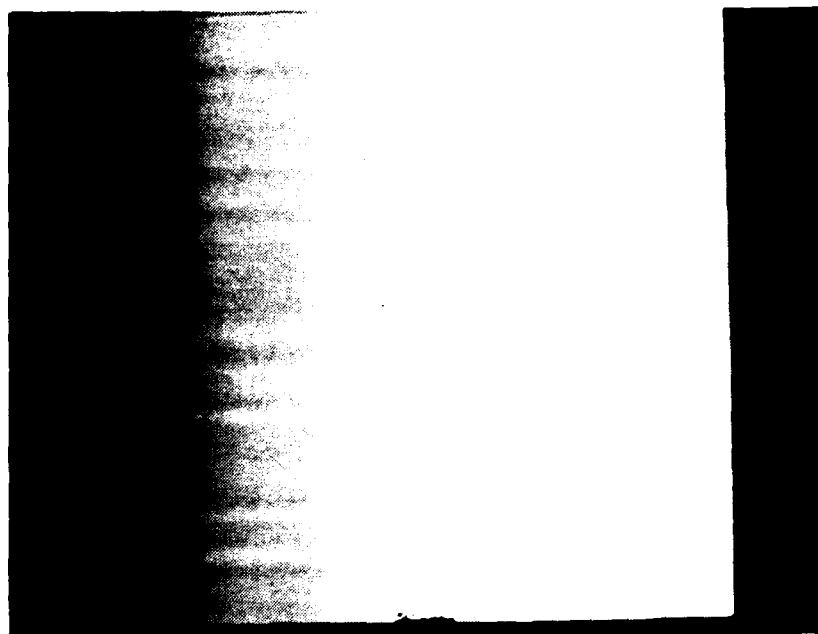
Figure 25.   Image and Plot of the Predistorted Gradient Used for
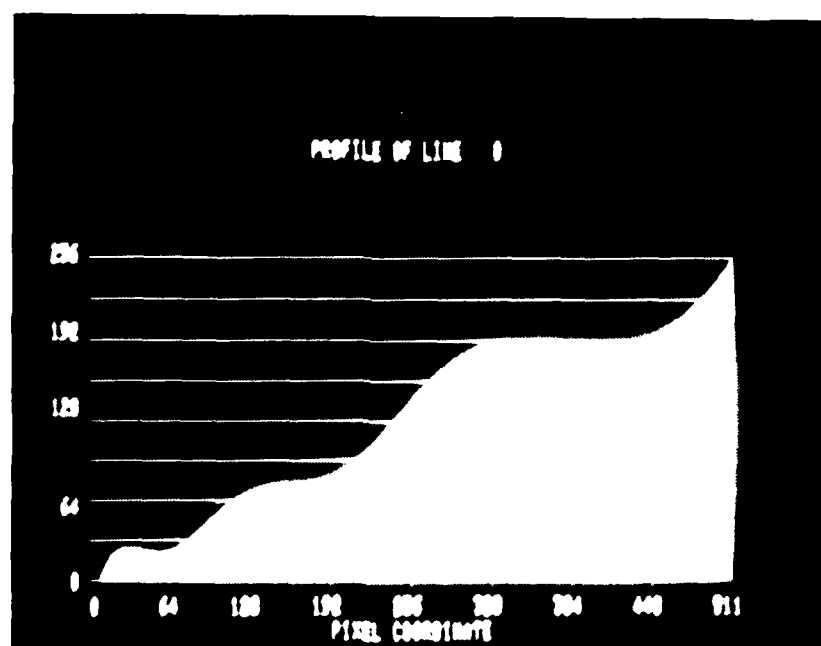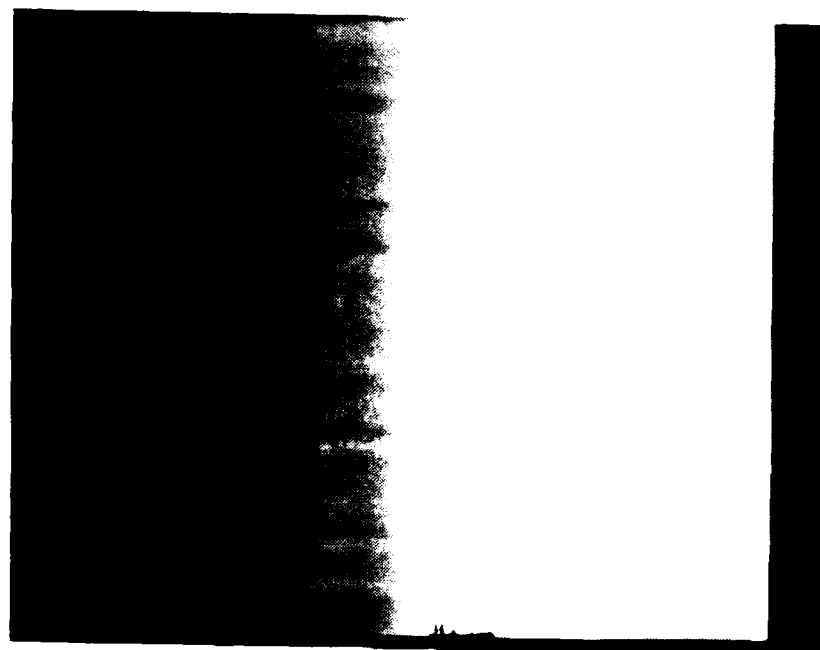   an Input

Figure 26. Image and Plot of the Output Transmission with
           Predistroted Input

spectral energy in the perfect signal is compared to the spectral energy in the distorted energy. The logical test input is a sine wave because it has a simple spectrum. This technique for testing nonlinearity is a long accepted method for electrical systems. The output power when the sinusoidal spectral components are removed, divided by the input power, is called the total harmonic distortion (THD).

Amplifiers for speech and music are typically rated by their THD specification. To measure THD in an audio amplifier, a sine wave is input to the amplifier but is filtered from the output. Any remaining power not at the input frequency is considered a result of harmonic distortion. Applied to the analysis of film, a sinusoid input is written to the film using the expression

$$s(x,y) = (1 + \cos kx)/2 \qquad (85)$$

where k determines the spatial frequency. Note that the expression is biased and scaled to provide an input between zero and one. This is further scaled to utilize the entire dynamic range of the film. The Fourier transform of equation (85) is

$$S(u,v) = 1/2 + 1/4 \, \delta (u-k) + 1/4 \, \delta (u+k), \qquad (86)$$

and the power spectral density is

$$I(u,v) = 1/4 + 1/16 \, \delta (u-k) + 1/16 \, \delta(u+k). \qquad (87)$$

There is a constant term and two spectral terms. The spectral terms in the intensity image are down from the constant term by 1/4. Now that the input spectrum is known, the film distortion is applied to the input and the effect on the output spectrum is analyzed. The sinusoidal input is shown in Figure 27. The film distorts the sinusoidal input and transmits the pattern shown in Figure 28. The spectrum of the distorted pattern recorded on the Agfa 8E75 plate is shown in Figure 29. Note that the spectrum is rich with harmonics. The input spectrum consisted of only the constant term and delta functions at plus and minus 8 cycles. The distorted spectrum clearly shows spectral components at the harmonic frequencies of plus and minus 16, 24, 32, and so on. These harmonic components account for the nonlinear behavior and
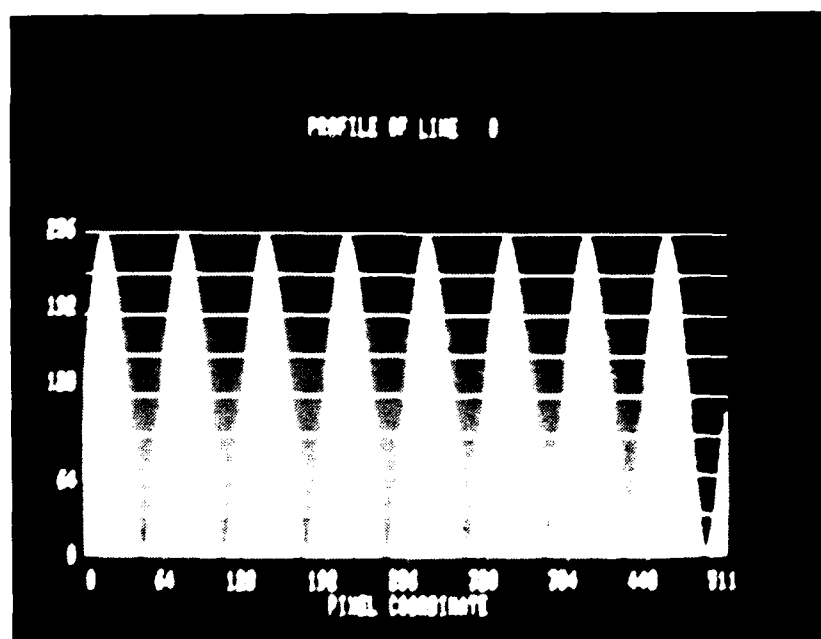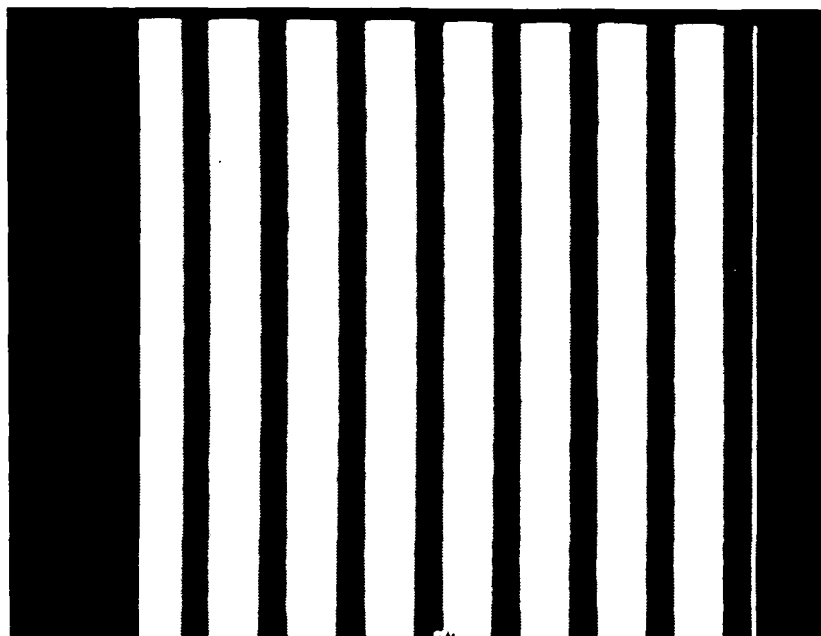
Figure 27. Image and Plot of a Sinusoidal Grating Pattern
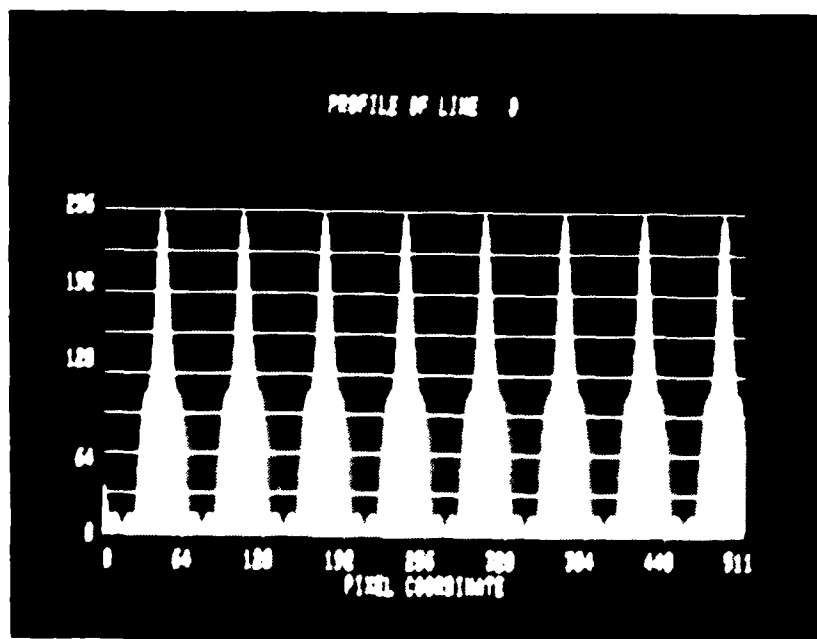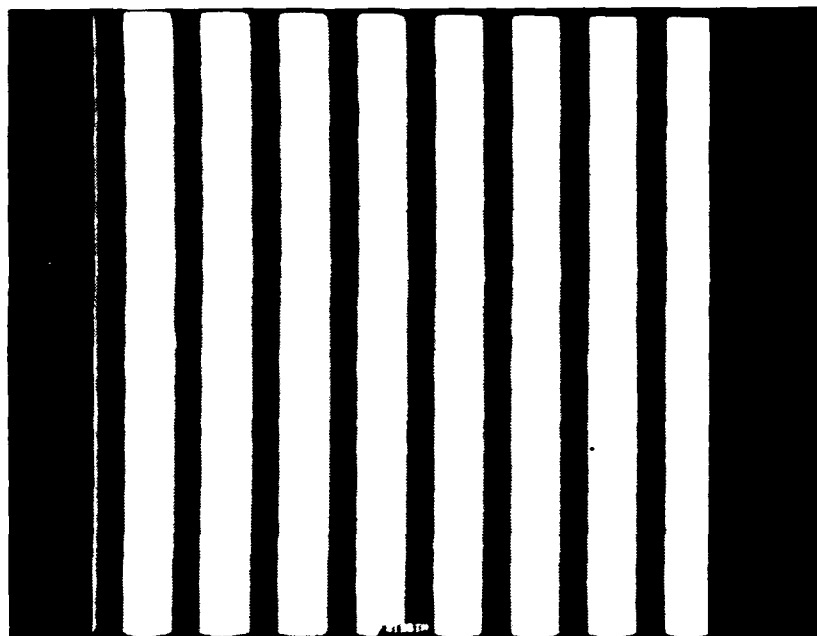Used for an Input

Figure 28. Image and Plot of the Output Transmission with
Sinusoidal Input

103

they represent noise in the recording. The THD is determined for the film by calculating the power in the output spectra with the constant and fundamental removed. This output power, normalized by the input power, is the THD for the 8E75 when exposed over its dynamic range. The THD for the 8E75 plate was 20 percent.

Now the input is predistorted and the resultant pattern used to expose an 8E75 plate. The predistorted image is shown in Figure 30 and the resultant transmission pattern on the plate shown in Figure 31. Although not perfect, this modified output is close to the desired result shown in Figure 27. The output spectrum is shown in Figure 32. The harmonic components, although not zero, are largely eliminated. The THD for the predistorted exposure was reduced to 4 percent, a factor of 5 improvement.

Significant improvement is possible by predistorting the desired pattern. The nonlinearity can be virtually eliminated by the use of a high order polynomial or spline fit. Once linearized, the film can be driven harder to take advantage of the entire dynamic range. This provides increased efficiency and permits a greater dynamic range in the input function to be recorded. Applied to continuous-tone CGH, this allows the simultaneous optimization of efficiency and signal-to-noise. Predistortion has no application to binary CGH because the response to a binary signal is inherently linear and no improvement is possible with predistortion.

In display holography, nonlinearity does not play an important role. The harmonic terms are diffracted at broad angles and do not significantly affect image quality. Because of increased efficiency, most display holograms are purposefully saturated. In a hologram for use as a matched filter, the linearity is very important. The harmonic terms contribute to noise and loss of efficiency. Even though some efficiency is gained by driving the film into saturation, light is diffracted to higher orders and this light is lost to the system. In an optical correlator with limited light budget, this loss may be intolerable. Such a system is the prime candidate for the advantages offered by the CGH utilizing computer linearization. Computer linearization is a vital step in optimizing the continuous-tone CGH matched filter.
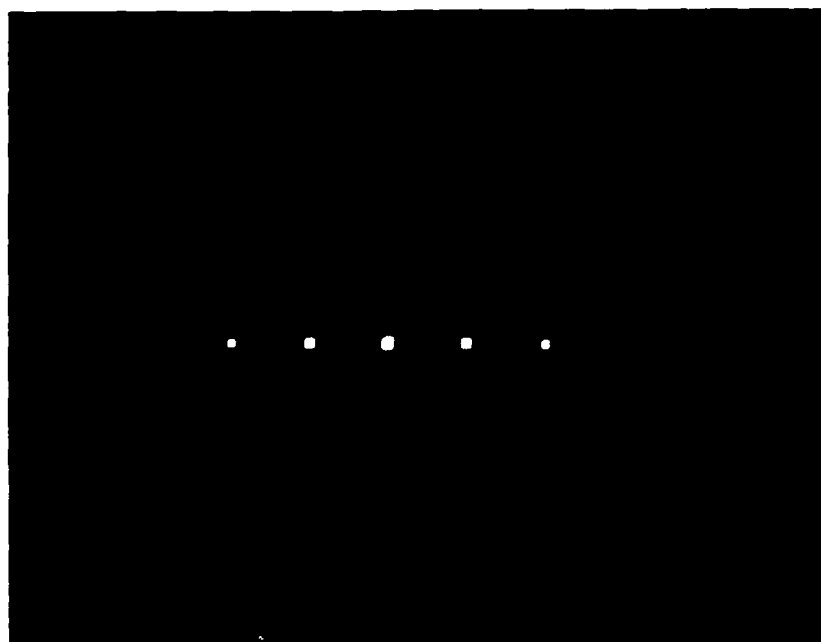
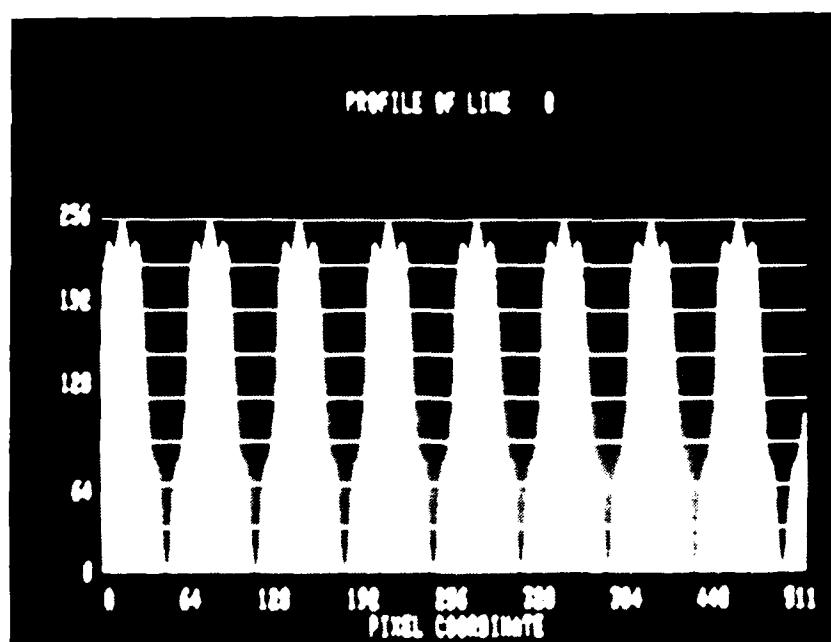Figure 29. Output Spectrum for a Sinusoidal Input

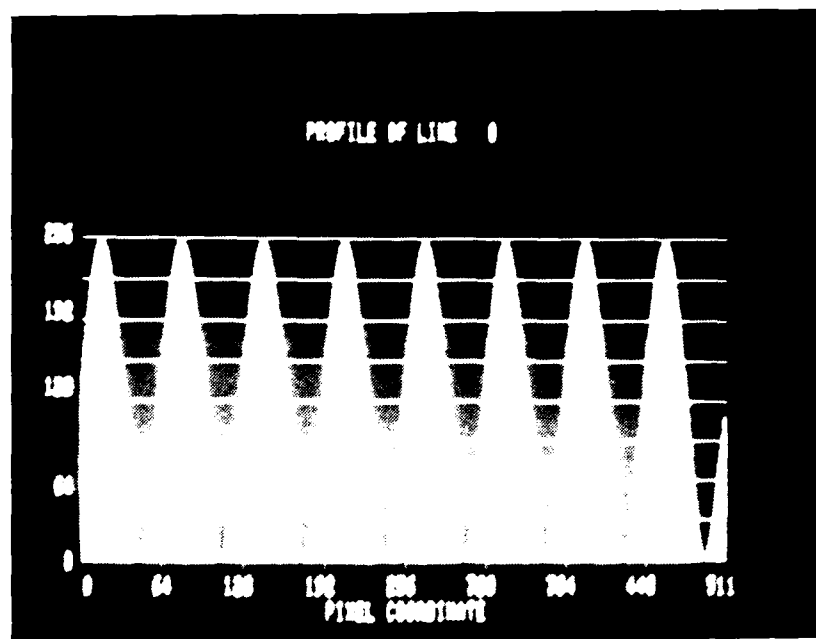Figure 30.   Image and Plot of the Predistorted Sinusoidal Grating
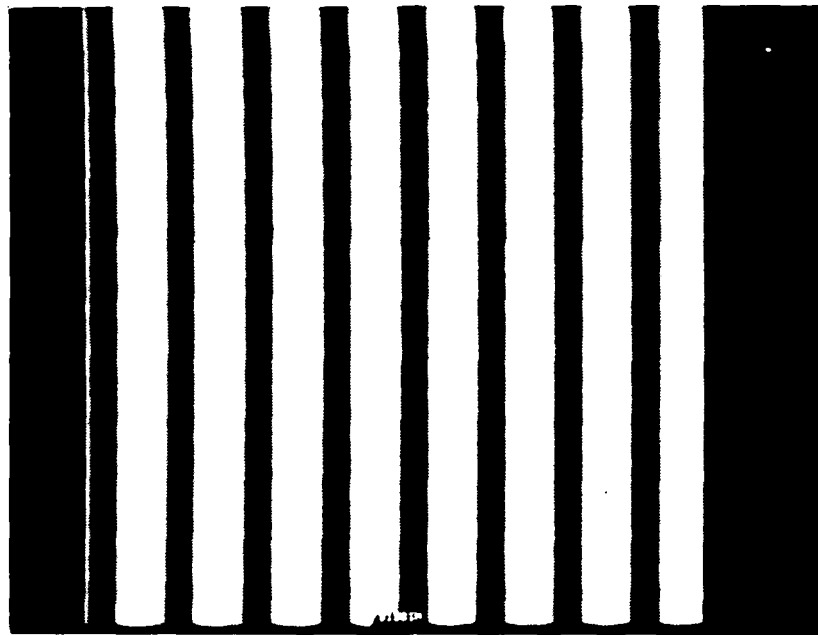Used for Input

Figure 31   Image and Plot of the Output Transmission for
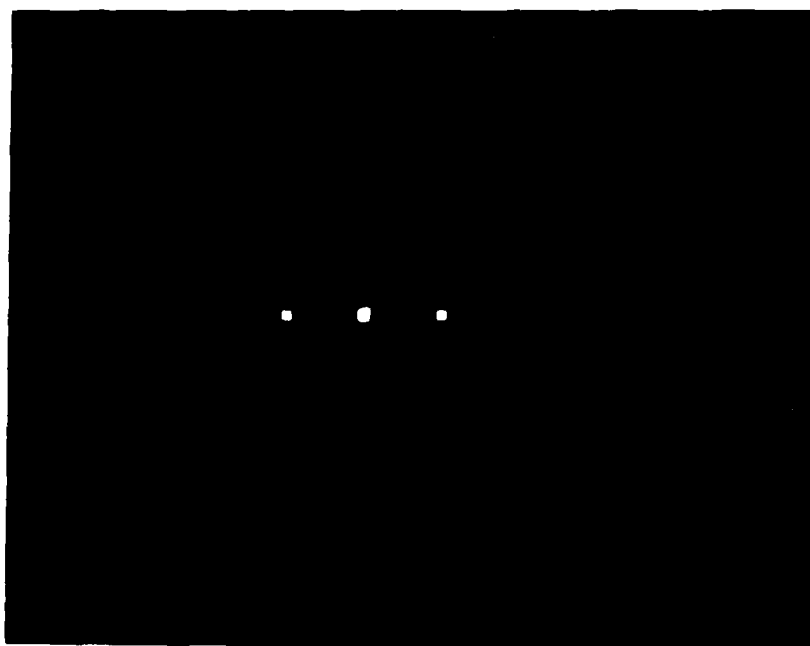   Predistorted Simusoidal Input

Figure 32. Output Spectrum for a Predistorted Grating Input

SECTION VII
SIMULATIONS

Hologram optimization has been described, based on such properties as signal-to-noise ratio and efficiency. The improvements due to some techniques are easily verified using analytical techniques. Other optimization techniques are based on filter content and thus the improvement is not determined until the actual filter response is applied. The filter function is typically complicated and defies analytical measurement. In those cases, it is possible to produce the hologram physically and test its actual response in an optical correlator. Though this truly represents the most realistic test environment, it is not convenient to produce a new hologram for each test case. In addition, with the many variables inherent in chemical development of film, it is difficult to repeat the same experiment and get the same result. With lengthy and careful efforts to standardize each experiment, meaningful results may be possible. Such an effort makes difficult the quick and exacting comparison necessary for hologram optimization.

Computers have provided a compromise between the analytical approach and the experimental approach. By simulating the analytical effects in the hologram and submitting digitized images to the simulation, a simulated experiment can take place. This simulation allows as much realism as desired, according to the extent of the simulation software. The advantages of digital simulation include the ease with which an experiment can be run, the ability to vary parameters, and repeatability. Assuming a sufficiently powerful computer is available, the cost of these digital experiments is small compared to their hardware implementation. A large number of computer experiments can be performed in the same time it takes to produce and process one optical hologram.

In the case of a computer-generated hologram, the digitized images and the hologram are already on the computer. It is a simple matter then to evaluate the hologram on the computer using simulation techniques. This step would naturally occur after the computation of the CGH but before physically writing the pattern to film. If the holographic pattern does not perform as expected in the simulation, it need not be produced until all problems are corrected. The computer allows many experiments to be used to optimize the CGH for its

specific application and to predict the performance before investing the time and expense of writing the pattern on film. This chapter describes these simulation techniques and shows specific examples and results for common CGH types.

1. TECHNIQUES FOR SIMULATING MATCHED FILTERS

Many of the digital processing techniques are described in Section III, but the simulations described here go beyond the creation of the hologram. The theory introduced in Section II shows that the correlation of two images can be accomplished by inverse transforming the product of the transforms of the two images. This approach is used in an optical correlator because of the ease with which Fourier transforms are accomplished using lenses. In each simulation, the effects are limited to those possible in a Vander Lugt optical correlator in which no active devices are used. That is, transmittance values cannot exceed one. All preprocessing is performed on the hologram, only to preserve the real-time capability of the optical correlator.

Figure 33 shows a flow chart of a simulation of an ideal correlation. In this case a hologram is not used to perform the correlation. Rather, the image transforms are multiplied directly in their complex form. Although this simulation could be used to represent the effects associated with an on-axis hologram capable of simultaneous amplitude and phase modulation, no hologram noise, film noise, or dynamic range limitations are included. This ideal correlation serves as a standard for latter comparison with off-axis hologram-derived correlations. There is an interaction between the preprocessing and the hologram that will affect the influence of the preemphasis. However, the ideal correlation simulation is useful for determining correlation properties of the images and preprocessing techniques independent of the influence of the hologram.

In Figure 33, the digitized reference image is stored in a matrix f_{ij}. Likewise, the digitized test image is stored in a matrix g_{ij}. The correlation image will contain more points than each of the input images. Specifically, the correlation image contains a number of points in each direction equal to the sum of the number of points of the two images in those directions. In general, the two images are equal in size and the correlation is twice the size

110

Reference Image   Test Image

| Embed in Zeroes | | Embed in Zeroes |

| Fast Fourier Transform | | Fast Fourier Transform |

| Reference Preemphasis |

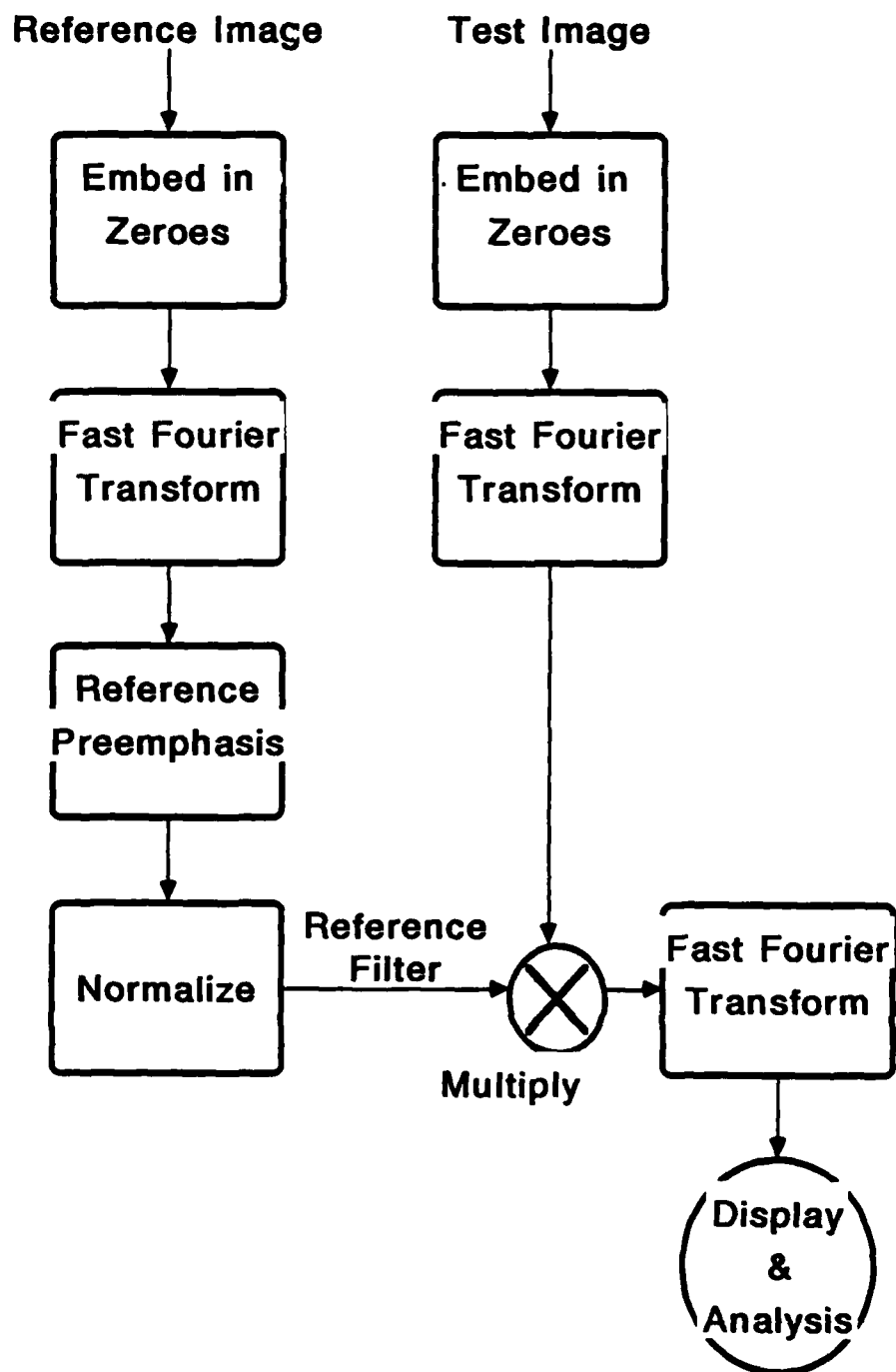| Normalize | Reference Filter   ⊗ | Fast Fourier Transform |

Multiply

(Display & Analysis)

Figure 33.   Computer Simulation of an Ideal Correlation

of the images in each direction. In order for the FFT routine to perform the
appropriate transformation resulting in a 2N by 2N image, the input images need
to be padded to create two images of size 2N by 2N. The padding consists of
placing the N by N image into a bed of zeros. That is, the matrix values in
the resultant image, f'_{ij}, are all zero except the N by N center portion which
contains the original image. This padded image, when Fourier transformed,
provides a matrix which is always smooth over two pixels in both directions. A
standard Cooley-Tukey Fast Fourier Transform (FFT) algorithm is used to provide
F'_{ij}, the Fourier transform of the reference image.

The normal correlation is performed by taking the product of F_{ij} and G_{ij}
point by point. This product is performed using real or double precision real
numbers and provides more than adequate dynamic range. In addition to a normal
correlation, the simulation software includes a preprocessing option. This
permits the reference image to be modified using a frequency emphasis or phase-
only filter. In the frequency emphasis option, the frequency plane values are
multiplied by a real valued coefficient based on the emphasis desired. In the
experiments shown in this report, the frequency emphasis applied is a gradient
filter. That is, the weight of each frequency component is equal to the radius
of that component. Because the reference image must account for the filter
effect for both the reference and the test image, the actual filter applied is
the square of the desired filter response. In this case the gradient squared
is known as a Laplacian filter or radius-squared weighting.

Another preprocessing technique possible in the ideal correlation simula-
tion is the phase-only filter. In this option, the magnitude at each location
is set to one while the angle of each complex value is left intact. This is
accomplished by dividing the real and imaginary components by the magnitude.
This normalizes the pattern such that the transmission is one at every location
except when the magnitude is zero. When the magnitude is zero, both components
are set to zero.

The Fourier transform of a test square is shown in Figure 34. The same trans-
form with high-frequency emphasized is shown in Figure 35. Note in Figure 12
that the edges are sharply accentuated in the modified image while broad areas
are dark. Despite the modified frequency, the image is still easily recog-
nized. The Fourier transform of a phase-only-filtered image has minimum

112

dynamic range, but the image is distorted by the extreme edge enhancement typical of phase-only filtering. When all preemphasis is completed, the reference and test patterns are normalized to possess continuous values between 0 and 1 inclusive. Normalization is performed to simulate the action of a transparency as would be used in an optical correlator and allows the determination of light efficiency.

Once the reference transform is modified according to the desired preprocessing technique, and the point by point product taken with the test image transform, the result is transformed again. Note that the inverse is not taken because the lenses which are simulated can only take forward Fourier transforms. Recall that the only difference between a forward and inverse transform is that the result will be inverted and perverted, equation (11). It is of no consequence in this case that the correlation is upside-down. Figures 36 through 38 show the auto-correlation of a square with no preemphasis, high-frequency emphasis, and phase-only filtering applied. Note that the correlation spikes with preemphasis are considerably sharper. This is to be expected as the correlation length is inversely related to the high-frequency content of the images.

Once the correlation image is obtained, an analysis program evaluates the signal-to-noise ratio and the efficiency of the correlation process. The signal-to-noise ratio is determined by the average of the correlation peak with its nearest six neighbors divided by the average of the entire correlation image.

$$\text{Signal-to-Noise} = \frac{\text{Correlation Peak}}{\text{Correlation Plane Average}} \qquad (88)$$

The efficiency is determined by dividing the energy in the correlation peak by the energy in the input test image. This efficiency is always less than or equal to 1 because the reference pattern is normalized and can only attenuate the test image energy. In the case of the phase-modulated reference, the efficiency can actually reach 1 if all of the diffracted light reaches the correlation peak. In this case, no attenuation occurs at each pixel but rather a phase shift occurs.
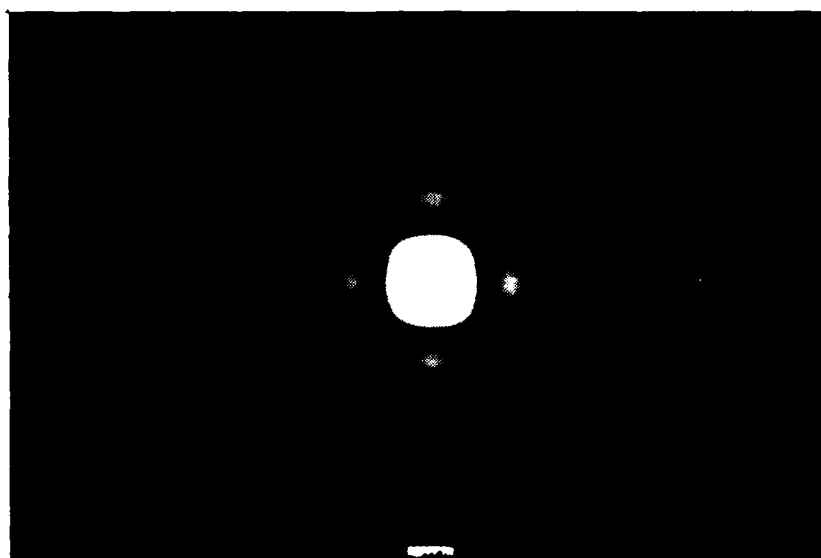
113

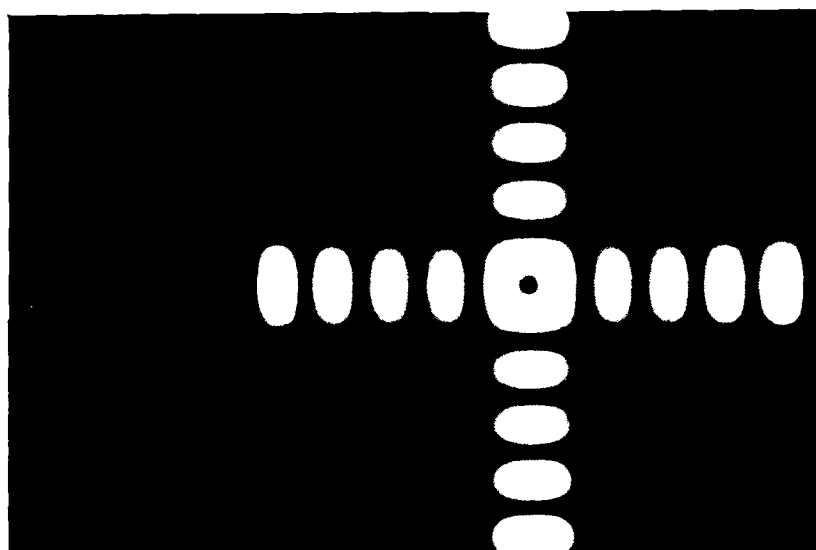Figure 34. Fourier Transform of a Reference Square

Figure 35. Fourier Transform of a Square with
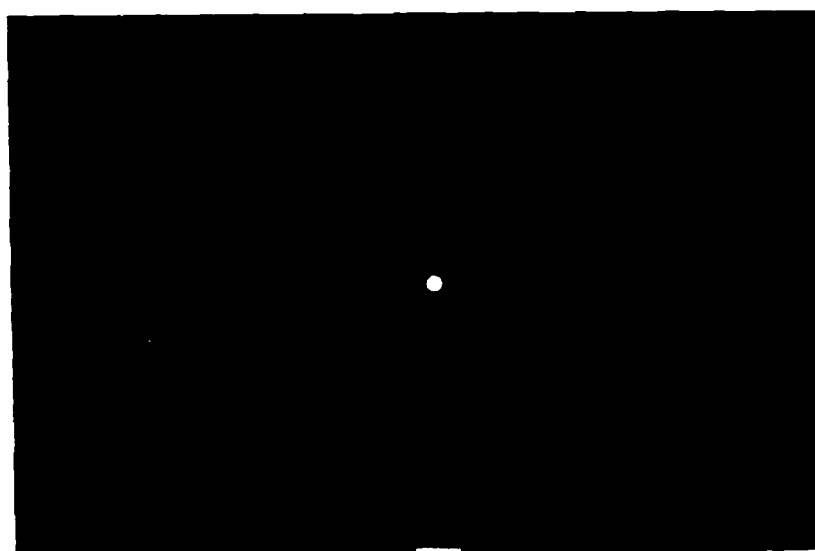High-Frequency Emphasis

Figure 36. Ideal Auto-Correlation of a Square
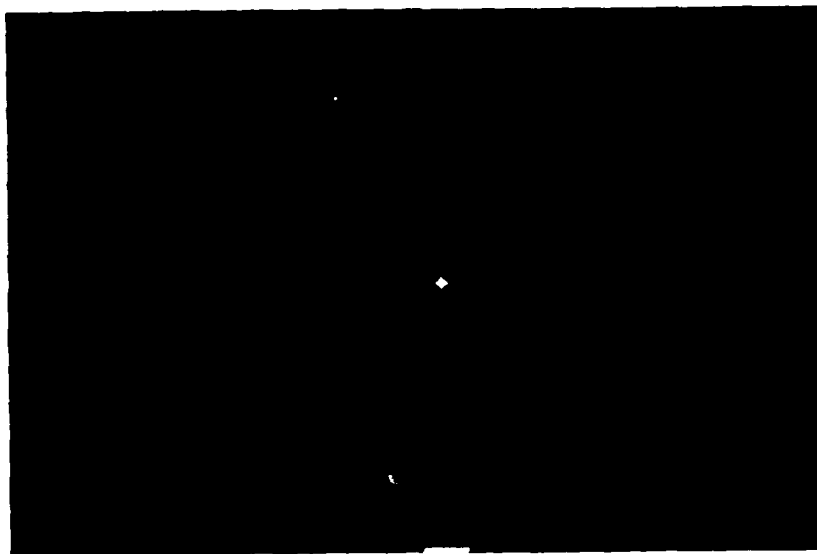with No Preemphasis

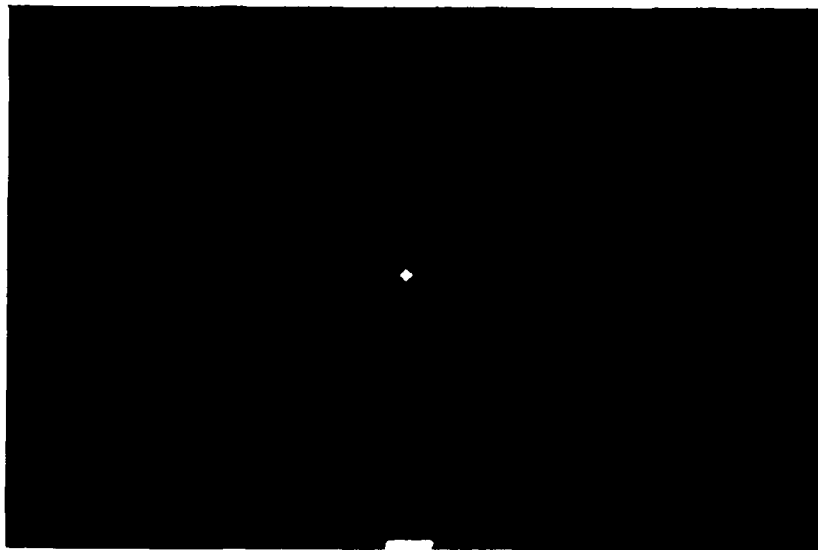Figure 37.   Ideal Auto-Correlation of a Square with
High-Frequency Emphasis

Figure 38. Ideal Auto-Correlation of a Square Using
Phase-Only Filtering

Total efficiency can be divided into two effects, medium efficiency and correlation efficiency. The medium efficiency depends on the attenuation by the reference pattern and can be computed by dividing the total light out of the reference pattern by the input energy.

$$\text{Medium Efficiency} = \frac{\text{Energy Leaving Hologram}}{\text{Energy in Test Image}} \qquad (89)$$

The medium efficiency goes to 1 for a phase-modulated hologram. The correlation efficiency is the energy in the correlation spike divided by the total energy output from the hologram. This efficiency term depends on how much of the output light is diffracted to the correlation spike rather than to other orders.

$$\text{Correlation Efficiency} = \frac{\text{Energy in Correlation Peak}}{\text{Energy Leaving Hologram}} \qquad (90)$$

The term Horner efficiency was originally applied to that which is here called the medium efficiency. (Reference 28). More recently, the Horner efficiency has been re-defined to describe the total efficiency including the correlation efficiency due to diffraction. (Reference 29). This new definition seems to describe more accurately Horner's original intent that the efficiency represent the entire correlation process including the transparency and diffraction effects.

$$\text{Total Efficiency} = \frac{\text{Energy in Correlation Peak}}{\text{Energy in Test Image}} \qquad (91)$$

The signal-to-noise and efficiency for the ideal auto-correlation of a square are tabulated in Table 1 for cases where no preprocessing, high-frequency emphasis, and phase-only filtering are applied.

2. SIMULATION OF A CONTINUOUS-TONE HOLOGRAM

The ideal correlation is useful for understanding the various preprocessing effects. In fact, for the on-axis hologram this ideal correlation model accurately predicts the actual results. However, holographic materials cannot

119

TABLE 1. SIGNAL-TO-NOISE RATIO AND EFFICIENCY OF AN IDEAL
AUTO-CORRELATION OF A SQUARE

	Signal-to-Noise Ratio	Medium Efficiency	Total Efficiency
No Preemphasis	661	44.9%	44.7%
High-Frequency Emphasis	967	0.1%	0.1%
Phase-Only Filtering	1631	100.0%	97.8%

record complex values, so off-axis techniques are required. The spatial modulation or other mapping from complex to real valued functions has a pronounced effect on the action of the reference filter. This step must be included to reflect the action of the hologram in an optical correlator. Figure 39 shows a block diagram of the simulation of an optical correlator using a continuous-tone hologram. Although quite similar to the ideal correlation model, this model includes a step which creates the computer-generated continuous-tone hologram.

This hologram, when used as the reference filter, provides a correlation similar to an ideal correlation. However, the differences can be quite pronounced. For example, the output not only includes the correlation, but also the convolution and other terms. These terms may, in some cases, overlap and cause degradation of the signal-to-noise ratio. In all cases, light will be lost to the convolution term and the on-axis term with a resulting loss of total efficiency. Obviously, the hologram has a significant effect on the result of the optical correlator and must be adequately modeled to obtain reasonable predictions of correlator performance.

The hologram produces not only the correlation plane but also other terms. As was discussed in Section III, the space-bandwidth requirement for the hologram is dependent on the spatial carrier frequency and the number of points in the reference image. Thus, based on these factors, the reference image must be padded in a field of zeros of the appropriate space-bandwidth. This requires greater padding than is needed in the ideal correlation and additional computing power for analyzing the same images.

In addition, there are modifications to the hologram which are possible when produced via computer generation. These modifications, discussed in the previous sections, are modeled here to predict their effect and usefulness. These hologram modifications include the use of nonlinear predistortion of the filter function to remove the distorting film response. The predistortion of the filter function would occur in the computer generation of the CGH. However, since the film is also modeled in this simulation, both the predistortion and film distortion must be included. To represent the overall effect, the predistortion must be the same as the one to be used in the production of the CGH. This predistortion model needs only to provide the accuracy desired in the CGH.

121

Reference Image   **Test Image**

```
┌─────────────┐            ┌─────────────┐
│  Embed in   │            │  Embed in   │
│   Zeroes    │            │   Zeroes    │
└─────────────┘            └─────────────┘
        │                          │
        ▼                          ▼
┌─────────────┐            ┌─────────────┐
│ Fast Fourier│            │ Fast Fourier│
│  Transform  │            │  Transform  │
└─────────────┘            └─────────────┘
        │                          │
        ▼                          │
┌─────────────┐                    │
│  Reference  │                    │
│ Preemphasis │                    │
└─────────────┘                    │
        │                          │
        ▼                          │
┌─────────────┐                    │
│   Encode    │                    │
│  Hologram   │                    │
└─────────────┘                    │
        │                          │
        ▼                          │
┌─────────────┐                    │
│ Pre – Distort│                   │
│   Hologram  │                    │
└─────────────┘                    │
        │        ┌─────────────┐   │
        │───────▶│  Model Film │   │
        │        │  Distortion │   │
        ▼        └─────────────┘   │
    ╭───────╮          │           │
   ╱  CGH    ╲         ▼           ▼
  │  Writer   │  ┌─────────────┐  ⊗ ──▶┌─────────────┐
   ╲         ╱   │ Amplitude or│       │ Fast Fourier│
    ╰───────╯    │Phase Modulat│──────▶│  Transform  │
                 └─────────────┘       └─────────────┘
                                              │
                                              ▼
                                         ╭───────╮
                                        ╱ Display ╲
                                       │     &     │
                                        ╲ Analysis ╱
                                         ╰───────╯
```
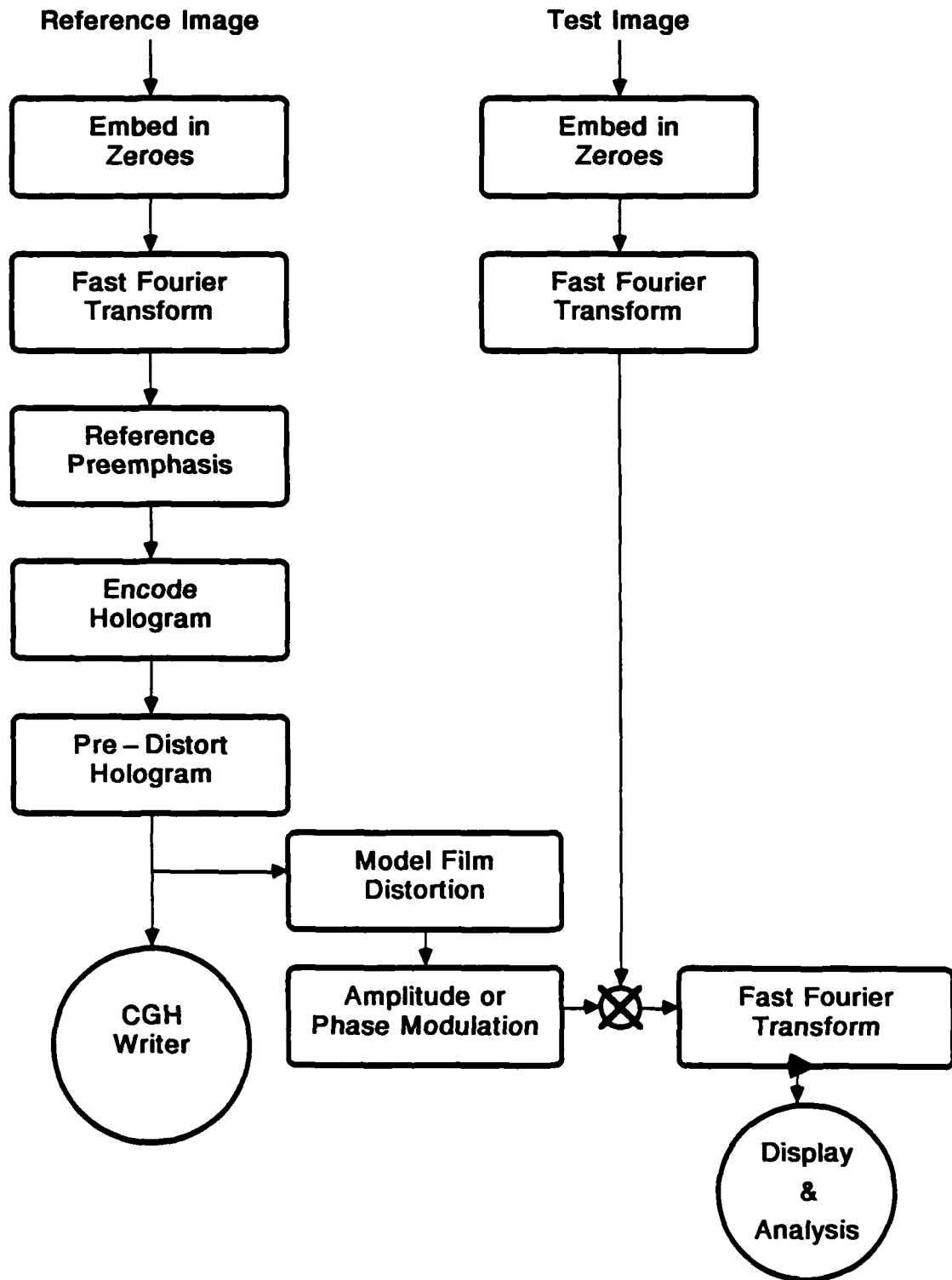
Figure 39.   Flow Chart for Continuous-Tone Hologram
            Simulation

However, the model of the film distortion must be as accurate as possible to provide an accurate simulation of the overall transfer function. Thus, while a third or fifth order fit may suffice for the predistortion, the model for the film distortion may require a higher order or a spline fit.

The type of modulation utilized in the hologram is also modeled. Recall that the preemphasis or phase-only filtering of the information used to create the hologram is independent of the modulation technique incorporated in the physical hologram. That is, the hologram may be produced as an amplitude hologram with constant phase but variable transmission, or produced as a phase hologram with constant amplitude but variable phase. The proper choice of hologram is modeled in the simulation by mapping the hologram pattern to transmission or phase. This modified amplitude or phase pattern is used as the reference and is multiplied by the Fourier transform of the test image. The following steps of the correlation and analysis are identical with the ideal correlation simulation with the exception that the correlation peak is no longer centered on-axis.

Figure 40 shows the continuous-tone hologram of a square, produced by the simulation. The carrier frequency was maximized in this case as half the sampling frequency to produce the greatest separation of the various output terms. The Fourier transform shown in Figure 34, a sinc function, is spatially modulated by a sinusoidal term diffracting the correlation and convolution terms off-axis. Figures 41 and 42 show the same hologram but with high-frequency emphasis and phase-only filtering. Figure 43 shows the auto-correlation of a square using a continuous-tone hologram. Figures 44 and 45 show the same auto-correlation with high-frequency emphasis and phase-only filtering. Table 2 lists the performance of the continuous-tone hologram when the various preprocessing techniques are applied.

3. SIMULATION OF A BINARY HOLOGRAM

The next case to be considered is the binary hologram. The binary hologram differs considerably from the continuous-tone hologram in that no gray-scale values are permitted. This requires that the dynamic range in the hologram be obtained through the use of additional points. This additional space-bandwidth

123
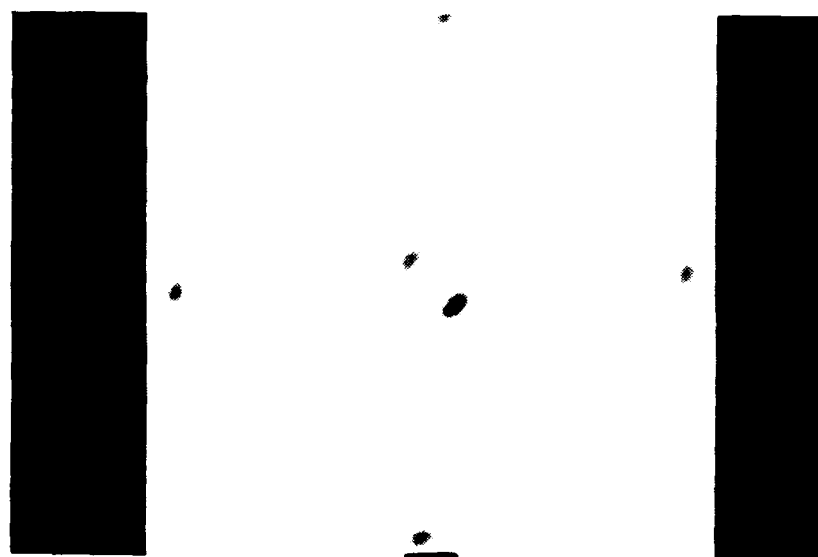
Figure 40. Continuous-Tone CGH of a Square

Figure 41. Continuous-Tone CGH of a Square With High-Frequency Emphasis
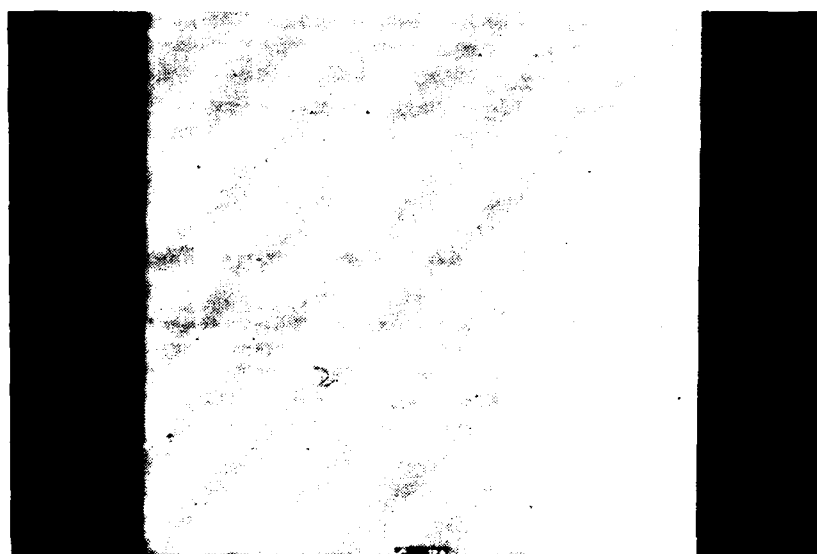
Figure 42. Continuous-Tone CGH of a Square with
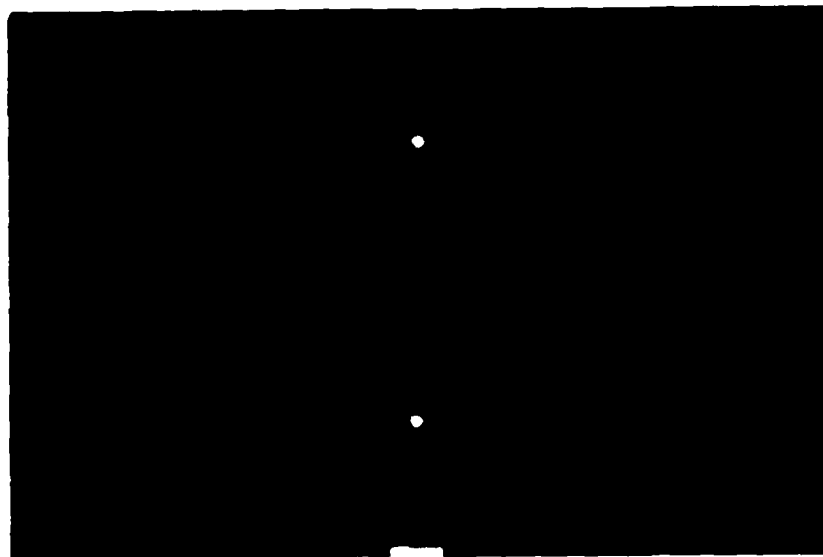Phase-Only Filtering

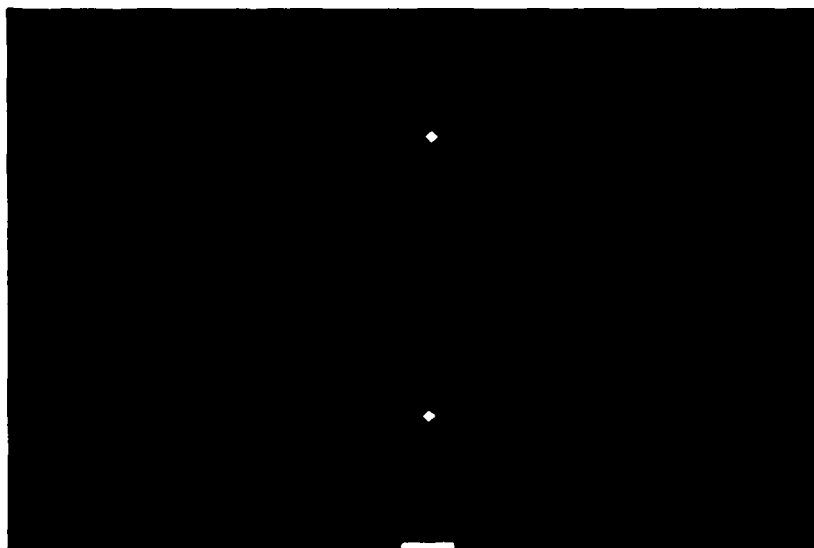Figure 43.   Auto-Correlation of a Square Using a
Continuous-Tone CGH

127

Figure 44. Auto-Correlation of a Square Using a
Continuous-Tone CGH with High-Frequency
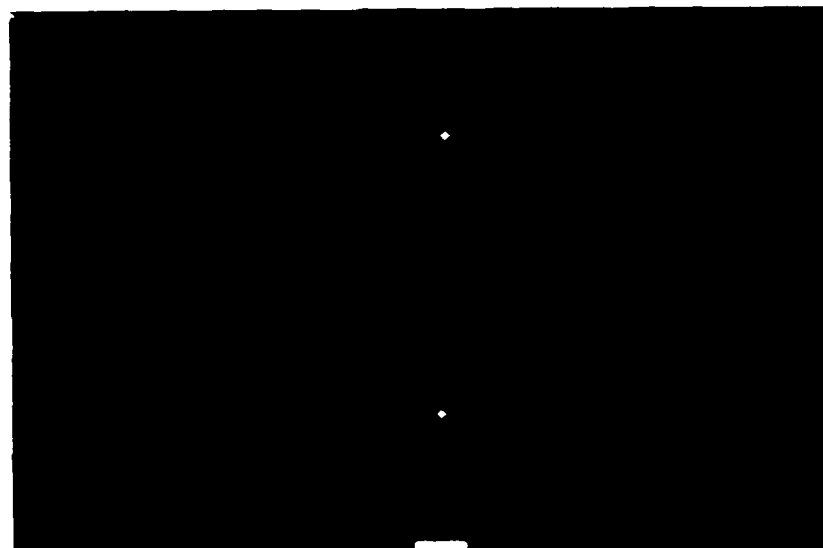Emphasis

128

Figure 45.   Auto-Correlation of a Square Using a
              Continuous-Tone CGH with Phase-Only Filtering

TABLE 2. SIGNAL-TO-NOISE RATIO AND EFFICIENCY FOR CONTINUOUS-TONE CGH

	Signal-to-Noise Ratio	Medium Efficiency	Total Efficiency
Absorption Hologram			
No Preemphasis	661	30.6%	2.56%
High-Frequency Emphasis	902	88.2%	0.04%
Phase-Only Filtering	1630	37.5%	5.87%
Phase Hologram			
No Preemphasis	661	100.0%	18.11%
High-Frequency Emphasis	902	100.0%	1.70%
Phase-Only Filtering	1630	100.0%	15.78%

requirement impacts the simulation and testing as well as the generation of the CGH. Figure 46 shows a block diagram of the binary hologram simulation. The structure is the same as the ideal and continuous-tone correlation with only small changes. The predistortion and distortion of the film is unnecessary for binary holograms since the two points can always describe a straight line and thus are always linear. It is meaningless to include the film response to the binary signal. The programming of the binary hologram simulation differs from the continuous-tone case in that the variables used to describe the holo-gram pixel values are not continuous, but rather, assume only integer values. When an amplitude hologram is selected, the transmission assumes values of zero or one. When a phase hologram is selected, the transmission is set to one but the phase assumes values of zero or pi.

The preprocessing and modulation options are the same as in the previous cases but there is an additional choice in the type of binary hologram. As was discussed in Section III, many mappings are possible to convert the complex filter function to a real binary pattern. Presently, Lohmann, Lee, and Allebach-Keegan (A-K) type holograms are available to the CGH algorithm and the simulation. Figures 47, 48, and 49 show the A-K hologram of the square using no preprocessing, frequency emphasis and phase-only filtering. Figures 50 and 51 show the auto-correlation of the square using the A-K hologram with fre-quency emphasis and phase-only filtering. Table 3 shows the signal-to-noise and the efficiency for the auto-correlation of the square using the A-K hologram.

4. AN EXAMPLE USING AN SDF AS A REFERENCE

The auto-correlation of the square is theoretically interesting and pro-vides a common tool by which various techniques can be compared. Additionally, the auto-correlation of a square is a problem which can be solved analytically for many of the types of holograms, lending credibility to the simulation results. However, the real power of the simulations occur when they are applied to more complicated imagery. Actual images with complicated shapes and patterns are impossible to correlate analytically but rather must be correlated by a computer. Figure 14 shows various images from a training set which was used to create the Synthetic Discriminant Function (SDF) shown in Figure 15. The SDF was created from 36 views of the object rotated 10 degrees between each
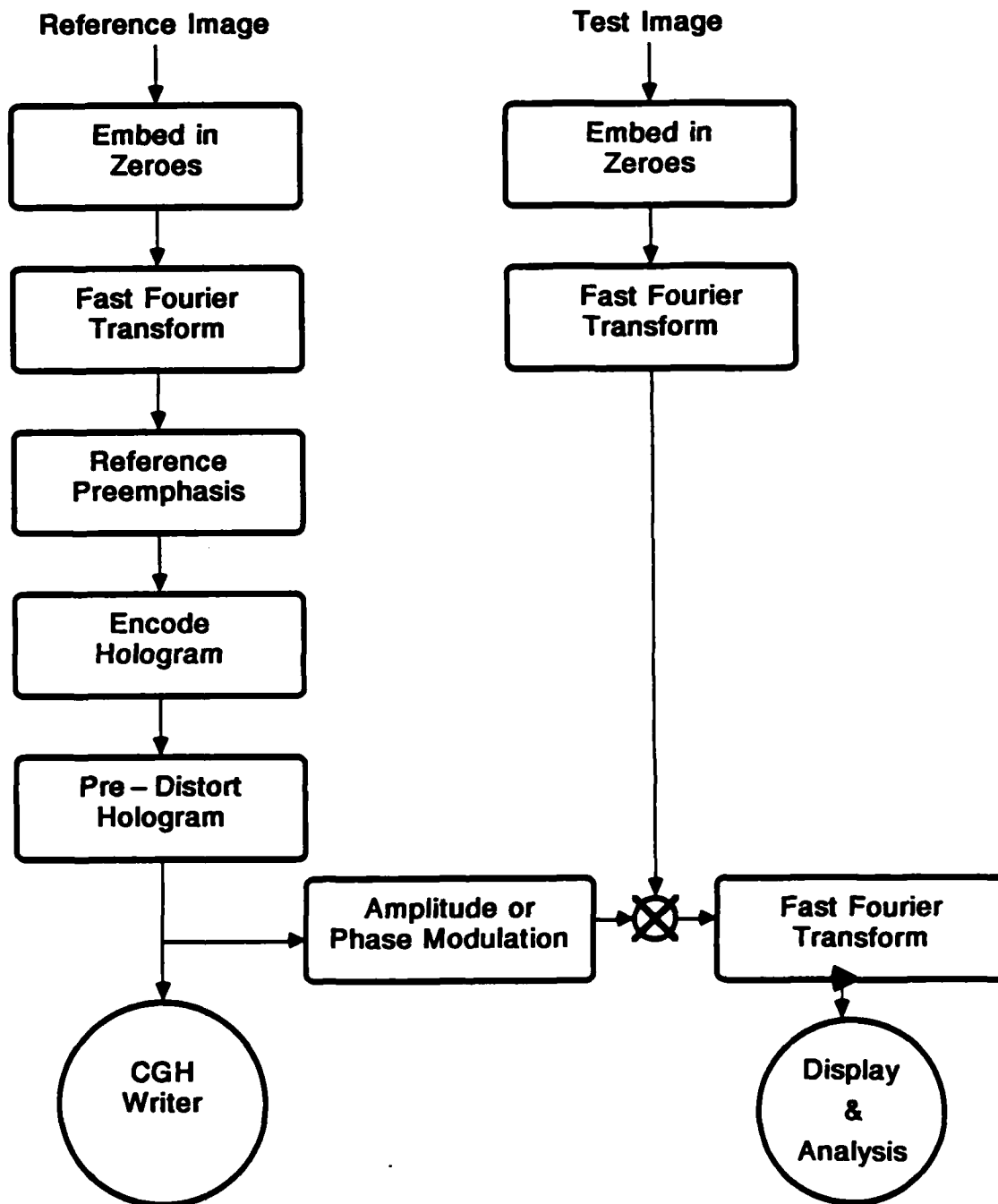
131

Figure 46. Flow Chart for the Binary Hologram Simulation

132

Figure 47. A-K Binary Hologram of a Square

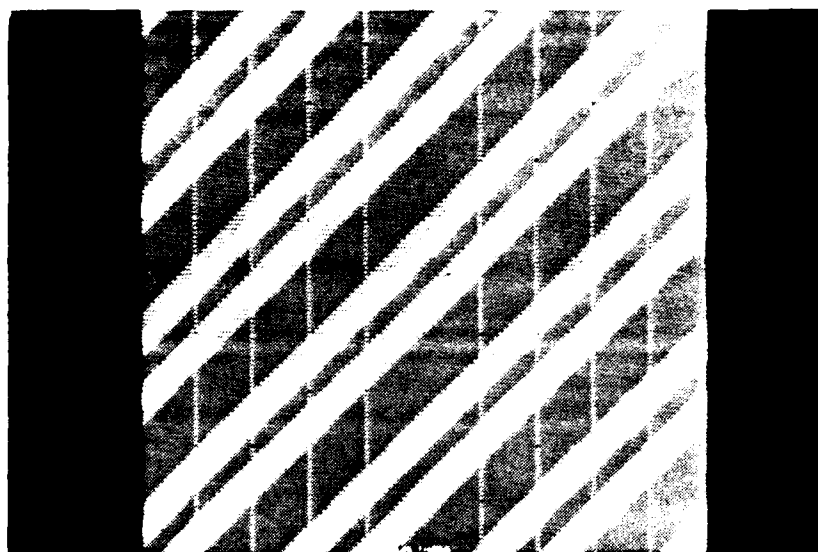Figure 48.   A-K Binary Hologram Using High-Frequency Emphasis

Figure 49.   A-K Binary Hologram of a Square with
Phase-Only Filtering

Figure 50. Auto-Correlation of a Square Using an A-K Binary Hologram with High-Frequency Emphasis
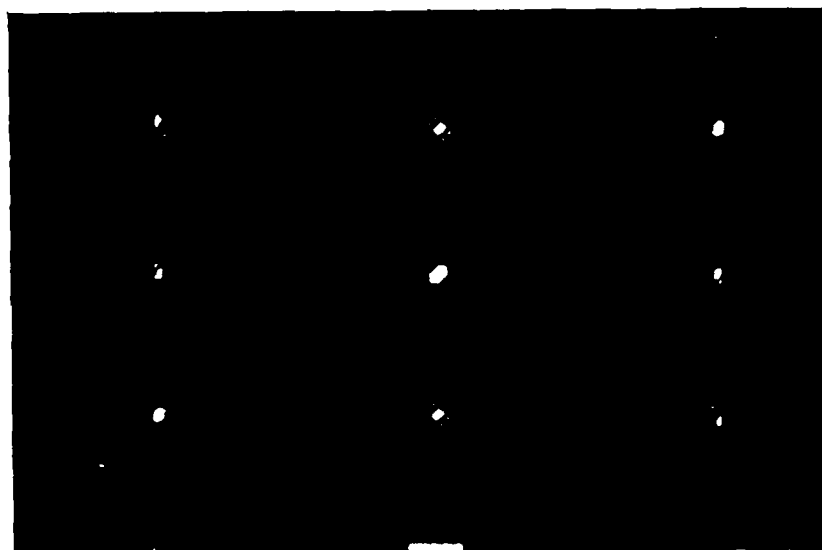
Figure 51   Auto-Correlation of a Square Using an A-K
Binary Hologram with Phase-Only Filtering

TABLE 3. SIGNAL-TO-NOISE RATIO AND EFFICIENCY FOR AN A-K HOLOGRAM
OF A SQUARE

	Signal-to- Noise Ratio	Medium Efficiency	Total Efficiency
Absorption Hologram			
No Preemphasis	66	15.6%	1.97%
High-Frequency Emphasis	287	42.5%	9.17%
Phase-Only Filtering	615	34.6%	7.22%
Phase Hologram			
No Preemphasis	66	100.0%	7.88%
High-Frequency Emphasis	287	100.0%	36.68%
Phase-Only Filtering	615	100.0%	28.87%

view. The SDF should therefore give reasonable cross-correlations with each of
the 36 views. An A-K hologram of the SDF using high-frequency emphasis is
shown in Figure 52. The cross-correlation of the SDF with the 40 degree view
is shown in Figure 53. The signal-to-noise and efficiencies for several repre-
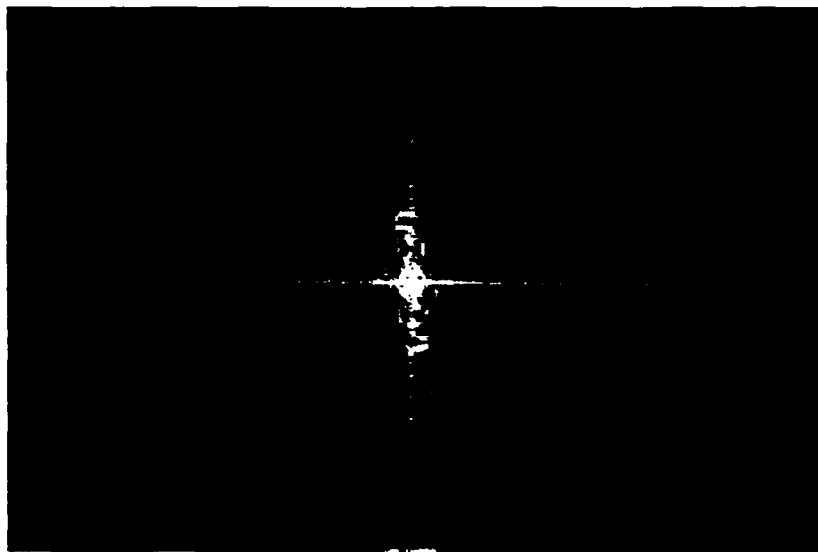sentative cases are shown in Table 4.

Figure 52. A-K Binary Hologram of the SDF Using
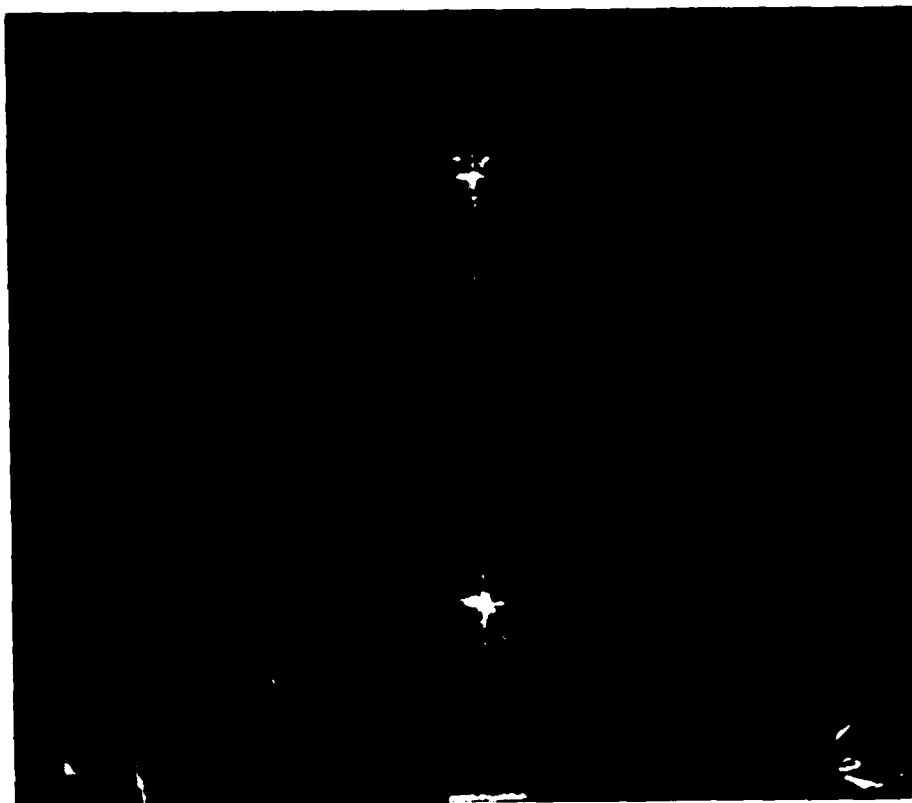High-Frequency Emphasis

Figure 53. Correlation of a Test Image at 30° and the
SDF Using an A-K Hologram with High-Frequency
Emphasis

TABLE 4. SIGNAL-TO-NOISE AND EFFICIENCY OF AN A-K HOLOGRAM OF AN SDF
CORRELATING WITH MEMBERS OF THE TRAINING SET

	Signal-to-Noise Ratio	Medium Efficiency	Total Efficiency
Absorption Hologram			
No Preemphasis			
30°	2.1	7.4%	0.03%
90°	2.1	8.1%	0.03%
130°	2.1	9.1%	0.04%
330°	2.1	8.2%	0.04%
High-Frequency Emphasis			
30°	7.2	34.0%	0.50%
90°	7.7	35.5%	0.52%
130°	8.0	37.6%	0.62%
330°	7.2	34.1%	0.51%
Phase-Only Filtering			
30°	13.4	27.7%	7.62%
90°	13.5	27.7%	6.85%
130°	14.9	27.2%	7.27%
330°	13.6	27.5%	6.42%
Phase Hologram			
No Preemphasis			
30°	2.1	100.0%	0.13%
90°	2.1	100.0%	0.13%
130°	2.1	100.0%	0.16%
330°	2.1	100.0%	0.14%
Emphasis			
30°	7.2	100.0%	2.00%
90°	7.7	100.0%	2.11%
130°	8.0	100.0%	2.49%
330°	7.2	100.0%	2.02%
Phase-Only Filtering			
30°	13.4	100.0%	3.05%
90°	13.5	100.0%	2.74%
130°	14.9	100.0%	2.91%
330°	13.6	100.0%	2.57%

SECTION VIII
OPTICAL IMPLEMENTATION

As shown in the previous sections, tremendous control over the matched filter hologram is possible when produced using computer-generation techniques. Figure 54 shows an interferometrically produced matched filter demonstrating the typical saturation in the dc and low frequency terms. The interferometric holograms include a magnitude squared term which produces a wide dynamic range signal to be recorded on the film along with the desired information. The non-linearity of the film limits the dynamic range of the information to be recorded interferometrically. In the matched filter shown in Figure 54, the useful, unsaturated information exists only in the higher harmonics shown enlarged. When produced as a computer-generated hologram, the matched filter need not include the magnitude squared term and permits a wider dynamic range to be recorded. The digital representations of the CGH are easily tested, analyzed, and modified to obtain the optimum pattern. To produce the hologram, a writing device is needed to convert the digital representation into a physical transparency for use in an optical correlator.

1. TECHNIQUES FOR OPTICAL IMPLEMENTATION

Certainly one of the major shortcomings of computer-generated holograms has been the limited space-bandwidth product which could be accommodated. This has been limited in part by the computational facilities required for encoding facilities required for encoding the holograms, but more importantly, the plotting devices have been the major bottleneck. The typical procedure for fabricating a computer generated hologram is to have the digitized interference pattern, which has been calculated and encoded by computer, drawn to a large scale by a computer-driven plotter. The drawing is then reduced photograph-ically onto high-resolution film to the desired final size. Unfortunately, errors are introduced in the plotting and photo-reduction processes. In addition, optical plotting devices are limited in spatial resolution and space-bandwidth product, typically to 106 pixels. The resolution is limited by the number of discrete points which can be placed on the paper and the accuracy of the copy process. This printing technique is strictly binary and requires binary mapping techniques as described in Section III.
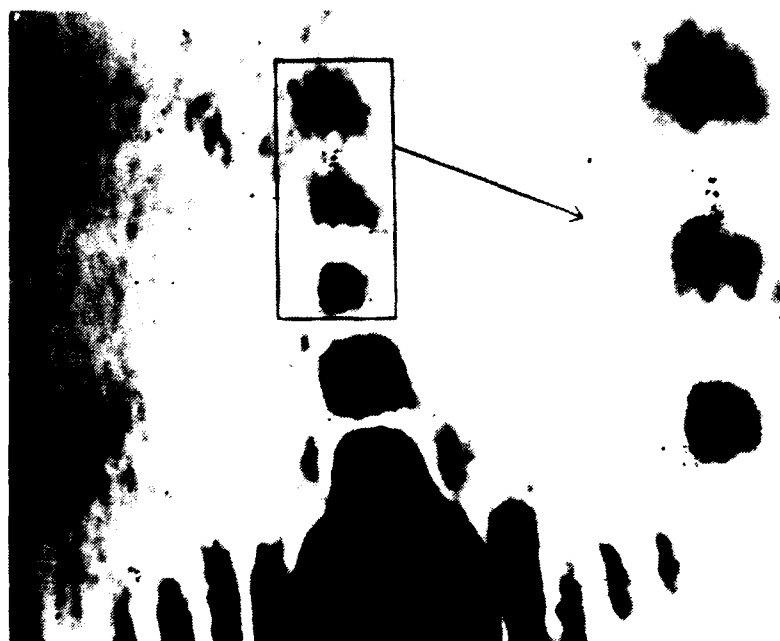
Figure 54. Photo of an Interferometrically Produced
Optical Matched Filter

To produce continuous-tone holograms, a gray-scale writing device is needed. With such a device, a pattern can be produced containing transmittance values ranging continuously from zero to one. One of the earliest examples is the rotating drum scanner. Such scanners have been used in the newspaper industry for years to send pictures electronically over phone lines. In the rotating drum scanner, a photographic film is wrapped about a drum rotating at a fixed speed. As the drum rotates, an incoherent light source is focused to a spot on the film and moves sideways along the length of the film. In this fashion, the light source scans a helix along the drum, providing a raster scan on the film. The light is intensity modulated according to the pattern to be written to the film. These drum writers are commonly available from a number of companies, along with interfaces for most common computers.

The disadvantages of such a system include spatial accuracy and film format limitations. The mechanical drum and light source drive have inherent position noise which limit the ultimate accuracy. The visible light source can, in general, be focused to a spot 10 micrometers in diameter. To produce a holo-gram on a glass plate or a film format not compatible with the drum fixture, a second copying process becomes necessary as with the plotters and printers. This step is tedious and introduces further noise.

To avoid this copy process, the plate or film can be mounted on a flat surface and the light source scanned in both horizontal and vertical direc-tions. This is usually accomplished using a fixed light source and scanning mirrors. A computer controls the source intensity and the mirror deflection. Such a system can quickly write onto standard film backs including glass plates. Such two dimensional scanners are not currently available from commer-cial sources but have been produced in several laboratories. These scanner writers produce position accuracies similar to that possible with the drum writers but provide greater writing speed and flexibility.

Another technique which is becoming popular is to use commercial image display systems coupled with a camera which convert the digital information from a computer to a raster-scan television image. This image, viewable on a normal television monitor, is easy to see and manipulate with the computer. The screen is then photographed with a camera mounted before the cathode-ray tube (CRT). Several cameras are specifically designed for this task. One in

particular contains a flat field CRT, a lens system, and a number of camera backs to accommodate most film formats, including glass plates. The CRT exhibits excellent positional stability and gray scale accuracy sufficient for 8-bit resolution. The camera, shown in Figure 55 is versatile, convenient and fast, but is currently limited to spot sizes of around 25 micrometers. This accuracy could probably be made comparable with the drum and scanner systems with modification to the camera optics. A more important limitation to the video based systems is the space-bandwidth product. The standard video format permits only computer images 512 by 512 points or smaller. This problem is overcome by a new digital camera capable of 4096 by 4096 points. However, this larger format precludes the use of common and convenient television equipment to view the entire image.

There is another device, which combines the CRT imaging system with a scanning capability and is shown in Figure 56. This device, created for the sole purpose of producing holograms by Caulfield et al (Reference 61), uses a CRT imaged onto film. However, the Holowriter translates the photographic plate such that many CRT fields can be placed together, side-by-side. In this way, the convenience of CRT imaging is combined with a capability to produce holograms of enormous space-bandwidth-products by interweaving frames. The Caulfield system is capable of 10 micrometer resolution with 6-bit amplitude accuracy. The system is also capable of blurring pixels together in order to gain better amplitude resolution at the expense of position resolution.

Each of the systems thus described is capable of continuous-tone response and uses visible sources to expose photographic films or plates. The use of visible sources causes a basic limitation in the ultimate spot size. A diffraction-limited system might be capable of 1 micrometer resolution at visible wavelengths but, as seen in each of these systems, 10 micrometers or larger spot sizes are typical. To produce CGHs of finer detail, shorter wavelengths are necessary. Some improvement is obtained by writing with ultraviolet light such as in the pattern generator, which uses an excimer laser. However, for significant improvement an electron-beam is used.

The electron-beam writing system pictured in Figure 57 was produced for large scale integrated circuits (Reference 67). This fine structure capability lends itself readily to the production of high resolution holograms. This

Figure 55.   Cathode-Ray Tube and Camera

Figure 56. Cathode-Ray Tube Imaged onto a Translation Table

system is capable of 0.5 micrometer resolution over several millimeters of work space. The primary advantage of the E-beam system is the fine detail size. With 0.5 micrometer resolution, high spatial frequencies are possible, thus providing large angular outputs from an optical correlator. This permits the incorporation of high space-bandwidth products in the hologram while minimizing the size of the optical correlator. As with each of the other systems, phase modulation holograms can be produced with the E-beam system by etching various depths into the glass. At this time, the E-beam device is strictly binary. When the resist is sufficiently exposed, the etchant removes the metal entirely. Thus, to maintain a large dynamic range, spatial resolution may need to be discarded in exchange for amplitude resolution using the techniques described in Section III.

The E-beam systems are still too expensive for most laboratories to own. These systems cater primarily to the integrated circuit industry, and hologram producers must wait in line. Turn-around times of over a month are typical, with cost per hologram exceeding $1000. Obviously, the E-beam holograms, while superior in their spatial resolution, must be supplemented with in-house holograms of lesser quality until research and testing indicate that a hologram is ready for production.

2. EXAMPLES OF CGH MATCHED FILTERS

The following paragraph does not attempt to quantitatively verify the simulations. However, the figures shown here indicate that indeed the appropriate patterns can be placed on film or glass plates, and that their reconstructions verify qualitatively results predicted by the appropriate simulations. In each case, the holograms shown here were produced using the Allebach-Keegan (A-K) algorithm and written to chrome-on-glass plates using the E-beam writer. These CGH matched filters are amplitude-modulated holograms and do not provide the high efficiency possible with phase-modulation relief holograms. However, the holograms do include examples of various preprocessing techniques, namely frequency emphasis and phase-only filtering.

Figure 58 shows magnified views of one of the E-beam written holograms indicating the fine detail possible. Figure 59 contains A-K CGH matched filters using no preprocessing, frequency emphasis and phase-only filtering.
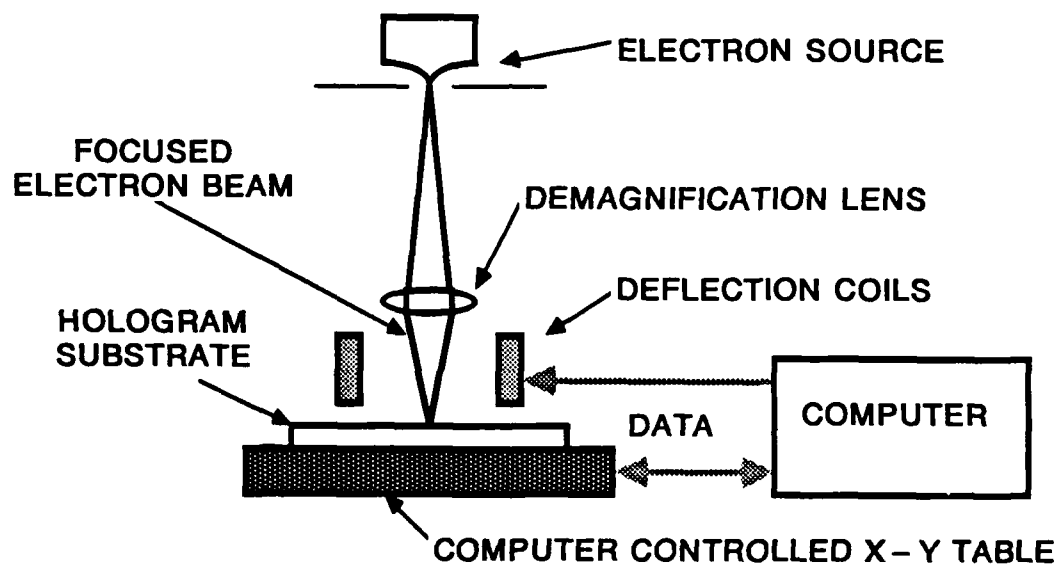
149

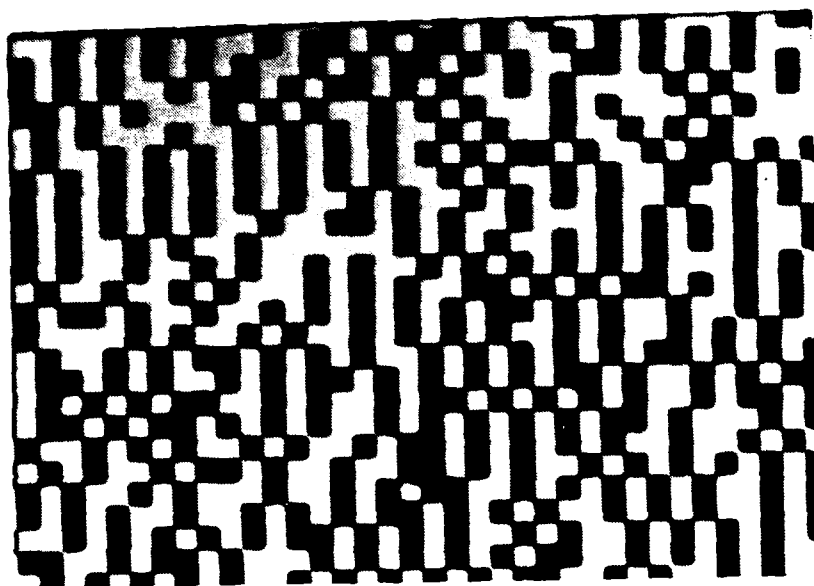Figure 57. Electron-Beam Writing System

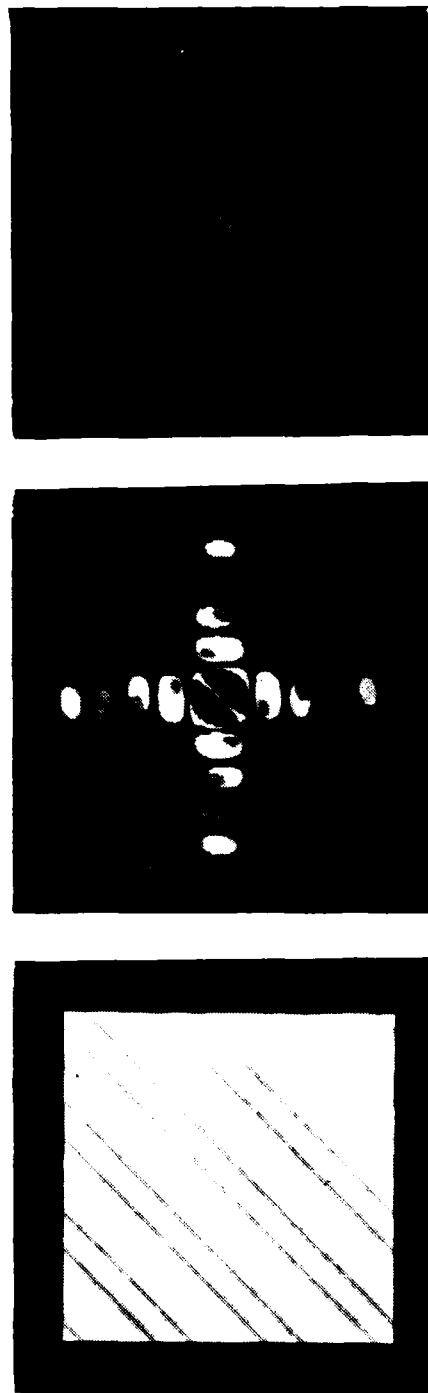Figure 58.   Magnified Views of a Binary Hologram

Figure 59. A-K CGH Matched Filters, Using a Square as a Reference,
Produced on the E-Beam Writer. The holograms shown
measure 1 mm along a side and include no preprocessing,
high-frequency emphasis, and phase-only filtering.

These photographs are taken of the actual holograms produced from the patterns shown in Section VII. Figures 60 through 62 show the reconstruction from each of the holograms in Figure 59. In each case, the hologram was placed in on-axis collimated light to produce the pattern in transmission. Figure 63 shows holograms of the letters AFATL using high-frequency emphasis and phase-only filtering. Figure 64 and 65 show the appropriate reconstructions of those holograms. The final hologram in Figure 66 is of a A-K CGH matched filter of an SDF using phase-only filtering. The reconstruction is shown in Figure 67 and should be compared to the original pattern shown in Figure 15.

Figure 60.   Reconstruction from an A-K CGH Matched Filter
of a Square Using no Preemphasis

154

Figure 61. Reconstruction from an A-K CGH Matched
Filter of a Square Using High-Frequency
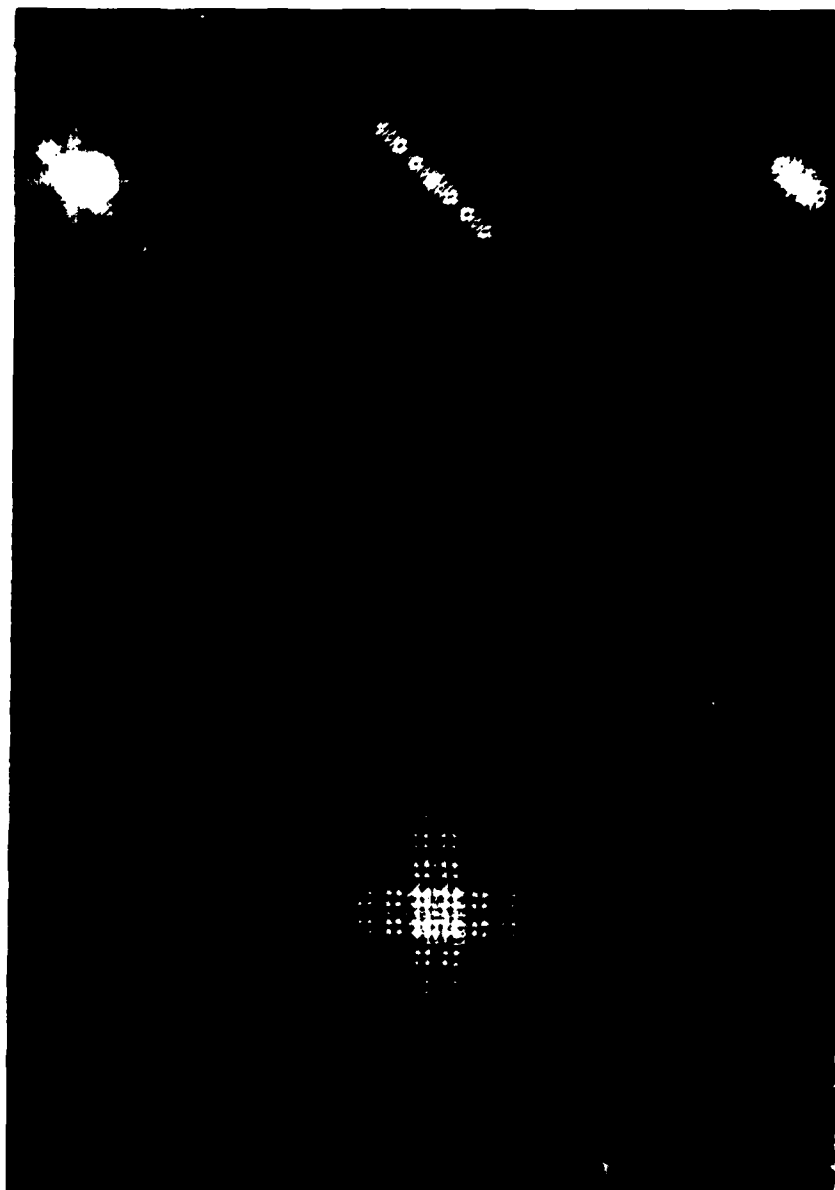Emphasis

155

Figure 62.   Reconstruction from an A-K CGH Matched
             Filter of a Square Using Phase-Only Filtering
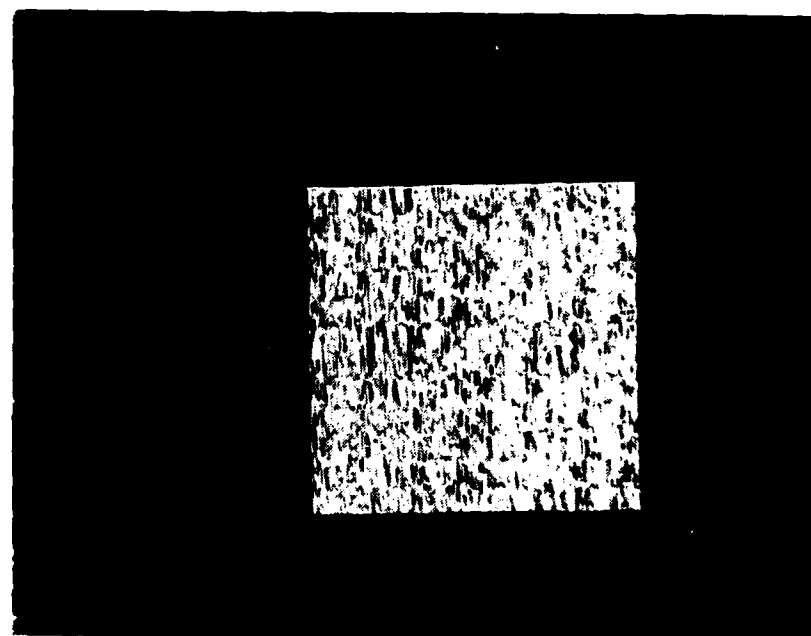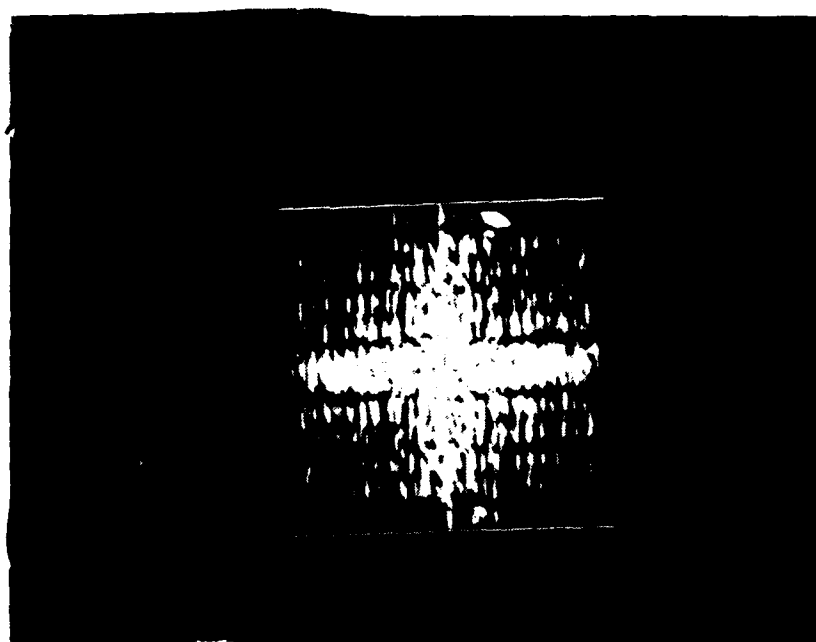
156

Figure 63.   A-K CGH Matched Filter of the Letters AFATL
             Using High-Frequency Emphasis and Phase-Only
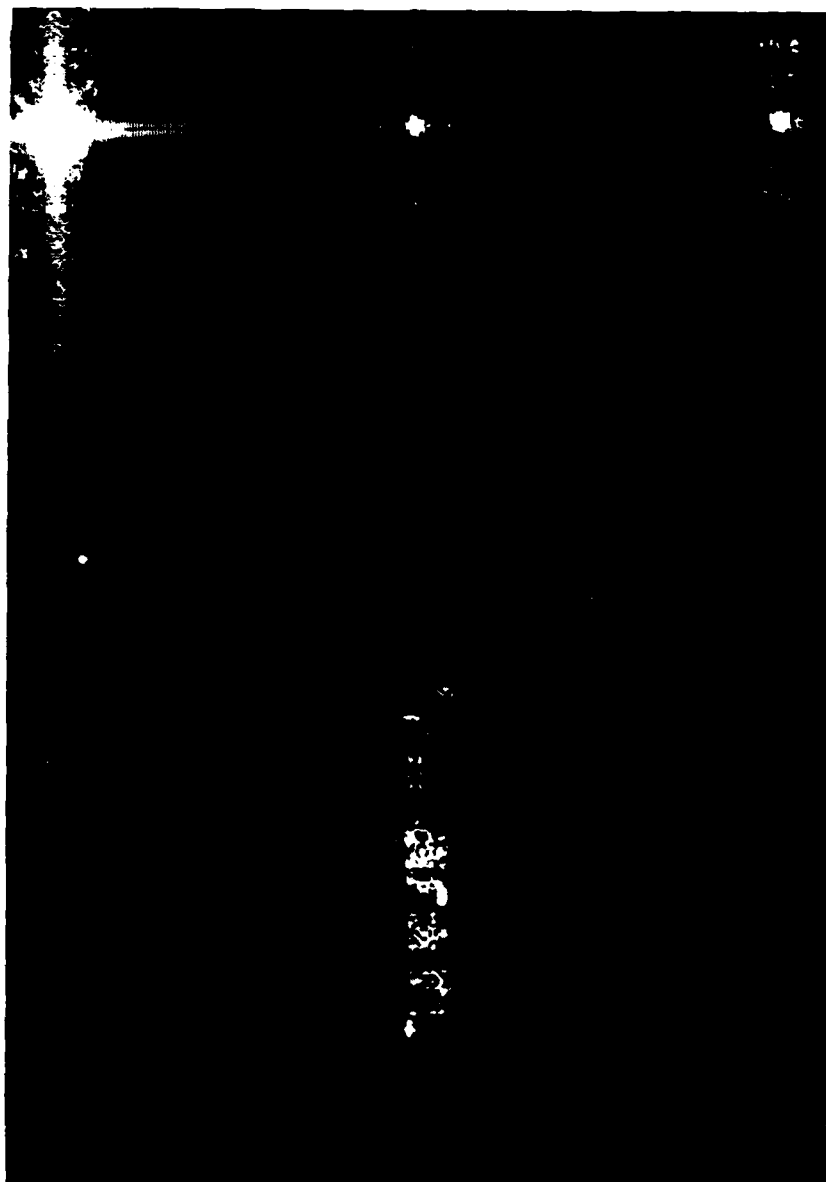   Filtering

Figure 64. Reconstruction from an A-K CGH Matched Filter
of the Letters AFATL Using High-Frequency Emphasis

Figure 65. Reconstruction from an A-K CGH Matched Filter
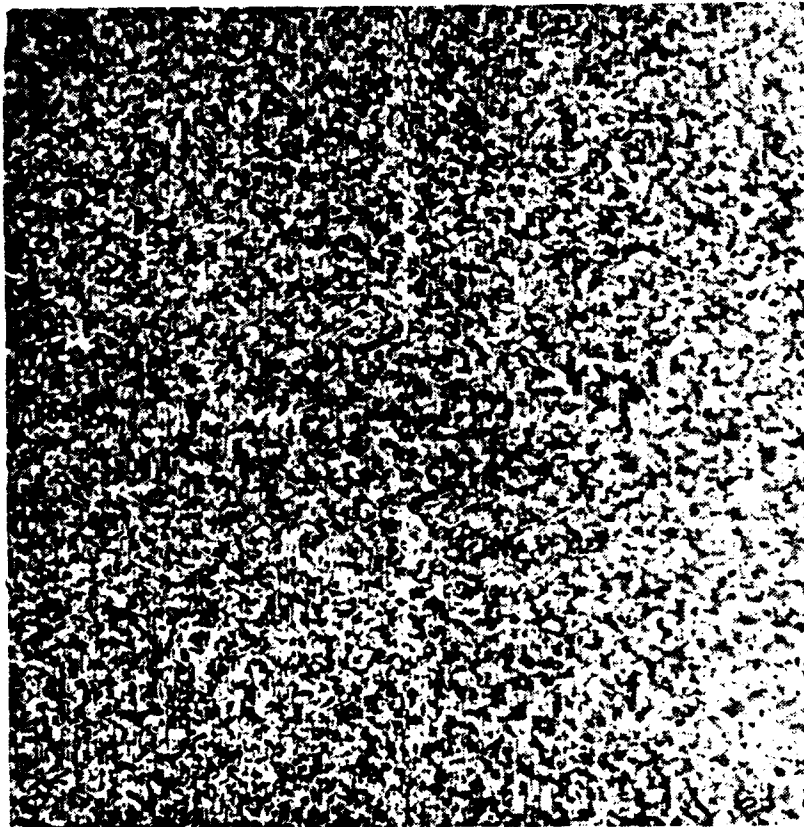of the Letters AFATL Using Phase-Only Filtering

159

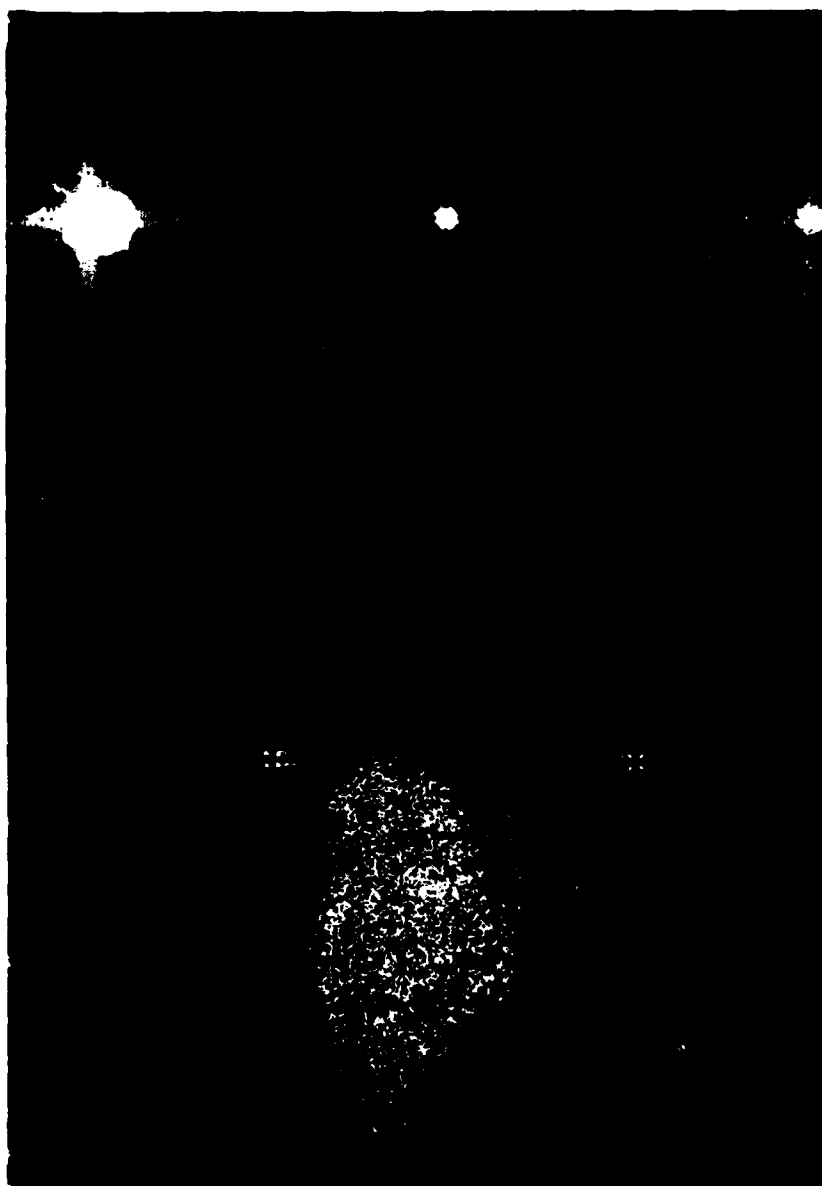Figure 66. A-K CGH Matched Filter of the SDF
Shown in Figure 14

Figure 67.   Reconstruction of an A-K CGH Matched Filter
of an SDF

161

SECTION IX
CONCLUSIONS AND RECOMMENDATIONS

By accounting for the changes in SBP, CGH techniques proposed by Lohmann, Lee, Allebach and others for reconstruction holography were used to create matched filters. These holograms, when used for matched filtering, require a larger space-bandwidth product to separate the correlation terms. Modifications to the conventional matched filter produced in this fashion improve the signal-to-noise and efficiency over that possible with conventional holography. Predistortion of the holographic pattern for minimizing the total harmonic distortion (THD) created by film nonlinearity provided a factor of five improvement. Each of these effects, along with various CGH types, were encoded in powerful computer simulations to test and demonstrate the benefits of the optimization techniques. The simulation showed that in all cases, high-frequency or phase-only filtering of the reference transform provided improved signal-to-noise ratios and efficiencies. Although phase-only filtering provided the lowest dynamic range, it did not in general provide results superior in overall performance. A-K CGH matched filters produced using E-beam lithography performed as predicted by the simulations.

This report shows that CGH matched filters can be produced and perform satisfactorily when preprocessing is utilized. However, further study is recommended in several areas. The A-K hologram can be improved by utilizing larger cell sizes and by making the cell dimension, N, an odd number as in the GBCGH. This would provide better representation of the complex value and render more pixels open, improving efficiency. Techniques for real-time dc elimination could permit programmable on-axis holograms, and thus, sequential single filters.

A better algorithm is needed to choose the appropriate frequency weighting for a given probability of detection and false-alarm rate. The nonlinear analysis would benefit from the incorporation of a spline model for the film. Corrections to the MTF should be modeled and included in the simulation to verify the belief that its effect is negligible. More simulation runs on a wide variety of images would better determine the interaction of image parameters with hologram performance.

Optical processors will provide low cost and high speed pattern recognition for specific tasks (Reference 68). It must be noted that Vander Lugt-type optical pattern recognizers are very limited in their application. The correlation process is hard-wired into the optical computer and the pattern recognition algorithm cannot be modified. This is in contrast to the slower digital processors which are software-controlled to perform a wide variety of image functions. The Vander Lugt correlator will fill a near-term need for a cheap processor of moderate performance in a controlled environment. This works well for machine vision application where the in-class and out-of-class patterns are well known. Performance of such a hard-wired system drops when the target is in a hostile environment.

The next generation of optical processors must be able to adapt and learn. Such a device must support a wide variety of algorithms at a speed compatible with the application. Future research in this area should address the need for optical processors which not only perform Fourier transforms and correlation, but many other functions as well. This could be accomplished using real-time spatial light modulators as holographic elements. The hologram might perform as a holographic lens for Fourier transforming or it might perform a coordinate transform depending on the command of a digital, high level, image understanding algorithm. Such a flexible processor should prove successful for the next generation of pattern or target recognizers.

REFERENCES

1. Hall, E. L., _Computer Image Processing and Recognition_ (Academic Press, New York, 1979), p. 158.

2. Tou, J. T., Gonzalez, R. C., _Pattern Recognition Principles_ (Addison-Wesley, Reading, MA 1974), p. 272.

3. Leib, K.G., _Optical Correlator Memory for a Terminal Homing Missile Scenario Against a Complex Target_ (Grumman Research Department Memorandum RM-660 August 1978), p. 31.

4. Casasent, D., Psaltis, D., Appl. Opt., 15 (1976), pp. 1795-1799.

5. Neff, J.A., Opt. Eng. 19-2 (1980) pp. 205-210.

6. Duthie, J.G., Upatnieks, J., Christensen, C.R., McKenzie, R.D., SPIE Proc. 231 (1980), pp. 281-290.

7. Vander Lugt, A., IEEE Trans. Inform. Theory, IT-10 (1964), p.139.

8. Lohman, A.W., Brown, B.R., Appl. Opt., 5 (1966), p. 976.

9. Lee, W.H., Appl. Opt., 9 (1970), p. 639.

10. Burckhardt, C.B., Appl. Opt., 9 (1970), p. 1949.

11. Haskell, R.E., J. Opt. Soc. Am., 63 (1973), p. 504.

12. Allebach, J.P., Appl. Opt., 20 (1981), p. 290.

13. Allebach, J.P., Keegan, J.J., J. Opt. Soc. Am., 68 (1978), p. 1440.

14. Horner, J.L., Appl. Opt., 21 (1982), p. 4511.

15. Oppenheim, A.V., Lim, J., IEEE Proc. 69-5 (1981), p. 529.

16. Thomas, J.B., _Statistical Communication Theory_ (Wiley, New York, 1969).

17. Yu, F.T., _Optics and Information Theory_ (Wiley, New York, 1976).

18. Chu, D.C., Fienup, J.R., Goodman, J.W., Appl. Opt., 11 (1972), p. 1386.

19. Goodman, J.W., _Fourier Optics_ (McGraw-Hill, New York, 1968).

20. Lee, W.H., "Computer Generated Optics: Techniques and Applications," _Progress in Optics_, E. Wolf, Ed. (North-Holland, Amsterdam, 1978), 16, pp. 121-232.

21. Fairchild, R., Fienup, J.R., SPIE Proc. 215 (1980), p. 1.

22. Ih, C.,Kong, N., Giriappa, T., Appl. Opt., 17 (1978), p. 1582.

23. Gallagher, N.C., Angus, J.C., Coffield, F.E., Edwards, R.V., Mann, J.A., Appl. Opt., 16 (1977), p. 413.

24. Caulfield, H.J., "Computer Generated Holography," Lasers and Applications (May 1983), pp. 59-64.

25. Strand, T.C., Opt. Eng., 13 (1974), p. 219.

26. Taub, H., Schilling, D.L., Communication Systems (McGraw-Hill, New York, 1971).

27. Max, J., IRE Trans. Inform Theory IT-6 (1960), pp. 7-12.

28. Caulfield, H.J., Appl. Opt., 21 (1982), p. 4391.

29. Butler, S.F., Riggins, J., SPIE Proc. 519 (1984), pp. 78-84.

30. Tescher, A., Tech. Rep USCIPI-510, Image Processing Institute, Univ. Southern California, Los Angeles, Dec. 1973.

31. Pearlman, W., Gray, R., IEEE Trans. Inform. Theory, IT-24, (1978), pp. 683-692.

32. Kirmisch, D., J. Opt. Soc. Amer., 60-1 (1970), pp. 15-17.

33. Psaltis, D., Paek, E., Venkatesh, S., Opt. Eng., 23-6 (1984), pp. 698-704.

34. Lesen, L.B., Hirsch, P.M., Jordan, J.A., Proc. Symp. Modern Optics, (New York, Polytechnic Instutute of Brooklyn 1967).

35. Kirk, J.P., Jones, A.L., J. Opt. Soc. Am., 61 (1971), p. 1023.

36. Chu, D.C., Fienup, J.R., Opt. Eng., 13 (1974), p. 189.

37. Benton, S.A., "Photographic Materials and Their Handling", Handbook of Optical Holography, J. Caulfield, Ed. (Academic Press, New York, 1979), pp. 349-366.

38. Cathey, W.T., Optical Information Processing and Holography, (Wiley, New York, 1974), p. 148.

39. Cathey, W.T., Appl. Opt., 9 (1970), pp. 1478-1479.

40. Vander Lugt, A., Rotz, F., Klooster, A., Optical and Electrooptical Information Processing, J. Tippett, Ed. (MIT Press, Cambridge, 1965).

41. Gara, A., Appl. Opt., 18 (1979), p.172.

42. Almeida, S., Eu, J., Appl. Opt., 15 (1976), p. 510.

43. Duthie, J.G., Upatnieks, J., Opt. Eng., 23-1 (1984), pp. 7-11.

44. Leger, J., Lee, S., Appl. Opt., 21 (1982), p. 274.

45. Casasent, D., Psaltis, D., Proc. IEEE 65-1 (1977), p. 77.

46. Bryngdahl, O., J. Opt. Soc. Am. 64-8 (1974), pp. 1092-1099.

47. Cederquist, J.N., Tai, A.M., Appl. Opt., 23 (1984) p. 3099.

48. Saito, Y., Komatsu, S., Ohzu, H., Opt. Comm. 47-1 (1983), pp. 8-11.

49. Bartelt, H., Case, S.K., Opt. Eng., 22-4 (1983), pp. 497-500.

50. Walkup, J.F., Opt. Eng., 19-3 (1980), pp. 339-346.

51. Anderson, R.C., Callary, P.R., Appl. Opt., 20-8 (1981), pp. 1272-1274.

52. Casasent, D., Psaltis, D., Appl. Opt., 16 (1977), p. 1472.

53. Leib, K.G., Bondurant, R.A., Hsiao, S., Wohlers, M.R., Herold, R.,Appl. Opt., 17 (1978), pp. 2892-2899.

54. Leib, K.G., Bondurant, R.A., Wohlers, M.R., Opt. Eng., 19-3 (1980), pp. 414-420.

55. Caulfield, H.J., Haimes, R.Horner, J.L., Israel Journal of Tech. 18, (1980), pp. 263-267.

56. Hester, C., Casasent, D., Appl. Opt., 19-11 (1980), pp. 1758-1761.

57. Hester, C., Casasent, D., SPIE Proc. 302 (1981), pp. 108-116.

58. Hester, C., Casasent, D., SPIE Proc. 292 (1981).

59. Casasent, D., Kumar, B., Sharma, V., SPIE Proc. 360 (1982).

60. Leger, J., Lee, S.H., Appl. Opt., 21-2 (1982), pp. 274-287.

61. Caulfield, H.J., Weinberg, M., Appl. Opt., 21-9 (1982), pp. 1699-1704.

62. Hayes, M., IEEE Trans. ASSP 30-2 (1982), pp. 140-154.

63. Casasent, D., J. Opt. Soc. Am., 23-10 (1984), pp. 1620-1627.

64. Yu, F.T., IEEE Trans. Inform. Theory IT-17 (1971), p. 524.

65. Lee, W.H., Greer, M.O., J. Opt. Soc. Am., 61-3 (1971), p. 402.

66. Thomas, C.E., Appl. Opt., 11-8 (1972), p. 175b.

67. Arnold, S.M., Opt. Eng., 24-5 (1985), pp. 803-807.

68. Casasent, D., Opt. Eng., 24-5 (1985), pp. 724-730.

END
FILMED

6-86

DTIC